Gaming the School Choice Mechanism

YINGHUA HE*

December 25, 2009

First version: October, 2009

JOB MARKET PAPER

Abstract

Many public school choice programs use centralized mechanisms to match students with schools in absence of market-clearing prices. Among them, the Boston mechanism is one of the most widely used. It is well-known that truth-telling may not be optimal under the Boston mechanism, which raises the concern that the mechanism may create a disadvantage to parents who do not strategize or do not strategize well. Using a data set from Beijing, this paper investigates parents' strategic behavior under the Boston mechanism and its welfare implications. School choice is modeled as a simultaneous game with parents' preferences being private information. The paper derives restrictions on parents' behavior under various assumptions on their information and sophistication, and the model is estimated by simulated maximum likelihood. The results suggest that parents' sophistication is heterogeneous; when parents have a greater incentive to behave strategically, they pay more attention to uncertainty and strategize better. There is no robust evidence that wealthier/more educated parents strategize better. If the Boston mechanism is replaced by the Deferred-Acceptance mechanism under which truth-telling is always optimal, among the sophisticated parents who always play a best response, the majority of them are worse off, and almost none of them are better off. The reform benefits half of the naive parents who are always truth-telling, while it also hurts about 20% of them.

JEL CODE: I28 C78 C15 D81

KEYWORDS: *School Choice, the Boston Mechanism, Bayesian Nash Equilibrium, Strategy-Proofness, Simulated Maximum Likelihood*

*Department of Economics, Columbia University. Email address: yh2165@columbia.edu. I am deeply indebted to Fang Lai for her generous help which makes this project possible. For their advice, constant encouragement and support, I am grateful to Kate Ho, W. Bentley MacLeod, Bernard Salanié, Miguel Urquiola and Eric Verhoogen. For their helpful comments and suggestions, I thank Atila Abdulkadiroğlu, Jushan Bai, Yeon-Koo Che, Pierre-André Chiappori, Navin Kartik, Wojciech Kopczuk, Dennis Kristensen, Hong Luo, Serena Ng, Brendan O'Flaherty, Michael Riordan, Jonah Rockoff, Johannes Schmieder, Herdis Steingrimsdottir, and participants at the 2009 Econometrics Society North American Summer Meeting in Boston. Financial supports from Columbia's Wueller and Vickrey research prizes and the Program for Economic Research are gratefully acknowledged.

1 Introduction

School choice programs are aimed at giving families more opportunities to choose the school for their children. An increasing number of countries around the world have adopted some form of school choice.[1] In the U.S., since the majority of students attend public schools, choosing among public schools may be the only option to many families, especially those with lower socio-economic status.[2,3] As more states enact school choice programs, more students have the opportunity to attend a public school of their choice.[4]

One popular form of school choice is known as *open enrollment*, which is used in all but four states in the U.S.[5] It offers a set of public schools from which families can choose, conditional on seat availability.[6] However, there is typically excess demand for good schools. In absence of market-clearing prices, a centralized mechanism is often necessary in order to determine who should be assigned to which school. Although the popularity of school choice is high and still growing, the question about which assignment mechanism should be used is still hotly debated.

At the center of the debate is the *Boston mechanism* (henceforth BM) which is one of the most widely used mechanisms. It was employed until 2004-05 by the Boston Public Schools, and is in practice in Cambridge, MA, Charlotte-Mecklenburg, NC, St. Petersburg, FL, Minneapolis, MN, Providence, RI and other school districts across the U.S.[7] It is also popular in other countries and in other context. For example, this is the mechanism used in China's college admissions.

The main criticism of BM is that it encourages parents to "game the system." Namely, it is not always in parents' best interests to report their true preferences when applying to schools (Abdulkadiroglu and Sonmez (2003)). Being concerned that "the need to strategize provides an advantage to families who have

[1]For example, Colombia (Angrist, Bettinger, Bloom, King, and Kremer (2002)), Chile (Hsieh and Urquiola (2006)), Finland (Seppanen (2003)), New Zealand, Denmark, and Sweden (Hepburn (1999)) are among the many other countries that have government policies promoting school choice.

[2]For example, in the U.S., 96% of grades 1-12 students from families in poverty are enrolled in public schools in 2003. Among all the students, the fraction is 89%. (Figure 2.2 in Tice, Chapman, Princiotta, and Bielick (2006)).

[3]On the other hand, many school choice programs encourages a switch from public to private schooling, for example vouchers, scholarships and tuition tax credits. A summary of types of school choice implemented in the United States is available at: http://www.heritage.org/research/education/schoolchoice/typesofschoolchoiceRD.cfm. Retrieved October 15, 2009.

[4]From 1993 to 2003, among all the grade 1-12 students in the U.S., those who were enrolled in a public school of their choice increased from 11% to 15% (Tice, Chapman, Princiotta, and Bielick (2006)). Among all students from families in poverty, this fraction increased from 14% to 18%. After the *No Child Left Behind* was enacted in 2002, students at lagging schools are allowed to attend other public schools in the district. This may also accelerate the increasing trend of public school choice.

[5]The District of Columbia, and four states, Alabama, Maryland, North Carolina, and Virginia, have not enacted any form of open enrollment. Source: the Education Commission of the States, 2008, "Open Enrollment: 50-State Report", available at http://mb2.ecs.org/reports/Report.aspx?id=268. Retrieved October 15, 2009.

[6]Cullen, Jacob, and Levitt (2006) investigates the high school admission in the Chicago Public Schools which is an example of open enrollment. Students can apply to gain access to public magnet schools and programs outside of their neighborhood school, but within the same school district.

[7]In Florida, many other school districts use the Boston mechanism as well, for example, Calhoun, Flagler, Lee, Madison, Palm Beach, Polk, and St. Lucie. A brief description of the mechanism in these districts and other mechanisms used in other districts are available at: https://app1.fldoe.org/flbpso/COEPSearch/. Retrieved October 15, 2009.

the time, resources and knowledge to conduct the necessary research (Payzant (2005))," the Boston School Committee voted in 2005 to replace the Boston mechanism with the student-proposing Deferred-Acceptance mechanism (henceforth DA, Gale and Shapley (1962)).[8] One main feature of DA is its strategy-proofness – reporting true preferences is always a dominant strategy (Dubins and Freedman (1981); Roth (1982)).

However, researchers have not reached a consensus on the welfare effects of such a reform. Experimental and empirical evidence in previous literature suggests that parents in BM have heterogeneous sophistication, or different ability to strategize.[9] If some parents do not understand the mechanism well, they may lose, while more sophisticated parents benefit (Pathak and Sonmez (2008)). Therefore, replacing BM by DA is welfare-enhancing, at least for less sophisticated parents. On the other hand, a recent strand of literature provides theoretical and experimental results in favor of BM.[10] Under certain circumstances, even less sophisticated parents can be better off in BM (Abdulkadiroglu, Che, and Yasuda (Forthcoming)). As the debate goes on, BM remains among the most popular mechanisms in practice.

Using data from Beijing where students are assigned to middle schools under BM, this paper provides new evidence on parents' strategic behavior in BM and its welfare implications. Unlike previous empirical studies which are either experimental or reduced-form, it takes a structural approach to estimating parents' preferences over schools and their degree of sophistication. Assuming that preferences do not change if switching from BM to DA, the paper simulates the outcomes under DA and compares them to those under BM. It answers two questions (1) whether poorer parents are less sophisticated and (2) whether BM harms less sophisticated parents relative to DA.

In BM, parents play a non-trivial preference revelation game. They submit a ranking of schools, and schools form a priority ordering of students according to some rules which usually entail lotteries as tie-breakers. In Beijing, schools' priority is only determined by a random lottery. At the beginning, each school considers the students who rank it first, and assigns seats in order of their priority at that school. Then, each school that *still has available seats* considers *unmatched* students who ranked it second, and assigns seats again in order of their priority. This process continues until the market is cleared. If a student ranks a popular school first and gets rejected, her chance of getting her second choice is greatly diminished because she can only be accepted after everyone who lists that school as the first choice. This can happen even when she has the highest priority at her second choice school. Therefore, the mechanism provides incentives for students to rank less popular schools higher.[11]

[8]See Abdulkadiroglu, Pathak, Roth, and Sonmez (2005) for a detailed description of the reform.

[9]For example Abdulkadiroglu, Pathak, Roth, and Sonmez (2006), Chen and Sonmez (2006), Lai, Sadoulet, and de Janvry (2009), and Pais and Pinter (2008).

[10]For example, Abdulkadiroglu, Che, and Yasuda(2008, Forthcoming), Featherstone and Niederle (2008), and Miralles (2008).

[11]In real life, this is well known to some parents. For instance, the West Zone Parents Group in Boston, recommends two types

DA differs from BM only in the admission process. At the beginning, each school considers students who rank it first and *tentatively assigns* seats in order of their priority at the school. Then, *every* school pools *those tentatively accepted* with unassigned students who rank it second, and again *tentatively assigns* seats in order of students' priority. This process repeats, with any unassigned students being considered by the next school on their list, and getting in if there are seats left, or bumping other students if they are higher on the priority list. When all students have been tentatively assigned, the process ends, and the assignments become final. In this mechanism, top-ranking a popular school and being rejected by that school do not sacrifice the chance of getting into lower choices. This makes truthful ranking a dominant strategy.

If schools have strict rankings over all students without random tie-breaking, DA produces the *student optimal stable matching* which is most preferred by every student among all stable matchings (Gale and Shapley (1962)).[12] This feature makes DA very popular in many contexts, for example, matching medical students with residency programs (Roth (1984)).[13] However, schools usually do not have strict rankings. For example, in the Boston Public Schools, students are given priority based on walk zone status and sibling's enrollment. There are many students in each priority class, and it is inevitable to use lotteries as tie-breakers. Whenever a school has to decide between two students in the same priority class, instead of their preferences, it is the lottery that matters. This incurs a potential efficiency loss under DA.[14,15]

To evaluate parents' welfare, this paper adopts the concepts of Bayesian Nash equilibrium and ex ante efficiency. Parents' preferences are private information, and lotteries are unknown when applying to schools.[16] In terms of ex ante efficiency, it is impossible to have a strategy-proof and efficient mechanism which treats the same type of parents equally (Zhou (1990)). DA imposes strategy-proofnees at the cost of efficiency.

When applying to schools, parents need to consider the uncertainty carefully. If the popularity of each

of strategies to its members in 2003, "One school choice strategy is to find a school you like that is undersubscribed and put it as a top choice, OR, find a school that you like that is popular and put it as a first choice and find a school that is less popular for a 'safe' second choice." Parents are also advised to do so. For example, in St. Petersburg, FL, a newspaper makes following suggestions: "Make a realistic, informed selection on the school you list as your first choice. It's the cleanest shot you will get at a school, but if you aim too high you might miss." (St. Petersburg Times; Sep 14, 2003; pg. 17)

[12]A matching is *stable* if no student or school can be strictly better off by leaving the current match and staying unmatched or rematching with some other participant without making him worse off.

[13]In this case, hospitals have strict preferences over residents/students. For a summary of theoretical properties of DA and its applications, please see Roth (2008).

[14]Using field data under DA, Abdulkadiroglu, Pathak, and Roth (2009) empirically documents that the potential efficiency loss associated with strategy-proofness is significant. But their comparison is not between DA and BM. Moreover, in this case, no strategy-proof mechanism can always produce a stable matching that is Pareto optimal (Erdil and Ergin (2008); Abdulkadiroglu, Pathak, and Roth (2009)). To recover some efficiency, Abdulkadiroglu, Che, and Yasuda (2008) propose a choice augmented deferred-acceptance (CADA) mechanism which brings some elements of BM into DA.

[15]Some parents viewed strategy-proofness as a loss of their influence on the outcome and opposed the adoption of DA in Boston. A parent at a Public Hearing by the Boston School Committee, 05/11/2004, stated, "I'm troubled that you're considering a system that takes away the little power that parents have to prioritize ... what you call this strategizing as if strategizing is a dirty word."

[16]In previous literature, some papers assume complete information, for example Ergin and Sonmez (2006) and Kojima (2008). They focus on Nash equilibrium and *ex post* efficiency. Recently, the ex ante view becomes popular, for example, Abdulkadiroglu, Che and Yasuda(2008, Forthcoming), Featherstone and Niederle (2008), Miralles (2008).

school is known, parents can figure out the probability of being accepted by each school when submitting different rankings. In BM, these probabilities serve as prices. If the probability is low even when the school is ranked first, then there is a huge excess demand for this school, and only those who can afford the price, i.e. having a high cardinal utility of the school, should rank it first. Otherwise, even if the school is the most preferred, one should hedge the risk of bad outcomes by listing other schools as first choice. In this way, BM creates a market for cardinal utilities. Particularly, if the schools' priority is only determined by lotteries and the market is large, the equilibrium in BM can be close to a competitive equilibrium (He (2009)).

On the contrary, DA eliminates the possibility of signaling the cardinal utility and hedging the risk of bad outcomes. If all students share the same ordinal preferences over schools but differ in cardinal preferences, DA performs very poorly relative to BM.[17,18] In this case, if schools' priority is only determined by lotteries, students would just be randomly assigned to schools under DA.

The above efficiency comparison assumes that every parent is sophisticated. It is still a concern that BM may harm less strategic parents relative to DA. Abdulkadiroglu, Pathak, Roth, and Sonmez (2006) provide some evidence of heterogeneous sophistication.[19] Allowing some naive parents who are always truth-telling in BM, Pathak and Sonmez (2008) find them worse off relative to DA in a complete information Nash equilibrium. On the contrary, when considering incomplete information, Abdulkadiroglu, Che, and Yasuda (Forthcoming) find that BM may not harm but rather benefit naive parents in terms of ex ante efficiency if students have the same ordinal preferences and schools' priority is only determined by lotteries.

To address these issues empirically, the major challenge is to estimate parents' true preferences when

[17]This efficiency comparison can be illustrated in the following example which first appeared in Abdulkadiroglu, Che, and Yasuda (2008). Suppose three student, $\{1, 2, 3\}$, are to be assigned to three schools, $\{A, B, C\}$, each with one seat. Schools have no priorities over students, and students' preferences are represented by the following von-Newmann Morgenstern utility values, where $u_{i,s}$ is student i's utility for school s:

	$v_{1,s}$	$v_{2,s}$	$v_{3,s}$
$s = A$	0.8	0.8	0.6
$s = B$	0.2	0.2	0.4
$s = C$	0	0	0

If DA is used, all students submit true ordinal preferences, and they are assigned to the schools with equal probabilities. Their expected utilities are $EU_1^{DA} = EU_2^{DA} = EU_3^{DA} = \frac{1}{3}$. If BM is used instead and preferences are common knowledge, in the unique equilibrium, students 1 and 2 still report truthfully, while student 3 submits (B, A, C) with B as the first choice. The assignment is that students 1 and 2 are randomly assigned between A and C and student 3 is assigned to B. The resulting expected utilities are $EU_1^{BM} = EU_2^{BM} = EU_3^{BM} = 0.4$. The assignment Pareto dominates the one in DA.

[18]See for example, Abdulkadiroglu, Che and Yasuda(2008, Forthcoming), Miralles (2008), and Featherstone and Niederle (2008).

[19]They investigate parents' strategic behavior in BM without estimating parents' preferences. Using school choice data from Boston, they define it as a mistake if a parent's second choice is an over-demanded school which has more students than its capacity ranking it first. Quite a few parents make this mistake and are adversely affected, which suggests the heterogeneity in parents' sophistication. The implicit assumption here is complete information, which means students know others' preferences. However, the decision is made *ex ante* when nobody knows for sure whether one school will be over-demanded or not. Under incomplete information, this kind of mistake may be consistent with optimal behavior.

they are not truth-telling in BM.[20] This paper provides a solution while allowing heterogeneous sophistication among parents. Formally, a parent is sophisticated if she correctly takes into account the uncertainty about others' preference and sophistication. Her subjective beliefs – the perceived probabilities of being accepted by each school when submitting different lists – are the same as the equilibrium beliefs. Less sophisticated parents have inaccurate beliefs, while naive parents disregard the uncertainty and always submit the true preferences. To recover meaningful identification in this context, restrictions are imposed on the preferences and sophistication. The distribution of utility is assumed to have a known functional form, and several cases of sophistication levels, from less to more restrictive, are considered.

In the first case, parents may not have accurate beliefs, but they understand the rules of BM. Therefore, beliefs must satisfy the properties imposed by the rules. For instance, moving a school toward the top of the list always (weakly) improves the probability of being accepted by that school. These properties lead to a set of restrictions on students' equilibrium behavior. For example, the first choice must be an acceptable school. They provide identifying conditions for the utility function, and the model is estimated by simulated maximum likelihood.[21] Given the estimates of the utility function, or the distribution of utilities, the subjective beliefs of each parent can be estimated as those which rationalize the observed behavior the best. In other words, the estimated subjective beliefs maximize the probability that the actual submitted list is a best response.

To determine the accuracy of subjective beliefs, a measure of the correct or equilibrium beliefs is necessary. When parents' sophistication is heterogeneous, it is difficult to fully characterize the equilibrium. However, when there are many participants in the game, the uncertainty in preferences will be averaged out to some extent. The equilibrium beliefs can be well approximated by the ex post beliefs calculated from the realization of the game.

The second case assumes that every parent is sophisticated and has correct beliefs in equilibrium. In this case, the possible difference in parents' beliefs only comes from the differences in their information set. When the number of parents is large, even if their information is slightly different, their beliefs will be very close to each other. The paper then focuses on the equilibrium when the beliefs are common to everyone. A method of simulated maximum likelihood with equilibrium constraints is proposed to jointly estimate the preferences and the equilibrium beliefs.

The second case is nested in the first one, and results from a model selection test reject the all-sophisticated

[20] Hastings, Kane, and Staiger (2008) estimate the demand for schools under the assumption that students are truth-telling in BM. Their data are from Charlotte-Mecklenburg Public School District in 2002 when BM was just implemented. Therefore, the truth-telling assumption may be more likely to be valid in their setting than others.

[21] This is similar to Berry (1992) who estimates the profit function of airline companies by features that are common to all equilibria when multiple equilibria exist.

case, suggesting that sophistication is heterogeneous. I then use the ex post beliefs as an estimate of equilibrium beliefs, and obtain several measures of sophistication, for example, the accuracy of parents' subjective beliefs which is measured by its Euclidean distance to the equilibrium beliefs. Given the actual submitted list, the paper also calculates the probabilities that it is the true preference or a best response. The mean and variance of utility achieved by the parents are estimated and compared to those associated with the list which maximizes the mean utility. Furthermore, to measure the incentive to be strategic, the paper calculates the probability of truth-telling being a best response, and the mean and variance of utility achieved when truth-telling, relative to when best responding.

Results show that on average, parents are 16.1% more likely to play a best response than to report truthfully. Among all participants, they achieve 94.4% of the maximal mean utility with a smaller variance.

The data also contains information on how much attention parents pay to school quality and uncertainty in the game. Exploiting this information which is not used in the estimation of preferences, the paper finds that parents who have a greater incentive to be strategic pay more attention to uncertainty but not to school quality. They strategize better in the sense that their beliefs are more accurate, they are also more likely to best respond, and they obtain a higher mean utility with a lower variance.

Paying more attention to uncertainty is associated with more accurate beliefs and a higher probability of playing a best response. Interestingly, it is negatively correlated with the variance of utility but not correlated with the mean utility. This indicates risk aversion may be important in the welfare evaluation. On the other hand, paying more attention to school quality has no significant correlation with these measures except a higher standard deviation in utilities. There is evidence that wealthier parents pay more attention to quality, but not to uncertainty. They strategize better on a few dimensions, but the results are not robust. This is consistent with the low incentive for them to "game the system." Their true preferences are more likely to be a best response because they have a better outside option.

Assuming that the preferences do not change across different mechanisms, the paper simulates the outcomes under DA with the estimated preferences. If the BM is replaced by the DA mechanism, almost all the sophisticated parent are weakly worse off, and the majority of them are strictly worse off. Among all the naive parents, about half of them are better off, and surprisingly, around 20% of them can be hurt by the reform.

Other Related Literature

The results found in this paper are consistent with Budish and Cantillon (2009). They use a data set on MBA students' strategically reported preferences as well as their true preferences to study strategic behavior in the course-allocation mechanism used at Harvard Business School. The mechanism is not strategy-proof and its ex-ante welfare is higher than under the random serial dictatorship which is strategy-proof. Implicitly, they assume that every student is sophisticated. In their model, the realization of students' preferences is common knowledge, and they focus on the Nash equilibrium. Their ex ante welfare is equivalent to the "interim" welfare in the current paper – the types of students are observed, but the lottery is unknown.

This study relates to the literature on testing if an equilibrium is played in real life games. For example, Chiappori, Levitt, and Groseclose (2002) and Kovash and Levitt (2009) investigate the case of professional sports. Hortacsu and Puller (2008) is a particularly relevant paper which looks at the strategic bidding in an electricity spot market auction. They characterize a Bayesian-Nash equilibrium model and compare actual bidding behavior to theoretical benchmarks. Evidence shows that large firms perform close to the benchmarks, while smaller firms deviate significantly. The difficulty in estimating a Bayesian-Nash equilibrium is to specify the beliefs. Under some technical assumptions, Hortacsu and Puller show the best response is also ex post optimal, i.e. seeing other players' behavior would not change one's behavior. Thus, they can just look at the ex post optimality without evaluating the beliefs. In contrast, the current study measures the subjective and objective beliefs, and thus provides more measures of sophistication.

This paper also relates to the literature on estimating simultaneous games of incomplete information. Most of previous studies rely on the condition of consistent beliefs to derive moment conditions or choice probabilities, for example, Seim (2006), Bajari, Hong, Krainer, and Nekipelov (2006), Aradillas-Lopez (2007a), and Aradillas-Lopez (2007b). Given the small number of players, these studies need multiple plays of the game for identification. It requires that the equilibrium beliefs are correct and stable across different plays of the game. The current study only uses one play of the game and allows beliefs to be inconsistent.

The remainder of the paper is organized as follows. Section 2 describes the two school choice mechanisms. Section 3 formalizes the school choice problem under the Boston mechanism as a Bayesian game. Restrictions on parents' behavior are derived under various assumptions. Section 4 describes how to use the restrictions to identify the model and proposes a method of simulated log-likelihood. Tests for overall sophistication are discussed. The data set is described in Section 5. Section 6 presents the estimation results. Section 7 shows the counterfactual analysis. The paper concludes in Section 8.

2 Background: Two School Choice Mechanisms

2.1 The Boston Mechanism and its Application in Beijing

In the following, "students" and "parents" are two interchangeable terms since that the school choice decision is mainly made by students' parents.

The Boston mechanism (BM) works as follows:

(i) Each school has a priority ordering of students which is determined by state or local law. For example, in Boston, there are four priority groups in the following order (a) students who have siblings at the school (siblings) and are in the school's reference area (walk zone), (b) siblings, (c) walk zone, (d) other students. Whenever needed, a lottery serves as a tie-breaker. In some other places, a test score or a lottery is the only determinant.

(ii) Schools announce enrollment quota and students submit a ranking of the schools. The strict priority ordering of students at each school is realized after the application is submitted.

(iii) Based on the priority ordering and submitted rankings, the matching process goes for several rounds:

Round 1. Each school considers all the students who list it as their first choice and assigns seats in order of their priority at that school until either there is no seat left or no such student left.

Generally, in

Round k. The kth choice of the students who have not yet been assigned is considered. Each school that still has available seats assigns the remaining seats to the students who list it as their kth choice in order of their priority at that school until either there is no seat left or no such student left.

The process terminates after any round k when every student is assigned a seat at a school, or if the only students who remain unassigned listed no more than k choices. Unassigned students are then matched with available seats randomly.

In 1998, the Eastern City District of Beijing adopted a version of BM for middle school admission. Students' priorities are solely determined by a random lottery (a single tie-breaker).[22] In 1999, the district was divided into 15 school neighborhoods based on students' elementary school enrollment. All schools were given a neighborhood specific enrollment quota by the Education Bureau. The best schools were available to more than one school neighborhood, while most lower-quality schools were only available to the school neighborhood of proximity. Students could submit a preference ranking and apply to all the middle schools available to their particular school neighborhood. After students submit the preference ranking, a computer-generated 10-digit number is randomly assigned to each student. The admission proceeds exactly as described above. This study uses data from the largest neighborhood which has access to 4 schools with a total quota of 960. Their submitted lists are observed, along with rich information on students and schools.

[22]Before the reform, elementary school students were admitted by public middle schools on a merit basis. BM was also used in middle school admission. The priority was only determined by test scores.

Students' outside option mainly consists of the 28 public schools in the district, including the four schools to which they can apply. In 1999, the private school system was not well developed and the new version of BM was implemented in all districts in Beijing. Moreover, the Eastern City District has a very good reputation in educational quality among all districts in Beijing, in addition to its advantage in location. Therefore, there was not much incentive for students to leave the public school system or to transfer out of this district. Students may choose the outside option – choose a public school outside of the Boston mechanism – through three channels. First, schools admit some students directly if their parents are employed in the school, if the students have received at least a city-level prize in academic or special skill achievements, or if a considerable direct payment is made to the school. Second, in addition to the quota announced, some top schools admit students by offering an admission exam. Third, schools admit some transfer students who are not satisfied with their assignment and make a considerable direct payment to the school.

2.2 The Deferred-Acceptance Mechanism

The DA mechanism also collects students' preference ranking and uses priorities in the admission process. It proceeds as follows:

Round 1. Every student applies to their first choice school. Each school rejects the lowest-priority students in excess of its capacity. Those who are not rejected are *temporarily held* by the schools.

Generally, in

Round k. Every student who is rejected in Round $k-1$ applies to the next choice on her list. Each school pools new applicants and those who are held from previous rounds together, and rejects the lowest-priority students in excess of its capacity. Those who are not rejected are *temporarily held* by the schools.

The process terminates after any round k when no rejections are issued. Each school is then matched with students it is currently holding.

If schools' priority is determined by the same criterion, for example, test scores or one single lottery, a student who has a high priority at one school also has a high priority at other school. In this case, the deferred acceptance mechanism is equivalent to the serial dictatorship (Abdulkadiroglu and Sonmez (1998)). Following their priority order, students sequentially choose the best among all available schools.

3 School Choice under Boston Mechanism as a Bayesian Game

In this section, the school choice problem under Boston mechanism is formalized as a Bayesian game. In a school choice problem, there are:

(i) a set of students/parents, $\{i\}_{i=1}^{I}$;

(ii) a set of schools, $\{s\}_{s=0}^{S}$, $S \geq 3$, where school zero is the outside option – home school, private schools and public schools through other channels;[23]

(iii) a capacity vector, $\boldsymbol{Q} = \{q_s\}_{s=1}^{S}$, $\sum_{s=1}^{S} q_s \geq I$ and $q_s < I$, $\forall q_s$.

(iv) a list of students' choices (preference rankings), $\{C_i\}_{i=1}^{I}$, where $C_i = \left\{c_i^1, ..., c_i^S\right\}$, $c_i^k \in \{s\}_{s=0}^{S}$, $\forall k = 1, ..., S$;

(v) and school's priorities which are only determined by random lottery numbers.

At the start of the game, each school announces its capacity, q_s. Usually, there are enough seats to accommodate all the students, i.e. $\sum_{s=1}^{S} q_s \geq I$, and any single school does not have enough seats to enroll all students, $q_s < I$, $\forall s = 1, ..., s$. Parents then submit their preference rankings or choice lists, $C_i = \left\{c_i^1, ..., c_i^S\right\}$ where c_i^1 is the first choice of parent i and c_i^S is Sth, or the last choice. If the outside option is not included in C_i, C_i is a full list, i.e. it includes all the schools in the system. Otherwise it is a partial list. Parents are allowed to submit partial lists.

After parents submit their applications, they are given random lottery numbers which determine their priority at all the schools. In other words, all students have the same *ex ante* priority. The analysis can be extended to other versions of BM where there are some pre-determined priority classes. With the applications and random lottery number, the admission process goes as described in the previous section. After students get their assignment, they can choose the outside option if they are not satisfied.

3.1 Utility Function

The von Neumann-Morgenstern (vNM, hereafter) utility function of student i attending s $(s = 1, .., S)$ is

$$u_{i,s} = u\left(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}\right)^{24}$$

where

[23] When the number of schools is less than 3, the game becomes trivial – everyone is truth-telling in the Boston mechanism.

[24] The vNM utility can be thought of as the expected utility of a future outcome. One may formulate the vNM utility function in the following way: Suppose the Bernoulli utility function of student i attending school s is

$$f_i\left(v_{i,s}, \xi_{i,s}\right),$$

where $f_i\left(\right)$ might be individual-specific; the value of $v_{i,s}$ is known by i, so is the distribution of $\xi_{i,s}$; but the realization of $\xi_{i,s}$ is unknown at the application stage. $v_{i,s}$ can be interpreted as the utility determined by student's own characteristics and observed school attributes. $\xi_{i,s}$ may include peer quality and other characteristics which are not observable at the application stage. The vNM utility function is the expectation of the Bernoulli utility function conditional on i's observed variables, $\boldsymbol{z}_s$, $\boldsymbol{X}_i$, and $\varepsilon_{i,s}$:

$$u_{i,s} = E\left[f_i\left(v_{i,s}, \xi_{i,s}\right) | \boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}\right] = u\left(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}\right).$$

• $\boldsymbol{z}_s \in \mathbb{R}^{K_1}$ are school attributes which are fixed and observed by parents when applying to schools;

• $\boldsymbol{X}_i \in \mathbb{R}^{K_2}$ are student/parent i's characteristics which are observed by researchers but may or may not be observed by other parents.

• $\varepsilon_{i,s} \in \mathbb{R}$ is the unobserved heterogeneity in the utility which is parent i's private information.[25]

• Both $\boldsymbol{X}_i$ and $\varepsilon_{i,s}$ are i.i.d. across students, and their distributions are common knowledge. Moreover, the distribution of $\varepsilon_{i,s}$ is continuous.

• $u\left(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}\right)$ is continuous and strictly monotonic in $\varepsilon_{i,s}$, and the variance of $max\left\{u_{i,s}, 0|\boldsymbol{X}_i\right\}$ is finite.

The utility when choosing the outside option is normalized to be zero.[26] If a school is worse than her outside option, or $u_{i,s} < 0$, it is an unacceptable school. Otherwise, it is an acceptable school.

Denote the vector of $\{\varepsilon_{i,s}\}_{s=1}^{S}$ as $\boldsymbol{\varepsilon}_i$, and assume that $\boldsymbol{\varepsilon}_i \perp \boldsymbol{X}_i$. $\boldsymbol{\varepsilon}_i$ is i.i.d. across i, while correlation between any $\varepsilon_{i,s}$ and $\varepsilon_{i,s'}$ is allowed. Part of the correlation is due to the normalization which introduces the error term in the utility of outside option into every $\varepsilon_{i,s}$. Another reason could be that there are unobserved heterogeneous effects of school attributes. For example, some schools are better at teaching science, while others are better in teaching art and humanities. Students who like science better will have a positive error term for one set of the schools and negative error terms for the other set of schools. $\overline{\boldsymbol{X}}_i \equiv (\boldsymbol{X}_i, \boldsymbol{\varepsilon}_i) \in \mathbb{R}^{K_2+S}$ is defined as students' "type" or the personal characteristics affecting the utility of attending each school. Hence, $\overline{\boldsymbol{X}}_i$ is i.i.d. across students and its distribution is common knowledge.

3.2 Information, Beliefs and Decision Making

The decision making process is to select an optimal element from the set of possible actions which is defined as $\mathcal{C} \equiv \left\{C^l\right\}_{l=1}^{L}$, where $L = S!\left(\frac{1}{S!} + \frac{1}{(S-1)!} + \frac{1}{(S-2)!} + ... + \frac{1}{1!}\right)$ is the total number of possible lists.[27] Each element in $\mathcal{C}$ is an ordered list of k different schools, $k = 0, ...S$. When choosing the optimal C, students process their information and form a belief system which specifies the perceived probability of

[25] There is a potential abuse of notation. To each parent, both $\boldsymbol{X}_i$ and $\varepsilon_{i,s}$ are observed, and thus they are constants. To the researcher and other parents, $\varepsilon_{i,s}$ and sometimes $\boldsymbol{X}_i$ are not observed, and thus they are random variables. To simplify the exposition, $\boldsymbol{X}_i$ and $\varepsilon_{i,s}$ denote both random variables and their realization.

[26] Equivalently, the utility function of attending any school $s \neq 0$ is the difference between the actual utility attending s and the utility when choosing the outside option.

[27] Usually, students are allowed to submit a list less than S – the number of total schools. Not submitting any school is also a choice. Any list containing $(S-1)$ schools is equivalent to list with S schools. Calculating the number of possible lists when submitting 0 to $(S-1)$ schools respectively, the total number of possible lists is the sum of S numbers:

$$\begin{aligned} & 1 + S + S * (S-1) + ... + S * (S-1) * (S-2) ... * 2 \\ = \; & S!\left(\frac{1}{S!} + \frac{1}{(S-1)!} + \frac{1}{(S-2)!} + ... + \frac{1}{1!}\right) \end{aligned}$$

Notice that those lists in which some school appears multiple times are excluded. They are obviously not optimal.

being accepted by each school given any submitted list C.

Let H_i denote i's set of information about other parents. $\mathcal{H}$ is the collection of all possible information sets. H_i contains two types of information: (1) the joint distribution of $\overline{\boldsymbol{X}}_{-i}$, H_{-i}, and other parents' beliefs, which is common knowledge; (2) realization of a subset of $\boldsymbol{X}_{-i}$. Parents may know realization of a subset of $\boldsymbol{X}_{-i}$ because they may know some other parents very well, for example their neighbors or those in their parent group. Therefore, the difference between H_i and H_j only comes from the different knowledge on the realization of $\boldsymbol{X}_{-i}$.

The belief system of parent i is how she processes the information. It is a probability measure, $B_i\left(C, H_i\right)$: $\mathcal{C} \rightarrow [0,1]^S$, such that $\forall C \in \mathcal{C}$:

$$B_i\left(C, H_i\right) = \left(P_1^A\left(C, i\right), ..., P_S^A\left(C, i\right)\right) \in [0,1]^S,$$

where $P_s^A\left(C, i\right)$ is the *subjective* probability of being assigned to school s when C is submitted. $B_i\left(\cdot, H_i\right)$ is dependent on the information set H_i which is used to form the expectation of other parents' behavior. Note that $B_i\left(\cdot, H_i\right)$ is individual specific because parents may have different ability to process the information.[28] Let $B\left(\cdot, H_i\right)$ denote the beliefs when parent i has the ability calculate the subjective probabilities correctly conditional on her information set. In other words, $B\left(\cdot, H_i\right)$ is the *objective* probability which can be calculated mathematically.

The strategy of parent i is a mapping from the space of types and beliefs to the set of possible choice lists, $\sigma_i\left(\overline{\boldsymbol{X}}_i, B_i\right) : \mathbb{R}^{K_2+S} \times [0,1]^{S\times L} \rightarrow \mathcal{C}$. The parent's decision is to maximize her subjective expected utility by choosing $C \in \mathcal{C}$:

$$\max_{C\in\mathcal{C}} V_i\left(C, B_i\right) \equiv \max_{C\in\mathcal{C}} \sum_{s=1}^{S} P_s^A\left(C, i\right) \max\left(u_{i,s}, 0\right),$$

where only $\max\left(u_{i,s}, 0\right)$ matters because parents can always choose the outside option whenever they find the assigned school unsatisfactory.[29]

Assume that students do not participate if all schools are unacceptable. The following analysis only considers the case where there is at least one acceptable school for each student, unless noted otherwise.

[28]In real-world plays of the incomplete information game, it is also possible that players have different beliefs because they have different information, although they are all sophisticated. In this paper, information sets are assumed to be "correct" in the sense that players have the same ability in obtaining the information, while they differ in the capability of processing the information. H_i can then be thought of as the set of information one can potentially obtain. Failure to obtain a subset of H_i is equivalent to ignoring that subset of information when process the whole set.

[29]Since the expected utility is in the objective function, the utility function is unique up to an affine transformation. Thus cardinal preferences are important in the Boston mechanism.

3.3 Restrictions on the Beliefs

Without any restriction on the beliefs, it might be possible to rationalize any observed behavior, even when the preferences are known. This section considers two cases which impose different structures on the beliefs. Correspondingly, restrictions on parents' behavior are discussed.

The first case to be considered is that every parent understands the rules of the game, while they may have different ability to form their beliefs. The rules impose a particular structure on the beliefs. Some dominated strategies will be identified.

The second case is to assume that every parent has the same ability to form the beliefs. In this case, the difference in beliefs only comes from the difference in the information set. Under some conditions, the beliefs will be common across parents, and symmetric Bayesian Nash equilibrium will be defined.

3.3.1 Reasonable Beliefs

A parent understands the rules of the game if she is aware of how the algorithm works and how lotteries are generated. In this case, her belief must have the following structure.

Proposition 1 *If a parent understands the rules of the game, the subjective belief,* $B_i(C, H_i)$*,* $\forall H_i \in \mathcal{H}$*, has the following properties:*

(i) A seat is guaranteed if participating: $\forall C \neq (0, ..., 0)$

$$B_i(C, H_i) \in \left\{ \left(P_1^A(C,i), ..., P_S^A(C,i)\right) | \sum_{s=1}^{S} P_s^A(C,i) = 1, P_s^A(C,i) \geq 0, \forall s = 1, ..., S \right\} = \Delta^{S-1}.$$

(ii) In any two lists, if a school is listed after a same ordering of schools, the probability of being accepted by that school is the same when submitting either of the two lists:

$$P_s^A(C,i) = P_s^A(C',i), \forall C, C' \in \mathcal{C} \text{ s.t. } c_K = c'_K = s, c_k = c'_k, \forall k \leq K, K \leq S.$$

(iii) Moving a school toward the top of the list, or including an otherwise omitted school in the list, weakly increases the probability of being accepted by that school:

$$P_s^A(C',i) \geq P_s^A(C,i), \forall C, C' \in \mathcal{C} \text{ s.t. } K' < K \leq S, c_K = c'_{K'} = s, c_k = c'_k, \forall k < K'.$$

(iv) If school s is ranked top, the probability of being accepted by that school is at least q_s/I:

$$P_s^A(C,i) \geq \frac{q_s}{I} > 0, \forall C \in \mathcal{C}, s.t.\ c_1 = s.$$

All proofs are in appendix. The intuition of the proposition is as follows. The last step of the algorithm assigns every remaining student to the available seats, which guarantees that each student gets one seat. The second and third properties are ensured by the priority-based admission process. A late choice is considered only when the student is rejected by her earlier choices. The last property describes the lower bound of the possibility of getting into the first choice school. When everyone else submits the same first choice, one can still get in with probability q_s/I due to the random lottery.

Notice that the properties are satisfied independent of students' information set as long as they understand the rules.

Let $\mathcal{B}_R$ denote the set of reasonable beliefs, $B_i(\cdot,\cdot)$, such that $\forall H_i \in \mathcal{H}$ and $\forall C \in \mathcal{C}$, $B_i(C,H_i)$ satisfies the properties in Proposition 1. With the definition of $\mathcal{B}_R$, understanding the rules of the game is equivalent to having a belief which is an element in $\mathcal{B}_R$. The objective beliefs, $B(\cdot,\cdot)$, are then in $\mathcal{B}_R$.

Remark 1 *The belief, $B_i(C,H_i)$, is defined for all possible lists. Given $B_i(C,H_i) \in \mathcal{B}_R$, the belief associated with any partial list can be calculated with those associated with corresponding full lists. For example, given a partial list, $C = (c_1, 0, ..., 0)$, $B_i(C,H_i)$ can be calculated by all $B_i(C',H_i)$ where C' is a full list with c_1 as first choice. Hence, one only needs to consider the beliefs associated with full lists when all beliefs are in $\mathcal{B}_R$.*

Let $\mathcal{C}^F \subset \mathcal{C}$ denote the set of all possible full lists. There are $S!$ lists in $\mathcal{C}^F$. For these $S!$ lists, there are $(L-1)$ probabilities being specified by beliefs in $\mathcal{B}_R$.[30] Each indicates the probability of being accepted by any school s when s is listed at any position and after any possible schools. When measuring the distance between subjective and objective beliefs, we can only look at the $(L-1)$ probabilities.

The following proposition derives some dominated strategies, given that the beliefs are reasonable.

Proposition 2 *Suppose there is at least one acceptable school, if students maximize their subjective expected utility, and $B_i(\cdot,\cdot) \in \mathcal{B}_R$, then*

[30] The beliefs specify each school being listed at each position and after any possible schools. Thus, S probabilities are needed for different first choices; $S*(S-1)$ needed for different second choices; etc. The total number of probabilities is:

$$L - 1 = S + S*(S-1) + ... + S*(S-1)*(S-2)...*2$$

Since a seat is guaranteed, the probability of being assigned to the last school is the residual probability.

(i) Listing an unacceptable school or the worst school as the first choice is strictly dominated.

(ii) Excluding an acceptable school from the list is weakly dominated.

(iii) Listing the worst school before any other school is weakly dominated.

(iv) Listing any unacceptable school before any acceptable school is weakly dominated.

The intuitions are as follows. Since listing a school as the first choice always gives her a strictly positive probability of being assigned to that school (part (iv) in Proposition 1), a parent should try better schools first. On the other hand, the worst outcome of participating is to be assigned to the worst school. By putting better school on top of the worst school, the parent increases her chance of being assigned to a better school. If a school is unacceptable, putting it at the bottom or excluding it from the list also increases the likelihood of getting into better schools.

Remark 2 *Results in Proposition 2 are independent of risk attitude. Dominated strategies are defined in terms of first-order stochastic dominance.*

Proposition 2 confirms the observation that truth-telling is a dominant strategy when there are only two schools. Proposition 2 puts some structure on the students' behavior.[31] It is enough to formulate the likelihood function and get consistent estimates of preferences if some further assumptions are imposed on the utility function. This will be discussed in detail when the econometric method is introduced.

3.3.2 Common Beliefs

For this part, it is assumed that all parents have the ability to calculate the beliefs correctly with their information set, and thus $B_i(C, H_i) = B(C, H_i)$, $\forall i$. Therefore, the differences in beliefs across students only come from the differences in their information sets. The following lemma describes a case when students have the same information set.

Lemma 1 *The information sets are common to every student, so are the subjective beliefs, i.e.* $H_i = H$, $B_i(C, H_i) = B(C, H)$, $\forall C \in \mathcal{C}$ $\forall i = 1, ..., I$, *if*

(i) Every student is sophisticated;

(ii) The realization of $\boldsymbol{X}_i$ *is private information of student* i.

[31] The results also provide a possible way to rationalize the seemingly "irrational" behavior found in Abdulkadiroglu, Pathak, Roth, and Sonmez (2006). Namely, some students submit a very short list and leave some choices blank. Proposition 2 predicts that this happens when some schools are unacceptable, or these students believe that the probability of being rejected by all the schools listed is zero. If information on the school assignment and actual attendance is available, we can distinguish these two explanations. For example, if a student is assigned to a school not in her list and attends it, the school must be acceptable.

The proof of Lemma 1 is omitted. The conditions in the lemma rule out any difference in the information sets. Thus, the beliefs are common to everyone.

However, the condition that parent i knows the distribution of $\overline{\boldsymbol{X}}_{-i}$ but not the realization of any $\overline{\boldsymbol{X}}_j$ might be too strong. One may argue that each student knows her classmates and friends relatively well. The following proposition considers this possibility.

Proposition 3 *Consider the following scenario:*

(i) Every student is sophisticated;

(ii) Student i also knows the realization of $\mathcal{X}^i = \{\boldsymbol{X}_{i_1}, ..., \boldsymbol{X}_{i_F}\}$. F is fixed and $\mathcal{X}^i$ may differ across students.

Given the number of schools, as the number of students becomes larger and the quotas grow at the same rate, the beliefs converge to a common belief, $B_i\left(C, H_i\right) = B\left(C, H_i\right) \rightarrow B\left(C, H\right)$, $\forall i$, $\forall C \in \mathcal{C}$.

In other words, the beliefs are almost common when the differences in students' information sets are small relative to the set of available information.

3.4 Bayesian Nash Equilibrium under Common Beliefs

Under the assumption that the beliefs are common, this subsection defines a symmetric Bayesian Nash equilibrium. Since the beliefs are common, the information set H can be suppressed. B and $B\left(C\right)$ is shorthand notation for $B\left(\cdot, H\right)$ and $B\left(C, H\right)$ respectively, when there is no confusion.

3.4.1 Consistent Beliefs

Given everybody's belief system B and strategy $\sigma_i\left(\overline{\boldsymbol{X}}_i, B\left(\cdot\right)\right)$, define the implied probabilities $\widetilde{B}\left(C, B\right)$: $\mathcal{B}_R \rightarrow \mathcal{B}_R$, which is a mapping from beliefs, B, to implied probabilities of student i being accepted by every school when submitting any list.

Figure 1 illustrates the mapping of $\widetilde{B}\left(\cdot, B\right)$ for an arbitrary student, say student 1 without loss of generality. Start with a belief system B and a list C. Since the action set is finite, the possible profile of other students' choice, C_{-1}, can be all written down. Let $M \equiv L^{(I-1)}$ be the total number of possible profiles. Index each profile by $m = 1, .., M$. For profile m, the algorithm can be run many times to calculate the probabilities, $b_m\left(C\right)$, that student 1 is assigned to each school when she submits C. The probability that the mth profile, $\left[C, C_{-1}^m\right]$ is realized, $p_m\left(B\right)$, can also be calculated given B, $\sigma_i\left(\overline{\boldsymbol{X}}_i, B\right)$ and the distribution of $\overline{\boldsymbol{X}}_i$. Then the implied probabilities of B is $\widetilde{B}\left(C, B\right) = \sum_{m=1}^{M} p_m\left(B\right) b_m\left(C\right), \forall C \in \mathcal{C}$.

The definition of consistent belief links the beliefs and the implied probabilities.

Definition 1 *Given everyone's strategy, the belief system,* B*, is consistent if* $\widetilde{B}(C, B) = B(C)$*,* $\forall C \in \mathcal{C}$*.*

3.4.2 Equilibrium Properties

A symmetric Bayesian Nash equilibrium can be defined as follows.

Definition 2 *A Bayesian Nash equilibrium of the school choice problem under Boston mechanism is defined as a strategy profile* $\{\sigma_i^*\}_{i=1}^I$*, and a belief system* $B^* \in \mathcal{B}_R$ *which is common to everyone, such that:*

(i) Given the belief system, $\forall i$*,*

$$\sigma_i^* \in \arg\max_{\sigma_i} \left\{ V_i(\sigma_i, B^*) = \sum_{s=1}^{S} P_s^A(\sigma_i) \max(u_{i,s}, 0) \right\}. \tag{1}$$

(ii) The belief system is consistent, i.e. $\widetilde{B}(C, B^*) = B^*(C)$*,* $\forall C \in \mathcal{C}$*.*

In the equilibrium, every parent has the same and correct beliefs.

Definition 3 *If a parent's beliefs are the same as the equilibrium beliefs, she is sophisticated.*

The existence of the equilibrium is characterized by the following proposition.

Proposition 4 *There always exists a Bayesian Nash equilibrium in the school choice game defined above if common belief conditions (in Lemma 1 or Proposition 3) are satisfied.*

To reiterate, given that students are maximizing their expected utility, the equilibrium is a belief system which is common and consistent under the common belief conditions. Thus, finding an equilibrium is equivalent to finding a consistent belief system. Since ε_i is i.i.d. across students and has a continuous distribution, any "extreme" outcome can happen with strictly positive probability in equilibrium. For example, there is a strictly positive probability that every student puts school 1 as their first choice. Thus, the equilibrium belief has the following property.

Lemma 2 *For any equilibrium belief,* B^**,* $\left\{P_s^A(C)\right\}_{s=1}^S \in (0,1)^S$ *if* $C \neq \{0\}$*, i.e. if a student chooses to participate, the probability of being assigned to each school is strictly positive and less than one.*

The following two propositions characterize the mixed strategy and pure strategy Bayesian Nash equilibrium of the game.

Proposition 5 *(i) If common belief conditions (in Lemma 1 or Proposition 3) are satisfied, and every school is acceptable, students play pure strategy in equilibrium almost surely.*

(ii) In any Bayesian Nash equilibrium in mixed strategies, student i plays mixed strategies only if $\exists s \neq s'$, such that

$$u_{i,s} < 0, \text{ and } u_{i,s'} < 0.$$

The student only mixes list $C_i = \left\{c_i^1, ..., c_i^S\right\}$ with $\widehat{C}_i = \left\{c_i^1, ..., c_i^K, \widehat{c}_i^{K+1}, ..., \widehat{c}_i^S\right\}$, such that

$$\begin{aligned} C_i &\in \arg\max_{\widehat{\sigma}_i\left(\overline{\mathbf{X}}_i, B^*\right)} \left\{V\left(\widehat{\sigma}_i\left(\overline{\mathbf{X}}_i, B^*\right)\right)\right\}, \\ u_i\left(c_i^k\right) &> 0 \; \forall k = 1, ..., K, \\ c_i^k &= 0, \forall k = K+1, ..., S, \text{ and} \\ \widehat{c}_i^{K+1}, ..., \widehat{c}_i^S &\in \left\{s \mid u_{i,s} < 0\right\}. \end{aligned}$$

Proposition 5 shows that students play mixed strategies only by including in their lists schools that they will never attend. This does not benefit them but hurts the other students. The following proposition formalizes the efficiency comparison between different equilibria.

Proposition 6 *If common belief conditions (in Lemma 1 or Proposition 3) are satisfied, then*

(i) There exists a continuum of Bayesian Nash equilibria in mixed strategies;

(ii) Assuming that the joint distribution function of $u_{i,s}$, $G\left(u_{i,1}, u_{i,2}, ..., u_{i,S}\right)$, has a density continuous function, and students never include unacceptable schools in the list, there exists a unique Bayesian Nash equilibrium in pure strategies;

(iii) In terms of students' welfare, the pure-strategy equilibrium in (ii) Pareto dominates all equilibria in mixed strategies in terms of ex ante efficiency.

3.4.3 Heterogeneous Ability to Form the Beliefs

Now suppose parents have different ability to form their beliefs. Some parents have the ability calculate the beliefs correctly, while others don't. Reorder the parents, such that parents $1, ..., I^*(I^* < I)$ have $B_i = B$, $\forall i = 1, ..., I^*$. Other parents have $B_i \neq B, \forall i = \left(I^* + 1\right), ..., I$.

Suppose everyone has the same information set, $H_i = H$. A Bayesian Nash equilibrium can be defined as follows:

Definition 4 *When parents have heterogeneous ability to process the information, a Bayesian Nash equilibrium of the school choice problem under the Boston mechanism is defined as a strategy profile* $\left\{\sigma_i^*\left(\overline{\boldsymbol{X}}_i, B_i\right)\right\}_{i=1}^{I}$*, and a belief system* $B_i \in \mathcal{B}_R$*, such that:*

(i) Given the belief system, $\forall i$*,*

$$\sigma_i^* \in \arg\max_{\sigma_i}\left\{V_i\left(\sigma_i, B_i\right)=\sum_{s=1}^{S} P_s^A\left(\sigma_i, i\right)\max\left(u_{i,s}, 0\right)\right\}.$$

(ii) For parents who have the full ability to process the information, the belief system is consistent, i.e. $B_i = B$ *and* $\widetilde{B}\left(C, B\right)=B\left(C, H\right)$*,* $\forall C \in \mathcal{C}$*,* $i = 1, ..., I^*$*.*

Remark 3 *In this equilibrium, parents who have the full ability to process the information have the correct equilibrium beliefs. Therefore, they are sophisticated.*

Since the information set contains information about other parents' beliefs, sophisticated parents take into account the distribution of sophistication of others. All the sophisticated parents have the same consistent beliefs, while others can have inconsistent but reasonable beliefs. Proving the existence of this equilibrium is then to prove that the consistent beliefs exist for parents who have the full ability to process the information. This is the same as the case when all parents have the full ability to process the information.

4 Econometric Framework

In the following, the utility function is further characterized, and the assumptions on the belief system are also formalized. Different sets of assumptions are considered. One requires the minimal assumptions on the beliefs (Proposition 1). The other two add two additional assumptions sequentially. A method of simulated log-likelihood is proposed to estimate the model when the choice list C_i is observed.

4.1 Utility Function and Beliefs

Suppose the utility is linear in the error term:

$$u_{i,s} = u\left(\boldsymbol{z}_s, \boldsymbol{X}_i; \boldsymbol{\beta}\right) + \varepsilon_{i,s}, s \neq 0; u_{i,0} = 0.$$

The functional form of $u\left(\boldsymbol{z}_s, \boldsymbol{X}_i; \boldsymbol{\beta}\right)$ is known. $\boldsymbol{\varepsilon}_i$ is assumed to have a multinomial normal distribution and i.i.d. across students.[32] The variance of $\varepsilon_{i,1}$ is normalized to be 1. One of the goals is to estimate $\boldsymbol{\beta}$ and the

[32]In many other context of discrete choice model, including Hastings, Kane, and Staiger (2008), error terms are usually assumed to be distributed i.i.d. extreme value. It suffers the independence of irrelevant alternatives problem, but the advantage is that the

covariance matrix of ε_i. Let $\boldsymbol{\theta}$ denote the collection of all parameters in the utility function to be estimated.

As discussed in Subsection 3.3, given the assumption that students or their parents understand BM, their beliefs must satisfy properties in Proposition 1. Their behavior must be consistent with Proposition 2. In this case, the belief system is not required to be common across students and thus cannot be identified without information on preferences.

The belief system has to be considered when solving the equilibrium under the common beliefs conditions. In subsection 3.4, the dimension of belief system $B(.,.)$ is $SL \times 1$. From Proposition 1, the dimension can be reduced to $(L-1)$, as long as $B(.,.) \in \mathcal{B}_R$. If the number of schools is small, B can be estimated non-parametrically. In the following application, $S = 4$, thus 40 probabilities are estimated.[33]

4.2 Necessary Equilibrium Conditions and Simulated Maximum Likelihood

4.2.1 Maintained Assumptions

Several assumptions are maintained in the following analysis.

The first is that students maximize their expected utility and understand the structure of the game. So their beliefs are always in $\mathcal{B}_R$, or satisfy the conditions in Proposition 1. This assumption will be tested against the case when everyone is truth-telling. The second assumption is that if a school is excluded from the list, it is an unacceptable school. Some (indirect) tests on these two assumptions will be provided.

The other maintained assumption is that if unacceptable schools are listed, these schools are listed after all acceptable schools and listed in the order of true preferences. This assumption rules out the possibility that students choose randomly among those indifferent lists. The last assumption is that students do not participate if all schools are unacceptable.

4.2.2 Necessary Equilibrium Conditions and Likelihood Functions

In last section, two equilibria are defined. One allows for the possibility that not all parents are sophisticated (Definition 4). The other assumes that all parents are sophisticated (Definition 2). The necessary conditions in each equilibrium will be described.

Reasonable Beliefs

When some students are not sophisticated, there are no common beliefs. But the beliefs are always in $\mathcal{B}_R$,

choice probability can be analytically written as multinomial logit probability which does not need integration. But this advantage is not applicable in this study. Given a list C, the error term of the expected utility from C is a linear combination of all error terms in ε_i. Thus, across different C, error terms are necessarily correlated.

[33] "The curse of dimensionality" applies here. For example, when $S = 7$, 8,659 different parameters have to be estimated.

since they understand the game. Hence, Proposition 2 which describes the dominated strategies holds true. GIven the maintained assumptions, the following conditions hold in equilibrium (henceforth reasonable beliefs conditions):

(i) If a full list $C_i = \left(c_i^1, ..., c_i^S\right)$ is submitted in equilibrium, then (a) $u_{i,c_i^1} > 0$; (b) $u_{i,c_i^S} = \min_s \left(u_{i,s}\right)$ or the probability of being assigned to c_i^S is zeros when listing it as $(S-1)$th or Sth choice.

(ii) If a partial list $C_i = \left(c_i^1, ..., c_i^K, 0, ..., 0\right)$ is submitted in equilibrium, then (a) $u_{i,c_i^1} > 0$; (b) $u_{i,s} \leq 0$, $\forall s \notin C_i$.

(iii) If a student does not participate, $C_i = (0, ..., 0)$, then $u_{i,s} \leq 0$, $\forall s = 1, ..., S$.

Notice that the above conditions allow a student to include unacceptable schools in her list. In part (i), c_i^S is listed as the last choice either because it is the worst school or the probability of being assigned to c_i^S is zero when c_i^S is at the lowest two position in the list. The second possibility is more likely to be true for good schools with small quota. These schools usually fill their seats in very early rounds. Thus, there is a almost-zero probability of being assigned to theses schools in very late rounds. Reasonably, some parents anticipate this and act accordingly. This will be taken into account and tested in the estimation.

The necessary conditions described above impose conditions on the utilities. Let $\omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)$ denote the set of $\boldsymbol{\varepsilon}_i$ satisfying all the conditions given any C_i. Thus, the event in which C_i is chosen by i is a subset of $\omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)$:

$$\{\text{Events in which } C_i \text{ is chosen}\} \subset \omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right).$$

The other way to interpret this is to recode the observed list. For example, if $C_i = \left(c_i^1, ..., c_i^K, 0, ..., 0\right)$ is observed, it is equivalent to recoding the information as $u_{i,c_i^1} > 0$ and $u_{i,s} \leq 0$, $\forall s \neq c_i^k$, $\forall k = 1, ..., K$. $\omega\left(C_i\right)$ contains the information after the recoding.

Given $\omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)$, the likelihood function is $P\left(\boldsymbol{\varepsilon}_i \in \omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)\right)$. Since the error term, $\boldsymbol{\varepsilon}_i$, in the utility function is assumed to be normally distributed, $P\left(\boldsymbol{\varepsilon}_i \in \omega_1 \left(C_i | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)\right)$ is an integral similar to those in multinomial probit models.[34] In the estimation, a Logit-kernel smoothed

[34] For example, if $C_i = \left(c_i^1, ..., c_i^K, 0, ..., 0\right)$ is observed, then

$$\begin{aligned} & P\left(\omega\left(C_i\right) | \boldsymbol{X}_i, \{\boldsymbol{Z}_s\}_{s=1}^S; \boldsymbol{\theta}\right) \\ = \; & P\left(\varepsilon_{i,c_i^1} > -u\left(\boldsymbol{Z}_{c_i^1}, \boldsymbol{X}_i; \boldsymbol{\beta}\right), \varepsilon_{i,s} \leq -u\left(\boldsymbol{Z}_s, \boldsymbol{X}_i; \boldsymbol{\beta}\right), \forall s \notin C_i | \boldsymbol{X}_i, \{\boldsymbol{Z}_s\}_{s=1}^S; \boldsymbol{\theta}\right) \end{aligned}$$

which is an integral of the joint density of normally distributed $\boldsymbol{\varepsilon}_i$.

acceptance-rejection simulator is used to calculate the integral.[35]

Thus a simulated maximum likelihood (SML) estimator is:

$$\widehat{\boldsymbol{\theta}}_0 \in \arg\max_{\theta} \sum_{i=1}^{I} \ln\left(\widetilde{P}\left(\boldsymbol{\varepsilon}_i \in \omega_1\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)\right)\right), \tag{2}$$

where $\widetilde{P}\left(\boldsymbol{\varepsilon}_i \in \omega_1\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)\right)$ is the simulated probability of $P\left(\boldsymbol{\varepsilon}_i \in \omega_1\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)\right)$. When I and the number of simulations R go to infinity and $\sqrt{I}/R \to 0$, the SML estimator is asymptotically equivalent to the maximum likelihood estimator.[36]

Additional Hypotheses for Equilibrium Selection

Additional hypotheses can be added and tested against the reasonable beliefs assumption. In particular, since there are multiple equilibria under the common beliefs, some hypotheses directly related to the equilibria selection will be considered.

For the selection between mixed-strategy and pure-strategy equilibrium, the key is whether to include unacceptable schools in the list. Start with the following condition: The second choice school in the list is acceptable: $u_{i,c_i^2} > 0$ if $c_i^2 \neq 0$.

Add this condition to the conditions under reasonable beliefs assumption. Let $\omega_2\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right) \subset \omega_1\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)$ denote the set of $\boldsymbol{\varepsilon}_i$ that satisfies the new set of conditions. Similarly, a simulated maximum likelihood can be used to estimate the parameters:

$$\widehat{\boldsymbol{\theta}}_1 \in \arg\max_{\theta} \sum_{i=1}^{I} \ln\left(\widetilde{P}\left(\boldsymbol{\varepsilon}_i \in \omega_2\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)\right)\right). \tag{3}$$

Since this new model is nested within the model under the reasonable beliefs conditions, it provides a test on the additional condition. If the additional condition is true, then $\widehat{\boldsymbol{\theta}}_1$ is a consistent and efficient estimator of $\boldsymbol{\theta}$, while $\widehat{\boldsymbol{\theta}}_0$ is consistent and inefficient. Otherwise, $\widehat{\boldsymbol{\theta}}_0$ is consistent, but $\widehat{\boldsymbol{\theta}}_1$ is not.[37] A Small-Hsiao test is used for the model selection in following analysis (Small and Hsiao (1985)).[38]

If the first additional condition is not rejected in the test, we can go on and add more conditions. For

[35] For a description of the simulator, see Chapter 5 of Train (2003). The analysis presented in the following uses a smoothing factor (or scale factor) equal 0.05. Values between 0.01 and 0.1 have been experimented, and results do not change very much.

[36] See for example, Gourieroux and Monfort (1997) Chapter 3 and Train (2003) Chapter 10.

[37] The Hausman test can be used in this scenario. But in practice, it may have a covariance matrix not invertible, or the covariance matrix not positive semi-definite.

[38] The Small-Hsiao test was originally proposed to test the independence of irrelevant alternatives (IIA) assumption in multinomial logit model. In the current context, the null hypothesis is that the estimates in both cases are consistent, and the estimates in the nested case is more efficient. The procedure goes as follows. Split the data randomly into two samples (A and B). Estimate the

example, the second and third choice schools in the list are acceptable if $c_i^2, c_i^3 \neq 0$. The same estimation and testing procedure can be used for model selection. If one accepts the condition that all listed schools except the last choice school are acceptable, she should look for a pure strategy Bayesian Nash equilibrium. Equilibria in mixed strategies have to be considered. In this case, it is crucial to identify how students mix the indifferent strategies. Proposition 5 predicts that students mix the indifferent strategies by including unacceptable schools at the bottom of their lists. The consistent and most efficient estimates obtained previously can be used to estimate the mixing pattern under the symmetric equilibrium assumption. Based on the submitted lists, one can calculate the probability of choosing each list among several lists, conditional on these lists being indifferent.

Bayesian Nash Equilibrium

Based on the result of equilibrium selection, either pure-strategy or mixed-strategy equilibrium is considered. Suppose we need to find a mixed-strategy equilibrium under the common beliefs. In addition to the conditions under reasonable beliefs assumption, we now can predict the probability of choosing C_i, given the beliefs system B, $P\left(\boldsymbol{\varepsilon}_i \in \underline{\omega}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; B; \boldsymbol{\theta}\right)\right)$, where $\underline{\omega}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; B; \boldsymbol{\theta}\right)$ is the set of $\boldsymbol{\varepsilon}_i$ such that C_i is chosen. Since $B \in \mathcal{B}_R$,

$$P\left(\boldsymbol{\varepsilon}_i \in \underline{\omega}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; B; \boldsymbol{\theta}\right)\right) \leq P\left(\boldsymbol{\varepsilon}_i \in \omega_1\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)\right),$$

which means the set of $\boldsymbol{\varepsilon}_i$ such that C_i is chosen in the Bayesian Nash equilibrium is a subset of $\omega_1\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)$. Again, the Small-Hsiao test can be used for model selection. The test result will show if the behavior of students as a whole is consistent with a Bayesian Nash equilibrium.

Simulation is also needed to calculate $P\left(\boldsymbol{\varepsilon}_i \in \underline{\omega}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; B; \boldsymbol{\theta}\right)\right)$. The simulated maximum

nested/more restrictive model for the two samples and obtain $\widehat{\boldsymbol{\theta}}_1^A$ and $\widehat{\boldsymbol{\theta}}_1^B$. A weighted average of the two estimates is:

$$\widehat{\boldsymbol{\theta}}_1^{AB} = \sqrt{2}\widehat{\boldsymbol{\theta}}_1^A + \left(1 - \sqrt{2}\right)\widehat{\boldsymbol{\theta}}_1^B.$$

Also estimate the less restrictive model for the two samples and obtain $\widehat{\boldsymbol{\theta}}_0^A$ and $\widehat{\boldsymbol{\theta}}_0^B$. The test statistic is:

$$-2\left(\sum_{i\in B} \ln\left(\widetilde{P}\left(\omega\left(C_i\right)|\boldsymbol{X}_i, \{\boldsymbol{Z}_s\}_{s=1}^S; \widehat{\boldsymbol{\theta}}_1^{AB}\right)\right) - \sum_{i\in B} \ln\left(\widetilde{P}\left(\omega\left(C_i\right)|\boldsymbol{X}_i, \{\boldsymbol{Z}_s\}_{s=1}^S; \widehat{\boldsymbol{\theta}}_0^B\right)\right)\right) \xrightarrow{d} \chi^2\left(K\right),$$

where K is the number of parameters in $\boldsymbol{\theta}$.

likelihood estimator is:

$$\left(\widehat{\boldsymbol{\theta}}_{BN};\widehat{B}\right)\in\arg\max_{\theta,B}\sum_{i=1}^{I}\ln\left(\widetilde{P}\left(\boldsymbol{\varepsilon}_i\in\underline{\omega}\left(C_i|\boldsymbol{X}_i,\{\boldsymbol{z}_s\}_{s=1}^{S};B;\boldsymbol{\theta}\right)\right)\right) \tag{4}$$

$$s.t.\overline{B}\left(C,B\right)=B,\ \forall C\in\mathcal{C}^F$$

where $\overline{B}\left(C,B\right)$ is an approximation of $\widetilde{B}\left(C,B\right)$. Instead of considering all the possible profiles of students' lists C_{-i}, draw a large number of profiles to calculate $\overline{B}\left(C,B\right)$. Appendix 2 describes how to simulate $\overline{B}\left(C,B\right)$. Notice that the beliefs are only considered for full lists, i.e. $C\in\mathcal{C}^F$. Thus, the number of parameters in B is $(L-1)$. As mentioned earlier, this is sufficient for calculating all the beliefs. The constraint is an equilibrium condition that the belief should be consistent. Thus, $\widehat{B}\left(C,B\right)$ implicitly is a function of $\boldsymbol{\theta}$.

In estimation, a penalty function approach is used.[39] Starting with a large $\pi_t>0$, the objective function is transformed as:

$$\max_{\theta,B}\left[\sum_{i=1}^{I}\ln\left(\widetilde{P}\left(\boldsymbol{\varepsilon}_i\in C_i|\boldsymbol{X}_i,\{\boldsymbol{z}_s\}_{s=1}^{S},B;\boldsymbol{\theta}\right)\right)-\pi_t\left|\left|B-\overline{B}\left(\cdot,B\right)\right|\right|^2\right], \tag{5}$$

where $\left|\left|B-\overline{B}\left(\cdot,B\right)\right|\right|$ is the Euclidean norm between B and $\overline{B}\left(\cdot,B\right)$. When π_t is sufficiently large, the solution to the unconstrained maximization problem (5) is numerically identical to the solution to the constrained maximization (4). In the estimation, one can increase π_t sequentially and use a stopping rule as follows: Increase π_t and estimate the model again unless $\left|\left|B-\overline{B}\left(\cdot,B\right)\right|\right|^2\leq Tol$ where Tol is a small positive number.[40]

4.3 Measuring Sophistication

The model selection tests described above provides information on the sophistication level of the players as a whole. Even when overall sophistication level is high, some players might be less sophisticated, and vice versa. Individual's sophistication is also more important for welfare evaluation. In this subsection, several measures are proposed to measure the level of sophistication.

[39] See Gill, Murray, and Wright (1982) and Belegundu and Chandrupatla (1999) for a discussion of the penalty function method. For an application in an econometric context, see Imbens, Spady, and Johnson (1998).

[40] Other approaches to the constrained maximization include the "nested fixed-point" procedure in Berry, Levinsohn, and Pakes (1996) and a mathematical program with equilibrium constraints (MPEC) as proposed by Dube, Fox, and Su (2009). Monte Carlo results show these two approaches don't perform better than the penalty function approach. And the penalty function approach takes less time.

4.3.1 Subjective and Objective Beliefs

An individual's sophistication can be measured by the distance between her subjective and objective beliefs. Given the estimates of the preferences, or more specifically, the distribution of preference, one can estimate the subjective beliefs by maximizing the likelihood that the observed behavior being a "subjective best response." That is,

$$\widehat{B}_i \in \arg\max_{B_i} P\left(C_i \text{ is a best response}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, B_i; \widehat{\boldsymbol{\theta}}\right), \tag{6}$$

where $\widehat{\boldsymbol{\theta}}$ is the consistent estimate of $\boldsymbol{\theta}$. In estimation, the probability is approximated by simulating the error term, ε_i, given the estimated distribution.

One limitation of this measure is that this is not feasible for students who submit a one-school list or do not participate. It is also less precisely measured if C_i is a partial list.

Measuring the objective beliefs differs across cases. When everyone is sophisticated, since the equilibrium beliefs are the objective beliefs, the estimated equilibrium beliefs, $\widehat{B}$, are the measure of the objective beliefs. However, if not every student is sophisticated, $\widehat{B}$ is not obtained in the process of estimating preferences. The equilibrium cannot be solved without knowing the distribution of sophistication. Fortunately, if the market is large, the ex post distribution of others' behavior will be close to its ex ante distribution.[41] Thus the probabilities calculated based on the ex post distribution of other players behavior can be used to approximate the objective beliefs.

4.3.2 Best Responding vs. Truth-Telling

Given the estimated objective beliefs, $\widehat{B}$, one can calculate the probability that the submitted list is a best response or true preference.

The probability that parent i plays a best response is:

$$p_i^{BR}(C_i) = P\left(C_i \text{ is a best response}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, \widehat{B}; \widehat{\boldsymbol{\theta}}\right).$$

The probability that parent i submits her true preference ranking is:

$$p_i^{TT}(C_i) = P\left(C_i \text{ is the true preference}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \widehat{\boldsymbol{\theta}}\right),$$

[41] As the market becomes large, the empirical distribution of others' behavior can be very close to the theoretical distribution. Since all that matters is the aggregate distribution of other players strategies, the ex post distribution provide a good approximation of ex ante distribution.

which is independent of the beliefs.

It is possible that truth-telling is a best response in some case. To measure how well the parents strategize, we need to control for this possibility. Consider the difference between the two probabilities, $p_i^{BR}(C_i) - p_i^{TT}(C_i)$. If a parent is sophisticated, the difference is more likely to be positive. Moreover, this can be measured by the probability that the parent plays a best response which is not truth-telling:

$$p_i^{BR \neq TT}(C_i) = P\left(C_i \text{ is a best response but not the true preference} | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \widehat{\boldsymbol{\theta}}\right).$$

4.3.3 Distance to the Maximum Expected Utility

Given the objective beliefs, one can calculate the expected utility from the best response:

$$EU_i^* = \max_C \sum_{s=1}^{S} P_s^A(C) \max(u_{i,s}, 0).$$

A measure of how well the parent strategizes is the difference between EU_i^* and the expected utility from the list submitted, C_i. Define the percentage of expected utility achieved as:

$$EU_i(C_i) = \frac{The\ expected\ utility\ from\ submitting\ C_i}{EU_i^*}$$

. If a parent plays a best response, then $EU_i(C_i) = 1$.

Since parents might be risk averse, the variance in the utility may also be informative.

$$VAR_i(C_i) = \text{(Variance of utility given } C_i\text{) - (Variance of utility given the best response).}$$

Conditional on the same $EU_i(C_i)$, a lower $VAR_i(C_i)$ is better since the risk is lower.[42]

4.4 Measuring the Incentives to Be Strategic

A parent has more incentives to be strategic if her truthful reporting is not a best response, or the payoff from truthful reporting is much lower than the payoff from best responding. If parents are sophisticated, they should optimize their strategy in response to these incentives.

The first measure of incentive to be considered is the probability that truth-telling is a best response:

$$p_i^{TT=BR} = P\left(\text{Truth-telling is a best response} | \boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}, \widehat{B}; \widehat{\boldsymbol{\theta}}\right).$$

[42] $VAR_i(C_i)$ is not calculated as a ratio, because variances are zero in many cases.

If $p_i^{TT=BR}$ is higher, the parent has less incentive to be strategic. Notice that this variable is constructed ex ante. Namely, the information contained in the submitted list is not used. In fact, the information in the submitted list will be used to identify how parents respond to the ex ante incentives.

The other measure is how close the truthful reporting is to the best response in terms of expected utility:

$$EU_i^{TT} = \frac{(\text{Expected utility if always truth-telling}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, \widehat{B}; \widehat{\boldsymbol{\theta}})}{\left(\text{Expected utility if always playing a best response}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, \widehat{B}; \widehat{\boldsymbol{\theta}}\right)}.$$

If EU_i^{TT} is high, truthful reporting is close to the best response. Therefore, the parent has less incentive to be strategic.

EU_i^{TT} only compares the mean of utility, the variance of utility may also be important if parents are risk averse.

$$Var_i^{TT} = \frac{(\text{Variance of utility if always truth-telling}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, \widehat{B}; \widehat{\boldsymbol{\theta}})}{\left(\text{Variance of utility if always playing a best response}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S, \widehat{B}; \widehat{\boldsymbol{\theta}}\right)}.$$

Given EU_i^{TT}, if Var_i^{TT} is higher, there is more incentive for the parent to be strategic. Similar to $p_i^{TT=BR}$, the submitted list is not used to construct EU_i^{TT} and Var_i^{TT}. All these incentives are measured ex ante.

5 Data and Descriptive Analysis

The data set is from Beijing Eastern City District which uses BM for the middle school admission. Students' priorities are solely determined by a random lottery number. This study uses data from the largest neighborhood in the district which has access to four schools with a total quota of 960 in 1999. There are two sources of the data. One is the administrative data which provides parents' actual choice lists, students' elementary school enrollment and test scores, and students' home address in 1999.

The other source is a survey in early 2002 conducted by the Education Bureau of Beijing's Eastern City District. The survey covered all students enrolled in the third and last year of middle schools in the district, as well as their parents and teachers. Dropping out or repeating grades is a negligible concern in these middle schools; inter-district transfers are extremely rare and can only be justified by parents changing jobs or moving away. Hence, the survey population is close to the population of students who entered middle school in 1999. Moreover, this allows researchers to observe most of the students who did not participate in the school choice and attended schools not available in their neighborhood. A questionnaire directed to parents collected information on household wealth, parents' education levels, and retrospective information on factors affecting parents' school choices and their preparedness for making school choice decisions in 1999.

5.1 Summary Statistics

There are four schools available to the students in the sample. Table 1 describes the quotas and school quality measured by past students performance. School 3 has the smallest quota and the highest average test score. The test score is the performance of the school's graduating class in the high school entrance exam in 1999. It is a city-wide and high-stake exam which is usually an important factor when parents choose schools. The third column provides the ranking of each school among 28 schools in the district based on the test score.

The number of observed students is 914. The difference between quota and observed students may come from two sources: 1) quota is usually set to be larger than actual number of students. 2) Students transferred to schools in other districts during 1999-2002. When estimating the Bayesian Nash equilibrium, we need all the students to solve the equilibrium beliefs. Thus, 46 students are imputed. 16 of them are random draws from non-participants. The other 30 are from randomly drawn from the observed population. In cases where full sample is not necessary, the 914-student sample is used to check the robustness. The results are not sensitive to the imputation as long as students are not imputed in an extreme way.

The distribution of submitted lists for observed and imputed samples are in Table 2. About 20 percent of the students did not participate. The majority submitted a full list with three or four schools, and only 7.7% submitted a partial list.

For the variables which will be used in the estimation, Table 3 presents the definition and summary statistics. In the estimation, logarithm of test score, income and distances are used. Variables are also normalized such that mean of log test score and log income is zero. School's average test score in 1999 will also be used in the estimation. Again, it is in logarithm and the lowest log average test score is normalized to be zero.

5.2 Understand the Game?

This subsection looks at the distribution of submitted lists to see if students' behavior is consistent with the truth-telling hypothesis.

The second column in Table 4 shows the distribution of students' first choice. 47% of the students choose school 1 as their first choice. School 1 has the second highest test score among the four schools and has a reasonably large quota. 25% listed school 3 as their first choice, while school 3 has the highest test score and the smallest quota.

In the 2002 survey, a question was asked, "Among schools to which you could apply, which school was the best?"[43] Among 696 valid responses, 83% of them list school 3 as the best school. And the distribution is consistent with the ranking based on average test score. Although the question does not ask about students' preferences, the results may still be informative. Comparing the submitted lists to claimed best school, the difference is significant, which suggests the truth-telling hypothesis may not hold true.

Under the hypothesis that every student understands the rules, the first choice school should never be the worst school (Proposition 2). This is consistent with the data in Table 4. Few students list school 4 as the first choice, while even fewer people claim it as the best school.

Proposition 2 also predicts that the last choice school should either be the worst school or a school which is almost impossible to get in if listing it as the third or fourth choice. The last four columns in Table 4 show how students rank the claimed best school. For schools 1, 2 and 4, only two students list the best school as the last choice. While for school 3, there are 36 or 6.2% students listing it as the last choice. Since school 3 has the smallest quota, in the observed realization, the only chance to be assigned to school 3 is to list it as the first choice. In that case, the probability is only 26.7%. Thus it is reasonable to have the belief that there is no chance to get into school 3 by listing it as the third or fourth choice. A test will be provided to see if the 36 students listing their best school as last choice are due to the zero probabilities.

Another way to see if students understand the rules is to examine their self-reported beliefs. Using the responses to a question in the 2002 survey: "On a scale of 0-10, how likely to to be accepted by your 1st/2nd choice?" Table 5 shows the means of the responses.[44] The observed probabilities are calculated from the data which is one realization of the game. The second column reports the mean reported probability when each school is listed as the first choice. The general pattern is consistent with the observed probabilities. For the reported probabilities when schools are listed as the second choice, the pattern is also consistent with the observed one.

Another property of the reasonable beliefs is that moving a school toward the top of the list increases the probability of being accepted by that school. Comparing the two columns of mean reported probability, the mean does increase when moving the school from the second to the first choice.

To summarize, Tables 4 and 5 provide evidence consistent with the hypothesis that students understand the rules of the game.

[43]Notice that it is not asking the favorite school of the student or the school most wanted by the student.

[44]Since these questions are asked after the assignment is realized, the results presented here may just reflect the ex post probability, or the fact that whether or not the student is accepted by that school.

6 Estimation and Results

In the following estimation, the utility function is specified as

$$\begin{aligned} u_{is} &= \alpha_s + \delta_1 TScore_i \times \overline{TScore99_s} + \delta_2 Distance_{is} \\ &\quad + \gamma_1 Income_i + \gamma_2 \#Awards_i + \gamma_3 TScore_i + \varepsilon_{is}, \end{aligned}$$

where α_s is the school fixed effect; $TScore_i$ is students' test score from elementary school; $\overline{TScore99_s}$ is the average test score of graduating class in school s in 1999. $Distance_{is}$ is the walking distance from student i's home to school s; $Income_i$ is student i's family income; $\#Awards_i$ is the number of student i's awards during the elementary school. The part $\gamma_1 Income_i + \gamma_2 \#Awards_i + \gamma_3 TScore_i$, which is constant for any inside school, captures the difference between inside schools and the outside option.[45]

$\varepsilon_{i1}, ..., \varepsilon_{i4}$ are assumed to have a multinomial normal distribution and i.i.d. across students, and are allowed to be correlated with each other. The variance of $\varepsilon_{i,1}$ is normalized to be 1.

6.1 Estimation and Tests of Overall Sophistication

In this section, I first estimate the model under the hypothesis that students understand the mechanism, and more conditions are added sequentially. The Small-Hsiao test is used to determine what kind of Bayesian Nash equilibrium should be estimated. Particularly, the following cases are considered.

Case 1 is the baseline case in which students understand the mechanism. This case allows some students to think that the probability of being assigned to school 3 is zeros when listing it as 3rd or 4th choice. The following conditions can be derived:[46]

(i) If a full list $C_i = \left(c_i^1, ..., c_i^S\right)$ observed, then: (a) $u_{i,c_i^1} > 0$; (b) $u_{i,c_i^S} = \min_s \left(u_{i,s}\right)$ if $c_i^S \neq 3$.

(ii) If a partial list $C_i = \left(c_i^1, c_i^2, 0, 0\right)$ observed, then: (a) $u_{i,c_i^1} > 0$; (b) $u_{i,s} \leq 0, \forall s \notin C_i$ and $s \neq 3$.

(iii) If a partial list $C_i = \left(c_i^1, 0, 0, 0\right)$ observed, then: (a) $u_{i,c_i^1} > 0$; (b) $u_{i,s} \leq 0, \forall s \notin C_i$.[47]

(iv) If a student doesn't participate, $C_i = (0, ..., 0)$, then $u_{i,s} \leq 0, \forall s = 1, ..., 4$.

[45] Other variables are included in the regression for robustness check, for example, parents' education, gender of the student, and student's elementary school. The coefficients are either not significant or marginally significant and very close to zero.

[46] An extra case is also considered where it requires the probability of being assigned to school 3 is strictly positive when it is the 3rd th or 4th choice. This case is a nested case of Case 1. Test result rejects this case in favor of Case 1. Notice that the truth-telling model is also a nested model of this case. These two are tested against each other, and the truth-telling hypothesis is strongly rejected. The results are not reported here.

[47] Conditions (ii) and (iii) predict that students won't attend schools which are not included in their list. The only except is if school 3 is excluded and the list has two schools. Among 70 students submitted partial lists, in the data only 2 students attend certain schools which violate the conditions in (ii) and (iii).

Case 2 assumes all the conditions in Case 1 and that the second choice school in the list is better than the outside option. Case 3 has one more condition than Case 2 – the third choice school in the list is better than the outside option.

Results are collected in Table 6.[48] The first test is Case 2 against Case 1. The null hypothesis that all conditions in Case 2 hold true is not rejected – P-value 0.30. Then test Case 3 against Case 1. Case 3 is marginally rejected – P-value is between 0.6. This means that for some students the third choice schools are worse than the outside option. For robustness check, both possibilities will be considered, in the following, the results are conditional on Case 3 being rejected.

Under the assumptions in Case 2, we have to consider Bayesian equilibrium in mixed strategies. In the following, the estimate from Case 2 will be used. The estimated average mean utility of each school is $[0.87, 0.73, 1.05, 0.57]$.[49]

Since the second choice school is always acceptable, we do not consider the mixing between one-school lists and other lists. Using the estimates and observed lists, we can calculate the probability of mixing a two-school list with a full list. For example, the probability of playing $(1, 2, 3, 4)$ while $u_{i,1}, u_{i,2} > 2$ and $u_{i,3}, u_{i,4} < 0$ can be estimated in this way:

$$\frac{P\left\{(1,2,3,4) \text{ submitted}; u_{i,1}, u_{i,2} > 2; u_{i,3}, u_{i,4} < 0\right\}}{P\left\{(1,2,3,4) \text{ or } (1,2,0,0) \text{ submitted}; u_{i,1}, u_{i,2} > 2; u_{i,3}, u_{i,4} < 0\right\}}$$

Given the mixed strategies estimated, a penalty function approach is used to estimate the beliefs and utility function. It start with a large π_t,

$$\max_{\theta, B}\left[\sum_{i=1}^{I}\ln\left(\widetilde{P}\left(C_i|\mathbf{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}, B; \boldsymbol{\theta}\right)\right) - \pi_t\left|\left|B - \overline{B}\left(\cdot, B\right)\right|\right|^2\right],$$

In the estimation, π_t increases sequentially and the stopping rule is: $\left|\left|B - \overline{B}\left(\cdot, B\right)\right|\right|^2 \leq 1 * 10^{-5}$.

The results are reported in column 3 of Table 6. At the estimated coefficients, the inconsistency in beliefs is $\left|\left|\widehat{B} - \overline{B}\left(\cdot, \widehat{B}\right)\right|\right|^2 = 1.84 * 10^{-8}$. The Small-Hsiao test is performed against Case 2. The null hypothesis is rejected – the results are inconsistent with students playing this particular Bayesian Nash equilibrium. One possible reason of the rejection is that not every student is sophisticated. Indeed, playing a Bayesian

[48] Standard errors are estimated by outer product of gradients. Robust/sandwich standard errors are in progress.

[49] The correlation between any two $\varepsilon_{i,s}$ is very high – from 0.95-0.99. This means conditional on all observables, parents have the same ordinal preferences over four schools. The utility shock only affects the preference ranking of the outside option. The plausible reason is that the variance of outside option is very high. After normalization, $\varepsilon_{i,s}$ captures the difference between school s and the outside option. The high-variance shock to the outside option appears in each $\varepsilon_{i,s}$. It dominates the school's own shock, then $\varepsilon_{i,s}$ may have a almost 1 correlation between each other.

Nash equilibrium without any coordination is hard to achieve. Another possibility is that students may be risk averse. The fact that students include unacceptable schools in their lists is one sign of risk aversion. While Case 2 is independent of risk attitude, the risk neutrality is imposed when estimating the Bayesian Nash equilibrium. This certainly calls for further investigation.

The Bayesian Nash equilibrium under Case 3 is also estimated. In this case, the equilibrium is unique and in pure strategies (Proposition 5). The Small-Hsiao test is performed against Case 3. Again, the null hypothesis that every parent is sophisticated is rejected – P-value is less than 0.01.

6.2 Sophistication and Strategic Behavior

Since the Bayesian Nash equilibrium is rejected, the probabilities calculated based on the ex post distribution of other players' behavior are used to approximate the objective beliefs.[50] Denote it as $\widehat{B}_0$.

The goal of this subsection is to find out who strategizes better in BM and how parents response to the incentives. Measures of sophistication and incentives are constructed based on the estimated preferences. Moreover, the 2002 survey asks about how much attention the parent pays to school quality and uncertainty in the game. Regressions are used to determine how the sophistication, the incentives and the attention to school quality and uncertainty are correlated.

6.2.1 Measures of Sophistication and Incentives to Be Strategic

To investigate the strategic behavior, the following measures of sophistication are constructed.

(i) Accuracy of the beliefs: $-\left\|\widehat{B}_i - \widehat{B}_0\right\|$

$\widehat{B}_i$ is the subjective belief and is estimated for students who submitted a full list – 663 students. It is the solution of maximizing the probability that C_i is a best response (equation 6) given $\widehat{\boldsymbol{\theta}}$. The probability is approximated by drawing 1000 times of $\boldsymbol{\varepsilon}_i$. The same draws are used in the estimation of following measures.

(ii) Probability that i played truthfully: $p_i^{TT}(C_i)$.

(iii) Probability that i played a best response: $p_i^{BR}(C_i)$.

(iv) Probability that i played a best response which is not truthful reporting: $p_i^{BR\neq TT}(C_i)$.

[50] A "smoothed' version of the empirical beliefs is also used. In the new version, the 0 or 1 probabilities in the empirical beliefs are deviated from original value by a small amount. But the results remain almost the same.

(v) The expected utility achieved by C_i as a fraction of the maximum: $EU_i(C_i)$.

The maximal expected utility is found by choosing one list with maximum expected utility among all 24 full lists. This list is also used to calculate the standard deviation of utility.

(vi) The variance of utility given C_i as a fraction of the variance of utility when expected utility is maximized: $Var_i(C_i)$.

In the estimation, one assumption is that a parent will not participate if and only if all schools are unacceptable. For those non-participants, they are playing a best response and truth-telling, i.e. $p_i^{BR}(C_i) - p_i^{TT}(C_i) = 0$. Thus, $p_i^{BR}(C_i) - p_i^{TT}(C_i)$ is better defined for participants. Similarly, $EU_i(C_i)$ and $Var_i(C_i)$ are better defined for participants as well. Otherwise, $EU_i(C_i) = 0$ and $Var_i(C_i) = 0$, because choosing the outside option gives the parents zero utility for sure. In the following, these three measures are used for the subsample of participants, unless noted otherwise.

Furthermore, the incentives to be strategic are measured by the following variables.

(i) Ex ante probability that truth-telling is a best response: $p_i^{TT=BR}$.

(ii) Ratio of the expected utility if truth-telling, to the expected utility if best responding: EU_i^{TT}.

(iii) Ratio of the variance of utility if truth-telling, to the variance of utility if best responding: Var_i^{TT}.

The following matrix shows the correlations between any two of the measures.

	Accuracy of beliefs	$p_i^{BR}(C_i) - p_i^{TT}(C_i)$	$p_i^{BR \neq TT}(C_i)$	$EU_i(C_i)$	$Var_i(C_i)$	$p_i^{TT=BR}$	EU_i^{TT}
Accuracy of beliefs	1						
$p_i^{BR}(C_i) - p_i^{TT}(C_i)$	0.736	1					
$p_i^{BR \neq TT}(C_i)$	0.465	0.795	1				
$EU_i(C_i)$	0.472	0.504	0.622	1			
$Var_i(C_i)$	-0.761	-0.645	-0.313	-0.341	1		
$p_i^{TT=BR}$	-0.319	-0.042	0.065	-0.300	0.118	1	
EU_i^{TT}	-0.022	-0.101	-0.290	0.127	-0.152	0.208	1
Var_i^{TT}	0.109	0.125	0.224	-0.029	0.077	-0.216	-0.747

The accuracy of beliefs are positively correlated with $p_i^{BR}(C_i) - p_i^{TT}(C_i)$ and $p_i^{BR \neq TT}(C_i)$ (probability of playing a non-truth-telling best response). It is also associated with a higher expected utility achieved

$(EU_i(C_i))$ with a lower variance $(Var_i(C_i))$.

The measures of incentives seem reasonable. Higher $p_i^{TT=BR}$ and EU_i^{TT} mean less incentive to be strategic. They are negatively correlated with the accuracy of beliefs and $p_i^{BR}(C_i) - p_i^{TT}(C_i)$. $p_i^{TT=BR}$ is positively correlated with $p_i^{BR\neq TT}(C_i)$, but the correlation is low. $p_i^{BR=TT}$ is also negatively correlated with the expected utility achieved $(EU_i(C_i))$ and positively correlated with the variance achieved $(Var_i(C_i))$, which means that more incentives to be strategic are associated with better outcomes. However, EU_i^{TT} is positively correlated with $EU_i(C_i)$ and negatively correlated with $Var_i(C_i)$. This is reasonable because the incentive measured by EU_i^{TT} is in terms of possible outcomes. When this incentive is low, the outcome will be very likely to be good. The third measure, Var_i^{TT}, is positively correlated with the incentive to be strategic. Its correlation between the sophistication measures is always opposite to the correlation between sophistication and EU_i^{TT}.

Table 7 shows the summary statistics of the measures. There is evidence that many parents strategize well. Given their submitted list, the probability that students are truth-telling is very low. For half of them, the likelihood is below 3.9%, while on average it is 16.1%. On the other hand, the probability of best responding is high, with a average at 32.3%. Overall, parents are more likely to be best responding – the mean of $\left(p_i^{BR}(C_i) - p_i^{TT}(C_i)\right)$ is 15.2%. There are only 138 (or 15.1%) parents who have a negative $\left(p_i^{BR}(C_i) - p_i^{TT}(C_i)\right)$. The probability of playing a non-truth-telling best response is also high. On average, the probability that each parent is playing a non-truth-telling best response is 24.5%. The probability is even higher among those who include at least one school in their list.

The fraction of maximum expected utility achieved, $EU_i(C_i)$, is high – the mean is 75.7% and the median is 92.8%. On the other hand, there are some parents who have very low EU_i. Those are the parents who choose the outside option. In fact, all parents who have $EU_i(C_i) < 44\%$ did not participate. $EU_i(C_i)$ for this group of parents might be under-estimated, because they know their realization of utilities when making the decision while $EU_i(C_i)$ is calculated with the distribution of utilities. In the model, those who choose the outside option play a best response, because it is assumed they do not participate if and only if every school is worse than the outside option. If this group is excluded, the mean of $EU_i(C_i)$ is 94.4%. The variance achieved, $Var_i(C_i)$, for the participating group is 91%, which means the chosen list gives a variance of utility 9% lower than the list maximizing the mean utility.

Among the incentive measures, the incentive to be strategic is not very high. On average, truth-telling is a best response with probability 42.9% ($p_i^{TT=BR}$). Always submitting the true preferences can achieve 94.5% of the expected utility that can be obtained if always best responding (EU_i^{TT}). However, always submitting the true preferences incurs a higher variance of utility – average Var_i^{TT} is 1.864, or 86.4%

higher.

6.2.2 Who Has More Incentives to Be Strategic?

When a parent has a decent outside option, there is less incentive to be strategic, or to misreport their preferences. She may just need to list the only school better than the outside option. Since the outside option is affected by income, test score, and awards, $p_i^{TT=BR}$ is expected to be correlated with these variables. The regression result is reported in column 1 of Table 8. After controlling for student's gender and elementary school, $p_i^{TT=BR}$ is positively correlated with income, parents' education and number of awards during the elementary school – 1% increase in income is associated with 0.034 percentage points increase in $p_i^{TT=BR}$; one year increase in parents' education is associated with 0.7 percentage points increase in $p_i^{TT=BR}$. However, it is negatively correlated with student's test score in elementary school. This may due to that student's test score determines the utility of attending "inside" schools because it is interacted with schools' average test score.

Similar regressions are run for EU_i^{TT} and Var_i^{TT}. Results are shown in columns 2-5 in Table 8. In the two regressions of EU_i^{TT}, one controls for Var_i^{TT}, while the other does not. The coefficients on other variables are similar in both of them. Income and awards have a significant negative effect on EU_i^{TT}, which means higher income parents whose child has more awards have more incentive to be strategic. However, the magnitude of the effects are small. 1% increase in income is related to 0.005 percentage points decrease in EU_i^{TT}. One more award only decrease EU_i^{TT} by 0.001 percentage points. Test score is negatively correlated with EU_i^{TT}; 1% increase in test score is associated with 0.086 percentage points decrease in EU_i^{TT} (0.055 if controlling for Var_i^{TT}).

In the two regressions of Var_i^{TT}, controlling for EU_i^{TT} makes a big difference. It may make more sense to look at the variance after controlling for the mean in this context, as the variance alone is not a good measure for welfare. After controlling for EU_i^{TT}, Var_i^{TT} is negatively correlated with income and test score, which means higher income and test score families have less incentive to be strategic. 1% increase in income is related to 0.033 percentage points decrease in Var_i^{TT}. 1% increase in test score is associated with 0.131 percentage points decrease in Var_i^{TT}.

Overall, parents with higher income and education have less incentives to be strategic. If the test score of the student is higher, there can be more incentives to be strategic if the parent is not too risk averse.

6.2.3 Parents' Attention on School Quality and Uncertainty

In the 2002 survey, there are questions asking about how much attention the parent pays to 12 factors. Most of them are about school quality, for example, teachers' quality, school facilities etc. Among these factors, three are about the uncertainty of the game: (1) Reducing the probability of being assigned to bad schools; (2) school quota and possibility of being accepted; (3) Other parents' application.[51] Parents give the answer on a scale of 1-5, where 5 means caring about it very much.

All 12 questions are summarized in Tables A-1 and A-2 in appendix which also report the coefficients from regressions of constructed sophistication and incentive measures on survey responses to each question. Their relationship will be investigated in more details shortly. In the following, two variables are created, $Attn\_Q$ and $Attn\_U$. $Attn\_Q$ is the average of responses to the questions on school quality; $Attn\_U$ is the average response to those on uncertainty questions. Higher $Attn\_Q$ ($Attn\_U$) means that patents say they pay more attention to school quality (uncertainty) on several dimensions. Table 9 reports the summary statistics of these two variables. Overall, parents pay more attention to school quality than to uncertainty, and the variance of parents' attention on uncertainty is higher than that on school quality. Parents who submit a full list pay more attention to uncertainty.

Similar questions are asked about parents' spending some effort collecting information on quality and uncertainty. These questions are asked at the school level. Construct Effort_Q1 and Effort_U1 as the efforts spent on quality and uncertainty of their first choice school. Due to the design of the questionnaire, more noise is expected in the effort variables than in the attention variables, even more noise in the effort variables about second or later choice schools.[52] Therefore, the effort variables will not be explored in the following.

6.2.4 Who Strategizes Better?

There are five measures of sophistication or how well the parent strategize: accuracy of the beliefs, $-\left\|\widehat{B}_i - \widehat{B}_0\right\|$, probability of best responding vs. truth-telling, $\left(p_i^{BR}\left(C_i\right) - p_i^{TT}\left(C_i\right)\right)$, probability of playing an non-truth-telling best response, $p_i^{TT \neq BR}\left(C_i\right)$, and mean utility achieved, $EU_i\left(C_i\right)$, and variance achieved, $Var_i\left(C_i\right)$. The following analysis examines how incentives and attention on school quality and uncertainty affect parents' strategy measured the by these five variables. Every regression includes fixed effects for student's gender and their elementary school.

[51] Paying attention to other parents' application may also relate to school quality in the sense that other parents' application reflects their preferences over school possibly in an untruthful way. The following results remain similar if this factor is excluded, although some coefficients becomes less significant.

[52] The 12 questions on how much the parents care about quality and beliefs are asked right before the effort questions. The 10 effort questions are asked at the school level. Given the great similarity among these 52 questions, the respondent may lose most of their attention when answering late questions.

Table 10 reports the regressions analysis of accuracy of the beliefs, $-\left\|\widehat{B}_i - \widehat{B}_0\right\|$. Higher $p_i^{TT=BR}$ is correlated with less accurate subjective beliefs. The effect of EU_i^{TT} have different signs if $p_i^{TT=BR}$ is included in the regression or not. Var_i^{TT} is positively correlated with more accurate beliefs. When $p_i^{TT=BR}$ is included in the regression, income, education and awards has no significant effect on the accuracy. When $p_i^{TT=BR}$ is excluded, all these variables have a negative effect. Moreover, test score has a significant negative effect in all specifications. Interestingly, more attention on uncertainty ($Attn\_U$) improves the accuracy of beliefs, while attention on school quality does not.

Table 11 reports the regression analysis of $\left(p_i^{BR}(C_i) - p_i^{TT}(C_i)\right)$. In this case, among three incentive measures, only Var_i^{TT} has a significant coefficient. Higher Var_i^{TT} is associated with more likely to play a best response than report truthfully. None of income, education, test score and awards has a significant effect. However, attention to uncertainty significantly increases $\left(p_i^{BR}(C_i) - p_i^{TT}(C_i)\right)$.

Table 12 shows regression results of the probability of playing a non-truth-telling best response, $p_i^{TT\neq BR}(C_i)$. Although $p_i^{TT=BR}$ has no significant effect, EU_i^{TT} and Var_i^{TT} have the expected signs and significant in most of the cases. Income has a significantly positive effect, but it becomes insignificant when controlling for EU_i^{TT}. Parents' education have a positive effect and always significant. One year increase in parents' education increases the probability of playing a non-truth-telling best response by 0.7-0.9 percentage points. Attention to uncertainty has a positive effect and significant at 15% level.

The analysis of parents' application outcome, $EU_i(C_i)$ and $Var_i(C_i)$, is in Tables 13 and 14. $p_i^{TT=BR}$ has a significantly negative effect on the expected utility but not on the variance of utility. EU_i^{TT} and Var_i^{TT} have an insignificant or a marginally significant effect. Income has a negative effect on the expected utility and sometime significant. Parent's education has no effect. Test score and awards are negatively correlated with expected utility and positively correlated with variance, which is significant in many cases. Interestingly, attention on school quality and uncertainty has no significant effect on expected utility, but has highly significant effect on variance. $Attn\_Q$ is positively correlated with variance, while $Attn\_U$ is negatively correlated.

To summarize the results, when there are more incentives to be strategic, parents tend to strategize better. $Attn\_U$ has a significantly positive effect on how well the parent strategize in many dimensions. There is no robust evidence that higher income and education help parent strategize better. Sometimes they are associated with worse outcomes.

Surprisingly, students' test score has a negative effect in several dimensions. High test scores are associated with inaccurate beliefs, lower expected utility and higher variance. One possible reason is that parents are still learning about the new system. Since test scores are the only determinant of priority under the old mechanism (two years before), they might think that test scores still play a role in the new system.

6.2.5 Behavioral Responses to the Incentives

The attention to school quality and uncertainty are behavioral responses when playing the game. It is plausible that a parent pays more attention to uncertainty when she has more incentives to be strategic. The same effect may also be possible in the case of attention on school quality.

Table 15 collects the results on the determinants of $Attn\_U$. The incentive measures, $p_i^{TT=BR}$ and EU_i^{TT}, have a significantly negative effect on the attention on uncertainty, while Var_i^{TT} does not. Since the survey questions used to construct $Attn\_Q$ and $Attn\_U$ are mixed with each other in the questionnaire, they may have correlated measurement errors. An IV regression is used to address this problem. Effort_Q which measures the effort spent on researching school quality is used as an IV for $Attn\_Q$. The first stage P-Value is between 0.056-0.062, but it may be invalid because it can have measurement errors which are correlated with those in $Attn\_U$. In the IV regression, the coefficients on incentive measures become insignificant. $Attn\_Q$ still has a significant positive effect. Another interesting result is that parents' education have a negative effect on $Attn\_U$, which is significant and quite robust. However, the magnitude is quite small. One year increase in parents' education decreases $Attn\_U$ by 0.027-0.034, while the mean of $Attn\_U$ is 3.844 and the standard deviation is 0.670.

Tables 16 reports the results about $Attn\_Q$. For the same reason as above, Effort_U which measures the effort spent on researching uncertainty is used as an IV for $Attn\_U$. The first stage P-Values are all less than 0.01.

$p_i^{TT=BR}$ and EU_i^{TT} have no significant effect on $Attn\_Q$. Var_i^{TT} is negatively correlated with $Attn\_Q$ and the correlation is significant. Income has a positive effect and sometimes significant, while parents' education has an insignificant effect.

In summary, there is evidence that when there are more incentives to be strategic, parents pay more attention to uncertainty but not to school quality. Higher income parents pay more attention to school quality but not to uncertainty. Highly educated parents pay less attention to uncertainty.

7 Efficiency Comparison

To evaluate the efficiency under the two mechanisms, the outcomes under DA and BM are simulated with the estimates of preference. Since truth-telling is the dominant strategy in DA, the outcome under DA is obtained by assuming every parent reports her true preferences. As mentioned earlier, the welfare criterion is the ex ante efficiency, so the probabilities of being assigned to each school when submitting any list should be calculated. These probabilities in DA are obtained by drawing 10,000 profiles of true preferences and simulating the outcome. In the case of BM, the equilibrium outcome has to be considered. Since parents' sophistication plays an important role in BM, the equilibrium is solved with different fractions of naive parents among all parents. Specifically, eleven cases with 0, 10%, 20%, ..., or 100% naive parents are investigated. Who should be assigned to naive parents in which case is determined by the constructed measure, $\left(p_i^{BR}\left(C_i\right) - p_i^{TT}\left(C_i\right)\right)$ – the difference between the probability of best responding and that of truth-telling. Higher $\left(p_i^{BR}\left(C_i\right) - p_i^{TT}\left(C_i\right)\right)$ means more sophisticated, and is assigned to be naive in later cases. In the first ten cases, the equilibrium beliefs are solved by finding a fixed point in a similar way as in the estimation of preference (Appendix A.2). The last case where all the parents are naive, the probabilities of being assigned to each school is simulated in the same manner as in DA.

The next step is to simulate profiles of parents and use the probabilities of being assigned to each school to calculate their welfare. 500 profiles of players are constructed by drawing the error terms, which essentially creates 960*500 distinctive parents. Each of them plays two types of games – BM and DA, and in BM there are 11 cases to be played. Moreover, parents have to play each case of BM as sophisticated parents and then as naive parents.

There are two dimensions of welfare comparison. One is the expected utility, and the other is the standard deviation of utility which reflects the uncertainty in the assignment process.[53] Figure 2 reports the average expected utility and standard error of utility. The results can be interpreted as the expected utility and the standard deviation of utility if a parent is randomly selected. The figure shows that a sophisticated parent can get higher expected utility and lower standard deviation in BM than what she can get in DA. As the fraction of naive parents increases, the welfare of a sophisticated parent also increases. For a naive parent, the difference between BM and DA is very small. When the fraction of naive parents is small (<40%), the expected utility that a naive parent can get in BM is very close to what she can obtain in DA. However, the standard deviation of utility is much higher in BM, which indicates a higher risk when becoming a naive parent. When the fraction of naive parents is large (>=40%), the standard deviation in BM becomes smaller than that in DA.

The above analysis of the average values implicitly assumes inter-personal comparison of cardinal utility. Another way to evaluate the welfare is to do the comparison within each parent. Figure **??** shows the

[53]Uncertainty is an important factor that parents care about. For example, in a presentation by Seattle Public Schools, it says, "*So the [school] district spent a lot of time p talking to families about what they wanted from the assignment process. They said predictability, equity, and ease of understanding.*" Available at http://www.seattleschools.org/area/newassign/nsap_mid_ws.pdf. Retrieved October 15, 2009.

welfare effects of replacing BM with DA which only uses intra-personal comparison. For a randomly selected sophisticated parent, there is about 80% of the chance that she is worse off in DA, while it is almost impossible to be better off. This result holds true regardless of the fraction of naive parents. On the other hand, for a naive parent, there is about 52% of the chance of she being better off. However, she can be worse off with probability 20%. There is a 20% probability that the parent obtains the same expected utility in both mechanisms, because these are the parents who choose the outside option.

The intra-personal comparison can be extended to standard deviation of utility. Figure 3 compares the welfare of a randomly selected parent in BM and DA, while the welfare is evaluated by both the expected utility and the standard deviation of utility. When there are no naive parents, a sophisticated parent has a 49% probability of being strictly worse off, while the probability of being better off is almost zero. When the fraction of other naive parents increases, the probability of being worse off increases, and the probability of having an ambiguous change in welfare (decreases in both the expected utility and the standard deviation) decreases. For a naive parent, the probability of being better off decreases with the fraction of naive parents at first. It reaches the lowest (26%) when 50% of other parents are naive, then it increases with the fraction of naive parents. The opposite pattern is observed for the possibility that a naive parent has a higher expected utility and a higher variance of utility. Adding these two probabilities together gives us the stable probability that a naive parent has a higher expected utility in DA. Moreover, the probability of a naive parent being worse off is between 16-20%.

A similar graph can be drawn by different levels of the incentive to be strategic which is measured by the probability that truth-telling is a best response ($p_i^{BR=TT}$). The parents with the smallest incentive are the 96 parents (10%) who have the highest probability that truth-telling is a best response. The parents with the greatest incentive are the 96 parents (10%) who have the lowest probability that truth-telling is a best response. The results for these two groups are plotted in Figure 4. For a sophisticated parent, she has a almost zero probability of being better off in DA regardless of her incentives. However, parents who have the greatest incentive are more likely to be worse off and less likely to have no change in welfare. Naive parents who have the smallest incentives are more likely to have no change in welfare, and have lower probabilities of being better off and being worse off. If naive parents are those who have the smallest incentives to be strategic, the welfare effects of replacing BM with DA can be better described by B and C in Figure 4. Only 23-36% of naive parents, or 2.3-3.6% of all parents, are strictly better off.

To summarize, if BM is replaced by DA, almost all the sophisticated parent are weakly worse off, and most of them are strictly worse off. Among all the naive parents, about half of them are better off, and supprisingly, around 20% of them can be hurt by the reform.

8 Concluding Remarks

On the efficiency comparison of BM and DA, previous literature provides an inconclusive answer. The theoretical results are mixed and depend on the sophistication level of players. The empirical evidence is either reduced-form or experimental. This paper uses a data set from Beijing to answer two questions: 1) How sophisticated are the parents in BM? 2) How efficient is BM comparing to DA in real life?

Assuming that students' preferences are private information, this paper models school choice under BM as a simultaneous game of incomplete information. Due to the lack of strategy-proofness, submitted preference lists are not necessarily students' true preferences. The paper describes a set of equilibrium conditions on parents' behavior which are then used to formulate the likelihood function. A simulated maximum likelihood method is used for the estimation.

The paper find that the behavior of parents as a whole is not consistent with a Bayesian Nash equilibrium, suggesting some parents are less sophisticated. Results also show that when parents have more incentive to be strategic, they tend to pay more attention to the uncertainty, which makes them strategize better. There is no robust evidence that wealthier or more educated parents strategize better. But it is easier for them to find the best response due to their better outside option.

Assuming that the preferences do not change across different mechanisms, the paper simulates the outcomes under DA with the estimated preferences. If the Boston mechanism is replaced by the DA mechanism, almost all the sophisticated parents are weakly worse off, and the majority of them are strickly worse off. Among all the naive parents, about half of them are better off, and surprisingly, around 20% of them are hurt by this kind of reform.

References

ABDULKADIROGLU, A., Y.-K. CHE, AND Y. YASUDA (2008): “Expanding "Choice" in School Choice,” *Mimeo*.

——— (Forthcoming): “Resolving Conflicting Preferences in School Choice: the Boston Mechanism Reconsidered,” *American Economic Review*.

ABDULKADIROGLU, A., P. A. PATHAK, AND A. E. ROTH (2009): “Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match,” *American Economic Review*, 99(5), 1954–78.

ABDULKADIROGLU, A., P. A. PATHAK, A. E. ROTH, AND T. SONMEZ (2005): “The Boston Public School Match,” *The American Economic Review, Papers and Procedings*, 95(2), 368–371.

——— (2006): “Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access,” *Mimeo*.

ABDULKADIROGLU, A., AND T. SONMEZ (1998): “Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems,” *Econometrica*, 66(3), 689–702.

——— (2003): “School Choice: A Mechanism Design Approach,” *American Economic Review*, 93(3), 729–747.

ANGRIST, J., E. BETTINGER, E. BLOOM, E. KING, AND M. KREMER (2002): “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment,” *The American Economic Review*, 92(5), 1535–1558.

ARADILLAS-LOPEZ, A. (2007a): “Pairwise Difference Estimation of Incomplete Information Games,” *Mimeo*.

——— (2007b): “Semiparametric Estimation of a Simultaneous Game with Incomplete Information,” *Mimeo*.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2006): “Estimating Static Models of Strategic Interaction,” Discussion paper, National Bureau of Economic Research, Inc, NBER Working Papers.

BARTLE, R. G. (1964): *The Elements of Real Analysis*. Wiley, New York.

BELEGUNDU, A. D., AND T. R. CHANDRUPATLA (1999): *Optimization Concepts And Applications In Engineering*. Prentice Hall, Upper Saddle River, New Jersey.

BERRY, S. T. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4), 889–917.

BERRY, S. T., J. LEVINSOHN, AND A. PAKES (1996): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841 – 890.

BHATTACHARYA, R. N. (1977): “Refinements of the Multidimensional Central Limit Theorem and Applications,” *The Annals of Probability*, 5(1), 1–27.

BUDISH, E., AND E. CANTILLON (2009): “Strategic Behavior in Multi-Unit Assignment Problems: Theory and Evidence from Course Allocation,” *Mimeo*.

CHEN, Y., AND T. SONMEZ (2006): “School Choice: An Experimental Study,” *Journal of Economic Theory*, 127(1), 202–231.

CHIAPPORI, P.-A., S. LEVITT, AND T. GROSECLOSE (2002): “Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer,” *American Economic Review*, 92(4), 1138–1151.

CULLEN, J. B., B. A. JACOB, AND S. LEVITT (2006): “The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries,” *Econometrica*, 74(5), 1191–1230.

DUBE, J.-P. H., J. T. FOX, AND C.-L. SU (2009): “Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation,” *NBER Working Paper Series*, No. 14991.

DUBINS, L. E., AND D. A. FREEDMAN (1981): “Machiavelli and the Gale-Shapley Algorithm,” *The American Mathematical Monthly*, 88(7), 485–494.

ERDIL, A., AND H. ERGIN (2008): “What’s the Matter with Tie-Breaking? Improving Efficiency in School Choice,” *American Economic Review*, 98(3), 669–89.

ERGIN, H., AND T. SONMEZ (2006): “Games of School Choice under the Boston Mechanism,” *Journal of Public Economics*, 90(1-2), 215–237.

FEATHERSTONE, C., AND M. NIEDERLE (2008): “Ex Ante Efficiency in School Choice Mechanisms: An Experimental Investigation,” *NBER Working Paper Series*, No. 14618.

FLANDERS, H. (1973): “Differentiation Under the Integral Sign,” *American Mathematical Monthly*, 80(6), 615–627.

GALE, D. E., AND L. S. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *The American Mathematical Monthly*, 69(1), 9–15.

GILL, P. E., W. MURRAY, AND M. H. WRIGHT (1982): *Practical Optimization*. Academic Press, London.

GOURIEROUX, C., AND A. MONFORT (1997): *Simulation-Based Econometric Methods*. Oxford University Press, Oxford.

GREENE, W. H. (1999): *Econometric Analysis*. Prentice-Hall, 4th edition edn.

HASTINGS, J., T. KANE, AND D. STAIGER (2008): “Heterogeneous Preferences and the Efficacy of Public School Choice,” *Mimeo*, Yale University.

HE, Y. (2009): “Random Assignment: Achieving Competitive Equilibrium from Equal Incomes via the Boston Mechanism,” *Mimeo*.

HEPBURN, C. R. (1999): “The Case For School Choice: Models from the United States, New Zealand, Denmark, and Sweden,” *Critical Issues Bulletins*, available at http://oldfraser.lexi.net/publications/critical_issues/1999/school_choice/.

HORTACSU, A., AND S. L. PULLER (2008): “Understanding strategic bidding in multi-unit auctions: a case study of the Texas electricity spot market,” *RAND Journal of Economics*, 39(1), 86–114.

HSIEH, C.-T., AND M. URQUIOLA (2006): “The effects of generalized school choice on achievement and stratification: Evidence from Chile’s voucher program,” *Journal of Public Economics*, 90(8-9), 1477–1503.

IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66(2), 333–357.

KOJIMA, F. (2008): "Games of School Choice under the Boston Mechanism with General Priority Structures," *Social Choice and Welfare*, 31(3), 357–365.

KOVASH, K., AND S. D. LEVITT (2009): "Professionals Do Not Play Minimax: Evidence from Major League Baseball and the National Football League," *NBER Working Papers*, No. 15347.

LAI, F., E. SADOULET, AND A. DE JANVRY (2009): "The Adverse Effects of Earents' School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program," *Economics of Education Review*, 28(4), 485–496.

MIRALLES, A. (2008): "School Choice: The Case for the Boston Mechanism," *Mimeo*.

PAIS, J., AND A. PINTER (2008): "School Choice and Information: An Experimental Study on Matching Mechanisms," *Games and Economic Behavior*, 64(1), 303–328.

PATHAK, P. A., AND T. SONMEZ (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 98(4), 1636–52.

PAYZANT, T. W. (2005): "Student Assignment Mechanics: Algorithm Update and Discussion," *Memorandum to Chairperson and Members of the Boston School Committee*.

ROTH, A. (2008): "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions," *International Journal of Game Theory*, 36(3), 537–569.

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

——— (1984): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *The Journal of Political Economy*, 92(6), 991–1016.

SEIM, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-type Choices," *The Rand Journal of Economics*, 37(3), 619–640.

SEPPANEN, P. (2003): "Patterns of 'Public-school markets' in the Finnish Comprehensive School from a Comparative Perspective," *Journal of Education Policy*, 18(5), 513 – 531.

SMALL, K. A., AND C. HSIAO (1985): "Multinomial Logit Specification Tests," *International Economic Review*, 26(3), 619–627.

TICE, P., C. CHAPMAN, D. PRINCIOTTA, AND S. BIELICK (2006): *Trends in the Use of School Choice 1993 to 2003*, NCES 2007-045. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics., Available at http://nces.ed.gov/pubs2007/2007045.pdf.

TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

ZHOU, L. (1990): "On A Conjecture by Gale about One-Sided Matching Problems," *Journal of Economic Theory*, 52(1), 123–135.

A Appendices

A.1 Proofs

Proof of Proposition 1.

(i) Suppose a participating student submits a full list, and she is rejected by all her choices but the last choice (s^*). Then in Round S, school s^* must have more available seats than students unassigned. If the total seats left at s^* is $\overline{q}$, then the number of students unassigned is $I - \sum_s q_s + \overline{q}$. Since $I \leq \sum_s q_s$, $I - \sum_s q_s + \overline{q} < \overline{q}$. Thus, the student must be assigned to her last choice.

Suppose a participating student submits a partial list and is rejected by all the schools in her list. After at most S rounds, she is still unassigned. The number of available seats then is at least $\sum_s q_s - I + 1 \geq 1$. Thus she will be assigned to some school.

(ii) Suppose C and C' have the same first K choices. In any realization of the game (any lottery number and any other players lists), if the student is assigned to one of the first K choices when submitting C, she will be assigned to that school when submitting C'. If she is not assigned to a school in the first K schools when submitting C, she will not be assigned to that school if submitting C' instead. This means she has the same probability to be assigned to any of the first K choices when she submits C or C'.

(iii) Suppose C and C' have the same first $K-1$ choices. School s is listed as Kth choice in C, but as K'th choice in C'. In any realization of the game (any lottery number and any other players lists), if the student is rejected by s when submitting C, she will not be accepted by s if submitting C'. If she is accepted by s when submitting C', in the same realization of the game, school s has more available seats than applying students in Round K. Thus, she will be assigned to s if submitting C. This implies the probability of being assigned to s weakly increases when moving it toward the top of the list. In the same manner, including an otherwise omitted school in the list has the same effect.

(iv) The number of students listing s as first choice is at most I. Since a lottery number is used to determine who will be accepted, among those who have the same first choice, everyone have the same probability being accepted by that school. The probability of being accepted by s is at least q_s/I if a student list s as first choice. ■

Proof of Proposition 2.

(i) Suppose the first choice in list C is unacceptable, or worse than the outside option. Construct a new list, C', such that the first school is the best school and all other choices in C' are the same as C. Then given any realization of the game (any lottery number and any profile of other players lists), if the student is accepted by an acceptable school when submitting C, she will be either accepted by the best school or that school. She is weakly better off in any realization. And there must exist cases such that she is matched with the first choice in C when submitting C, and she will be matched with the best school when submitted C' instead. Thus, C is dominated by C'. In fact, C' first-order stochastically dominates C.

In the same manner, if the first choice in C is the worst school, C is dominated by C' which is the same as C except the first choice in C' is replaced by the best school.

(ii) Since including an otherwise omitted school always weakly increases the probability of being accepted by that school (Proposition 1), adding an acceptable school to the bottom of a partial list always weakly improves the expected utility.

(iii) Suppose the submitted list of i is $C_i = \left\{c_i^1, ..., c_i^S\right\}$ such that $c_i^K = \widehat{s}$, $1 \leq K < S$, such that $u_{i,\widehat{s}} = \min_{t=1,...,S}\{u_{i,t}\}$. Consider an alternative list, $\widehat{C}_i = \left\{c_i^1, ...c_i^{K-1}, 0, c_i^{K+1}, ..., c_i^S\right\}$, i.e., delete the worst school from the original list and leave that position blank. Given any realization of the game, if the student is accepted by any school of $c_i^1, ...c_i^{K-1}$ and $c_i^{K+1}, ..., c_i^S$ when submitting C, she will be still accepted by that school when submitting C' instead. If the student is accepted by $\widehat{s}$ when submitting C, she is likely to be accepted by one of $c_i^{K+1}, ..., c_i^S$ when submitting C'. This implies:

$$\begin{aligned} P_s^A(C_i, i) &= P_s^A\left(\widehat{C}_i, i\right), s = c_i^1, ...c_i^{K-1}; \\ P_s^A(C_i, i) &\leq P_s^A\left(\widehat{C}_i, i\right), s = c_i^{K+1}, ..., c_i^S; \\ P_{\widehat{s}}^A(C_i, i) &\geq P_{\widehat{s}}^A\left(\widehat{C}_i, i\right). \end{aligned}$$

Thus, the difference in expected utility between two lists is:

$$\begin{aligned} & \sum_{s=1}^{S}\left[P_s^A(C_i, i) - P_s^A\left(\widehat{C}_i, i\right)\right]\max\{u_{i,s}, 0\} \\ = & \sum_{s\in\left\{c_i^{K+1},...,c_i^S\right\}}\left[P_s^A(C_i, i) - P_s^A\left(\widehat{C}_i, i\right)\right]\max\{u_{i,s}, 0\} \\ & + \left[P_{\widehat{s}}^A(C_i, i) - P_{\widehat{s}}^A\left(\widehat{C}_i, i\right)\right]\max\{u_i(\widehat{s}), 0\} \\ \leq & \sum_{s\in\left\{c_i^{K+1},...,c_i^S\right\}}\left[P_s^A(C_i, i) - P_s^A\left(\widehat{C}_i, i\right)\right]\max\{u_i(\widehat{s}), 0\} \\ & + \left[P_{\widehat{s}}^A(C_i, i) - P_{\widehat{s}}^A\left(\widehat{C}_i, i\right)\right]\max\{u_i(\widehat{s}), 0\} \\ = & \ 0 \end{aligned}$$

The last equality uses the condition that

$$\sum_{s\in\left\{c_i^{K+1},...,c_i^S\right\}}\left[P_s^A(C_i, i) - P_s^A\left(\widehat{C}_i, i\right)\right] = -\left[P_{\widehat{s}}^A(C_i, i) - P_{\widehat{s}}^A\left(\widehat{C}_i, i\right)\right].$$

Thus, C is dominated by C'.

(iv) In the same manner as in (iii), we can delete the unacceptable schools in the list. The new list weakly dominants the original list. ■

Proof of Proposition 3.

Denote the total number of possible lists as L, thus $L = 1 + S!\left(\frac{1}{(S-1)!} + \frac{1}{(S-2)!} + ... + \frac{1}{1!}\right)$. Every student has a multinomial distribution over L outcomes. Since at least one of the L outcomes must be chosen, we focus on L outcomes. Fix the order of all the lists, and let $\boldsymbol{Y}_i = (Y_{i,1}, ..., Y_{i,L})' \in \{0,1\}^L$, $Y_{i,l} = 1$, if and only if the lth list is chosen by student i. Thus, $\sum_{l=1}^{L} Y_{i,l} = 1$.

In the following, the distribution of the sum of $\boldsymbol{Y}_i$ is shown to converge to a normal distribution.

Without loss of generality, student 1 is considered in the following. Her perceived probability of other students' choices is a function of her information set – $\overline{\boldsymbol{X}}_1$, $\mathcal{X}^1$ and the distribution of $\overline{\boldsymbol{X}}_i$, $i > 1$. Reorder the students such that $\mathcal{X}^1 = \{\boldsymbol{X}_2, ..., \boldsymbol{X}_{F+1}\}$.

Let $(\pi_{i,1}, ..., \pi_{i,L})$ be student 1's belief about the probability that each list is being chosen by student i. Then $\sum_{l=1}^{L} \pi_{i,l} = 1, \forall i > 1$ and $\forall i = 2, ..., I$,

$$E(\boldsymbol{Y}_i) = (\pi_{i,1}, ..., \pi_{i,L})', Var(\boldsymbol{Y}_i) = \begin{bmatrix} \pi_{i,1}(1-\pi_{i,1}) & ... & \pi_{i,1}\pi_{i,L} \\ \pi_{i,1}\pi_{i,2} & .. & \pi_{i,2}\pi_{i,L} \\ ... & ... & ... \\ \pi_{i,1}\pi_{i,L} & ... & \pi_{i,L}(1-\pi_{i,L}) \end{bmatrix}.$$

$\forall i=2,...,F+1$, the beliefs is a function of the realization of $\mathcal{X}_1$,

$$\left(\pi_{i,1},...,\pi_{i,L}\right)=\left(\pi_{i,1}\left(\mathcal{X}_1\right),...,\pi_{i,L}\left(\mathcal{X}_1\right)\right).$$

Given that $\left\{\overline{\mathbf{X}}_i\right\}_{i=1}^{I}$ are i.i.d. across students, then $\forall i=F+2,...,I$,

$$\left(\pi_{i,1},...,\pi_{i,6}\right)\equiv\left(\overline{\pi}_1,...,\overline{\pi}_6\right),$$

which is not a function of $\mathcal{X}_1$.

Consider a vector of random variables, $\boldsymbol{N}=\left(N_1,N_2,...,N_L\right)'\in\mathbb{N}^L$, which are the numbers of students submitting each list, i.e.,

$$\begin{aligned} N_l &= \sum_{i=2}^{I} Y_{i,l}, l=1,...,L. \\ \sum_{l=1}^{L} N_l &= I-1, \text{ thus } N_l\geq 0 \text{ and } N_l\leq I-1. \end{aligned}$$

The mean and variance of $\boldsymbol{N}$ are

$$\begin{aligned} \boldsymbol{\mu}_1 &= \frac{1}{I-1}\left(\sum_{i=2}^{F+1}\pi_{i,1}+\left(I-F-1\right)\overline{\pi}_1,...,\sum_{i=2}^{F+1}\pi_{i,L}+\left(I-F-1\right)\overline{\pi}_L\right), \\ Q_1 &= \frac{1}{I-1}\begin{bmatrix} \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,1}\left(1-\pi_{i,1}\right)\\+\left(I-F-1\right)\overline{\pi}_1\left(1-\overline{\pi}_1\right)\end{array} & ... & \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,1}\pi_{i,L}\\+\left(I-F-1\right)\overline{\pi}_1\overline{\pi}_L\end{array} \\ \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,1}\pi_{i,2}\\+\left(I-F-1\right)\overline{\pi}_1\overline{\pi}_2\end{array} & .. & \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,2}\pi_{i,L}\\+\left(I-F-1\right)\overline{\pi}_2\overline{\pi}_L\end{array} \\ ... & ... & ... \\ \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,1}\pi_{i,L}\\+\left(I-F-1\right)\overline{\pi}_1\overline{\pi}_L\end{array} & ... & \begin{array}{c}\sum_{i=2}^{F+1}\pi_{i,L}\left(1-\pi_{i,L}\right)\\+\left(I-F-1\right)\overline{\pi}_L\left(1-\overline{\pi}_L\right)\end{array} \end{bmatrix}, \end{aligned}$$

And

$$\lim_{I\to\infty}\boldsymbol{\mu}_1=\left(\overline{\pi}_1,...,\overline{\pi}_6\right)\equiv\mu,$$

$$\lim_{I\to\infty}Q_1=\begin{bmatrix}\overline{\pi}_1\left(1-\overline{\pi}_1\right) & ... & \overline{\pi}_1\overline{\pi}_L \\ \overline{\pi}_1\overline{\pi}_2 & .. & \overline{\pi}_2\overline{\pi}_L \\ ... & ... & ... \\ \overline{\pi}_1\overline{\pi}_L & ... & \overline{\pi}_L\left(1-\overline{\pi}_L\right)\end{bmatrix}\equiv Q.$$

$$\lim_{I\to\infty}\left(IQ_1\right)^{-1}Var\left(\mathbf{Y}_i\right)=\lim_{I\to\infty}\left(\sum_{j=2}^{I}Var\left(Y_j\right)\right)^{-1}Var\left(\mathbf{Y}_i\right)=0,\forall i=2,...,S.$$

By Lindeberg–Feller Central Limit Theorem (See for example, Greene (1999), page 117),

$$\sqrt{I-1}\left(\frac{\boldsymbol{N}}{I-1}-\boldsymbol{\mu}_1\right)\xrightarrow{d}N\left(0,Q\right), as\ I\to\infty.$$

Also notice that

$$\lim_{I\to\infty}\boldsymbol{\mu}_1=\left(\overline{\pi}_1,...,\overline{\pi}_6\right)\equiv\boldsymbol{\mu}.$$

Thus

$$\sqrt{I-1}\left(\frac{\boldsymbol{N}}{I-1}-\boldsymbol{\mu}\right)\xrightarrow{d}N\left(0,Q\right),as\ I\to\infty.$$

By definition, the beliefs are

$$P_s^A\left(C_1,1\right)\equiv\sum_{C_{-i}\in\mathcal{C}^{(I-1)}}\Pr\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)=C_{-1}|\mathcal{X}^1\right)p_s^A\left(C_1,C_{-1}\right),\forall C_i\in\mathcal{C}.$$

where $p_s^A\left(C_1,C_{-1}\right)$ is the probability that student 1 is accepted by school s given $\left(C_1,C_{-1}\right)$.

Given the mechanism, if C_{-1} and C'_{-1} are such that they have the same $\boldsymbol{N}=(n_1,n_2,...,n_L)$, then $p_s^A\left(C_1,C_{-1}\right)=p_s^A\left(C_1,C'_{-1}\right),\forall C_i\in\mathcal{C}$.

For $s=1,...,S$, the beliefs can be rewritten as

$$P_s^A\left(C_1|\overline{\mathbf{X}}_1\right)=\sum_{j=1}^{\overline{N}}\left[Prob\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)\ s.t.\ \boldsymbol{N}=\boldsymbol{n}_j\ |\mathcal{X}^1\right)\sum_{C_{-1}\text{ s.t. }\boldsymbol{N}=\boldsymbol{n}_j}p_s^A\left(C_1,C_{-1}\right)\right],$$

$\overline{N}$ is the total number of different $\boldsymbol{N}_j$,

$$\overline{N}=\left(\begin{array}{c}I-1+L-1\\L\end{array}\right)=\left(\begin{array}{c}I+L-2\\L\end{array}\right).$$

Since $\sqrt{I-1}\left(\frac{\boldsymbol{N}}{I-1}-\boldsymbol{\mu}\right)\xrightarrow{d}N\left(0,Q\right)$, as $I\to\infty$,

$$\begin{aligned}&Prob\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)\ s.t.\ \boldsymbol{N}=\boldsymbol{n}_j|\mathcal{X}^1\right)\\=\ &Prob\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)\ s.t.\ \frac{\boldsymbol{n}_j-\frac{1}{2}}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1}<\left(\frac{\boldsymbol{N}}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1}\right)<\frac{\boldsymbol{n}_j+\frac{1}{2}}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1}|\mathcal{X}^1\right)\\\to\ &\Phi_Q\left(\left(\frac{\boldsymbol{N}}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1}\right)\in B\left(\frac{\boldsymbol{n}_j}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1},\frac{1}{2\sqrt{I-1}}\right)\right),\end{aligned}$$

where Φ_Q is the distribution function for $N\left(0,Q\right)$; $B\left(\frac{\boldsymbol{n}_j}{I-1}-\boldsymbol{\mu},\frac{1}{2}\right)$ is an open ball with center $\left(\frac{\boldsymbol{n}_j}{I-1}-\boldsymbol{\mu}\right)$ and radius $\frac{1}{2\sqrt{I-1}}$. Thus, $\forall j=1,...,\overline{N}$,

$$\Psi_1\left(\boldsymbol{n}_j\right)\equiv Prob\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)\ s.t.\ \mathbf{N}<\boldsymbol{n}_j\ |\mathcal{X}^1\right)\to\Phi_Q\left(\frac{\boldsymbol{n}_j+\frac{1}{2}}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1}\right).$$

where $\Psi_1\left(\boldsymbol{n}\right)$ is the distribution function of $\boldsymbol{N}$.

Further, construct the following function,

$$f\left(\boldsymbol{z}\right)=\sum_{j=1}^{\overline{N}}\mathbf{1}\left(\boldsymbol{z}\in B\left(\frac{\boldsymbol{n}_j}{\sqrt{I-1}}-\boldsymbol{\mu}\sqrt{I-1},\frac{1}{2\sqrt{I-1}}\right)\right)\sum_{C_{-1}\text{ s.t. }\boldsymbol{N}=\boldsymbol{n}_j}p_s^A\left(C_1,C_{-1}\right).$$

$f_1\left(\boldsymbol{z}\right):\mathbb{R}^L\to[0,1]$ is then a real-valued, simple function and thus bounded Borel measurable function whose points of discontinuity form a set of measure zero with respect to $d\Phi_Q$.

That $\sqrt{I-1}\left(\frac{\boldsymbol{N}}{I-1}-\boldsymbol{\mu}\right)$ converges in distribution to $\boldsymbol{N}\left(0,Q\right)$ means (see for example Bhattacharya

(1977)),

$$\begin{aligned}
P_s^A\left(C_1|\overline{\mathbf{X}}_1\right) &= \sum_{j=1}^{\overline{N}}\left[Prob\left(\sigma_{-1}\left(\overline{\mathbf{X}}_{-1},B_{-1}\right)\ s.t.\ \boldsymbol{N}=\boldsymbol{n}_j\ |\mathcal{X}^1\right)\sum_{C_{-1}\text{ s.t. }\boldsymbol{N}=\boldsymbol{n}_j} p_s^A\left(C_1,C_{-1}\right)\right] \\
&= \int_{\mathbb{R}^L} f\left(\boldsymbol{z}\right)d\Psi_1\left(\boldsymbol{z}\right) \\
&\rightarrow \int_{\mathbb{R}^L} f\left(\boldsymbol{z}\right)d\Phi_Q\left(\boldsymbol{z}\right) = P_s^A\left(C_1\right),
\end{aligned}$$

which is independent of student $1'$s information.

Since this can be proved this for any other student, thus beliefs converge to the same one as $I\rightarrow\infty$. ■

Proof of Proposition 4.

The existence can be proved by Brouwer's Fixed Point Theorem.

Fix the strategies, $\sigma_i\left(\overline{\mathbf{X}}_i,B\right)$, as maximizing the (subjective) expected utility. To find an equilibrium is equivalent to find a fixed point of B, such that $\widetilde{B}\left(C,B\right)=B\left(C\right),\forall C\in\mathcal{C}$. Notice that $\widetilde{B}\left(\cdot,B\right):\mathcal{B}_R\rightarrow\mathcal{B}_R$. Since each element in $\mathcal{B}_R$ is L points in a $(S-1)$-simplex $\boldsymbol{\Delta}^{(S-1)}$, $\mathcal{B}_R$ can be enlarged to $\left(\boldsymbol{\Delta}^{(S-1)}\right)^L$ and $\widetilde{B}\left(\cdot,B\right):\left(\boldsymbol{\Delta}^{(S-1)}\right)^L\rightarrow\left(\boldsymbol{\Delta}^{(S-1)}\right)^L$ is well defined. $\left(\boldsymbol{\Delta}^{(S-1)}\right)^L$ is a convex and compact set, to apply Brouwer's Fixed Point Theorem, we only have to show that $\widetilde{B}\left(\cdot,B\right)$ is continuous in B.

Notice from Figure 1, for student 1, $\forall C\in\mathcal{C}$

$$\widetilde{B}\left(C,B\right)=\sum_{m=1}^{M}p_m\left(B\right)b_m\left(C\right), \tag{7}$$

where $p_m\left(B\right)$ is probability of mth profile of C_{-1} is realized; $b_m\left(C\right)$ is probabilities that student 1 is assigned to each school given $\left(C,C_{-1}\right)$. Since $b_m\left(C\right)$ is mechanically determined by the lottery process and is not a function of B, what remains to show is $p_m\left(B\right)$ is continuous in B.

$$p_m\left(B\right)=\prod_{i=2}^{I}Prob\left(C_i=C_i^m|\sigma_i\left(\overline{\mathbf{X}}_i,B\right)\right),$$

where $Prob\left(C_i=C_i^m|\sigma_i\left(\overline{\mathbf{X}}_i,B\right)\right)$ is the probability i chooses C_i^m. Further, without loss of generality, suppose that there is at least one acceptable school, and C_i^m is a full list and is compared to other full lists. The probability C_i^m being chosen is the sum of probabilities of C_i^m being the single best response, being one of the two best responses, and etc.

$$\begin{aligned}
& Prob\left(C_i=C_i^m|\sigma_i\left(\overline{\mathbf{X}}_i,B\right)\right) \\
= & Prob\left(C_i^m\text{ is the unique best response}\right)+\frac{1}{2}Prob\left(C_i^m\text{ and }C_i'\text{ are the only two best responses}\right) \\
& +\frac{1}{3}Prob\left(C_i^m,\ C_i'\text{ and }C_i''\text{ are the only three best responses}\right)+....
\end{aligned}$$

When there are more than one best responses, students may play mixed strategies. Suppose the probability distribution among the mixed strategies is determined exogenously to B. The above specification assumes each best response is equally likely to be chosen.

It suffices to show each of the above probabilities is continuous in B or $P_s^A\left(C_i\right),\forall C_i\in\mathcal{C}^F$, $s=1,...S$.

First consider when there is C_i^m is the unique best response.

$$\begin{aligned}
& Prob\left(C_i^m \text{ is the unique best response}\right) \\
= \quad & Prob\left(\sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)-P_s^A\left(C_i\right)\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right)>0,\forall C_i\in\mathcal{C}^F,C_i\neq C_i^m\right). \quad (8)
\end{aligned}$$

Given $C_i \neq C_i^m \in \mathcal{C}^F$ and any sequence $B^n = \left\{\left\{P_s^A\left(C\right)^n\right\}_{s=1}^S, \forall C\in\mathcal{C}\right\} \rightarrow B$, one can show, by Chebychev's inequality

$$\begin{aligned}
& \sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)^n-P_s^A\left(C_i\right)^n\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right) \\
& \xrightarrow{p}\sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)-P_s^A\left(C_i\right)\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right),
\end{aligned}$$

since $V\left(\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right)\right)<+\infty$. Thus,

$$\begin{aligned}
& \sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)^n-P_s^A\left(C_i\right)^n\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right) \\
& \xrightarrow{d}\sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)-P_s^A\left(C_i\right)\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right).
\end{aligned}$$

Notice that when $\left\{P_s^A\left(C_i^m\right)\right\}_{s=1}^S = \left\{P_s^A\left(C_i\right)\right\}_{s=1}^S$, the limiting distribution is degenerate at 0. Convergence in distribution means

$$\begin{aligned}
& Prob\left(\sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)^n-P_s^A\left(C_i\right)^n\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right)>0,\forall C_i\in\mathcal{C}^F,C_i\neq C_i^m\right) \\
\rightarrow \quad & Prob\left(\sum_{s=1}^{S}\left[P_s^A\left(C_i^m\right)-P_s^A\left(C_i\right)\right]\max\left(u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right),0\right)>0,\forall C_i\in\mathcal{C}^F,C_i\neq C_i^m\right).
\end{aligned}$$

Hence, the probability in equation (8) is then continuous in B.

In the same manner, we can show that the probability is continuous in B when there are more than one best responses. $Prob\left(C_i = C_i^m|\sigma_i\left(\overline{\boldsymbol{X}}_i,B\right)\right)$ and thus $\widetilde{B}\left(C,B\right)$ is continuous in B.

By Brouwer's Fixed Point Theorem, $\widetilde{B}\left(\cdot,B\right):\left(\boldsymbol{\Delta}^{(S-1)}\right)^L\rightarrow\left(\boldsymbol{\Delta}^{(S-1)}\right)^L$ has a fixed point B^*, such that $\widetilde{B}\left(\cdot,B^*\right)=B^*$, and $B^*\in\left(\boldsymbol{\Delta}^{(S-1)}\right)^L$. Next we show $B^*\in\mathcal{B}_R$. This can be seen from the definition of $\widetilde{B}\left(\cdot,B\right)$ in equation (7). Feed in any candidate belief B, $\widetilde{B}\left(C,B\right)$ is a probability weighted sum of $b_m\left(C\right)$ which is the assignment probabilities given a particular profile of $[C,C_{-1}]$. It is shown $b_m\left(\cdot\right)$ is always in $\mathcal{B}_R$ because of the rules of the game. Thus, $\widetilde{B}\left(\cdot,B\right)\in\mathcal{B}_R$. This proves the fixed point B^* is in $\mathcal{B}_R$ and is the equilibrium belief. ■

Proof of Lemma 2.

Since $\boldsymbol{\varepsilon}_i$ is continuously distributed and $u\left(\boldsymbol{z}_s,\boldsymbol{X}_i,\varepsilon_{i,s}\right)$ is strictly monotonic in $\varepsilon_{i,s}$, given any $\boldsymbol{z}_s$, $\boldsymbol{X}_i$ and equilibrium beliefs B^*, the probability of choosing any list is strictly positive. From Figure 1, any profile of lists, C^m_{-i}, is associated with a positive probability. $\forall s$ and C, $P^A_s(C)$ is always in $[0,1]$; and it is strictly positive and less than 1 when some C^m_{-i} is realized. Thus, the equilibrium beliefs are always between zero and one. ■

Proof of Proposition 5.

Assume that every school is acceptable to student i. To show she plays pure strategy almost surely, it suffices to show that her discrete choice problem (1) has a unique solution almost surely, given $\boldsymbol{X}_i$ and $\boldsymbol{\varepsilon}_i$.

If student i plays a mixed strategy, there are at least two choice lists being the solutions of (1) for student i of any type in a set with a positive measure. Suppose $\bar{C}_i$ and $\hat{C}_i$ are solutions given any type $\overline{\mathbf{X}}_i \in \mathcal{E}$, where $\mathcal{E}$ has positive probability measure.

Since the belief system only depends on student's choices, C_i, and doesn't depend on students' own type, the probabilities associated with $\bar{C}_i$ and $\hat{C}_i$ are thus independent of $\boldsymbol{\varepsilon}_i$. Let $(u_i(1),...,u_i(S))'$ be the vector of utilities of student i getting into each school give her type. Then $u_{i,s} > 0, \forall s$. Denote $\left\{P^A_s(C_i)\right\}^S_{s=1}$ the probabilities of being accepted by school s when C_i is submitted. Given the mechanism, we must have $\left\{P^A_s(\bar{C}_i)\right\}^S_{s=1} \neq \left\{P^A_s\left(\hat{C}_i\right)\right\}^S_{s=1}$. Otherwise $\bar{C}_i$ and $\widetilde{C}_i$ must be the same.

Since $\bar{C}_i$ and $\widetilde{C}_i$ are two solutions to the discrete choice problem, we have the following equation:

$$V\left(\bar{C}_i\right) = V\left(\widetilde{C}_i\right),\ \forall \overline{\mathbf{X}}_i \in \mathcal{E}$$

which can be rearranged as:

$$\sum_{s=1}^{S}\left(P^A_s\left(\bar{C}_i\right) - P^A_s\left(\widetilde{C}_i\right)\right)\max\left\{u_{i,s},0\right\} = \sum_{s=1}^{S}\left(P^A_s\left(\bar{C}_i\right) - P^A_s\left(\widetilde{C}_i\right)\right)u_{i,s} = 0, \forall \overline{\mathbf{X}}_i \in \mathcal{E}.$$

Since $\left(P^A_s\left(\bar{C}_i\right) - P^A_s\left(\widetilde{C}_i\right)\right)$ is independent of $\boldsymbol{\varepsilon}_i$, the equations above is a homogenous system of linear equations in $u_{i,s}$. There are at most S types of the student such that $\bar{C}_i$ and $\widetilde{C}_i$ are both solutions. But $\mathcal{E}$ has positive probability measure and there are infinite different $\overline{\mathbf{X}}_i \in \mathcal{E}$. Contradiction. Thus, students must play pure strategy almost surely, when every school is acceptable.

If for student i, $\exists s \neq s'$, such that $u_{i,s} \leq 0$, and $u_{i,s'} \leq 0$. C_i described in the proposition is optimal strategy. But the student is also indifferent between C_i and $\widehat{C}_i$, i.e.,

$$V\left(C_i\right) = V\left(\widehat{C}_i\right).$$

Thus, students' mixed-strategy only mixes C_i with $\widehat{C}_i$. ■

Proof of Proposition 6.

(i) From Proposition 5, as long as there are $\exists s \neq s'$, such that $u_{i,s} \leq 0$, and $u_{i,s'} \leq 0$, we can construct a continuum of mixed strategies which have the same expected utility give any set of beliefs. Fix any one mixed strategy, following the same proof in Proposition 4, we can show a Bayesian Nash equilibrium exists.

(ii) Under the assumption that students never include unacceptable schools in their list, and all the elements in equilibrium beliefs are in $(0,1)$ (Lemma 2), there is only one best response for every student.

What needs to prove is there exists only one set of equilibrium beliefs. Suppose B_1, $B_2 \in \mathcal{B}_R$ are two equilibrium beliefs, and $B_1 \neq B_2$. Thus

$$\widetilde{B}\left(\cdot, B_1\right) = B_1, \widetilde{B}\left(\cdot, B_2\right) = B_2.$$

$\widetilde{B}(\cdot, B)$ is continuous as shown in the proof of Proposition 4. When the distribution of $u(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s})$ is twice differentiable, $\widetilde{B}(\cdot, B)$ is also differentiable in B if B is a convex combination of B_1 and B_2. Thus $B \in \mathcal{B}_R$ and every element of B is in $(0,1)$.

To show this, without loss of generality, consider C_i^m which is a full list, thus all the schools in C_i^m except the last, c_i^S, are acceptable. Let $C\left(c_i^S\right)$ denote all the lists with c_i^S as the last choice. Thus, the probability of C_i^m is preferable to $C_i$is:

$$
\begin{aligned}
& Prob\left(\begin{array}{c} u(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}) > 0, \forall s \neq c_i^S; c_i^S = \min_s\left[u(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s})\right]; \\ \sum_{s=1}^S \left(P_s^A\left(C_i^m\right) - P_s^A\left(\widetilde{C}_i\right)\right) \max\left\{u(\boldsymbol{z}_s, \boldsymbol{X}_i, \varepsilon_{i,s}), 0\right\} > 0, \forall \widetilde{C}_i \in C\left(c_i^S\right) \backslash C_i^m \end{array}\right) \\
= & \int_{b_S(B)}^{a_S(B)} \ldots \int_{b_s(B)}^{a_s(B)} \ldots \int_{b_1(B)}^{a_1(B)} dG\left(u_{i,1}, u_{i,2}, \ldots, u_{i,S}\right),
\end{aligned}
$$

where $G(u_{i,1}, u_{i,2}, ..., u_{i,S})$ is the joint distribution function of utilities which can be derived from the joint distribution of $\boldsymbol{X}_i$ and $\boldsymbol{\varepsilon}_i$. By assumption, $G(u_{i,1}, u_{i,2}, ..., u_{i,S})$ has a continuous density function. $a_s(B)$ and $b_s(B)$ are the upper and lower limit of multidimensional integrals, which are both functions of B. Since they are derived from the linear conditions in the probability, and $a_s(B)$ and $b_s(B)$ are continuous in B. The derivative of the above probability with respect to B then can be obtained by the Leibniz integral rule in the context of multidimensional integrals (see for example Flanders (1973)). Thus $\widetilde{B}(\cdot, B)$ is differentiable in B.

Fix a vector $\boldsymbol{d} \in \mathcal{B}_R$, define a new function $\boldsymbol{d}' \cdot \widetilde{B}(\cdot, B) : \mathcal{B}_R \to \mathbb{R}$. By a version of mean value theorem (see Bartle (1964), p. 239),

$$
\boldsymbol{d}' \cdot (B_1 - B_2) = \boldsymbol{d}' \cdot \left(\widetilde{B}(\cdot, B_1) - \widetilde{B}(\cdot, B_2)\right) = \boldsymbol{d}' \cdot \boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d})) \cdot (B_1 - B_2),
$$

where $\boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d}))$ is the $SL \times SL$ Jacobian matrix of $\widetilde{B}(\cdot, B)$ at $B = B(\boldsymbol{d})$. $B(\boldsymbol{d})$ is a convex combination of B_1 and B_2, and thus $B(\boldsymbol{d}) \in \mathcal{B}_R$, since $\mathcal{B}_R$ is convex. The value of $B(\boldsymbol{d})$ depends on $\boldsymbol{d}$.

Rearrange the equation, we have

$$
\boldsymbol{d}' \cdot \left(I - \boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d}))\right) \cdot (B_1 - B_2) = 0, \forall \boldsymbol{d} \in \mathcal{B}_R. \tag{9}
$$

Generally, $\boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d}))$ is not an identical matrix. Given any list, C, when increasing the probability of being accepted by s which is the first choice in C, it will be more likely for students to submit C and thus $P_s^A(C)$ decreases. Thus the diagonal of $\boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d}))$ contains negative terms. Besides, the off-diagonal elements are usually not zero. $\boldsymbol{D}\widetilde{B}(\cdot, B(\boldsymbol{d}))$ is also changes when $\boldsymbol{d}$ changes. Hence, for equation (9) being satisfied $\forall \boldsymbol{d} \in \mathcal{B}_R$, it must be $B_1 = B_2$. This proves the uniqueness of the pure strategy equilibrium.

(iii) Suppose $\left(\sigma^*\left(\overline{\boldsymbol{X}}_i, B^*\right), B^*\right)$ is the pure-strategy equilibrium and $\left(\sigma^{**}\left(\overline{\boldsymbol{X}}_i, B^{**}\right), B^{**}\right)$ is a mixed-strategy equilibrium. Then,

$$
\begin{aligned}
\sigma^*\left(\overline{\boldsymbol{X}}_i, B^*\right) &= \arg\max_{\widehat{\sigma}_i\left(\overline{\boldsymbol{X}}_i, B^*\right)} \left\{V\left(\widehat{\sigma}_i\left(\overline{\boldsymbol{X}}_i, B^*\right), B^*\right)\right\} \\
\text{and } \sigma^{**}\left(\overline{\boldsymbol{X}}_i, B^{**}\right) &\in \arg\max_{\widehat{\sigma}_i\left(\overline{\boldsymbol{X}}_i, B^{**}\right)} \left\{V\left(\widehat{\sigma}_i\left(\overline{\boldsymbol{X}}_i, B^{**}\right)\right), B^{**}\right\}.
\end{aligned}
$$

Fix $\overline{\boldsymbol{X}}_i$ for student i, and suppose $\sigma^*\left(\overline{\boldsymbol{X}}_i\right) = C^*$ and $C^{**} \in \sigma^{**}\left(\overline{\boldsymbol{X}}_i, B^{**}\right)$ is one of the strategy played with positive probability in equilibrium. Then,

$$V\left(C^{*},B^{*}\right) = \sum_{s=1}^{S} P_{s}^{A}\left(C^{*}\right)^{*}\max\left\{u_{i,s},0\right\} \geq V\left(C,B^{*}\right), \forall C\in\mathcal{C};$$
$$V\left(C^{**},B^{**}\right) = \sum_{s=1}^{S} P_{s}^{A}\left(C^{**}\right)^{**}\max\left\{u_{i,s},0\right\} \geq V\left(C,B^{**}\right), \forall C\in\mathcal{C}.$$

Since B^{**} considers other students submitting their unacceptable schools, we have

$$P_{s}^{A}\left(C\right)^{*} \geq P_{s}^{A}\left(C\right)^{**}, \forall s\in C, \forall C\in\mathcal{C},$$

with strict inequality for some s in some C. Hence,

$$V\left(C^{*},B^{*}\right) \geq V\left(C^{**},B^{*}\right) = \sum_{s=1}^{S} P_{s}^{A}\left(C^{**}\right)^{*}\max\left\{u_{i,s},0\right\} \geq V\left(C^{**},B^{**}\right),$$

with strict inequalities for some i.

In other words,

$$\max_{\widehat{\sigma}_{i}\left(\overline{\boldsymbol{X}}_{i},B^{*}\right)}\left\{V\left(\widehat{\sigma}_{i}\left(\overline{\boldsymbol{X}}_{i},B^{*}\right),B^{*}\right)\right\} \geq \arg\max_{\widehat{\sigma}_{i}\left(\overline{\boldsymbol{X}}_{i},B^{**}\right)}\left\{V\left(\widehat{\sigma}_{i}\left(\overline{\boldsymbol{X}}_{i},B^{**}\right)\right),B^{**}\right\}, \forall i,$$

with strict inequality for some i. This proves the *ex ante* Pareto dominance. ■

A.2 Simulating the Implied Probabilities

This section describes how to find equilibrium beliefs when the parameters in the utility function are given. The following procedure is used in the estimation of the Bayesian Nash equilibrium and finding the equilibrium in the Monte Carlo experiment.

Since everyone has the same beliefs, they should also have the same implied probabilities. It is suffice to just look at Student 1's probabilities of being admitted by the schools in her list.

The simulation of the implied probabilities has seven steps as following:

1. Draw 90% of NC (= 50,000) profiles of choice lists, $\left\{C_i = \left(c_i^1, ..., c_i^S\right)\right\}_{i=2}^{I}$,[54] from the distribution of observed lists[55]. Given any profile, $\left\{C_i = \left(c_i^1, c_i^2, c_i^3, c_i^4\right)\right\}_{i=2}^{960}$, student 1 tries all ($S$!) full choice lists. Combine them together, we get $S! \times NC$ profiles of $\{C_i\}_{i=1}^{I}$. Other 5% of the draws are "extreme" cases where 300 students are only assigned with the same first choice. For the rest 5% draws, students are assigned with each list with the same probability.

2. Given any one profile of lists, $\left\{C_1, \{C_i\}_{i=2}^{I}\right\}$, draw a random lottery number for each student, $L_{I\times 1}$, and then run the admission process to see which school admits student 1, i.e., get the values for the following indicator functions:

$$\mathbf{1}\left(Student\ 1\ assigned\ to\ s|C_1, L, \{C_i\}_{i=2}^{I}\right), s = 1, ..., S.$$

3. Repeat Step 2 NL times with different lottery number draws and calculate the probabilities of Student 1 being admitted by every s respectively.

$$\begin{aligned}&\widetilde{\Pr}\left(Student\ 1\ assigned\ to\ s|C_1, L, \{C_i\}_{i=2}^{I}\right)\\ =\ &\frac{1}{NL}\sum_{l=1}^{NL}\mathbf{1}\left(Student\ 1\ assigned\ to\ s|C_1, L^{(l)}, \{C_i\}_{i=2}^{I}\right), s = 1, ...S.\end{aligned}$$

4. Repeat Steps 2 and 3 for all S! profiles lists with $\{C_i\}_{i=2}^{I}$ fixed and Student 1 selecting each of all S! choice lists.

   These four steps are independent of the belief system and the error terms in the utility functions. Thus they are only simulated once.

5. Simulate the probability of choosing each list by logit-smoothed accept-reject simulator.
   Given the utility functions, simulate R draws of $\{\eta_i\}_{i=2}^{I}$. Given the candidate belief, $\boldsymbol{B}$, the simulated probability of student i choosing full list C_i is

$$\widetilde{P}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right) = \frac{1}{R}\sum_{r=1}^{R} p^{(r)}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right),$$

   where $p^{(r)}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)$ is probability of C_i being a best response which specified in equation (4). If C_i is a partial list, $\widetilde{P}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right)$ is similarly simulated. When considering

[54] c_i^s can be the outside option. If $c_i^s = 0$, then $c_i^t = 0, \forall s < t \leq S$.

[55] There are 9 lists not observed. They are added into the distribution of observed lists as 9 observations with different lists.

an equilibrium in mixed strategies, $p^{(r)}\left(C_i|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^S; \boldsymbol{\theta}\right)$ is then weighted by the corresponding probability which is estimated based on the observed outcomes.

6. Calculate the probability of the profiles $\left\{\left\{C_i^{(t)}\right\}_{i=2}^S\right\}_{t=1}^{NC}$ simulated in Step 1 being realized, i.e., if $\left\{C_i^{(t)}\right\}_{i=2}^I = \left(C_2^{(t)}, C_3^{(t)}, ..., C_I^{(t)}\right)$, then

$$\widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^I \; realized\right) = \frac{1}{K}\prod_{i=2}^I \widetilde{\Pr}\left(C_i^{(t)} \; chosen\right),$$

where K is a normalization term,

$$K = \sum_{t=1}^{NC} \widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^I \; realized\right).$$

7. Calculate the implied probability of Student 1 being admitted by school s as follows, $\forall s = 1, ..., S$:

$$\begin{aligned}
&\widetilde{\Pr}\left(Student\ 1\ assigned\ to\ s|C\right) \\
= \; & \frac{1}{NC}\sum_{t=1}^{NC}\widetilde{\Pr}\left(Student\ 1\ assigned\ to\ s|C, \left\{C_i^{(t)}\right\}_{i=2}^I\right) \times \widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^I \; realized\right) \\
= \; & \frac{1}{NC}\frac{1}{K}\sum_{t=1}^{NC}\widetilde{\Pr}\left(Student\ 1\ assigned\ to\ s|C, \left\{C_i^{(t)}\right\}_{i=2}^I\right) \times \prod_{i=2}^I \widetilde{\Pr}\left(C_i^{(t)} \; chosen\right) \\
= \; & \frac{1}{NC}\frac{1}{K}\sum_{t=1}^{NC}\left\{\begin{array}{c} \frac{1}{NL}\sum_{l=1}^{NL}\mathbf{1}\left(Student\ 1\ assigned\ to\ s|C, L^{(l)}, \left\{C_i^{(t)}\right\}_{i=2}^I\right) \\ \times\prod_{i=2}^I\left[\frac{1}{NU}\sum_{s=1}^{NU}\frac{\exp\left(V\left(C_i^{(t)}|B,\eta_i^{(s)}\right)/\lambda\right)}{\sum_{C_i'\in\mathcal{C}}\exp\left(V\left(C_i'|B,\eta_i^{(s)}\right)/\lambda\right)}\right]\end{array}\right\}.
\end{aligned}$$

This is calculated for all $S!$ possible C. All the probabilities together are the simulated implied probabilities, $\widehat{B}(\cdot, B)$.

A.3 Additional Tables

Table A-1: Correlation between the Sophistication Measures and Survey Responses: Univariate Regression

	Accuracy of beliefs $-\left\|\widehat{B}_i - \widehat{B}_0\right\|$ (1)	Played truthfully $p_i^{TT}(C_i)$ (2)	Best responded $p_i^{BR}(C_i)$ (3)	(3)-(2) (4)	Non-truth-telling BR $p_i^{BR\neq TT}(C_i)$ (5)
Schools' reputation and past performance	-0.030 (0.025)	0.018 (0.012)	0.018* (0.009)	0.001 (0.016)	0.009 (0.010)
Quality of Schools' neighborhood	0.026 (0.020)	-0.007 (0.010)	0.008 (0.008)	0.015 (0.013)	0.019** (0.008)
Teachers' quality	-0.018 (0.030)	0.025* (0.015)	0.031*** (0.012)	0.006 (0.019)	0.010 (0.012)
School's equipment	-0.008 (0.019)	0.005 (0.009)	0.017** (0.007)	0.012 (0.012)	0.017** (0.008)
Possible peers	-0.001 (0.018)	0.004 (0.009)	0.015** (0.007)	0.011 (0.012)	0.017** (0.007)
School's atmosphere	-0.002 (0.029)	0.032** (0.014)	0.027** (0.011)	-0.005 (0.019)	0.016 (0.012)
Distance to the school	0.022 (0.014)	-0.008 (0.007)	0.004 (0.005)	0.012 (0.009)	0.003 (0.006)
Parents know someone at that school	0.000 (0.012)	0.000 (0.006)	-0.001 (0.005)	0.000 (0.008)	0.000 (0.005)
Child's own ability	-0.006 (0.021)	0.008 (0.010)	0.006 (0.008)	-0.001 (0.013)	-0.002 (0.009)
Other parents' submitted lists	0.028** (0.013)	-0.022*** (0.006)	-0.007 (0.005)	0.015* (0.008)	0.002 (0.005)
Enrollment quota & prob. of being accepted	0.009 (0.018)	-0.006 (0.008)	0.012* (0.007)	0.018 (0.011)	0.016** (0.007)
Reducing the prob. of going to a bad school	0.026 (0.022)	-0.004 (0.010)	0.023*** (0.008)	0.027** (0.013)	0.028*** (0.008)

Notes: All the variables are from survey response. Each question asks how much attention the parent pays to the stated factor (on a scale of 1-5, 1=not at all). Each cell reports a univariate regression of a constructed measure on the survey responses of each question (a constant is included).
*, **, *** denote significance at 10%, 5%, and 1% level, respectively. Standard errors in parentheses.
All measures are evaluated ex post, or given the list they submitted, C_i. The sample sizes are 914, except for the Accuracy of Beliefs (sample size 663 – those who submitted a full list).

Table A-2: Correlation between the Sophistication Measures and Survey Responses: Univariate Regression

	Expected Utility Achieved[a] $EU_i(C_i)$ (1)	Var. of Utility Achieved[a] $Var_i(C_i)$ (2)	Truth-Telling= Best Response[b] $p_i^{TT=BR}$ (3)	EU Achieved by Truth-Telling[b,c] EU_i^{TT} (4)	Var Achieved by Truth-Telling[b,c] Var_i^{TT} (5)
Schools' reputation and past performance	0.009 (0.020)	0.084** (0.036)	0.010** (0.005)	-0.002** (0.001)	0.012 (0.007)
Quality of Schools' neighborhood	0.045*** (0.016)	0.001 (0.029)	-0.008** (0.004)	0.000 (0.001)	-0.002 (0.006)
Teachers' quality	-0.009 (0.024)	0.038 (0.044)	0.015*** (0.006)	-0.003** (0.001)	0.005 (0.009)
School's equipment	0.016 (0.015)	0.010 (0.028)	0.007* (0.004)	-0.001 (0.001)	-0.004 (0.006)
Possible peers	0.021 (0.014)	0.022 (0.026)	-0.002 (0.003)	-0.001 (0.001)	0.001 (0.005)
School's atmosphere	0.012 (0.024)	0.078* (0.042)	0.009 (0.006)	-0.003** (0.001)	0.007 (0.009)
Distance to the school	-0.004 (0.012)	-0.013 (0.021)	-0.005** (0.003)	0.000 (0.001)	0.003 (0.004)
Parents know someone at that school	-0.005 (0.010)	0.019 (0.018)	-0.003 (0.002)	-0.001 0.000	0.000 (0.004)
Child's own ability	-0.017 (0.017)	-0.001 (0.030)	0.007* (0.004)	-0.001 (0.001)	-0.005 (0.006)
Other parents' submitted lists	0.007 (0.010)	-0.031* (0.018)	-0.009*** (0.002)	0.000 0.000	0.000 (0.004)
Enrollment quota & prob. of being accepted	0.024* (0.014)	0.014 (0.026)	-0.002 (0.003)	-0.001* (0.001)	0.008 (0.005)
Reducing the prob. of going to a bad school	0.031* (0.016)	0.002 (0.029)	0.001 (0.004)	-0.002*** (0.001)	0.005 (0.006)

Notes: All the variables are from survey response. Each question asks how much attention the parent pays to the stated factor (on a scale of 1-5, 1=not at all). Each cell reports a univariate regression of a constructed measure on the survey responses of each question (a constant is included). Sample size = 914.
*, **, *** denote significance at 10%, 5%, and 1% level, respectively. Standard errors in parentheses.
a. Measured ex post given the submitted list and relative to the list which maximizes the expected utility.
b. Measured ex ante – for each realization of error terms, calculate the truth-telling and best response.
c. Measured relative to always best responding.

Given $\sigma_i\left(\overline{\mathbf{X}}_i, B\right)$, start with beliefs $B \in \mathcal{B}_R, \forall C \in \mathcal{C}$

$\Downarrow$

	Possible Profiles: $\left[C, C_{-1}^m\right]$	Prob(accepted by each school $\left|C_{-1}^m\right.$)	Prob(C_{-1}^m Chosen)
(1)	$\left[C, C_2^1, C_3^1, ..., C_I^1\right]$	$b_1\left(C, C_{-1}^1\right)$	$p_1(B)$
(2)	$\left[C, C_2^2, C_3^2, ..., C_I^2\right]$	$b_2\left(C, C_{-1}^2\right)$	$p_2(B)$
...	...	...	...
(M)	$\left[C, C_2^M, C_3^M, ..., C_I^M\right]$	$b_M\left(C, C_{-1}^M\right)$	$p_M(B)$

$\Downarrow$

Implied Probabilities: $\widetilde{B}(C, B) = \sum_{m=1}^{M} p_m(B)\, b_m(C), \forall C \in \mathcal{C}$

Figure 1: **Mapping from Beliefs to the Implied Probabilities for Student 1**

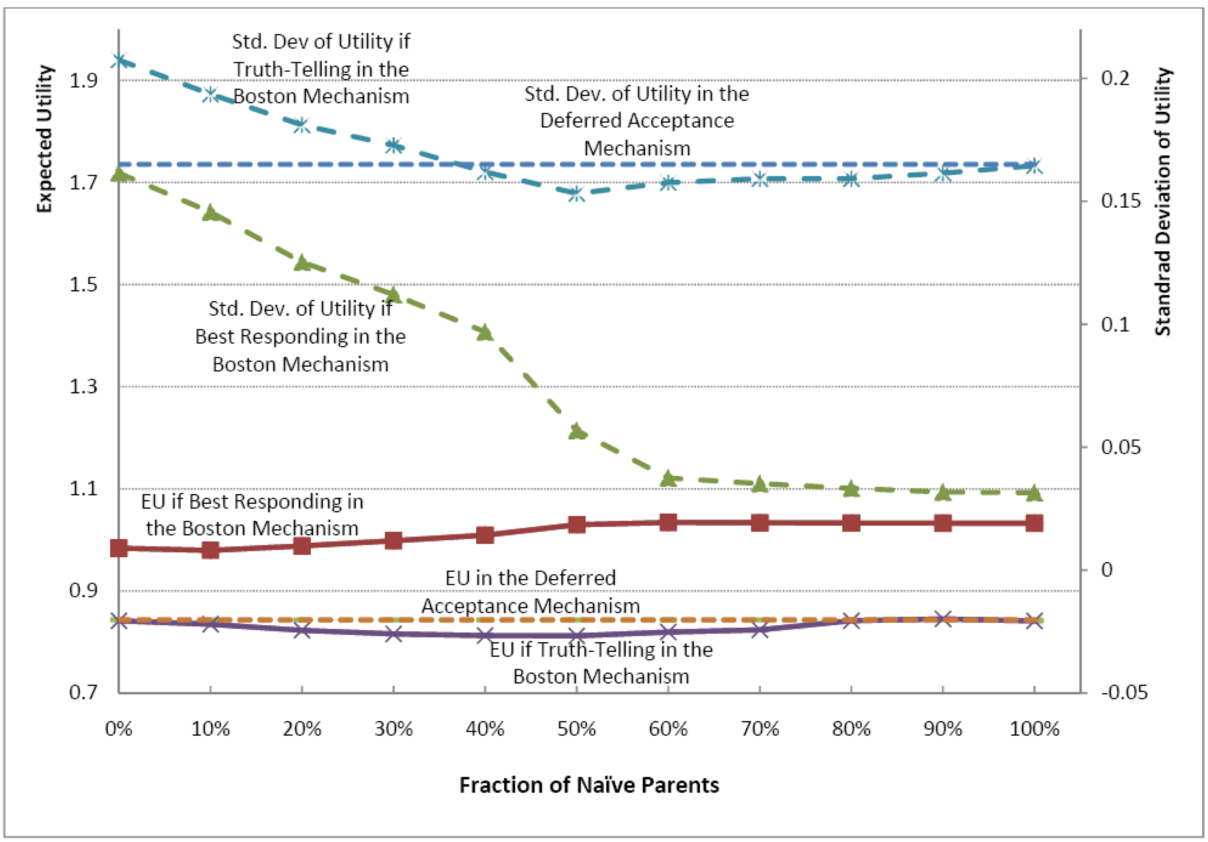

Figure 2: **The Boston Mechanism vs the Deferred Acceptance Mechanism**
The graph reports the expected utility and the standard deviation of utility for a randomly chosen parent who always plays a best response or always reports truthfully in the Boston Mechanism. The fraction of naive parents is the percentage of other parents who are always truth-telling. In the deferred acceptance mechanism, truth-telling is optimal, and the outcome is independent of the fraction of naive parents.

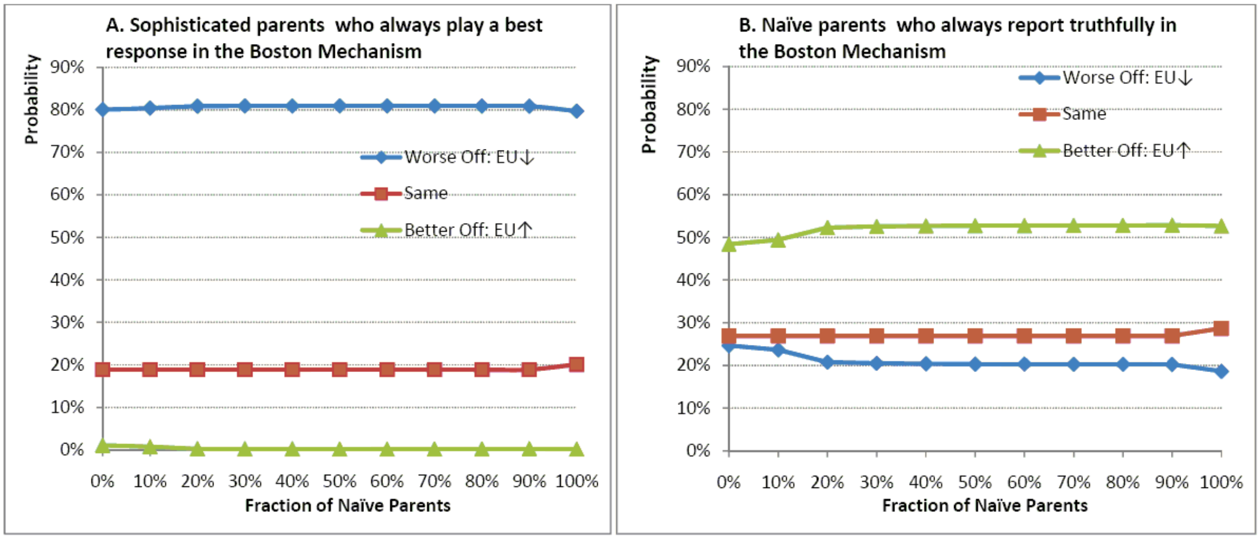

Figure 3: **Welfare Effects of Replacing the Boston Mechanism with the Deferred Acceptance Mechanism**
The graphs report the probabilities of a randomly selected and sophisticated/naive parent being better off, worse off, or having no change in her expected utility. The fraction of naive parents is the percentage of other parents who are always truth-telling. In the deferred acceptance mechanism, truth-telling is optimal, and the outcome is independent of the fraction of naive parents.

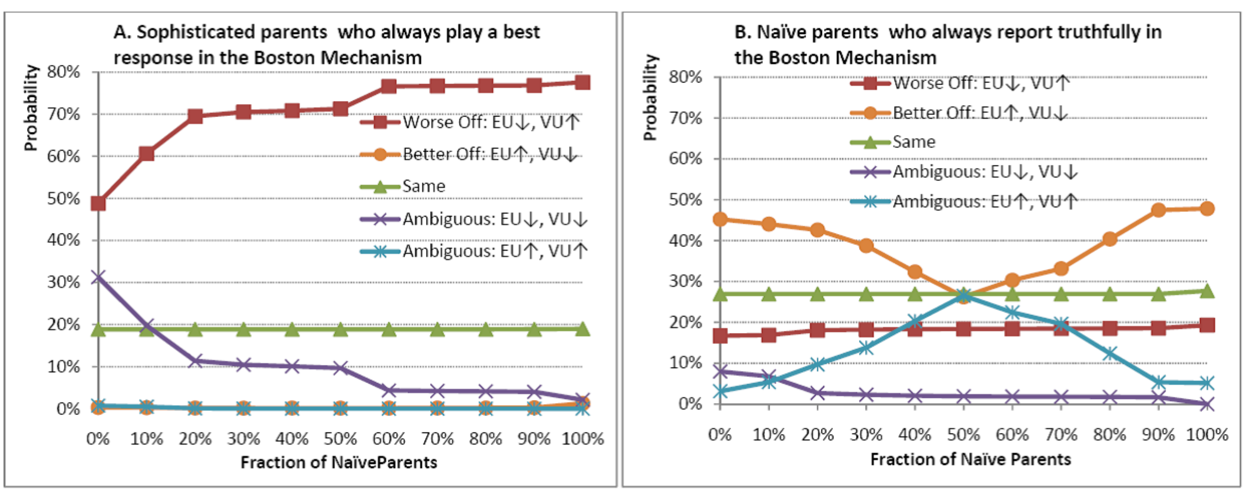

Figure 4: **Welfare Effects of Replacing the Boston Mechanism with the Deferred Acceptance Mechanism**
The graphs report probabilities of a randomly selected, sophisticated/naive, and risk averse parent being better off, worse off, having no change, or having an ambiguous change in her welfare measured by expected utility (EU) and variance of utility (VU). The fraction of naive parents is the percentage of other parents who are always truth-telling. In the deferred acceptance mechanism, truth-telling is optimal, and the outcome is independent of the fraction of naive parents.

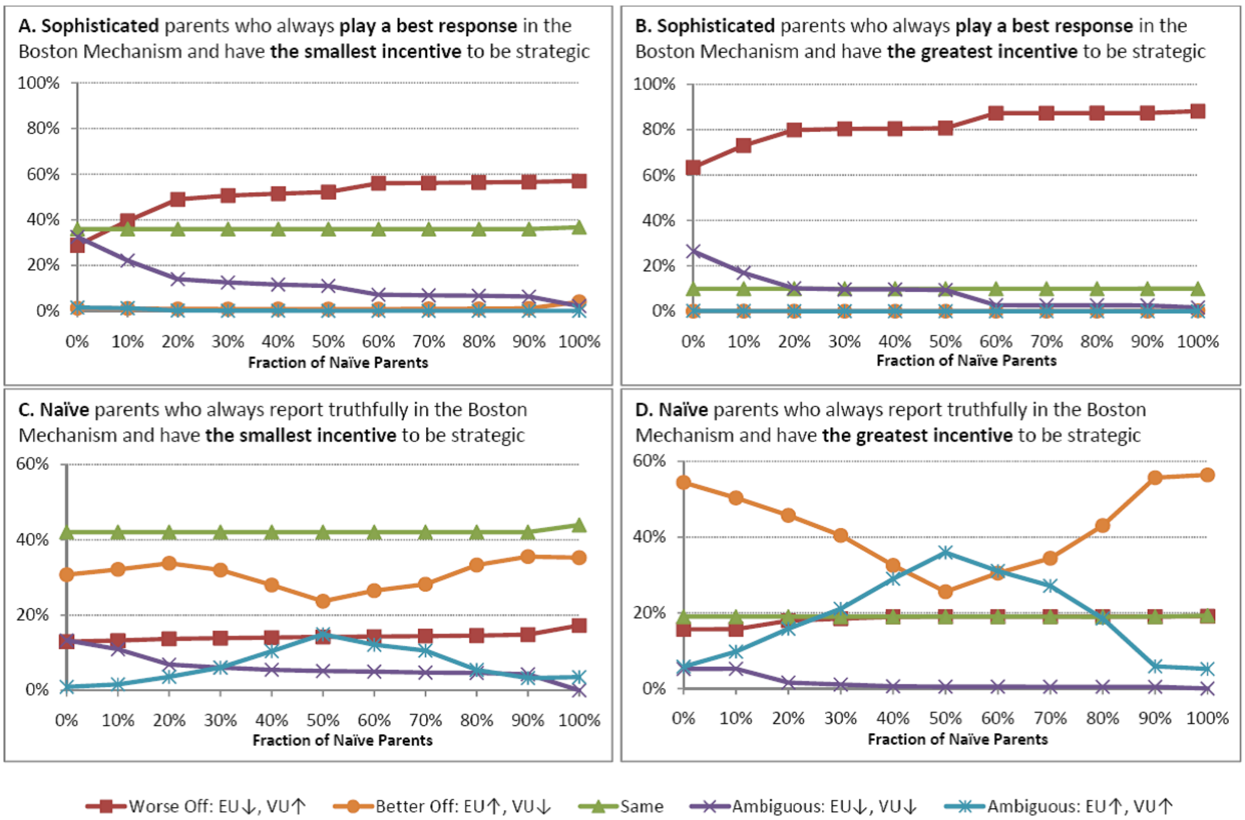

Figure 5: **Welfare Effects of Replacing the Boston Mechanism with the Deferred Acceptance Mechanism by Levels of the Incentives to Be Strategic**

The graphs report the probabilities of a sophisticated/naive and risk averse parent who is randomly selected within a certain group being better off, worse off, having no change, or having an ambiguous change in her welfare measured by expected utility (EU) and variance of utility (VU). The incentives to be strategic is measured by the probability that truth-telling is a best response (See Table 7). The parents with the smallest incentive are the 96 parents (10%) who have the highest probability that truth-telling is a best response. The parents with the greatest incentive are the 96 parents (10%) who have the lowest probability that truth-telling is a best response. The fraction of naive parents is the percentage of other parents who are always truth-telling. In the deferred acceptance mechanism, truth-telling is optimal, and the outcome is independent of the fraction of naive parents.

Table 1: Schools: Quota and Quality

Schools	Quota	Average Test Score[a]	Ranking[b]
1	227	522.91	7
2	310	508.47	14
3	63	559.27	1
4	360	470.13	28
Total	960		

a. Out of 600. Average test score of the graduating class in high school entrance exam 1999.

b. Ranking of average test score among all 28 schools in the district.

Table 2: Distribution of Submitted Lists: Observed vs. Imputed

	Observed		Imputed	
List	Number	Percentage	Number	Percentage
Not Participating	181	19.8	200	20.8
4 Schools	561	61.4	581	60.5
3 Schools	102	11.2	106	11.0
2 Schools	60	6.6	63	6.6
1 School	10	1.1	10	1.0
Total	914		960	

Notes: From administrative data. Observed sample is calculated from students' actual application. Imputed sample includes all the actual applications observed and 46 imputed students.

Table 3: Summary Statistics

	Mean	Std. Dev	Definition	Source
Test Score	183.20	13.52	Elementary Chinese + Math	Administrative data
Awards	0.74	1.08	District level awards in elementary school	Survey in 2002
Income	3610.30	3446.49	Parents' income Yuan/month in 2002	Survey in 2002
Parents' education	13.43	2.23	Parents' average years of education	Survey in 2002
Distance 1	2.22	2.29	Walking distance to school 1, Km in 1999	Administrative data
Distance 2	2.52	2.21	Walking distance to school 2, Km in 1999	Administrative data
Distance 3	2.32	2.27	Walking distance to school 3, Km in 1999	Administrative data
Distance 4	2.42	2.20	Walking distance to school 4, Km in 1999	Administrative data

Notes: The number of observations is 960.

Table 4: Students' First Choice and Claimed Best School

School	Quota	# Stud. Listing it #1		# Stud Claiming it as the Best [a]		List the Claimed Best as[b] #1	#2	#3	#4
1	227	431	(47%)	58	(8%)	49	5	0	0
2	310	66	(7%)	26	(4%)	11	9	4	1
3	63	228	(25%)	579	(83%)	186	107	163	36
4	360	8	(1%)	3	(0%)	0	1	0	1
Null		181	(20%)						
Other				33	(5%)				
Total	960	914		696					

a. Source: Responses to a 2002 survey question: "Among schools to which you could apply, which school was the best?"
b. Among all the students claim a school as the best school, these four columns show how they rank it in the application.

Table 5: Ex Post Beliefs and Students' Self-Reported Beliefs

	Listed as #1				Listed as #2			
	Ex post	Survey Responses[b]			Ex post	Survey Responses[b]		
School	Beliefs[a]	Mean	Std. Dev	# Response	Beliefs[a]	Mean	Std. Dev	# Response
1	50.7%	6.72	2.39	290	0	5.13	2.52	189
2	1	8.11	2.05	82	1	6.53	2.23	206
3	26.7%	4.35	2.93	249	0	3.00	2.24	112
4	1	8.32	2.06	22	1	7.63	2.52	40

a. Calculated from the actual submitted lists. Each entry means when the corresponding school is ranked as first or second choice, what is the probability being accepted by that school, conditional on all other students' actual submitted lists.
b. Responses to the survey question: "On a scale of 1-10, what is the likelihood of being accepted by your first choice? Second choice?" (1 = 10%; 10 = 100%). The first and second choice schools in this question are self-reported. Thus, the numbers of responses are not the same as the numbers of the actual submitted lists.

Table 6: Preferences over Schools: Estimation Results for Different Cases

Variable	**Case 1** Understand the Rules	**Case 2** Case 1 + 2nd Choice>0	**Bayesian N. Equ.** Case 2 + All Sophisticated	**Case 3** Case 2 + 3rd Choice>0	**Bayesian N. Equ.** Case 3 + All Sophisticated
FE_1	0.923	0.976	0.983	0.981	0.911
	(0.055)	(0.054)	(0.084)	(0.055)	(0.053)
FE_2	0.809	0.856	0.745	0.915	0.703
	(0.081)	(0.064)	(0.034)	(0.091)	(0.048)
FE_3	1.252	1.167	1.132	1.157	0.855
	(0.165)	(0.144)	(0.512)	(0.152)	(0.255)
FE_4	0.669	0.682	0.550	0.752	0.556
	(0.073)	(0.068)	(0.035)	(0.081)	(0.074)
$TScore_i \times \overline{Score}99_s$	1.653	1.969	4.483	2.197	3.142
	(0.716)	(0.748)	(2.125)	(0.923)	(1.351)
$Distance_{is}$	-0.055	-0.063	-0.095	-0.054	-0.074
	(0.015)	(0.019)	(0.064)	(0.016)	(0.031)
$Income_i$	-0.171	-0.156	-0.158	-0.146	-0.084
	(0.048)	(0.047)	(0.063)	(0.042)	(0.034)
$\#Awards_i$	-0.397	-0.378	-0.387	-0.404	-0.300
	(0.066)	(0.059)	(0.085)	(0.066)	(0.109)
$TestScore_i$	-2.810	-3.061	-3.397	-2.919	-0.697
	(0.826)	(0.901)	(0.565)	(0.811)	(0.328)
Small-Hsiao Test Statistics $\chi^2\,(18)$	Baseline	Against Case 1 16.017 (P=0.60)	Against Case 2 35.657 (P<0.01)	Against Case 1 28.426 (P=0.06) Against Case 2 23.055 (P=0.19)	Against Case 3 43.164 (P<0.01)

Note: The above coefficients are parameters in the utility function:
$u_{is} = \alpha_s + \delta_1 TScore_i \times \overline{TScore}99_s + \delta_2 Distance_{is} + \gamma_1 Income_i + \gamma_2 \#Awards_i + \gamma_3 TScore_i + \varepsilon_{is}$,
where α_s is school fixed effect.
Cases 1-3 are estimated by simulated maximum likelihood. The two cases of Bayesian Nash equilibrium is estimated by simulated maximum likelihood with equilibrium constraints – the beliefs are consistent. An adaptive penalty function approach is used to estimate the constrained problem. Each case also estimates the covariance matrix of the error terms which includes 9 more parameters but not reported here. Small-Hsiao test is to determine if the nested case is consistent (same idea as in Hausman test). All tests are based on all 18 parameters. Standard errors are in parentheses.

Table 7: Measures of Sophistication and Incentives to Be Strategic: Summary Statistics

Measure and Definition	Obs.	Median	Mean	Std. Dev.	Min	Max
$-\left\|\widehat{B}_i - \widehat{B}_0\right\|$: Accuracy of the beliefs	663	-0.680	-0.866	0.395	-1.780	-0.373
$p_i^{TT}(C_i)$: Prob. C_i is the true preference ranking	914	0.039	0.312	0.402	0.000	1.000
– among those who participate[a]	733	0.010	0.142	0.235	0.000	0.867
$p_i^{BR}(C_i)$: Prob. C_i is a best response	914	0.393	0.473	0.309	0.024	1.000
– among those who participate[a]	733	0.317	0.343	0.183	0.024	0.873
$p_i^{BR}(C_i) - p_i^{TT}(C_i)$	914	0.169	0.161	0.295	-0.664	0.808
– among those who participate[a]	733	0.240	0.201	0.317	-0.664	0.808
$p_i^{BR\neq TT}(C_i)$: Prob. C_i is a non-truth-telling best response	914	0.202	0.245	0.190	0.000	0.808
– among those who participate[a]	733	0.258	0.305	0.164	0.021	0.808
$EU_i(C_i)$: (Achieved expected utility, EU(C_i))/(max(EU(C)))	914	0.972	0.955	0.047	0.750	1.000
– among those who participate[a]	733	0.947	0.944	0.046	0.750	1.000
$Var_i(C_i)$: (Var(U|C_i))/(Var(U|C) s.t C=argmax(E(U|C))	914	1.000	0.928	0.557	0.000	3.353
– among those who participate[a]	733	0.879	0.910	0.621	0.000	3.353
$p_i^{TT=BR}$: Prob(truth-telling is a best response), ex ante[b]	914	0.414	0.429	0.086	0.244	0.885
EU_i^{TT}: (E(U) if truth-telling)/(E(U) if best responding)[b]	914	0.943	0.945	0.017	0.894	0.998
Var_i^{TT}: (Var(U) if truth-telling)/(Var(U) if best responding)[c]	914	1.870	1.864	0.136	1.311	2.521

Notes: All the measures are calculated based on the estimated utility function.

a. A parent participates in the Boston mechanism if she submits a partial list or a full list, i.e. her list includes at least one school which is in the specified choice set.

b. Ex ante measure of incentive to be strategic. Higher value means less incentive to be strategic.

c. Ex ante measure of incentive to be strategic. Lower value means less incentive to be strategic.

Table 8: Determinants of Incentive to Be Strategic: Regression Analysis

	$p_i^{TT=BR}$	EU_i^{TT}		Var_i^{TT}	
ln(income)	0.034***	-0.005***	-0.005***	-0.003	-0.033***
	(0.003)	(0.001)	(0.000)	(0.005)	(0.004)
Parents' Edu	0.007***	-0.000	-0.000	0.002	0.001
	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)
ln(TestScore)	-0.078***	-0.086***	-0.055***	0.390***	-0.131***
	(0.024)	(0.005)	(0.004)	(0.048)	(0.039)
# of Awards	0.041***	-0.001***	-0.001**	0.008*	0.000
	(0.002)	(0.000)	(0.000)	(0.004)	(0.003)
Var_i^{TT}			-0.078***		
			(0.003)		
EU_i^{TT}					-6.083***
					(0.215)
Obs.	907	907	907	907	907
R-Squared	0.56	0.46	0.71	0.32	0.64

Notes: Other controls are dummies for student's gender and elementary school. Definition of incentive measures can be found in Table 7.

$p_i^{TT=BR}$ and EU_i^{TT} are negatively correlated with the incentive to be strategic. Var_i^{TT} is positively correlated with the incentive.

*, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses.

Table 9: Attendtion on School Quality and Uncertainty: Summary Statistics

	Attention	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Full sample	$Attn\_U$	862	3.825	0.694	1	5
	$Attn\_Q$	862	4.096	0.460	1.889	5
Non-participants	$Attn\_U$	166	3.743	0.781	1	5
	$Attn\_Q$	166	4.080	0.461	2.889	5
Participants: All	$Attn\_U$	696	3.844	0.670	1	5
	$Attn\_Q$	696	4.100	0.460	1.889	5
Participants: Partial List	$Attn\_U$	66	3.783	0.763	1	5
	$Attn\_Q$	66	4.160	0.464	3.111	5
Participants: Full List	$Attn\_U$	630	3.851	0.660	1	5
	$Attn\_Q$	630	4.094	0.460	1.889	5

Notes: $Attn\_U$ is how much attention the parent pays to uncertainty. $Attn\_Q$ is how much attention the parent pays to school quality. Both of them are calculated from survey responses. Higher value means more attention. There are 52 parents who did not respond. A parent is a non-participant if she didn't submit any list. A partial list is a list that has less than 3 schools included. A full list has 3 or 4 schools listed.

Table 10: Who Strategizes Better: Regression Analysis of Accurate Beliefs

	Accuracy of beliefs: $\left(-\left\|\widehat{B}_i-\widehat{B}_0\right\|\right)$						
$p_i^{TT=BR}$	-1.418***	-1.453***	-1.173***				
	(0.306)	(0.316)	(0.252)				
EU_i^{TT}	3.258*	3.450*		-1.433	-2.208*		
	(1.849)	(1.897)		(1.598)	(1.173)		
Var_i^{TT}	0.300*	0.29		0.126		0.233*	
	(0.171)	(0.177)		(0.176)		(0.129)	
ln(income)	0.022	0.025	-0.001	-0.043**	-0.047**	-0.036**	-0.036**
	(0.024)	(0.024)	(0.019)	(0.020)	(0.019)	(0.018)	(0.018)
Parents' Edu	0.002	0.003	0.001	-0.009	-0.008	-0.009	-0.008
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ln(TestScore)	-0.526**	-0.529**	-0.701***	-0.805***	-0.824***	-0.722***	-0.634***
	(0.224)	(0.231)	(0.199)	(0.227)	(0.225)	(0.207)	(0.202)
# of Awards	-0.013	-0.01	-0.025	-0.072***	-0.072***	-0.069***	-0.066***
	(0.021)	(0.022)	(0.019)	(0.018)	(0.017)	(0.017)	(0.017)
$Attn\_Q$		-0.052	-0.054	-0.051	-0.053	-0.051	-0.056
		(0.037)	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)
$Attn\_U$		0.052**	0.053**	0.057**	0.057**	0.057**	0.058**
		(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Obs.	661	628	628	628	628	628	628
R-Squared	0.18	0.19	0.19	0.16	0.16	0.16	0.16

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ and $Attn\_U$ are survey responses and measure how much attention parents pay to school quality and uncertainty respectively. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses. Parents in the sample are those who submit a full list.

Table 11: Who Strategizes Better: Regression Analysis of Best Responding vs. Truth-Telling

	Prob(i played a best response)-Prob(i played truthfully): $p_i^{BR}(C_i) - p_i^{TT}(C_i)$						
$p_i^{TT=BR}$	-0.203	-0.214	-0.218				
	(0.248)	(0.255)	(0.208)				
EU_i^{TT}	0.556	0.964		0.266	-1.073		
	(1.497)	(1.535)		(1.290)	(0.949)		
Var_i^{TT}	0.240*	0.248*		0.223		0.202*	
	(0.144)	(0.149)		(0.145)		(0.107)	
ln(income)	0.022	0.025	0.02	0.015	0.008	0.014	0.013
	(0.020)	(0.021)	(0.016)	(0.017)	(0.016)	(0.015)	(0.015)
Parents' Edu	0.002	0.003	0.004	0.002	0.002	0.002	0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ln(TestScore)	-0.178	-0.178	-0.163	-0.203	-0.228	-0.217	-0.142
	(0.161)	(0.165)	(0.144)	(0.162)	(0.162)	(0.148)	(0.143)
# of Awards	0.012	0.015	0.016	0.005	0.005	0.005	0.008
	(0.018)	(0.018)	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)
$Attn\_Q$		-0.029	-0.032	-0.029	-0.031	-0.029	-0.032
		(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
$Attn\_U$		0.044**	0.045**	0.045**	0.045**	0.045**	0.046**
		(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Obs.	731	694	694	694	694	694	694
R-Squared	0.04	0.05	0.05	0.05	0.05	0.05	0.050

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ and $Attn\_U$ are survey responses and measure how much attention parents pay to school quality and uncertainty, respectively. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses. Parents in the sample are those who participate, i.e. submitting a full list or a partial list.

Table 12: Who Strategizes Better: Regression Analysis of Playing a Non-Truthful Best Response

	Prob(i played a best response which is not truth-telling): $p_i^{BR \neq TT}(C_i)$						
$p_i^{TT=BR}$	0.131	0.159	-0.055				
	(0.122)	(0.125)	(0.102)				
EU_i^{TT}	-1.617**	-1.535**		-1.018	-1.457***		
	(0.735)	(0.750)		(0.631)	(0.464)		
Var_i^{TT}	0.05	0.054		0.073		0.151***	
	(0.071)	(0.073)		(0.071)		(0.052)	
ln(income)	0.007	0.007	0.021***	0.014*	0.012	0.020***	0.019***
	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Parents' Edu	0.007**	0.007**	0.009***	0.008**	0.008**	0.008**	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ln(TestScore)	0.052	0.058	0.180**	0.076	0.068	0.129*	0.185***
	(0.079)	(0.081)	(0.071)	(0.079)	(0.079)	(0.072)	(0.070)
# of Awards	0.008	0.007	0.019**	0.014**	0.013*	0.015**	0.017**
	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
$Attn\_Q$		-0.004	-0.006	-0.004	-0.004	-0.003	-0.006
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$Attn\_U$		0.016	0.017	0.015	0.015	0.016	0.017
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Obs.	731	694	694	694	694	694	694
R-Squared	0.14	0.14	0.13	0.14	0.14	0.14	0.13

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ and $Attn\_U$ are survey responses and measure how much attention parents pay to school quality and uncertainty, respectively. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses. Parents in the sample are those who participate, i.e. submitting a full list or a partial list.

Table 13: Who Strategizes Better: Regression Analysis of Mean Utility Achieved Given the Submitted List

	Mean Utility Achieved by the Actual Submitted List C_i: $EU_i\left(C_i\right)$						
$p_i^{TT=BR}$	-0.167***	-0.161***	-0.108***				
	(0.032)	(0.033)	(0.027)				
EU_i^{TT}	0.498**	0.496**		-0.033	-0.006		
	(0.193)	(0.199)		(0.170)	(0.125)		
Var_i^{TT}	0.015	0.015		-0.004		-0.002	
	(0.019)	(0.019)		(0.019)		(0.014)	
$Var_i\left(C_i\right)$	-0.019***	-0.018***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ln(income)	-0.001	-0.001	-0.005**	-0.009***	-0.009***	-0.009***	-0.009***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Parents' Edu	0.001	0.001	0.001	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln(TestScore)	-0.030	-0.028	-0.055***	-0.045**	-0.045**	-0.044**	-0.044**
	(0.021)	(0.021)	(0.019)	(0.021)	(0.021)	(0.020)	(0.019)
# of Awards	0.000	-0.001	-0.004*	-0.008***	-0.008***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$Attn\_Q$		0.002	0.003	0.003	0.003	0.003	0.003
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$Attn\_U$		0.001	0.001	0.001	0.001	0.001	0.001
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Obs.	731	694	694	694	694	694	694
R-Squared	0.24	0.23	0.22	0.21	0.21	0.21	0.21

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ and $Attn\_U$ are survey responses and measure how much attention parents pay to school quality and uncertainty, respectively. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses. Parents in the sample are those who participate, i.e. submitting a full list or a partial list. $Var_i\left(C_i\right)$ is the variance of utility when submitting C_i. Excluding this variable does not change the results significantly

Table 14: Who Strategizes Better: Regression Analysis of Variance of Utility Achieved Given the Submitted List

	Variance of Utility Achieved by the Actual Submitted List C_i: $Var_i\left(C_i\right)$						
$p_i^{TT=BR}$	0.247	0.333	-0.198				
	(0.465)	(0.480)	(0.388)				
EU_i^{TT}	-3.530	-4.550		-3.462	-3.244*		
	(2.763)	(2.853)		(2.382)	(1.749)		
Var_i^{TT}	-0.040	-0.076		-0.036		0.228	
	(0.264)	(0.274)		(0.268)		(0.197)	
$EU_i\left(C_i\right)$	-3.793***	-3.692***	-3.822***	-3.765***	-3.764***	-3.771***	-3.781***
	(0.513)	(0.530)	(0.527)	(0.520)	(0.519)	(0.520)	(0.520)
ln(income)	-0.052	-0.065*	-0.026	-0.050	-0.049	-0.031	-0.032
	(0.037)	(0.038)	(0.030)	(0.031)	(0.030)	(0.028)	(0.028)
Parents' Edu	0.009	0.005	0.010	0.008	0.008	0.008	0.009
	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
ln(TestScore)	0.737**	0.769**	1.044***	0.803***	0.808***	0.982***	1.066***
	(0.296)	(0.305)	(0.269)	(0.301)	(0.300)	(0.275)	(0.266)
# of Awards	0.047	0.043	0.072**	0.057**	0.057**	0.063**	0.066**
	(0.032)	(0.033)	(0.029)	(0.027)	(0.027)	(0.026)	(0.026)
$Attn\_Q$		0.134**	0.131**	0.134**	0.134**	0.135**	0.131**
		(0.057)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)
$Attn\_U$		-0.098**	-0.096**	-0.098**	-0.098**	-0.096**	-0.095**
		(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Obs.	731	694	694	694	694	694	694
R-Squared	0.16	0.17	0.17	0.17	0.17	0.17	0.17

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ and $Attn\_U$ are survey responses and measure how much attention parents pay to school quality and uncertainty, respectively. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses. Parents in the sample are those who participate, i.e. submitting a full list or a partial list. $EU_i\left(C_i\right)$ is the expected utility when submitting C_i. Excluding this variable does not change the results significantly

Table 15: Determinants of Attention on Uncertainty/Beliefs: Regression Analysis

	How much attention the parent pays to the uncertainty: $Attn\_U$								
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
$p_i^{TT=BR}$	-0.846**	-0.587*	-0.089						
	(0.407)	(0.354)	(0.634)						
EU_i^{TT}				-3.536*	-3.359**	-3.033	-5.092**	-3.152	0.414
				(1.823)	(1.583)	(2.311)	(2.522)	(2.193)	(4.130)
Var_i^{TT}							-0.255	0.034	0.565
							(0.286)	(0.249)	(0.533)
ln(income)	0.031	-0.006	-0.077	-0.015	-0.043	-0.093*	-0.024	-0.042	-0.075
	(0.034)	(0.030)	(0.068)	(0.033)	(0.029)	(0.056)	(0.034)	(0.030)	(0.050)
Parents' Edu	-0.027**	-0.029***	-0.032*	-0.034***	-0.034***	-0.033**	-0.034***	-0.034***	-0.033**
	(0.013)	(0.011)	(0.016)	(0.012)	(0.011)	(0.015)	(0.012)	(0.011)	(0.015)
ln(TestScore)	-0.068	-0.193	-0.434	-0.292	-0.424	-0.665	-0.323	-0.42	-0.597
	(0.304)	(0.264)	(0.428)	(0.337)	(0.293)	(0.461)	(0.339)	(0.294)	(0.447)
# of Awards	0.033	0.016	-0.017	-0.006	-0.012	-0.024	-0.006	-0.012	-0.024
	(0.030)	(0.026)	(0.045)	(0.025)	(0.022)	(0.033)	(0.025)	(0.022)	(0.033)
$Attn\_Q$		0.745***	2.174**		0.748***	2.122**		0.748***	2.123**
		(0.045)	(1.026)		(0.045)	(1.017)		(0.045)	-1.018
IV for $Attn\_Q$			√			√			√
Obs.	855	855	855	855	855	855	855	855	855
R-Squared	0.03	0.27	–	0.03	0.27	–	0.03	0.27	–

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_Q$ is survey responses and measures how much attention parents pay to school quality. The instrumental variable for $Attn\_Q$ is from a survey question and measures the effort spent on researching first-choice school quality. The first stage P-values are all between 0.056-0.062. *, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses.

Table 16: Determinants of Attention on School Quality: Regression Analysis

	How much attention the parent pays to school quality: $Attn\_Q$								
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
$p_i^{TT=BR}$	-0.348	-0.07	-0.312						
	(0.270)	(0.236)	(0.332)						
EU_i^{TT}				-0.237	0.932	-0.056	-2.594	-0.916	-2.269
				(1.212)	(1.055)	(1.454)	(1.673)	(1.457)	(2.014)
Var_i^{TT}							-0.386**	-0.302*	-0.370*
							(0.189)	(0.165)	(0.191)
ln(income)	0.050**	0.040**	0.048**	0.037*	0.042**	0.038*	0.024	0.032	0.026
	(0.023)	(0.020)	(0.023)	(0.022)	(0.019)	(0.021)	(0.023)	(0.020)	(0.022)
Parents' Edu	0.002	0.011	0.003	0.000	0.011	0.001	0.000	0.011	0.002
	(0.008)	(0.007)	(0.010)	(0.008)	(0.007)	(0.011)	(0.008)	(0.007)	(0.011)
ln(TestScore)	0.169	0.191	0.171	0.176	0.272	0.191	0.129	0.236	0.15
	(0.202)	(0.175)	(0.196)	(0.224)	(0.195)	(0.228)	(0.225)	(0.195)	-0.229
# of Awards	0.023	0.012	0.022	0.009	0.011	0.009	0.009	0.011	0.009
	(0.020)	(0.017)	(0.021)	(0.017)	(0.014)	(0.016)	(0.017)	(0.014)	(0.016)
$Attn\_U$		0.329***	0.043		0.330***	0.051		0.329***	0.064
		(0.020)	(0.240)		(0.020)	(0.245)		(0.020)	(0.240)
IV for $Attn\_U$			√			√			√
Obs.	855	855	855	855	855	855	855	855	855
R-Squared	0.03	0.26	0.08	0.02	0.26	0.09	0.03	0.27	0.11

Notes: Other controls are dummies for student's gender and elementary school. Definition of constructed incentive measures (first 3 independent variables) can be found in Table 7. $Attn\_U$ is survey responses and measures how much attention parents pay to the uncertainty. The instrumental variable for $Attn\_U$ is from a survey question and measures the effort spent on researching the uncertainty. Every first stage of IV regression has a P-value less than 0.01.

*, **, ***, significant at 10%, 5%, 1% level. Standard errors are in parentheses.