TSE WP working paper series

12-345

Research Group: Econometrics and Statistics

May 2012

"Gaming the Boston School Choice Mechanism in Beijing"

Yinghua He

Toulouse School of Economics (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31015 TOULOUSE Cedex 6
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.tse-fr.eu** - contact@tse-fr.eu

Toulouse School of Economics

Gaming the Boston School Choice Mechanism in Beijing

Yinghua He[1]

May 2012

(First Version: November, 2009)

[1]This paper is previously distributed as "Gaming the School Choice Mechanism". Toulouse School of Economics. Email address: yinghua.he@gmail.com. I am deeply indebted to Fang Lai for her generous help that has made 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, Greg Lewis, Sanxi Li, Hong Luo, Thierry Magnac, Serena Ng, Brendan O'Flaherty, Parag Pathak, Michael Riordan, Jonah Rockoff, Johannes Schmieder, Herdis Steingrimsdottir, Priscilla K. Yen, and seminar participants at Autonoma de Barcelona, Bristol, Columbia, Copenhagen Business School, CUHK, LSE, Nanyang, Paris School of Economics, Toulouse, Western Ontario, Wisonsin-Madison, WUSTL, WZB, Zhejiang, the 2009 Econometric Society NASM in Boston, 2010 NBER Market Design Meeting, 2010 China Economics Summer Institute, and 2010 Econometric Society World Congress in Shanghai. Financial support from Columbia's Wueller and Vickrey research prizes and the Program for Economic Research are gratefully acknowledged.

Abstract

The Boston mechanism is criticized for its poor incentive and welfare performance compared to the Gale-Shapley deferred-acceptance mechanism (DA). Using school choice data from Beijing, I investigate parents' behavior under the Boston mechanism, taking into account parents' possible mistakes when they strategize. Evidence shows that parents are overcautious as they play "safe" strategies too often. Wealthier/more educated parents are less overcautious and perform slightly better because they have better outside options while not being any more adept at strategizing. Parents who are always truth-telling experience a utility gain in switching from the Boston mechanism to the DA, equivalent to a 7.1% decrease in the distance to a school. Among them, 44.2% are better off under the DA, while 35.5% are worse off.

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

Given that monetary transfers are usually precluded in the allocation of students to public schools, a centralized assignment mechanism is often necessary. Despite the increasing popularity of school choice programs, the question about which assignment mechanism should be used is still debated among researchers and policy makers.

At the center of the debate is the Boston mechanism, one of the most popular mechanisms in practice. It was used by the Boston Public Schools (BPS) from 1989 to 2005 before being abolished.[1] The main criticism of the Boston mechanism is that it encourages parents to "game the system." Namely, parents may have incentives to misreport their preferences when submitting rank-ordered lists of schools (Abdulkadiroglu and Sonmez (2003)). Schools also form a strict priority ordering of students, usually with lotteries as tie-breakers. Each school first considers 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 rank it second. 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 their first choice.[2]

Since the mechanism is not strategy-proof the ability to strategize, or the level of sophistication, might affect parents' or students' welfare. Experimental and empirical evidence in previous literature suggests that parents strategize not at the same level (e.g., Abdulkadiroglu, Pathak, Roth, and Sonmez (2006), Chen and Sonmez (2006), Lai, Sadoulet, and de Janvry (2009), and Pais and Pinter (2008)). In a theoretical paper, Pathak and Sonmez (2008) consider two types of parents: sincere (or naive) parents who always reveal their preferences truthfully, and sophisticated parents who always play a best response against others. They show that the mechanism may give an advantage to sophisticated parents.

[1]There are many school districts that still use the mechanism, e.g., Cambridge, MA; Charlotte-Mecklenburg, NC; St. Petersburg, FL; Minneapolis, MN; and Providence, RI. The mechanism is also popular in other countries and in other contexts, for example, China's college admissions.

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

These results were instrumental in the BPS' decision. 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)), which is strategy-proof: reporting true preferences is a weakly dominant strategy (Dubins and Freedman (1981); Roth (1982)). A description of the reform can be found in Abdulkadiroglu, Pathak, Roth, and Sonmez (2005).

One of the key arguments for the reform is that the Boston mechanism might penalize less sophisticated parents, while the DA protects them. For example, the BPS Strategic Planning Team claimed in 2005 that "*a strategy-proof algorithm 'levels the playing field' by diminishing the harm done to parents who do not strategize or do not strategize well*." More importantly, policy makers are worried that poor and/or less educated parents are less sophisticated. Therefore, under the Boston mechanism, "*the need to strategize provides an advantage to families who have the time, resources and knowledge to conduct the necessary research*," as stated by Thomas Payzant then BPS Superintendent (Payzant (2005)).

Researchers have not, however, reached a consensus on these arguments. There is no evidence relating parents' sophistication level to family background, and there are mixed, mainly theoretical and experimental, results on naive parents' welfare. A recent strand of literature provides results in favor of the Boston mechanism (e.g. Featherstone and Niederle (2008), Miralles (2008)). In particular, **?** show that some naive parents can even be better off under the Boston mechanism. Using field data from Beijing, this paper fills the gap by answering two questions: (i) whether poorer/less educated parents are more likely to be naive and (ii) whether the Boston mechanism harms naive parents relative to the DA.

In the data, 914 students apply to four middle schools under a version of the Boston mechanism, and schools use a random lottery to rank students without pre-determined priorities. To evaluate parents' welfare, I use concepts of Bayesian Nash equilibrium and *ex ante* efficiency.[3] At the time of application, the lottery is unknown, and parents' preferences

[3]In terms of *ex ante* efficiency, Zhou (1990) shows that it is impossible to have a strategy-proof and efficient mechanism that treats the same type of parents equally. Therefore, the DA is not *ex ante* efficient, because it satisfies the other two properties.

are private information.[4] Parents maximize expected utility by selecting a rank-ordered list of schools under uncertainties from two sources: other parents' behavior and the lottery.

The data contain parents' submitted lists of schools and family background, but not their true preferences. The challenge is to estimate true preferences when parents are not necessarily truth-telling.[5] I assume a random utility model for parents' preferences over schools, with normally distributed errors as in a multinomial probit model. Parents' potentially heterogeneous sophistication is explicitly considered.

Under the assumption that everyone understands the uncertainty from the lottery, a parent's sophistication depends on her assessment of other parents' behaviors, which are determined by the joint distribution of their preferences and sophistication. A parent is sophisticated if she assesses correctly; her subjective beliefs – the perceived probabilities of being accepted by each school when submitting different lists – therefore match what are implied by the true distribution. Less sophisticated parents have inaccurate beliefs, while naive ones disregard the uncertainty and are always truth-telling.

While probably wrong, beliefs must satisfy the properties imposed by the mechanism, e.g., moving a school upward in a list (weakly) increases the probability of being accepted by that school. These properties lead to a set of dominated strategies, for instance, ranking an unacceptable school first. Assuming these dominated strategies are not played in equilibrium, I group certain lists together and, loosely speaking, the necessary equilibrium conditions become necessary and sufficient with respect to the new choice set.

In practice, indeterminacy regarding parents' behavior arises because (i) there are schools worse than the outside option and (ii) some probabilities in parents' beliefs might be zero. I provide evidence that indeterminacy exists in the data and propose solutions while main-

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

[5]Hastings, Kane, and Staiger (2008) estimate the demand for schools under the assumption that students are truth-telling under the Boston mechanism. They use data from Charlotte-Mecklenburg Public School District in 2002, where the mechanism had just been implemented. The truth-telling assumption may be more likely to be valid in their setting than others. I also estimate the model under the truth-telling assumption, and it is rejected when tested against the model with strategic behavior.

taining the model's point identification. A method of simulated maximum likelihood, similar to that in a multinomial probit, is used for estimation.

Results reject the hypothesis that everyone is naive, and also reject that everyone is sophisticated. Parents understand the rules well, but they are overcautious, as they avoid top ranking the school with best quality and smallest quota more often than their best responses would prescribe.[6] Income and education offset overcautiousness slightly, because wealthier/more educated parents' true preference order is more likely to be a best response. There is no evidence of these parents being more sophisticated, as these findings are driven by the fact that they have a better outside option.

The data include information on how much attention parents pay to uncertainty in the game. Poorer parents pay more attention, which implies that they try to find a best response. However, paying more attention does not mitigate, and sometimes even worsens, their overcautiousness.

To evaluate the effect of replacing the Boston mechanism with the DA, I simulate outcomes under both mechanisms, assuming preferences do not change across mechanisms. If other parents are overcautious and behave as in the data, both naive and sophisticated parents suffer a significant utility loss under the DA, amounting to a tripling of the distance to a school. For naive parents, only 8% are better off under the DA, while 71.5% are worse off. The negative effects are larger for sophisticated parents, and decrease with parents' income and education because of the outside option.

If everyone is either sophisticated or naive and no one is overcautious, switching from the Boston to the DA has mixed effects. Sophisticated parents suffer a utility loss equivalent to increasing the distance to a school by 90.6%. Among them, only 11.5% are better off, while about 68.0% are worse off. Naive parents on average have a utility gain under the DA, although the gain amounts to decreasing the distance to a school by merely 7.1%. The DA helps about 44.2% of naive parents but hurts 35.3% of them.

[6]This overcautiousness is related to, but different from, the "small school bias" found in experimental studies (Chen and Sonmez (2006), Calsamiglia, Haeringer, and Klijn (2010)). Namely, schools with fewer slots are ranked at lower positions than those in the true preference. Instead of focusing on the true preferences, overcautiousness compares observed behaviors with best responses.

Other Related Literature There is a growing literature using real life data to study assignment problems. For example, **?** compare a strategy-proof mechanism with a non-strategy-proof one using a data set on MBA students' course-allocations. Braun, Dwenger, and Kubler (2010) study the strategic behavior in the centralized university admissions in Germany, and Carvalho and Magnac (2009) investigate the college admission with exams in Brazil.

This study also relates to the literature on testing whether an equilibrium is played in real life games. For example, Chiappori, Levitt, and Groseclose (2002) and Kovash and Levitt (2009) study professional sports, and Hortacsu and Puller (2008) examine the strategic bidding in an electricity spot market auction. Hortacsu and Puller characterize a Bayesian-Nash equilibrium model and compare actual bidding behavior to theoretical benchmarks. The difficulty in estimating a Bayesian-Nash equilibrium lies in specifying the beliefs. Under some technical assumptions, they 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 allows players to make mistakes and derives identification independent of beliefs.

Another related strand of literature is the estimation of simultaneous games of incomplete information. Most studies need the condition of consistent beliefs to derive moment conditions or choice probabilities, e.g., Seim (2006), Bajari, Hong, Krainer, and Nekipelov (2010), Aradillas-Lopez (2007a), and Aradillas-Lopez (2007b). Given the small number of players, identification requires multiple game plays and equilibrium beliefs which are correct and stable across game plays. I relax these assumptions and allow inaccurate beliefs.

In the following, Section 2 describes the two school choice mechanisms and the data from Beijing. Section 3 formalizes the school choice problem under the Boston mechanism as a Bayesian game. Restrictions on parents' behavior are derived under various assumptions, and I also characterize choice probabilities and propose a method of simulated log-likelihood. Section 4 presents reduced-from results, while Section 5 shows the model estimation. In particular, I present the correlation between sophistication and family background in both sections. Section 6 shows the counterfactual analysis for replacing the Boston with the DA mechanism. The paper concludes in Section 7.

1 The Two Mechanisms, Background, and Data

1.1 Deferred-Acceptance Mechanism

The DA mechanism works as follows:

(i) Each school forms a strict priority ordering of students with rules which are determined by state or local laws. In the Boston schools, for example, it depends on sibling enrollment, distance to schools, and a lottery.

(ii) Schools announce their enrollment quota and students submit rank-ordered lists of schools.

(iii) With priority orderings and submitted lists, the matching process has several rounds:

Round 1. Every student applies to her first choice. Each school rejects the lowest-priority students in excess of its capacity and *temporarily holds* the other students.

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 Round $(k-1)$ 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 use the same factor, e.g., the same test score or lottery, and rank students in the same way, the DA is equivalent to the serial dictatorship mechanism (Abdulkadiroglu and Sonmez (1998)). Following their priority order, essentially, students sequentially choose their favorite among schools which still have available seats.

1.2 Boston Mechanism

Similarly, the Boston mechanism asks students to submit rank-ordered lists, uses predefined rules to determine schools' ranking over students, and has multiple rounds:

Round 1. Each school considers all the students who rank it first 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 students who rank it as 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.

1.3 Boston Mechanism in Beijing

I study school choice in the largest neighborhood of Beijing's Eastern City District in 1999. Students could apply to four middle schools with a total quota of 960, as determined by the Education Bureau. To be included in this neighborhood, a student must be enrolled as a 6th grader in one of seven given elementary schools in 1999. A more detailed description of the education system and the matching is available in Lai, Sadoulet, and de Janvry (2009).

The neighborhood adopted a version of the Boston mechanism in which schools' ranking over students was solely determined by a random lottery (single tie-breaker). Students could submit a list ranking up to four schools. Upon submission, a computer-generated 10-digit number was randomly assigned to each student, and then the admission proceeded as previously described.[7]

Students' outside option was mainly the 28 public schools in the district, including the

[7]The same mechanism was used in all Beijing's neighborhoods in 1999, including those in other districts.

four to which they could apply through the mechanism. They might attend a public school without going through the mechanism in three ways. (i) Schools admit students directly if their parents are employed by the school, if they have received at least a city-level prize in academic or special skill achievements, or if a considerable payment is made to the school.[8] (ii) Besides the quota announced, some top schools admit additional students by offering an admission exam. (iii) Schools admit some transfer students who are not satisfied with their assignment and make a payment to the accepting school.

Other possible outside options were not very relevant at that time. Specifically, private schools were not well developed in 1999. Besides, there was no strong incentive for students to transfer out of the district, because the Eastern City District had both an advantageous location and a very good reputation for educational quality. Such transfers were only possible when there was a formal relocation of parents or an even higher payment made to the out-of-district accepting school.

1.4 Data

The data in this study come from two sources: submitted lists, elementary school enrollment, grade 6 test scores, and home addresses in 1999 are provided from administrative data, and all other information is from a district-level survey in early 2002.

Chinese middle schools provide three years of education – grades 7-9 – so the survey covered all students in the district enrolled in the last year of middle school, as well as their parents. Dropping out or repeating grades was negligible in these schools, and inter-district transfers were extremely rare as discussed above. Hence, the survey population is close to the population of students who entered middle schools in the district in 1999. A questionnaire directed to parents collected information on their educational attainment

[8] Such payments, or "*ze xiao fei*" (literally, "fees for choosing a school"), may depend on the student's ability and parents' connection. Unfortunately, information on these payments is not publicly available. Since 2008, the education authority of Beijing has regulated that such fees cannot be more than 30,000 yuan (Source: http://www.bjedu.gov.cn/publish/portal0/tab67/info11554.htm). This is slightly above the average disposable income among urban residents of Beijing in 2008, 24,725 yuan. The out-of-pocket cost for parents may easily exceed this limit. For example, a blog post claims that some people paid 250,000 yuan to get into a very good elementary school in 2011 (Source: http://blog.sina.com.cn/s/blog_6ce3959f0102dr2x.html).

and income, as well as retrospective information on their preparedness for making school choice decisions in 1999.

1.4.1 Heterogeneity among the Middle Schools

The four schools are highly differentiated on two dimensions: enrollment quota and quality. Table 1 shows that School 1 has the smallest quota, 63 seats. Note that this does not imply that its size is small because it also enrolls students from other neighborhoods. School 4 has the largest quota, 360 seats.

Table 1: Middle Schools: Quota and Quality

Schools	Quota	$School\_Score_s$: Average Test Score[a]	Ranking in the district[b]
1	63	559.27	1
2	227	522.91	7
3	310	508.47	14
4	360	470.13	28
Total	960		

a. Average test score of the graduating class in the high school entrance exam in 1999, out of 600.
b. Ranking based on average test score among all 28 public schools in the district.

School 1 also has the best quality as measured by the performance of the school's graduating class in the high school entrance exam in 1999. The exam is city-wide and high-stakes, and thus it is a factor that parents weigh heavily in determining school quality. As column 3 shows, these schools span the quality distribution of the 28 schools in the district, with better schools having smaller quotas.

1.4.2 Students' Characteristics and Behavior: A First Look

Using their elementary school enrollment, I identify 914 students as qualified applicants in this neighborhood in 1999. The 46 "missing" students, i.e., the difference between the total quota (960) and the number of observed students (914), may have come from three sources: (i) enrollment quota is usually larger than the number of students; (ii) some may have skipped the mechanism and gone to schools outside the district in 1999; and (iii) some may have transferred to schools outside the district after 1999.

Students in (ii) and (iii) may have made that decision because they were unsatisfied with the expected or realized school assignment, and thus sample selection may arise. However, as discussed above, (ii) and (iii) are plausibly negligible, although this cannot be verified. I therefore focus on the 914 observed students.[9]

The distribution of submitted lists in 1999 and middle school enrollment in 2002 are shown in Table 2. About 20% did not participate in the centralized mechanism and took their outside option directly, while 60.77% of the non-participants were still enrolled in one of the four schools in 2002. The majority submitted a full list with three or four schools; only 7.44% submitted a partial list which ranks one or two schools. Overall, in 2002, only 10.07% of the students were enrolled in a school other than the four schools.[10] The best in the district, School 1, enrolled 147 of the 914 students, more than double its quota. Enrollment at any other school was lower than its announced quota.

Table 2: Distribution of Submitted Lists in 1999 and Middle School Enrollment in 2002

Submitted Lists in 1999			Middle School Enrollment in 2002				
	Freq.	Percent	School 1	School 2	School 3	School 4	Other[a]
Not Participating	181	19.80%	58	20	25	7	71
Full Lists	665	72.76%	73	185	237	152	18
4 Schools	558	(61.05%)	57	155	203	126	17
3 Schools	107	(11.71%)	16	30	34	26	1
Partial Lists	68	7.44%	16	19	22	8	3
2 Schools	58	6.35%	10	18	20	7	3
1 School	10	1.09%	6	1	2	1	0
Total	914	100%	147	224	284	167	92

a. "Other" means one of the other 24 public middle schools in the district.

In the estimation, I focus on students' family background ($Parent\_Inc_i$, $Parent\_Edu_i$), ability (Own_Score_i, $Awards_i$), gender ($Girl_i$), and distance to each school ($Distance_{i,s}$). Table 3 presents their definitions.

Table 4 further summarizes these variables. In the estimation, most of the variables are expressed in logarithms and de-meaned. I present summary statistics of the raw data

[9] As a robustness check, I impute 46 additional students by drawing from the observed 914 students to "complete" the data. A few experiments show that the results are not sensitive to the imputation as long as the imputation is not extreme.

[10] This is another piece of indirect evidence that transferring to another district is rare, as it is already unusual for students to choose a within-district school other than the four "inside" ones.

Table 3: Definitions of Main Variables

Variables	Definition	Source
$Parent\_Inc_i$	Parents' income yuan/month in 2002	Survey in 2002
$Parent\_Edu_i$	Parents' average years of education	Survey in 2002
$Girl_i$	=1 if student i is a girl	Survey in 2002
Own_Score_i	Elementary Chinese + math, out of 200	Administrative data
$Awards_i$	District level awards in elementary school	Survey in 2002
$Distance_{i,s}$	Walking distance to School s in 1999, km	Administrative data

for the full sample and 3 subsamples – non-participants, participants submitting partial lists, and participants submitting full lists. Non-participants have richer and more educated parents than average, and they have higher test scores and have earned more awards. This is consistent with the earlier discussion that parents' income and students' ability increase the quality of their outside option. The same pattern of parental income and education is observed for participants submitting partial lists, although these students have lower test scores and have earned fewer awards than average.

Table 4: Summary Statistics

	Full Sample			Non-Participant	Partial List	Full List
	Used in Estimation		Raw Data	Raw Data		
Variables	Transformation	Mean	Mean	Mean	Mean	Mean
$Parent\_Inc_i$	Log, de-mean[a]	0.00	3664.01	4249.07	4191.12	3450.87
		(0.82)	(3468.85)	(2457.23)	(2069.04)	(3782.76)
$Parent\_Edu_i$	De-mean	0.00	13.44	14.28	14.19	13.14
		(2.24)	(2.24)	(2.57)	(2.13)	(2.07)
$Girl_i$	None	0.52	0.52	0.51	0.49	0.53
		(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Own_Score_i	Log, de-mean	0.00	183.56	187.09	178.41	183.12
		(0.08)	(11.66)	(7.84)	(15.55)	(11.83)
$Awards_i$	De-mean	0.00	0.73	1.12	0.51	0.65
		(1.00)	(1.00)	(1.29)	(0.82)	(0.90)
$Distance_{i,1}$	Log, de-mean[b]	-0.02	2.31	2.64	1.94	2.27
		(0.76)	(2.27)	(2.31)	(1.25)	(2.32)
$Distance_{i,2}$	Log, de-mean[b]	-0.12	2.22	2.55	1.89	2.16
		(0.86)	(2.29)	(2.31)	(1.41)	(2.35)
$Distance_{i,3}$	Log, de-mean[b]	0.13	2.51	2.95	2.21	2.42
		(0.66)	(2.21)	(2.17)	(1.34)	(2.28)
$Distance_{i,4}$	Log, de-mean[b]	0.02	2.41	2.84	1.84	2.35
		(0.82)	(2.20)	(2.14)	(1.43)	(2.28)
# of Obs.		914	914	181	665	68

Standard deviations in parentheses. a. More precisely: "log(Parent_Inc$_i$+1), de-mean";
b. The mean here is that of all 4 distances.

2 Model: Boston School Choice as A Bayesian Game

In this section, the school choice problem under the Boston mechanism is formalized as a Bayesian game. 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 0 is the outside option;[11]

(iii) a capacity vector, $\{q_s\}_{s=1}^{S}$; $\sum_{s=1}^{S} q_s \geq I$, $\sum_{s=1}^{S} q_s - q_{s'} < I$, and $q_{s'} < I, \forall s' \neq 0$.

(iv) students' rank-ordered lists, $\{C_i\}_{i=1}^{I}$, where $C_i = \left(c_i^1, ..., c_i^S\right)$, $c_i^k \in \{s\}_{s=0}^{S}$, $\forall k = 1, ..., S$;

(v) schools' priorities over students, determined solely by a random lottery.

At the start of the game, each school announces its capacity, q_s. There are enough seats to accommodate all the students, i.e. $\sum_{s=1}^{S} q_s \geq I$; no school has enough seats to enroll all students, $q_{s'} < I, \forall s'$; and every school is significantly big, in the sense that not all students can be accommodated by the other $(S-1)$ schools, $\sum_{s=1}^{S} q_s - q_{s'} < I$.

Parents or students submit their choice lists, $C_i = \left(c_i^1, ..., c_i^S\right)$ where c_i^k is the kth choice. C_i is a full list if it ranks all S schools; otherwise, it is a partial list. They may submit partial lists or submit $(0, ..., 0)$. In the latter case, the student is not considered in the mechanism.

After collecting $\{C_i\}_{i=1}^{I}$, the mechanism assigns each student a random number which determines her priority at all schools. In this case, all students have the same *ex ante* priority, although pre-determined priorities can be considered as well. With the lists and the random lottery, the admission proceeds as described in the previous section. After students receive their assignments, they can choose the outside option if they are not satisfied.

In the following, "student" and "parent" are used interchangeably. I first present the set-up and the benchmark case where everyone is (equally) sophisticated and shares a common prior. I then extend the model to the case where parents have heterogeneous levels of sophistication. The definition of sophistication is formalized in due course.

[11] When the number of schools is less than 3, truth-telling is a dominant strategy.

2.1 Set-Up

The utility of student i attending school s ($s = 1, .., S$) is defined as:

$$u_{i,s} = \alpha_s + \boldsymbol{X}_i \beta_X + \mathbf{Z}_{i,s} \beta_Z + \varepsilon_{i,s}$$

where α_s is school s's fixed effect; $\boldsymbol{X}_i \in \mathbb{R}^{K_1}$ are i's characteristics, such as test score, parents' income, and parents' education, etc.; $\mathbf{Z}_{i,s} \in \mathbb{R}^{K_2}$ are student-school specific attributes, e.g., the distance from i's home to s, and $\mathbf{Z}_i \equiv \{\mathbf{Z}_{i,s}\}_{s=1}^{S}$; $\varepsilon_{i,s} \in \mathbb{R}$ includes all other factors, and $\boldsymbol{\varepsilon}_i \equiv \{\varepsilon_{i,s}\}_{s=1}^{S}$.

The utility when choosing the outside option is normalized to zero. Equivalently, the utility of attending any school $s \neq 0$ should be interpreted as the difference between attending s and choosing the outside option. If a school is worse than the outside option, $u_{i,s} < 0$, it is defined as unacceptable. Otherwise, it is acceptable.

The following assumptions are maintained throughout the paper:

AM.1. Parents are expected utility maximizers who know their own preferences, $\{u_{i,s}\}_{s=0}^{S}$, as well as the function of $u_{i,s}$ and its parameters.

AM.2. $(\boldsymbol{X}_i, \mathbf{Z}_i)$ are i.i.d. over i with C.D.F. $G(\boldsymbol{X}, \mathbf{Z})$ which is common knowledge, while $(\boldsymbol{X}_i, \mathbf{Z}_i)$ is private information of i.

AM.3. $\boldsymbol{\varepsilon}_i \perp (\boldsymbol{X}_i, \mathbf{Z}_i)$ and $\boldsymbol{\varepsilon}_i \sim N(0, \Sigma)$ i.i.d. over i, with C.D.F. $F_\Sigma(\boldsymbol{\varepsilon}_i)$ and $Var(\varepsilon_{i,1}) = 1$. $\boldsymbol{\varepsilon}_i$ is private information of i, while its distribution is common knowledge.

AM.4. A parent does not participate, or submits $(0, ..., 0)$, if and only if no school is acceptable.

The assumption that $(\boldsymbol{X}_i, \mathbf{Z}_i)$ is private information is made for ease of exposition. When I is large, similar results hold if $(\boldsymbol{X}_i, \mathbf{Z}_i)$ is common knowledge, or if i knows a fixed number of others' $(\boldsymbol{X}_j, \mathbf{Z}_j)$. Appendix A.3 discusses this in detail.

AM.3 allows an arbitrary correlation between any $\varepsilon_{i,s}$ and $\varepsilon_{i,s'}$. For example, some schools are better at sciences, while others are better at arts. Students who like sciences more than arts have positive shocks for some schools, and negative shocks for others.

AM.4 is somewhat restrictive. It requires that the outside option does not change after parents observe the matching outcome, and it also rules out the possible uncertainty aversion of parents. Appendix A.4 discusses this in detail.

2.2 Benchmark: Homogeneous Sophistication

In the following, everyone is (equally) sophisticated and is endowed with a common prior. Namely, they have the same information and correctly use this information in the same way.

2.2.1 Strategy, Payoff, and Decision Making

A strategy $\sigma_i(\boldsymbol{X}_i, \mathbf{Z}_i, \varepsilon_i)$, possibly in mixed strategies, is a mapping from i's "type" space to the set of all probability distributions over possible lists: $\mathbb{R}^{K_1+SK_2+S} \rightarrow \Delta(\mathcal{C})$. The total number of pure strategies or possible lists in $\mathcal{C}$ is finite, $L \equiv S!\left(\frac{1}{S!} + \frac{1}{(S-1)!} + ... + \frac{1}{1!}\right)$.[12] Each element in $\mathcal{C}$ is a rank-ordered list of k different schools, where $k = 0, ...S$.

The payoff to i can be characterized in two steps: (i) other parents' actions, C_{-i}, are given; and (ii) instead of C_{-i}, other parents' strategies, σ_{-i}, are given.

Given C_{-i}, if $\sigma_i = C$ is a pure strategy, the expected payoff to i is:

$$V_i(C, C_{-i}) \equiv \sum_{s=1}^{S} \left[a_s(C, C_{-i}) \max(u_{i,s}, 0)\right],$$

where only $\max(u_{i,s}, 0)$ matters because parents choose the outside option whenever the assigned one is unacceptable; and $a_s(C, C_{-i})$ is the probability of student i being accepted by s given (C, C_{-i}). $a_s(C, C_{-i})$ is completely determined by the random lottery, and the following lemma summarizes its properties.

Lemma 1 *Given any C and C_{-i}, $a_s(C, C_{-i})$ has the following properties:*

[12] Notice that those lists in which one school appears multiple times are excluded, as they are obviously not optimal. Other obviously dominated lists are also not considered.

(i) A seat is guaranteed if participating: $\forall C \neq (0, ..., 0)$, $\sum_{s=1}^{S} a_s(C, C_{-i}) = 1$;

(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: $a_s(C, C_{-i}) = a_s(C', C_{-i})$, $\forall C, C' \in \mathcal{C}$, *s.t.,* $c_K = c'_K = s$ *and* $c_k = c'_k$, $\forall k \leq K \leq S$.

(iii) Moving a school up (or including an otherwise omitted one) in the list weakly increases the probability of being accepted by that school: $a_s(C', C_{-i}) \geq a_s(C, C_{-i})$, $\forall C, C' \in \mathcal{C}$, *s.t.,* $c_K = c'_{K'} = s$, $K' < K \leq S$, *and* $c_k = c'_k$, $\forall k < K'$.

(iv) If school s is top ranked, the probability of being accepted by that school is at least q_s/I: $a_s(C, C_{-i}) \geq q_s/I, \forall C \in \mathcal{C}$, *s.t.,* $c^1 = s$.

Proofs are collected in Appendix A.1; these properties can, however, be easily verified given the mechanism. Similarly, since σ_i is a probability distribution over pure strategies, $a_s(\sigma_i, C_{-i})$ shares the above properties and $V_i(\sigma_i, C_{-i})$ can be defined in the same way as $V_i(C, C_{-i})$.

Now, instead, suppose that (σ_i, σ_{-i}) is given. i's expected payoff is defined as:

$$\begin{aligned} V_i(\sigma_i, \sigma_{-i}) &\equiv \sum_{n=1}^{L^{(I-1)}} \left\{ \Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\right) V_i\left(\sigma_i, C_{-i}^n\right) \right\} \\ &= \sum_{s=1}^{S} \sum_{n=1}^{L^{(I-1)}} \left[\Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\right) a_s\left(\sigma_i, C_{-i}^n\right) \right] \max\left(u_{i,s}, 0\right), \end{aligned}$$

where the probability that other parents choose C_{-i}^n, $\Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\right)$, is:

$$\int \Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\left(\mathbf{X}_{-i}, \mathbf{Z}_{-i}, \boldsymbol{\varepsilon}_{-i}\right)\right) dG\left(\mathbf{X}_{-i}, \mathbf{Z}_{-i}\right) dF_{\Sigma}\left(\boldsymbol{\varepsilon}_{-i}\right);$$

Given that others play σ_{-i}, i's probability of being accepted by s when playing σ_i can be written:

$$A_s(\sigma_i, \sigma_{-i}) \equiv \sum_{n=1}^{L^{(I-1)}} \Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\right) a_s\left(\sigma_i, C_{-i}^n\right),$$

which may be individual-specific because $\Pr\left(C_{-i}^n \text{ played under } \sigma_{-i}\right)$ might differ across i.

The expected payoff is simplified as $V_i(\sigma_i, \sigma_{-i}) = \sum_{s=1}^{S} A_s(\sigma_i, \sigma_{-i}) \max(u_{i,s}, 0)$.

Furthermore, denote $B(\sigma_i, \sigma_{-i}) \equiv (A_1(\sigma_i, \sigma_{-i}), ..., A_s(\sigma_i, \sigma_{-i})) : \Delta(\mathcal{C}) \to [0,1]^S$ as i's beliefs. By definition, $A_s(\sigma_i, \sigma_{-i})$ is a probability weighted sum of $a_s(C, C_{-i})$, $\forall C$ and C_{-i}. Therefore, it is straightforward to verify that the properties of $a_s(C, C_{-i})$ in Lemma 1 still hold for $A_s(\sigma_i, \sigma_{-i})$. With the beliefs, I define sophistication as follows.

Definition 1 *Given homogeneous sophistication, i is sophisticated if her beliefs are $B(\sigma_i, \sigma_{-i})$.*

Given her beliefs, parent i chooses a strategy to maximize her expected utility:

$$\sigma_i(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i) \in \arg\max_{\hat{\sigma}_i \in \Delta(\mathcal{C})} \sum_{s=1}^{S} A_s(\hat{\sigma}_i, \sigma_{-i}) \max(u_{i,s}, 0). \tag{1}$$

i's optimal strategy may not be unique: (i) the operator $\max(u_{i,s}, 0)$ creates multiple payoff-equivalent lists if some schools are unacceptable; and (ii) additional payoff-equivalent lists arise if $A_s(\sigma_i, \sigma_{-i})$ is zero for some s. This indeterminacy presents a challenge for empirical analysis, since it complicates the characterization of choice probabilities.

2.2.2 Symmetric Bayesian Nash Equilibrium

To mitigate the indeterminacy problem, I consider a symmetric equilibrium in which all parents employ the same strategy, i.e., $\sigma_i(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i) = \sigma_j(\mathbf{X}_j, \mathbf{Z}_j, \boldsymbol{\varepsilon}_j)\ \forall i \neq j$, if $u_{i,s} = u_{j,s}\ \forall s$. Given that everyone is an expected-utility maximizer, the symmetry only requires that, when there are multiple solutions to their maximization problem, parents all use the same rule to choose one strategy, pure or mixed.

Definition 2 *A mixed-strategy symmetric Bayesian Nash equilibrium in the Boston school choice game with homogeneous sophistication is a common strategy $\sigma^* \in \Delta(\mathcal{C})$, s.t.,*

$$\sigma^*(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i) \in \arg\max_{\sigma \in \Delta(\mathcal{C})} \sum_{s=1}^{S} A_s\left(\sigma, \sigma^*_{-i}\right) \max(u_{i,s}, 0), \textit{ given } (\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i), \forall i;$$

and there are common equilibrium beliefs, $B^*(C, \sigma^*) \equiv B\left(C, \sigma^*_{-i}\right)$, $\forall i$ and C.

The existence and a characterization of such an equilibrium is presented in Proposition 1.

Proposition 1 *There always exists a symmetric Bayesian Nash equilibrium in the Boston school choice game. In any symmetric equilibrium,*

(i) equilibrium beliefs are such that $A_s\left(C, \sigma^*_{-i}\right) \in (0,1)\ \forall s, \forall C \neq (0,...,0)$*;*

(ii) if at most one school is unacceptable, i *plays a pure strategy with probability one;*

(iii) if i *plays mixed strategies, with probability one she has at least two unacceptable schools; furthermore she only mixes over lists in which the unacceptable schools are excluded, or included after the acceptable ones.*

This paper estimates a large game played once, and thus multiplicity of equilibrium is not a concern for estimation, as there is only one equilibrium being played in the data. However, it matters for counterfactual analyses where an equilibrium must be selected. In such cases, I focus on the equilibrium that has been played in the data.

Proposition 1 leaves some indeterminacy: Parents may rank unacceptable schools unsystematically in equilibrium, as long as all unacceptable schools are ranked after the acceptable schools. This claim is formalized later in Proposition 2, and I make the following assumption:

Assumption UNACCEPTABLES In a symmetric Bayesian Nash equilibrium, if some or all of the unacceptable schools are included in the list, they are ranked according to their true preference order among themselves after the acceptable schools. Moreover, the excluded unacceptable schools are always less desirable than those included.

It is not implausible that parents follow this strategy in real life. The symmetric equilibrium also requires that parents play the same mixed strategy. The following assumption further clarifies the possible pure strategies in equilibrium mixed strategies:

Assumption MIXING When playing mixed strategies, everyone follows the same mixing rule in equilibrium. Namely, if, without loss of generality, $u_{i,1} > u_{i,2} > ...u_{i,K} > 0 > u_{i,K+1} >$

$... > u_{i,S}$, where $K < (S-2)$, i only submits the following lists with positive probabilities: $(c_i^1, ..., c_i^K, 0, ..., 0)$ where all included schools are acceptable; $(c_i^1, ..., c_i^K, (K+1), 0, ..., 0)$; $(c_i^1, ..., c_i^K, (K+1), (K+2), 0, ..., 0)$; ...; and $(c_i^1, ..., c_i^K, (K+1), (K+2), ..., S)$; where every one of them is a best response. Let $m_{K,l}$, $l \geq K$, denote the probability that an l-school list is submitted while only K schools are acceptable. $m_{K,l}$ is common to everyone, and is independent of the identities of the acceptable schools.

One should ideally show the existence of such a symmetric equilibrium. Unfortunately, I have not found a proof or a disproof, and thus leave it to future research.

2.2.3 Estimation

With Proposition 1 and Assumptions UNACCEPTABLES and MIXING, I characterize the probabilities that each list is played in equilibrium. I assume equilibrium beliefs, B^*, are known and set S equal to 4 in the following.

Characterization of Choice Probabilities Given $(\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the unknown parameters, the conditional probability of i choosing C_i, $\Pr(C_i|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$, is:

(i) if $C_i = (0,0,0,0)$, $\Pr(u_{i,s} < 0, \text{ for all } s|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$;

(ii) if $C_i = (c^1, 0, 0, 0)$, $m_{1,1} * \Pr(u_{i,c^1} > 0 > u_{i,s}, \text{ for } s \neq c^1|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$;

(iii) if $C_i = (c^1, c^2, 0, 0)$,

$$m_{2,2} * \Pr(C_i \text{ is a best response; } u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$$
$$+m_{1,2} * \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1 \neq c^2|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta});$$

(iv) if $C_i = (c^1, c^2, c^3, c^4)$,

$$\Pr(C_i \text{ is a best response; } u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$$
$$+m_{1,4} * \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$$
$$+m_{2,4} * \Pr(C_i \text{ is a best response; } u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta}).$$

Part (i) says that the probability of not participating equals the probability that all schools are unacceptable. The probability of submitting a one-school list, by part (ii), is the mixing probability $m_{1,1}$ times the probability that only one school is acceptable, because parents may submit a two-school or a full list ($m_{1,2}, m_{1,4} \geq 0$).

Part (iii) shows that the likelihood of submitting a two-school list comes from two scenarios: (a) there are two acceptable schools, and (b) there is only one acceptable school. In (a), students may submit either a two-school or a full list,[13] and thus the probability of optimally ranking two acceptable schools is weighted by the mixing probability, $m_{2,2}$. In (b), the first choice must be acceptable and the second choice unacceptable. The omitted schools are unacceptable and worse than the second choice.

Similarly, in part (iv), parents submit a full list in three cases: (a) there are at least three acceptable schools, (b) there are two acceptable schools, and (c) there is only one acceptable schools. Again, the last two cases contribute to the likelihood because of the mixing assumption, while in case (a) there is no possibility of mixing.

The mixing probabilities are independent of $(\boldsymbol{X}_i, \mathbf{Z}_i, B^*)$ due to Assumption MIXING. Moreover, $m_{1,1} + m_{1,2} + m_{1,4} = 1$ and $m_{2,2} + m_{2,4} = 1$.

Simulated Maximum Likelihood Estimation Since the equilibrium beliefs, B^*, are unknown, I use the empirical beliefs, $\hat{B}$, as an approximation.[14] The model is estimated by the following (simulated) maximum likelihood:

$$\max_{\boldsymbol{\theta}} \sum_{i=1}^{I} \ln \left[\Pr \left(C_i | \boldsymbol{X}_i, \mathbf{Z}_i, \hat{B}; \boldsymbol{\theta} \right) \right] \tag{2}$$

The choice probabilities are simulated by the smoothed logit-smoothed accept-reject simulator (Chapter 5, Train (2009)) which is described in Appendix A.6.[15]

[13] It is also possible to submit a three-school list, but it is equivalent to submitting a four-school/full list.

[14] The distribution of equilibrium strategies is approximated by the 914 observations, plus 9 rank-ordered lists which are not played by anyone in the data. 5,000 samples of random draws from the distribution are created. Each sample consists of 914 random draws from the 923 data points, with replacement. Fixing other parents' submitted lists in each sample, I then calculate $\hat{B}^n$ for parent 1. Namely, parent 1 experiments the 24 full lists. The probability of being accepted by each school given any list are calculated by drawing 1,000 independent sets of lotteries and running the mechanism 1,000 times. It is sufficient to consider the full lists only, because either the beliefs associated with partial lists can be derived from those associated with the full lists, or the partial lists are dominated. After repeating this for the 5,000 samples, I calculate $\hat{B} = \sum_{n=1}^{5000} \hat{B}^n / 5000$. Note that $\hat{B}$ may have many elements equal to 0 or 1 given the observed data. This contradicts the results in Proposition 1, and as a remedy I perturb the system a little so that all elements fall within $(0, 1)$ – the maximum absolute difference between the original and the perturbed is 7.282×10^{-8}.

[15] In Appendix A.2, I discuss an alternative approach which uses equilibrium constraints and solves B^* as a function of $\boldsymbol{\theta}$, a fixed point.

2.3 General Case: Heterogeneous Sophistication

In the following, I relax the sophistication assumption, and allow parents to make mistakes when forming their beliefs. Mistakes may be due to information differences and/or different abilities to process information. Under the maintained assumptions, a particular structure on the beliefs and some dominated strategies are identified.

2.3.1 Equilibrium Concept, Definition of Sophistication, and Dominated Strategies

To highlight the heterogeneity in beliefs, denote i's belief as $B_i(C, \sigma_{-i}) \equiv \{A_{i,s}(C, \sigma_{-i})\}_{s=1}^{S} \in [0,1]^S$, $\forall C$, where $A_{i,s}(C, \sigma_{-i}) \equiv \sum_{n=1}^{L(I-1)} \Pr_i\left(C_{-i}^n \text{ played under } \sigma_{-i}\right) a_s\left(C, C_{-i}^n\right)$ and the probability measure $\Pr_i()$ is i's subjective assessment of an event's likelihood.

By the above notation, the extent to which parents can make mistakes is limited. They may be wrong when assessing others' behavior and thus $\Pr_i()$ is individual specific; however they know the rules of the game and thus know $a_s\left(C, C_{-i}^n\right)$ precisely.

Since $A_{i,s}(\sigma_i, \sigma_{-i})$ is a probability weighted average of $a_s(C, C_{-i})$, the properties of $a_s(C, C_{-i})$ still hold for $A_s(\sigma_i, \sigma_{-i})$, and the proof of the following lemma is omitted.

Lemma 2 *Given σ_{-i}, $A_{i,s}(\sigma_i, \sigma_{-i})$ has the same properties as $a_s(C, C_{-i})$ in Lemma 1.*

I rewrite i's strategy as an explicit correspondence of beliefs, $\sigma_i[\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i; B_i(\cdot, \sigma_{-i})]$ and again consider a symmetric equilibrium.

Definition 3 *A mixed-strategy symmetric Bayesian Nash equilibrium in the Boston school choice game with heterogeneous sophistication is a common strategy $\sigma^* \in \Delta(\mathcal{C})$ s.t.,*

$$\sigma^*\left[\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i; B_i\left(\cdot, \sigma_{-i}^*\right)\right] \in \arg\max_{\sigma \in \Delta(\mathcal{C})} \sum_{s=1}^{S} A_{i,s}\left(\sigma, \sigma_{-i}^*\right) \max(u_{i,s}, 0), \textit{ given } (\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i), \forall i.$$

In this definition, the only requirement is that everyone is a subjective expected utility maximizer; there is no restriction on subjective beliefs. The existence of such an equilibrium is thus guaranteed. This definition also provides a measure of sophistication in terms of how correct one's prediction of the game play is.[16]

Definition 4 *With heterogeneous sophistication, in equilibrium σ^*, i is sophisticated if*

$$A_{i,s}\left(C,\sigma^*_{-i}\right)=\sum_{n=1}^{L^{(I-1)}}\Pr\left(C^n_{-i}\text{ played under }\sigma^*_{-i}\text{ given }B_{-i}\right)a_s\left(C,C^n_{-i}\right),\text{ for all }s\text{ and }C,$$

where $\Pr\left(C^n_{-i}\text{ played under }\sigma^*_{-i}\text{ given }B_{-i}\right)$ *is the objective (correct) probability of C^n_{-i} being played under σ^*_{-i} given B_{-i}.*

The above definition implies that if i is sophisticated, she plays a best response against others with knowledge of their beliefs and of the distribution of their preferences.

Given the properties of beliefs in Lemma 2, a set of dominated strategies can be identified. More importantly, these dominated strategies are independent of the beliefs as long as they satisfy the properties in Lemma 2.

Proposition 2 *Suppose i has at least one acceptable school, given beliefs $B_i\left(\cdot,\sigma^*_{-i}\right)$,*

(i) listing an unacceptable or the worst school as the first choice is strictly dominated;

(ii) listing an unacceptable or the worst before an acceptable school is weakly dominated;

(iii) excluding an acceptable school from the list is weakly dominated;

(iv) if $A_{i,s}\left(C,\sigma^*_{-i}\right)\in(0,1)$, $\forall s$ and $C\neq(0,...,0)$, moving s upward in the list strictly increases the probability of being accepted by s and the dominances in (ii) and (iii) become strict.

Intuitively, ranking a school first always gives a strictly positive probability of being assigned to that school, and a parent should thus list better schools first. Besides, the worst outcome of participation is being accepted by the worst school. By ranking better schools before the worst school, a parent increases her child's chance of being assigned to a better school. If a school is unacceptable, putting it at the bottom or omitting it also

[16] One common approach to model levels of sophistication is the level-k model introduced by Crawford and Iriberri (2007). In that model no player plays best response and therefore no one is fully sophisticated.

increases the likelihood of getting into better schools. The above results also hold in the case of homogeneous sophistication, and a truth-telling strategy is not dominated according to Proposition 2.

Assuming parents do not play dominated strategies, I characterize the choice probabilities. The term "choice probabilities" is defined in a broad sense. With heterogeneous beliefs, the model cannot predict the probability of each parent choosing a particular list; it can only predict the probability that a choice falls into a group of lists.

Since weak dominance creates more indeterminacy, given the results in part (iv) of Proposition 2, I consider two cases: (a) for any parent, all elements in her beliefs fall within $(0,1)$; and (b) some elements in some parents' beliefs may be zero.

2.3.2 All Elements in Beliefs Are Positive

Given $S = 4$, I assign the lists into 15 groups, g_n, $n = 1, ..., 15$. The criteria of grouping are the number and identities of schools included in the list while the order among the listed schools does not necessarily matter. The groups are of three types: (a) 5 groups in which the lists include no more than one school; (b) 6 groups which include only two-school lists; and (c) 4 groups of full lists. Appendix A.5 details the groupings and the characterization of choice probabilities, the outline of which is given below.

For type-(a) groups, the choice probabilities are the same as those in the Bayesian Nash equilibrium, since they are independent of beliefs. For the 6 groups of type (b), the grouping is only based on which two schools are included in the list but not on their ranking. For example, $(s, s', 0, 0)$ and $(s', s, 0, 0)$ are in the same group, but not $(s, s'', 0, 0)$, given $s \neq s' \neq s''$. The choice probabilities for these groups have two sources: either the two included schools are the only acceptable schools, or only one of the two is acceptable, and the other is unacceptable but better than the two excluded schools. Similar to those in the Bayesian Nash equilibrium, the contributions of both sources to the choice probabilities are weighted by mixing probabilities.

The remaining 4 groups of type (c) are differentiated by their last school. Namely,

(s, s', s'', s''') and (s', s, s'', s''') are in the same group, while (s, s', s''', s'') is in a different one, given that s, s', s'', and s''' are different. The choice probabilities for these groups come from three sources: either the top three schools are all acceptable; two of the top three are acceptable; or only one of the top three is acceptable. Again, the contributions of last two sources are weighted by the mixing probabilities.

Three points should be highlighted here: (i) none of the choice probabilities involves beliefs; (ii) after grouping, the model is complete, as it implies a unique distribution of groups given a distribution of preferences;[17] and (iii) the model implies a unique distribution of preferences given a distribution of groups, under the maintained assumptions AM.1-AM.4 and Assumptions UNACCEPTABLES and MIXING. The model is thus estimated by the (simulated) maximum likelihood:

$$\max_{\boldsymbol{\theta}} \sum_{i=1}^{I} \ln \left[\Pr \left(C_i \in g_n | \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\theta} \right) \right], n = 1, ..., 15.$$

As for a multinomial probit, given the assumptions, there is a unique solution to the above maximization problem, and thus the model is identified. This is best illustrated in full-list groups. Given any of these groups one knows which school is the worst; similarly, in a multinomial probit the best school is known. The mixing probabilities present a further complication. Fortunately, they are also identified since the choice probability of the group $(0, 0, 0, 0)$ is independent of the mixing probabilities.

2.3.3 Some Elements in Beliefs May Be Zero

When parents are allowed to make mistakes, some elements in their beliefs are likely to be zero. However, the more zeros are permitted in the beliefs, the less tractable the characterization of choice probabilities becomes. Facing this trade-off, I consider the following:

Assumption ZERO-PROB Some parents may expect that the probabilities of being assigned to School 1 are zero if it is ranked third, fourth, or omitted, while other elements in the beliefs

[17] The choice probabilities of the 15 groups add up to 1.

are always in $(0,1)$ for all parents.

School 1 has the smallest quota, 63 seats and is also the "best" school with the highest average test score of graduating students. In the data, 228 parents rank School 1 first, and it is impossible to get into School 1 unless it is ranked first.

I assume that all parents assign positive values to the probabilities of being assigned to School 1 when it is ranked 2nd because: (i) 157 parents rank School 1 second, implying that many have assigned a positive value to the probability; and (ii) it would otherwise require possibly too restrictive assumptions to characterize the choice probabilities. For similar reasons, I do not consider zero probabilities for School 2, the second "best" school.[18]

Under Assumption ZERO-PROB, if a parent does have these zero probabilities, it does not matter if School 1 is ranked third or fourth as long as it is optimal not to rank it in the top two.

Lemma 3 *Under Assumption ZERO-PROB, for parent i with at least three acceptable schools, there exists a cutoff, $\bar{u}_i \geq 0$, which is a function of i's beliefs and preferences but not of $u_{i,1}$, such that School 1 is not ranked as a top two choice if and only if $(u_{i,1} - u_{i,s'}) < \bar{u}_i$, where s' is the second best school among schools $s \neq 1$.*

If $\bar{u}_i = 0$, Lemma 3 is consistent with truth-telling behavior. When $\bar{u}_i > 0$, i takes precautions by ranking School 1 low if it is not significantly better than the others.

Since the zero probabilities create many payoff-equivalent lists, to simplify the analysis, I make the following assumption regarding the mixed strategies in equilibrium.

Assumption ZERO-PROB-MIXING Given Assumption ZERO-PROB, in addition to the mixed strategies specified in Assumption MIXING, if i's preferences are such that the list $(c_1, c_2, 1, c_4)$ is a best response, and that $u_{i,c_4} = \min\{u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4}\}$ and $u_{i,s} > 0$ for $s \neq c_4$, then i mixes among:

[18] School 2 has the second smallest quota, 227 seats. In the data, there are 431 parents ranking School 2 first. However, there are 206 parents ranking School 2 second, while 75 ranking it 3rd or 4th. Further discussion of this is found in Subsection 3.2.

(i) $(c_1, c_2, 1, c_4)$, $(c_1, c_2, c_4, 1)$, and $(c_1, c_2, 0, 0)$, if $u_{i,c_4} > 0$;

(ii) $(c_1, c_2, 1, c_4)$ and $(c_1, c_2, 0, 0)$, if $u_{i,c_4} < 0$.

If i has zeros in her beliefs, as specified in Assumption ZERO-PROB, $(c_1, c_2, c_4, 1)$ and $(c_1, c_2, 0, 0)$ are also best responses whenever $(c_1, c_2, 1, c_4)$ is a best response. Therefore, the mixing patterns should be interpreted as a combination of parents without zeros in their beliefs and those with zeros and playing these mixed strategies. The common strategy assumption then implicitly requires that the group of parents with zeros be exogenously determined. This may not be too restrictive if all parents expect a very small probability of being accepted by School 1 when it is ranked third or fourth.

Moreover, if no one has zeros in their beliefs, the mixing probabilities should be such that $(c_1, c_2, 1, c_4)$ is always being played with probability one given i's preferences. This provides a test for Assumption ZERO-PROB.

Choice Probabilities and Estimation Putting together Proposition 2, Assumptions UNACCEPTABLES, MIXING, ZERO-PROB, and ZERO-PROB-MIXING, I re-assign the lists into 18 groups. The grouping now depends on how School 1 is ranked.

The new groups can be summarized by six types: (a) 5 groups in which the lists include no more than one school; (b) 3 groups in which the lists only include School 1 and another school; (c) 3 groups in which the lists only include two schools and exclude School 1; (d) 3 groups where the lists rank all four schools and School 1 is ranked top two; (e) 3 groups where the lists rank all four schools and School 1 is ranked third; (f) 1 group where the lists rank all four schools and School 1 is ranked fourth.

The detailed characterization is again shown in Appendix A.5. The choice probabilities for groups of types (a) and (b) can be formulated the same as in the previous case. The main difference is that for types (c)-(f), one has to consider how School 1 is ranked. Namely, based on Lemma 3 and Assumption ZERO-PROB-MIXING, the choice probability takes into account weather School 1 is optimally ranked top two or not.

For example, for groups of type (c), there is now a possibility that School 1 is acceptable but is optimally not ranked top two. Thus, it may be excluded from the list because some parents expect the probability of being accepted by School 1 to be zero if it is ranked third. The characterization takes this into account for groups of types (c)-(f).

I further assume the cutoff in Lemma 3, $\bar{u}_i$, has the following form:

$$\bar{u}_i = \exp\left(\eta_0 + \boldsymbol{X}_i\eta_X + \left(\mathbf{Z}_{i,2}, \mathbf{Z}_{i,3}, \mathbf{Z}_{i,4}\right)\eta_Z + \boldsymbol{Y}_i\eta_Y\right),$$

where $\boldsymbol{Y}_i$ is correlated with beliefs but not with preferences; and the exponential function ensures $\bar{u}_i$ is non-negative. The characteristics of School 1 are excluded due to Lemma 3, and together with $\boldsymbol{Y}_i$, preferences and the cutoff are separately identified.

Let $\bar{\boldsymbol{\theta}}$ be the set of parameters including the additional mixing probabilities and the coefficients in $\bar{u}_i$. Similar to previous cases, the model is estimated by (simulated) maximum likelihood: $\max_{\bar{\boldsymbol{\theta}}} \sum_{i=1}^{I} \ln\left[\Pr\left(C_i \in \bar{g}_n | \boldsymbol{X}_i, \mathbf{Z}_i, \boldsymbol{Y}_i; \bar{\boldsymbol{\theta}}\right)\right]$, $n = 1, ..., 18$.

2.4 Relationship among the Different Cases

The relationship among the above three cases is such that: $BNE \subset Positive\_Belief \subset Zero\_Belief$. The Bayesian Nash equilibrium (BNE) under homogeneous sophistication is nested in the other two cases, where sophistication may be heterogeneous. The case where all elements in everyone's beliefs are positive ($Positive\_Belief$) is nested in the other case, where I allow some elements to be zero ($Zero\_Belief$).

Case TT, where everyone is truth-telling, is also considered in estimation. It is nested in the two cases with heterogeneous sophistication: $TT \subset Positive\_Belief \subset Zero\_Belief$. However, there is no clear nesting structure between TT and BNE. Some model selection tests will be presented along with the estimation results.

2.5 Sophistication and Incentives

To see who is more strategic in the game, it is necessary to measure parents' sophistication. Measures of individual sophistication are ideal; however, since our estimates of preferences can only tell us the distribution of preferences conditional on $(\boldsymbol{X}_i, \mathbf{Z}_i)$, the sophistication can only be measured conditional on $(\boldsymbol{X}_i, \mathbf{Z}_i)$ as well.

With either heterogeneous or homogeneous beliefs, a parent is defined as sophisticated if her beliefs are the (correct) equilibrium beliefs. With the large number of players, the empirical beliefs, $\hat{B}$, as discussed in Section 2.2.3, provide a good approximation of the equilibrium beliefs. In particular, with heterogeneity in sophistication, it is impossible to solve for the equilibrium given that the joint distribution of preference and sophistication is unknown and not estimated. All the following measures can be calculated by replacing equilibrium beliefs B^* with $\hat{B}$, and $\boldsymbol{\theta}$ with estimates $\hat{\boldsymbol{\theta}}$.

2.5.1 Probability of Observing A Given Action

Under the assumption that everyone with $(\boldsymbol{X}_i, \mathbf{Z}_i)$ plays a best response, the model can predict the choice probability for each list, $P_{i,k}^{BR} \equiv \Pr\left(C_i = C^k | \boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta}\right)$, $k = 1, ..., 41$, for $S = 4$. Given i chooses C_i, define $d_{i,k}$, such that $d_{i,k} = 1$ if $C_i = C^k$, and 0 otherwise. If i always plays a best response,

$$E\left[d_{i,k} | \boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta}\right] - P_{i,k}^{BR} = 0, \forall k.$$

One may test the hypothesis that i plays a best response by running the 41 regressions:

$$d_{i,k} - P_{i,k}^{BR} = \delta_0 + \boldsymbol{X}_i \delta_X + \mathbf{Z}_i \delta_Z + \mathbf{W}_{i,k} \delta_W + \nu_{i,k}, \forall k, \tag{3}$$

where $\mathbf{W}_{i,k}$ is a vector of variables other than $\boldsymbol{X}_i$ and $\mathbf{Z}_i$. Under the null, all coefficients $(\delta_0, \delta_X, \delta_Z, \delta_W)$ should be zero.

If C^k is a one-school list or is $(0, 0, 0, 0)$, all the coefficients should always be zero, because the model assumes parents do not make mistakes when playing these strategies. I therefore use these five regressions as placebo tests.

Under the assumption that i is always truth-telling, C^k's choice probability is:

$$P_{i,k}^{TT} \equiv \Pr\left(C^k \text{ is chosen under truth-telling } | \boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta}\right).$$

Similarly, I regress $\left(d_{i,k} - P_{i,k}^{TT}\right)$ on $(\boldsymbol{X}_i, \mathbf{Z}_i, \mathbf{W}_{i,k})$ and test the truth-telling hypothesis.

2.5.2 Incentives to Be Strategic

In real life, it is not implausible that it is costly to find a best response. The incentive to be strategic, or to play a best response, would thus affect parents' behavior.

The first incentive measure is the probability that truth-telling is a best response:

$$P_i^{TT=BR} \equiv \Pr\left(\text{truth-telling is a best response}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta}\right).$$

I assume that the cost of finding a best response is lower if truth-telling itself is a best response. Therefore, a high $P_i^{TT=BR}$ means a greater incentive for i to play a best response.

The second measure is the expected utility gain if i changes from truth-telling to best responding:

$$Gain_i \equiv \left(V_i^{BR} - V_i^{TT}\right)/V_i^{BR},$$

where V_i^{BR} is the expected utility if i always plays a best response, and V_i^{TT} is the one when she is always truth-telling.[19] If $Gain_i$ is higher, i has a greater incentive to find her best response.

I later include $P_i^{TT=BR}$ and $Gain_i$ in the 41 regressions in (3) to test if parents' behavior is affected by these incentives.

3 Reduced-Form Results

Before reporting the model estimates, I present evidence from the data which is consistent with the assumptions and model predictions.

[19]More precisely, V_i^{BR} is defined as follows: Given any realization of ε_i, i plays a best response. I calculate the expected utility given ε_i and then integrate it over all possible ε_i. Namely,

$$V_i^{BR} = \int_{\varepsilon_i} \max_{\sigma \in \Delta(\mathcal{C})} \sum_{s=1}^{S} A_s\left(\sigma, \sigma_{-i}^*\right) \max\left(u_{i,s}, 0\right) dF_\Sigma\left(\varepsilon_i\right).$$

V_i^{TT} is similarly defined.

3.1 Understanding the Rules of the Game

One of the important assumptions is that parents understand the rules of the game, and therefore their beliefs follow the structure specified in Lemma 2. I examine parents' responses to two questions in the 2002 survey: "On a scale of 0-10, what is the probability that your child is admitted into your 1st (2nd) choice?" Table 5 shows the summary statistics.[20] The empirical beliefs are calculated from the submitted lists. The empirical beliefs and self-reported beliefs share the same pattern, although they do not exactly match.

Table 5: Empirical Beliefs and Parents' Self-Reported Beliefs

	Ranked as 1st Choice				Ranked as 2nd Choice			
	Empirical	Survey Responses[b]			Empirical	Survey Responses[b]		
School	Beliefs[a]	Mean	Std. Dev	# Obs[c]	Beliefs[a]	Mean	Std. Dev	# Obs.[c]
1	26.7%	4.35	2.93	249	0%	3.00	2.24	112
2	50.7%	6.72	2.39	290	0%	5.13	2.52	189
3	100%	8.11	2.05	82	100%	6.53	2.23	206
4	100%	8.32	2.06	22	100%	7.63	2.52	40

a. Calculated from the actual submitted lists. Each entry shows the probability being accepted by the school when that school is ranked 1st or 2nd, given all other students' submitted lists.
b. Responses to the survey question: "On a scale of 0-10, what is the probability that your child is admitted into your 1st (2nd) choice?"
c. The 1st and 2nd choices are self-reported and thus are not necessarily the submitted ones.

Consistent with Lemma 2, parents on average expect that moving a school up in the list increases the probability of being accepted by that school.

3.2 Undominated Strategies, Truth-Telling, and Zero Probabilities

Lemma 2 leads to the dominated strategies in Proposition 2, and parents should not play these strategies in equilibrium. Table 6 shows the distribution of parents' first choice: 24.9% rank School 1 first, while 47.0% rank School 2 first.

Another survey question asks, "Among the schools to which you could apply, which school was the best?"[21] Among 699 valid responses, 82.8% claim School 1 as the best.

[20] Since these questions are asked after the assignment is realized, the results may just show the *ex post* probability, namely, whether or not the student has been accepted by that school. For this reason, I use their self-reported top two choices, which are not necessarily the ones they submitted.

[21] This question is not necessarily asking the parent's favorite or her most preferred school.

Table 6: Parents' First Choices and Claimed Best Schools

School	Quota	# Parents Rank It #1		# Parents Claim It as the Best[a]		Rank the Claimed Best[b] #1	#2	#3	#4	Omitted
1	63	228	(24.8%)	579	(82.8%)	186	107	163	36	25
2	227	431	(47.0%)	58	(8.3%)	49	5	0	0	0
3	310	66	(7.2%)	26	(4.3%)	11	9	4	1	0
4	360	8	(0.9%)	3	(0.4%)	0	1	0	1	0
Non-Particip.		181	(19.8%)							
Other[c]				33	(4.7%)					
Total	960	914	(100%)	699	(100%)					

a. Responses to a survey question: "Among those to which you could apply, which school was the best?"
b. Among all the parents who claim a given school as the best school, these five columns show how they rank it in the application, conditional on participating.
c. "Other" means schools other than the four schools. This may be due to misreporting/misunderstanding.

Comparing the first-choice school in submitted lists with the most recognized school, the difference is significant. This is evidence against the truth-telling hypothesis.

If everyone understands the rules, the first-choice school should never be the worst school (Proposition 2). This is consistent with the data in Table 6: only 8 parents top rank School 4, while even fewer people claim it as the best school.

Proposition 2 also predicts that the last-choice school (or the omitted school, conditional on participating) should either be an unacceptable/the worst, or a school which is impossible to get in if it is ranked low. The last five columns in Table 6 show how parents rank the claimed best school. For Schools 2, 3 and 4, only two parents rank their claimed best school fourth or omit it. Following the discussion in Section 2.3.3, this implies that zero probability is less of a concern for these schools.

However, there are 36 (6.2%) parents ranking School 1 fourth, while another 25 (4.3%) participants exclude School 1 altogether. Since School 1 has the smallest quota, only those who top rank it have a chance of getting in; and even then the probability of success is merely 26.7%. It is highly plausible that a parent might expect that there is no chance of getting into School 1 when ranking it third or fourth. This is consistent with the discussion in Section 2.3.3.

3.3 Attention on Uncertainty

Several survey questions consider parents' perceptions of the importance of 12 different factors in the choice process. Parents rate them on a scale of 1-5, with 5 being very important. Three factors are related to the game's uncertainty: (i) admission quota and the possibility of being accepted; (ii) the probability of being assigned to bad schools; and (iii) consideration of other parents' applications. Since (iii) may also be correlated with school quality because other parents' applications reveal their preferences over schools, I create $Attn\_U_i$ (attention on uncertainty) as the average of responses to the first two factors and use the third for $Attn\_Others_i$ (attention on others' application).

The nine other factors are about school quality: teachers' quality, peer quality, etc. I define $Attn\_Q_i$ (attention on quality) as the average of responses to these questions.

A sophisticated parent understands the uncertainty of other parents' behavior. This implies a positive correlation between sophistication and $Attn\_U_i$. Before I investigate of the correlation between $Attn\_U_i$ and parents' performance in Section 4.2, I first explore how family background is correlated with $Attn\_U_i$.

Table 7 presents regression results of $Attn\_U_i$ on family background and student characteristics, while controlling for $Attn\_Q_i$ and $Attn\_Others_i$. Column 1 shows family background has no significant correlation with $Attn\_U_i$ in the full sample. I exclude non-participants (column 2) and then those who submitted partial lists (columns 3 and 4). The negative coefficient on parents' income becomes significant at the 10% level and larger in magnitude, particularly in the subsample of parents who submitted full lists. As a comparison, this coefficient is significantly positive (at the 5% level) for the sample of non-participants (column 5).

The pattern of the coefficients of parents' income across subsamples is consistent with the model prediction that parents submitting partial lists are insensitive to the uncertainty in the game because of their better outside option.

The above negative correlation between $Attn\_U_i$ and parents income is robust if $Attn\_U_i$ is broken into attention on admission quota and attention on probability of getting into bad

Table 7: Attention on Factors Related to Uncertainty: Regression Analyses

	Dependent Variable: Attention on Uncertainty				
	Full Sample[a]	Participant[a]	≥2 Schools[a]	Full List[a]	Non-Participant[a]
	(1)	(2)	(3)	(4)	(5)
Mean(Dep V)	4.339	4.357	4.361	4.350	4.190
Std Dev(Dep V)	0.743	0.721	0.708	.698	0.900
$Parent\_Edu_i$	0.005	0.007	0.006	0.012	-0.015
	(0.013)	(0.014)	(0.014)	(0.015)	(0.054)
$Parent\_Inc_i$	-0.037	-0.053	-0.056*	-0.069*	0.550**
	(0.035)	(0.034)	(0.034)	(0.035)	(0.228)
Own_Score_i	0.231	0.288	0.312	0.944*	1.739
	(0.414)	(0.403)	(0.402)	(0.512)	(3.567)
$Awards_i$	0.020	0.038	0.038	0.061*	0.048
	(0.026)	(0.029)	(0.029)	(0.032)	(0.080)
$Girl_i$	-0.047	-0.036	-0.026	-0.023	-0.248
	(0.050)	(0.051)	(0.051)	(0.057)	(0.205)
$Attn\_Others_i$	-0.003	-0.011	-0.013	-0.016	0.073
	(0.021)	(0.021)	(0.021)	(0.024)	(0.093)
$Attn\_Q_i$	0.823***	0.822***	0.798***	0.750***	0.755***
	(0.056)	(0.056)	(0.056)	(0.063)	(0.237)
Constant	-0.011	-0.169	-0.143	-3.230	-12.600
	(2.138)	(2.085)	(2.078)	(2.649)	(18.430)
Observations	676	605	597	457	71
R-squared	0.270	0.294	0.281	0.279	0.364

a. The full sample includes every parent whose relevant variables are not missing. Participants are those who submits a list which is not (0,0,0,0). The subsample (>= 2 schools) includes participants whose submitted lists have at least 2 schools. And the subsample with full list are those who submit a full list. Elementary school fixed effects included. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

schools, as shown in Table A-1 in the appendix. The coefficient on income is negative in both regressions, although in one of them it is not significant. The same regression is run for $Attn\_Others_i$: there the coefficient on parents' education is significantly negative, although the one on parents' income is insignificant. I also regress $Attn\_Q_i$ on the same set of variables. Results in Table A-1 show that, contrary to those from the $Attn\_U_i$ regressions, parents' income is significantly positively correlated with $Attn\_Q_i$.

In short, there is a negative correlation between $Attn\_U_i$ and parents' income; how $Attn\_U_i$ affects parents' performance in the game is investigated in Section 4.2.

4 Model Estimation and Test Results

This section presents the estimates from four cases: (i) BNE: symmetric Bayesian Nash equilibrium with homogeneous sophistication; (ii) $Positive\_Belief$: heterogeneous sophistication with all belief elements being positive; (iii) $Zero\_Belief$: heterogeneous sophistication with some belief elements being possibly zero; and (iv) TT: truth-telling. The utility function is specified as:

$$\begin{aligned} u_{i,s} \; = \; & \alpha_s + \beta_{X,1} Own\_Score_i + \beta_{X,2} Parent\_Inc_i + \beta_{X,3} Parent\_Edu_i + \beta_{X,4} Girl_i \\ & + \beta_{X,5} Awards_i + \beta_{Z,1} Distance_{i,s} + \beta_{Z,2} Own\_Score_i \times School\_Score_s \\ & + \beta_{Z,3} Parent\_Inc_i \times School\_Score_s + \beta_{Z,4} Parent\_Edu_i \times School\_Score_s \\ & + \beta_{Z,5} Awards_i \times School\_Score_s + \beta_{Z,6} Girl_i \times School\_Score_s + \varepsilon_{is}, \end{aligned}$$

where α_s is the middle school fixed effect; $School\_Score_s$ is the (log) average test score of school s; and other variables are defined in Table 3. The part $\beta_{X,1} Own\_Score_i + \beta_{X,2} Parent\_Inc_i + \beta_{X,3} Parent\_Edu_i + \beta_{X,4} Awards_i + \beta_{X,5} Girl_i$, which is constant for any inside school, captures the quality of outside option. $(\varepsilon_{i1}, ..., \varepsilon_{i4}) \sim N(0, \Sigma)$ are i.i.d. across students and the variance of $\varepsilon_{i,1}$ is 1.

In Case $Zero\_Belief$, the cutoff $\bar{u}_i \geq 0$ as in Lemma 3 is specified as:

$$\bar{u}_i = \exp\left(\begin{array}{c} \eta_{i,e} + \eta_{X,1} Own\_Score_i + \eta_{X,2} Parent\_Inc_i + \eta_{X,3} Parent\_Edu_i \\ + \eta_{X,4} Awards_i + \eta_{X,5} Girl_i + \sum_{s=2}^{4} \eta_{Z,s} Distance_{i,s} \end{array} \right), \quad (4)$$

where, $\eta_{i,e}$ is the elementary school e's fixed effect. Elementary schools may matter as teachers usually help them with filling out applications. I also estimate the model with $Attn\_U_i$ and $Attn\_Q_i$ in $\bar{u}_i$, as these variables are possibly correlated with beliefs.

In all four cases, the choice probabilities are simulated as in a multinomial probit model by the logit-smoothed accept-reject simulator with 300 draws (details in Appendix A.6).

4.1 Estimation Results and Model Selection

Table 8 presents the coefficients of the main variables in the utility function for all four cases. Standard errors are calculated by the robust asymptotic approximation, as in McFadden and Train (2000). In $Zero\_Belief$, results from excluding and including $Attn\_U_i$ and $Attn\_Q_i$ in $\bar{u}_i$ are reported in columns (1) and (2) respectively.

As discussed previously, to distinguish between $Zero\_Belief$ and $Positive\_Belief$, the mixing probabilities which are unique to $Zero\_Belief$ provide a test. Precisely, based on Assumption ZERO-PROB-MIXING, $(c_1, c_2, c_4, 1)$ and $(c_1, c_2, 0, 0)$ are also best responses whenever $(c_1, c_2, 1, c_4)$ is a best response, while this is not true in $Positive\_Belief$. As column (1) shows, when the worst school is acceptable ($u_{c^4} > 0$), parents submit $(c_1, c_2, 0, 0)$ with probability 40.3% and submit $(c_1, c_2, c_4, 1)$ with probability 59.7%.[22] The 95% confidence intervals for these two probabilities are quite far away from zero. Column (2) shows similar results. Therefore, the assumptions in $Positive\_Belief$ are rejected. Moreover, since $BNE \subset Positive\_Belief$ and $TT \subset Positive\_Belief$, both BNE and TT are thus rejected in favor of $Zero\_Belief$. More test results will be presented shortly.

Table 9 presents marginal effects of variables in the interaction terms, as calculated with the two sets of estimates from $Zero\_Belief$. For the five variables except $Girl_i$, all have a negative effect on the utility of any school, since wealthier and more educated parents, and students with better achievements, have better outside options.

In the following, I use estimates from $Zero\_Belief$ without attention measures because (i) the estimates are not very different, as shown in Tables 8 and 9; and (ii) the original attention measures are missing for 26% observations, and results may be sensitive to imputation.

[22] Interestingly, when the worst school is unacceptable, $u_{c^4} < 0$, parents do not mix; rather they play $\left(c^1, c^2, 1, c^4\right)$ wth a probability close to one, or play $\left(c^1, c^2, 0, 0\right)$ with almost zero probability, as columns (1) and (2) show. Therefore, $Zero\_Belief$ is not an issue when $u_{c^4} < 0$. From the estimates in Column (1) of Table 8, on average, the probability that School 1 is optimally ranked out of top two and $u_{c^4} < 0 < u_1$ is 21.27%. The probability that School 1 is optimally ranked below top two and $0 < u_{c^4} < u_1$ is 6.92%. This means that ignoring $Zero\_Belief$ and using $Positive\_Belief$ would misinterpret 63 parents' revealed preferences.

Table 8: Preferences over Schools: Model Estimation Results from Different Cases

	$Zero\_Belief$		$Posit\_Belief$	BNE	TT
	(1)	(2)	(3)	(4)	(5)
$Distance_{i,s}$	-0.254***	-0.246***	-0.242***	-0.026	-0.096***
	(0.017)	(0.007)	(0.018)	(0.016)	(0.005)
$Own\_Score_i \times School\_Score_s$	11.516***	19.600***	65.545***	21.008***	21.073***
	(0.182)	(0.021)	(0.152)	(0.023)	(0.034)
$Parent\_Inc_i \times School\_Score_s$	-0.047***	0.329***	0.663***	0.195***	0.013
	(0.001)	(0.023)	(0.015)	(0.002)	(0.014)
$Parent\_Edu_i \times School\_Score_s$	-0.256***	-0.263***	-0.450***	0.159***	-0.179***
	(0.006)	(0.004)	(0.005)	(0.002)	(0.003)
$Girl_i \times School\_Score_s$	0.266	0.271***	-0.359***	-0.492***	-0.765***
	(0.164)	(0.012)	(0.063)	(0.009)	(0.038)
$Awards_i \times School\_Score_s$	0.631***	0.545***	0.688***	0.855***	0.991***
	(0.018)	(0.002)	(0.017)	(0.003)	(0.065)
Own_Score_i	-74.678***	-126.712***	-409.774***	-131.443***	-131.892***
	(0.682)	(0.176)	(0.811)	(0.159)	(0.338)
$Parent\_Inc_i$	-0.019	-2.402***	-4.537***	-1.288***	-0.378***
	(0.044)	(0.142)	(0.107)	(0.011)	(0.080)
$Parent\_Edu_i$	1.538***	1.626***	2.728***	-1.005***	1.065***
	(0.038)	(0.024)	(0.028)	(0.009)	(0.011)
$Awards_i$	-4.198***	-3.550***	-4.512***	-5.406***	-6.403***
	(0.104)	(0.020)	(0.091)	(0.025)	(0.406)
$Girl_i$	-1.647	-1.572***	2.260***	3.112***	4.979***
	(1.025)	(0.074)	(0.362)	(0.040)	(0.224)
Mixing Probabilities:					
$m_{1,4}$	0.766	0.794	0.752	0.766	0.563
	[0.689, 0.833]	[0.725, 0.852]	[0.696, 0.803]	[0.677, 0.835]	[0.368, 0.734]
$m_{1,2}$	0.179	0.156	0.212	0.190	0.369
	[0.119, 0.254]	[0.102, 0.223]	[0.165, 0.267]	[0.122, 0.272]	[0.205, 0.576]
$m_{2,4}$	1.000	1.000	1.000	0.965	0.980
	[0.005, 1.000]	[0.111, 1.000]	[1.000, 1.000]	[0.912, 0.987]	[0.030, 1.000]
Given $u_{c^4} \leq u_s \; \forall s$, Prob. submitting:[a]					
$(c^1, c^2, 0, 0)$ if $u_{c^4} > 0$	0.403	0.308			
	[0.277, 0.536]	[0.130, 0.556]			
$(c^1, c^2, c^4, 1)$ if $u_{c^4} > 0$	0.597	0.499			
	[0.464, 0.723]	[0.222, 0.719]			
$(c^1, c^2, 0, 0)$ if $u_{c^4} < 0$	0.000	0.077			
	[0.000, 0.005]	[0.035, 0.173]			
LR Test: $\chi^2_{(11)}$	213.560	217.999	127.698	681.616	244.072
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

In Case $Zero\_Belief$, column (1) are estimates without attention measures, while column (2) are those which include attention measures in the cutoff function, $\bar{u}_i$. Middle school fixed effects are included in all cases.
$m_{k,l}$, $l \geq k$, denotes the probability that a l-school is submitted when only k schools are acceptable.
a. These are the probabilities of submitting each list when $(c^1, c^2, 1, c^4)$ is a best response in $Zero\_Belief$.
The likelihood test is for the hypothesis that coefficients of the 11 individual characteristics equal zero.
95% confidence intervals in brackets for mixing probabilities, as logistic functions are used in the estimation to ensure all values in [0,1]. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Marginal Effects of Individual Characteristics: Estimates from Case Zero_Belief

	$Parent\_Inc_i$		Own_Score_i		$Parent\_Edu_i$		$Awards_i$		$Girl_i$	
School	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1	-0.318	-0.321	-1.822	-2.713	-0.079	-0.040	-0.203	-0.105	0.034	0.146
	(0.042)	(0.009)	(0.535)	(0.217)	(0.007)	(0.006)	(0.020)	(0.013)	(0.026)	(0.019)
2	-0.315	-0.343	-2.597	-4.030	-0.062	-0.023	-0.263	-0.141	0.009	0.127
	(0.042)	(0.008)	(0.523)	(0.216)	(0.007)	(0.006)	(0.020)	(0.013)	(0.022)	(0.019)
3	-0.313	-0.352	-2.919	-4.579	-0.055	-0.015	-0.263	-0.157	0.009	0.120
	(0.042)	(0.007)	(0.519)	(0.215)	(0.007)	(0.006)	(0.019)	(0.013)	(0.022)	(0.019)
4	-0.310	-0.378	-3.822	-6.116	-0.035	0.005	-0.312	-0.199	-0.012	0.098
	(0.042)	(0.007)	(0.506)	(0.214)	(0.007)	(0.006)	(0.019)	(0.013)	(0.027)	(0.019)

For each variable, columns (1) are calculated using estimates from the $Zero\_Belief$ case without attention measures, while columns (2) are calculated using those from the $Zero\_Belief$ case including attention measures as independent variables. For $Parent\_Inc_i$ and Own_Score_i, the table reports the change in each school's utility, in percentage points, if there is a 1% increase in the variable. For the other 3 variables, it reports the change in utility when the variable is increased by 1 unit. Standard errors in parentheses.

4.2 Sophistication and Incentives to Be Strategic

This section investigates who strategizes better and how parents response to incentives. Measures of sophistication and incentives are constructed using estimated preferences and empirical equilibrium beliefs.

4.2.1 Deviations from Best Responding and Truth-Telling: Overcautiousness

As Section 2.5.1 shows, deviations from best responding are on average zero if everyone plays best responses, as are deviations from truth-telling if everyone is truth-telling. For the 24 full lists, Table 10 presents summary statistics on how observed behaviors deviate from best responding and truth-telling. I run a t-test for the null hypothesis that each mean independently equals zero, and in only 4 cases I fail to reject the null. This is true for deviations both from best responding and truth-telling.

Table 10 highlights the importance of distinguishing between best responding and truth telling, as they lead to different predictions. For example, $(1, 2, 3, 4)$ is the most common true preference order, and it is played by 13.89% of the parents. The truth-telling hypothesis predicts that 25.19% should choose that list, but the best-responding hypothesis only predicts 10.65%. A similar discrepancy is found for the most under-used list, $(1, 3, 2, 4)$.

Table 10: Deviation from Best-Responding and Truth-Telling Predictions: Full Lists

Rank-Ordered Lists	Observed Data Percent[a]	Deviation from the Prediction of Best Responding: Mean[b] (10^{-2})	Best Responding: Std. Dev.	Truth-Telling: Mean[b] (10^{-2})	Truth-Telling: Std. Dev.
$(1,2,3,4)$	13.89%	3.24***	0.346	-11.30***	0.356
$(1,2,4,3)$	2.41%	-2.31**	0.153	-2.72***	0.153
$(1,3,4,2)$	0.77%	-0.15	0.087	-0.15	0.087
$(1,3,2,4)$	5.14%	-30.20***	0.235	-17.70***	0.223
$(1,4,3,2)$	0.00%	-0.21***	0.003	-0.21***	0.003
$(1,4,2,3)$	0.22%	-0.30**	0.047	-0.30*	0.047
$(2,1,3,4)$	12.58%	10.76***	0.331	10.91***	0.331
$(2,1,4,3)$	2.74%	2.24***	0.163	2.42***	0.163
$(2,3,1,4)$	21.33%	18.50***	0.406	19.92***	0.408
$(2,3,4,1)$	4.70%	4.01***	0.211	4.29***	0.211
$(2,4,3,1)$	1.09%	0.93***	0.103	1.01***	0.104
$(2,4,1,3)$	0.77%	0.60**	0.087	0.70**	0.087
$(3,1,4,2)$	0.22%	-0.03	0.047	-0.18	0.047
$(3,1,2,4)$	0.88%	-2.55***	0.094	-6.18***	0.096
$(3,2,4,1)$	0.88%	-2.12***	0.093	-1.56***	0.093
$(3,2,1,4)$	4.16%	-3.96***	0.199	-1.26*	0.199
$(3,4,1,2)$	0.22%	0.09	0.047	0.12	0.047
$(3,4,2,1)$	0.22%	-0.23	0.047	-0.16	0.047
$(4,1,3,2)$	0.00%	0.00***	0.000	-0.01***	0.000
$(4,1,2,3)$	0.00%	-0.01***	0.000	-0.02***	0.001
$(4,2,3,1)$	0.00%	-0.03***	0.001	-0.03***	0.001
$(4,2,1,3)$	0.00%	-0.02***	0.001	-0.02***	0.001
$(4,3,1,2)$	0.00%	-0.01***	0.000	-0.01***	0.000
$(4,3,2,1)$	0.55%	0.51**	0.074	0.52**	0.074

A t-test is run for the null hypothesis that each mean equals zero. *** p<0.01, ** p<0.05, * p<0.1.
a. It is the percentage out of the total of 914 students.
b. The means should be interpreted as percentages points, i.e., 3.24 means 3.24 percentage points.

I repeat this exercise for the partial lists, including $(0,...,0)$. Among the 12 two-school lists, the null is not rejected in 5 cases. More importantly, as the model predicts, deviations for the 4 one-school lists and $(0,...,0)$ have means which are not significantly different from zero, since by assumption parents do not make mistakes when playing these strategies. Details are collected in the appendix Table A-2.

These results provide additional evidence that parents are neither all-best-responding nor all-truth-telling in the game. Instead, parents are overcautious because they low rank School 1 and top rank School 2 too often. In the data, 24.51% of the parents top rank

School 1, while the model predicts that 54.88% should optimally do so given other parents' behavior. For School 2, 47.16% rank it first, while the model predicts only 6.69% should do so. To top rank School 2 while avoiding School 1 is *ex ante* rational, because School 1 is the best and only has 63 slots, while School 2 is still a very good school and has 227 slots. However, as many parents choose a "safe" strategy, the overcautiousness leads to a coordination failure.

4.2.2 Incentives to Be Strategic

Table 11 shows summary statistics of the incentive measures and how they are correlated with individual characteristics. The main result is that wealthier and more educated parents, and students with better achievements, have a lower cost of finding best responses (higher $P_i^{TT=BR}$) and a lower incentive to move away from always truth-telling (lower $Gain_i$).

The probability that truth-telling is a best response, $P_i^{TT=BR}$, is relatively high at 77.4%. However, the variation is not low: min = 42.9%, max = 99.4%, and standard deviation 9.4%. In both regressions, $Parent\_Inc_i$, $Parent\_Edu_i$, Own_Score_i, and $Awards_i$ are positively correlated with $P_i^{TT=BR}$, weather controlling for $Gain_i$ or not.

$Gain_i$ is the utility gain when changing from always truth-telling to always best responding. The mean gain is 0.029, equivalent to reducing the distance to a school by 10.8%, and its variation is also high: min = 0.000, max = 0.167 (reducing the distance by 48.2%), and standard deviation 0.019. When not controlling $P_i^{TT=BR}$, $Gain_i$ is negatively correlated with $Parent\_Inc_i$, $Parent\_Edu_i$, Own_Score_i, and $Awards_i$. Conditional on $P_i^{TT=BR}$, however, all the correlations become positive and those of $Parent\_Edu_i$, Own_Score_i, and $Awards$ are significant as well.

Below, regressions are used to investigate how the incentives affect parents' behavior. It is tempting to include both $P_i^{TT=BR}$ and $Gain_i$ in the same regression; however, this might cause multicollinearity, as the correlation between $P_i^{TT=BR}$ and $Gain_i$ is -0.945.

Table 11: Determinants of Incentive to Be Strategic: Regression Analysis

	$P_i^{TT=BR}$		$Gain_i$	
Mean(Dep V)	0.774		0.029	
Std Dev(Dep V)	0.094		0.019	
$Gain_i$		-3.469***		
		(0.138)		
$P_i^{TT=BR}$				-0.230***
				(0.005)
$Parent\_Inc_i$	0.046***	0.013***	-0.009***	0.001
	(0.004)	(0.004)	(0.002)	(0.002)
$Parent\_Edu_i$	0.013***	0.005***	-0.002***	0.001***
	(0.001)	(0.001)	(0.000)	(0.000)
Own_Score_i	0.334***	0.193***	-0.041***	0.036***
	(0.024)	(0.016)	(0.008)	(0.004)
$Awards_i$	0.039***	0.019***	-0.006***	0.003***
	(0.002)	(0.001)	(0.000)	(0.000)
$Girl_i$	-0.000	0.001	0.000	0.000
	(0.003)	(0.001)	(0.001)	(0.000)
Constant	0.792***	0.872***	0.023***	0.206***
	(0.004)	(0.003)	(0.001)	(0.004)
Obs.	914	914	914	914
R-Squared	0.797	0.959	0.654	0.930

$P_i^{TT=BR}$: probability that truth-telling is a best response.
$Gain_i$: utility gain if changing from always truth-telling to best responding.
Elementary school fixed effects are included in all regressions.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.2.3 Who Strategizes Better?

Table 12 reports how family background affects deviations from best responding. I focus on two lists, the most under-used list $(1, 3, 2, 4)$ and the most over-used $(2, 3, 1, 4)$. Neither of them is bad *ex ante* since they both rank a popular school first and a safe one second. Parents are overcautious, however; they choose $(2, 3, 1, 4)$ too often – 18.50 percentage points more often than what best responding parents would do, and the list, $(1, 3, 2, 4)$, is under-used by 30.20 percentage points.

I regress the deviations on 5 sets of regressors, with or without controlling incentive measures and/or attention measures (columns 1-5 in Table 12). For the most under-used list, not controlling for incentive and attention measures, wealthier and more educated parents and students with better achievements play $(1, 3, 2, 4)$ more often. However, family

Table 12: Who Strategizes Better: Regression Analysis of Sophistication Measures

	Most Under-Used List: (1,3,2,4)					Most Over-Used List: (2,3,1,4)				
	Deviation from Best-Responding Prediction					Deviation from Best-Responding Prediction				
	Depend. V – Mean: -0.302; Std Dev: 0.235					Depend. V – Mean: 0.185; Std Dev: 0.406				
$Parent\_Inc_i$	0.026***	0.001	-0.001	-0.003	-0.004	-0.012	0.022	0.019	0.032	0.028
	(0.008)	(0.010)	(0.010)	(0.012)	(0.011)	(0.021)	(0.027)	(0.021)	(0.029)	(0.021)
$Parent\_Edu_i$	0.017***	0.008	0.008	0.005	0.005	-0.004	0.005	0.003	0.010	0.007
	(0.004)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)
Own_Score_i	0.389***	0.242***	0.310***	0.278***	0.364***	-0.352	-0.101	-0.243	0.028	-0.166
	(0.088)	(0.092)	(0.081)	(0.106)	(0.096)	(0.223)	(0.248)	(0.222)	(0.268)	(0.237)
$Awards_i$	0.037***	0.013	0.016	0.009	0.013	-0.028**	0.001	-0.011	0.007	-0.008
	(0.009)	(0.015)	(0.013)	(0.017)	(0.014)	(0.012)	(0.017)	(0.013)	(0.018)	(0.014)
$Girl_i$	-0.001	-0.001	-0.000	-0.007	-0.006	0.023	0.023	0.022	0.041	0.041
	(0.014)	(0.014)	(0.014)	(0.016)	(0.016)	(0.027)	(0.027)	(0.027)	(0.028)	(0.028)
$P_i^{TT=BR}$		0.482**		0.563***			-0.880*		-1.189**	
		(0.188)		(0.206)			(0.482)		(0.520)	
$Gain_i$			-2.379***		-2.731***			4.631**		6.176***
			(0.667)		(0.741)			(2.104)		(2.272)
$Attn\_U_i$				-0.006	-0.006				0.039**	0.039**
				(0.012)	(0.012)				(0.017)	(0.017)
$Attn\_Q_i$				-0.002	-0.002				-0.026	-0.026
				(0.017)	(0.017)				(0.035)	(0.035)
$Attn\_Others_i$				-0.008	-0.008				0.018	0.018
				(0.006)	(0.006)				(0.012)	(0.012)
$P_{i,k}^{TT}$		-0.510	-0.677*	-0.603	-0.781*		-1.788	-3.317	-3.153	-5.060
		(0.372)	(0.385)	(0.399)	(0.413)		(2.856)	(3.022)	(3.009)	(3.110)
Obs.	914	914	914	810	810	914	914	914	810	810
R-Squared	0.154	0.161	0.164	0.172	0.175	0.048	0.053	0.055	0.060	0.063

Definitions of $P_i^{TT=BR}$ and $Gain_i$ are in Table 11. $P_{i,k}^{TT}$ is the probability of the list (in the dependent variable) being the true preference order. When including $P_{i,k}^{TT}$ in the first regressions, the coefficients remain similar, and the coefficient of $P_{i,k}^{TT}$ is similar to the one in other regressions. Elementary school fixed effects are included. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

background cannot eliminate the underutilization of the strategy – increasing $Parent\_Inc_i$ and $Parent\_Edu_i$ at the same time by two standard deviations would only reduce the deviation by 11.88 percentage point.

When including incentive measures, other variables become insignificant, although Own_Score_i remains significantly positive. $P_i^{TT=BR}$ reduces the deviation, $Gain_i$ increases the deviation, and $P_{i,k}^{TT}$ (the probability that $(1, 3, 2, 4)$ is the true preference order) increases the deviation marginally. When including attention measures, the coefficients on these three variables are insignificant, while the other coefficients do not change.

This implies that family background offsets some underutilization of the strategy because wealthier and more educated parents' true preference order is more likely to be a best response (higher $P_i^{TT=BR}$). This is not because they are more sophisticated: they respond to incentives in the same way as others.[23] Indeed, when deviations from truth-telling are regressed on family background, wealthier and more educated parents' report true preferences at a similar or marginally higher rate (Table A-3 in the appendix).

The last 5 columns in Table 12 show the results for the most over-used list $(2, 3, 1, 4)$. Without controlling incentive and attention measures, only $Awards_i$ has a significant effect. The two incentive measures still have a significant effect. Similar to the previous results, $P_i^{TT=BR}$ reduces the deviation, $Gain_i$ extends the deviation, and $P_{i,k}^{TT}$ has negative but insignificant coefficients. Surprisingly, attention on uncertainty, $Attn\_U_i$, increases the deviation (significant at 5%).

The same regressions are run for all other lists as well. The coefficients on family background and incentive and attention measures are mostly insignificant. For the most likely true preference order, $(1, 2, 3, 4)$, results are presented in Table A-4 in the appendix.

As a placebo test, the same regressions are run for the one-school lists and $(0, 0, 0, 0)$. By assumption, nobody makes mistakes when playing these strategies, and thus all the coefficients should be zero. Indeed, very few of them are significant.[24]

[23]If $Gain_i$ is interacted with $Parent\_Inc_i$ and $Parent\_Edu_i$, the two terms have insignificant coefficients – negative and positive respectively. Separately, if $P_i^{TT=BR}$ is interacted with $Parent\_Inc_i$ and $Parent\_Edu_i$, the two terms are insignificant as well – positive and negative respectively.

[24]Among the 25 regressions, $Parent\_Inc_i$ and $Parent\_Edu_i$ never have a significant coefficient;

To summarize, these results show that, on average, parents are overcautious as they rank School 1 low more often than they should. Paying more attention to uncertainty does not help, and sometimes exacerbate the overcautiousness. Wealthier and more educated parents are not more sophisticated, but they do slightly better because they their true preferences happen to be best responses more often.

5 Counterfactual Analyses

To analyze the welfare effects of replacing the Boston mechanism by the DA, I consider two experiments: (i) I take the empirical beliefs as the equilibrium beliefs, or equivalently, assume that other parents behave as in the data; and (ii) assume that parents are either naive or sophisticated.

5.1 Simulating Outcomes under the Two Mechanisms

Every parent reports her true preferences in the DA. The probabilities of being assigned to each school when submitting a list are obtained by drawing 20,000 profiles of preferences rankings, simulating the outcomes, and weighting them by the probability of obtaining each profile.

Under the Boston mechanism in the second experiment, the equilibrium needs to be solved as a fixed point. Eleven cases, each with 0, 10%, ..., or 100% naive parents, are considered. The naive parents in each case are randomly chosen, and the remaining parents are sophisticated. A naive parent always reports her true preferences, while a sophisticated parent plays a best response as if she knows the joint distribution of others' preferences and beliefs. In the first ten cases, the equilibrium beliefs are solved as described in Appendix A.6. When all parents are naive, the probabilities of being assigned to each school are calculated similar to that in the DA. During these calculations, the mixing probabilities are

$Attn\_U_i$, $P_i^{TT=BR}$, and $Gain_i$ are significant only in 2 regressions; and $P_{i,k}^{TT}$ has 6 significant coefficients out of 25.

held constant, as estimated from the data.

After solving the equilibrium, I simulate parents' behavior and use the equilibrium beliefs to calculate their welfare. For each parent, 1,000 profiles of preferences are constructed using random draws of errors. Each of them plays two types of games – the DA and the Boston; in the latter, parents play each case as sophisticated and as naive.

5.2 Other Parents Behave as in the Data

The first experiment considers the equilibrium in the data, so the empirical beliefs, $\hat{B}$, are taken as the equilibrium beliefs. The following results measure the effect of changing from the Boston mechanism to the DA in Beijing, while parents behave as they do in the data under the Boston mechanism. Table 13 reports the results.

Switching to the DA hurts both naive and sophisticated parents on average (columns (1) and (2)), and the utility loss is sizable – for naive parents, the average utility loss is equivalent to increasing the distance to a school by 183.9%; and similarly by 218.2% for sophisticated parents. Regression results show that this loss is smaller for wealthier and more educated parents.

Inter-personal welfare comparison is implicitly assumed when calculating utility losses, so in columns (3)-(8) I consider the probabilities of being better off, indifferent, or worse off. On average, 20.5% of naive or sophisticated parents achieve the same level of welfare in either case, as they do not participate in either mechanism. This probability is positively correlated with family background, because the outside option increases with it.

Among naive parents, only 8.0% are better off under the DA, while, surprisingly, 71.5% are worse off. Among sophisticated parents, 6.2% are better off and 73.3% are worse off under the DA. For any parent, family background reduces the probabilities of both being better off and being worse off, but it reduces the latter more quickly. This is again due to the outside option.

Table 13: Welfare Effects of Replacing the Boston Mechanism with the DA: Regression Analyses Given the Observed Equilibrium

	Mean Utility Diff[a]		Prob(Better off)		Prob(Indiff.)[b]	Prob(Worse off)	
	Naive	Sophist.	Naive	Sophist.	Naive/Sophist.	Naive	Sophist.
mean(Dep V)	-0.265	-0.294	0.080	0.062	0.205	0.715	0.733
S.D(Dep V)	(0.059)	(0.071)	(0.049)	(0.035)	(0.104)	(0.074)	(0.082)
$Parent\_Inc_i$	0.029***	0.038***	-0.013***	-0.007***	0.046***	-0.033***	-0.040***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.006)	(0.005)	(0.004)
$Parent\_Edu_i$	0.008***	0.011***	-0.006***	-0.004***	0.016***	-0.011***	-0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Own_Score_i	0.236***	0.277***	-0.128***	-0.078***	0.374***	-0.246***	-0.296***
	(0.018)	(0.018)	(0.020)	(0.014)	(0.025)	(0.021)	(0.020)
$Awards_i$	0.023***	0.029***	-0.016***	-0.012***	0.049***	-0.033***	-0.037***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$Girl_i$	-0.001	-0.002	0.001	0.001	0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Constant	-0.275***	-0.298***	0.063***	0.049***	0.209***	0.728***	0.742***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Obs	914	914	914	914	914	914	914
R-squared	0.786	0.850	0.510	0.472	0.858	0.776	0.842

Under the DA, everyone is truth-telling. Under the Boston mechanism, naive parents are truth-telling, and sophisticated ones play a best response given the empirical beliefs, $\hat{B}$.
a. The mean utility difference is defined as the average expected utility obtained under the DA minus the one obtained under the Boston mechanism.
b. The probability of being indifferent under the Boston mechanism and under the DA is the same for both naive and sophisticated parents, because these are parents who do not participate at all under either mechanism.
Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.3 Equilibrium with only Naive and Sophisticated Parents

The second experiment considers that there are only two types of parents – naive and sophisticated, with the percentage of naive parents ranging from 0%, 10%, to 100%. Given the equilibrium in the Boston mechanism in each case, every parent plays the Boston school choice game two times: truth-telling and best-responding.

In Figure 1, I report the average expected utility under both mechanisms and probabilities of different welfare changes when switching to the DA. Overall, the results are not sensitive to the fraction of naive parents, although the Boston mechanism delivers better outcomes when there are fewer naive parents.

The first subfigure shows average expected utilities under both mechanisms. If the

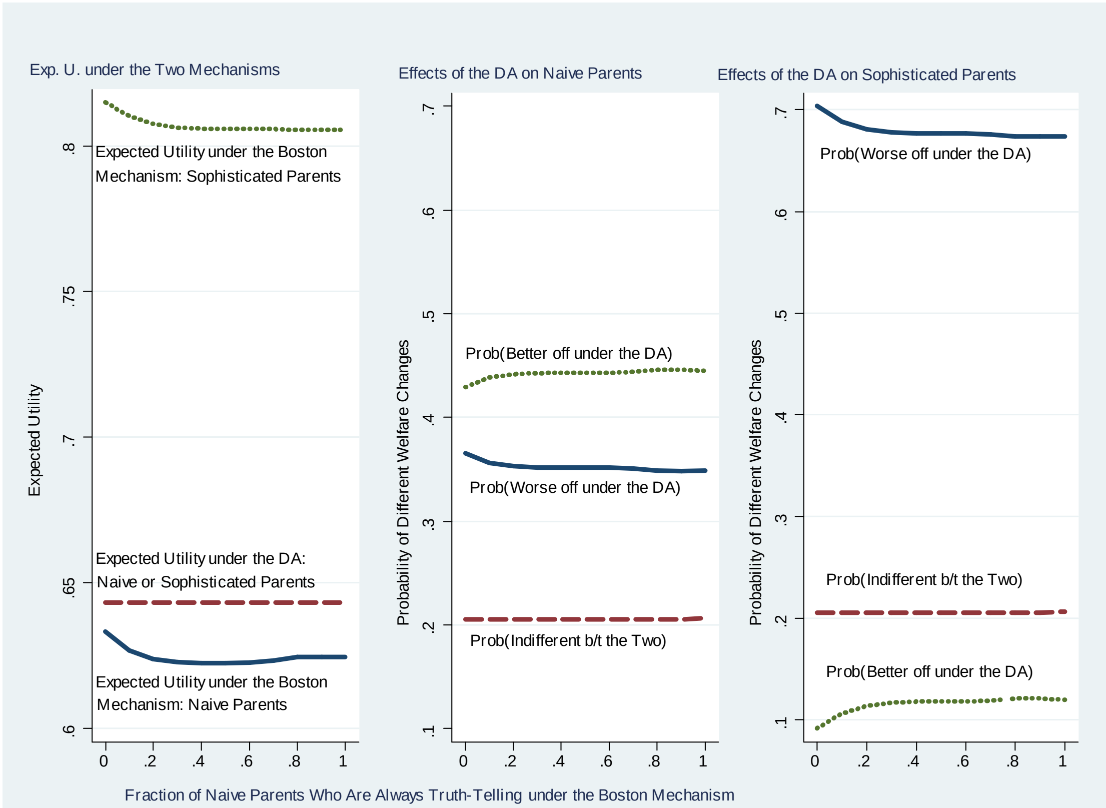

Figure 1: Welfare Effects of Replacing the Boston with the DA: Naive and Sophisticated Parents
In all simulations, parents play the DA and then the Boston mechanism. Everyone is truth-telling under the DA. Under the Boston mechanism, naive parents are always truth-telling, and sophisticated parents are always best-responding. The equilibrium is solved with fractions of naive parents from 0% to 100%. Given each equilibrium, all parents behave naively and then sophisticatedly under the Boston mechanism. Welfare effects in each case are calculated for everyone.

Boston mechanism is replaced by the DA, for sophisticated parents, their utility loss is on average 0.164, equivalent to increasing the distance to a school by 90.6%. Unlike in the first experiment, naive parents on average enjoy a utility increase in the DA, although the gain only amounts to decreasing the distance to a school by 7.1%.

The probability of being better off in the DA is about 44.2% for naive parents, but more importantly, about 35.3% of them are hurt by the DA (subfigure 2). For sophisticated ones, on average 11.5% are better off, while about 68.0% are worse off (subfigure 3).

The different results from the two experiments highlight the significance of parents' overcautiousness. Given the presence of severe overcautiousness, being naive actually helps, as it offsets some of the deviation from best responses.

6 Concluding Remarks

This paper uses data from Beijing on school choice under the Boston mechanism to answer two questions: (i) whether poor and/or less educated parents are more likely to be naive, and (ii) whether the Boston mechanism harms naive parents relative to the DA.

Assuming that students' preferences are private information, I model school choice under the Boston mechanism as a simultaneous game of incomplete information. Due to the lack of strategy-proofness, submitted choice lists are not necessarily true preferences. While allowing parents to make mistakes in the game, I derive a set of dominated strategies and assume they are not played in equilibrium. I group some lists together and characterize choice probabilities. A simulated maximum likelihood method is used for the estimation.

Results reject two hypotheses that everyone is naive and that everyone is sophisticated. Parents are revealed to be overcautious, in the sense that they avoid top ranking the most popular school more often than what they should do if they played best responses. Income and education offset the overcautiousness slightly; however, this is because wealthier/more educated parents' true preference order is more likely to be a best response. There is no evidence of them being more sophisticated. These findings are driven by the fact that such parents have a better outside option. Poorer parents pay more attention to uncertainty in the game, and this indicates that they try to find a best response. However, paying more attention to uncertainty does not help and sometimes even worsens their overcautiousness.

Given parents' behavior, especially their overcautiousness, when replacing the Boston mechanism by the DA, both naive and sophisticated parents suffer an average utility loss roughly equivalent to tripling the distance to a school. For naive parents, only 8% of them are better off under the DA, while 71.5% are worse off. The negative effects are larger for sophisticated parents, and the effects decrease with parents' income and education because of the outside option.

If every parent is either sophisticated or naive and no one is overcautious, switching from the Boston to the DA has mixed effects. For sophisticated parents, on average, their utility loss is equivalent to increasing the distance to a school by 90.6%. Among them, only 11.5% are better off, while about 68.0% are worse off. Naive parents on average have a utility gain under the DA, although the gain is only equivalent to decreasing the distance to a school by 7.1%. The probability of being better off is on average 44.2% for naive parents, but about 35.3% of them are hurt by the DA.

These results suggest that instead of replacing the Boston mechanism by the DA, to improve welfare, it may be more effective to help parents find best responses. This is especially important for poorer/less educated parents who have no good outside option. How to help parents find best responses is therefore a worthwhile avenue of future research.

References

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.

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 (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics*, 28(4), 469–482.

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

BRAUN, S., N. DWENGER, AND D. KUBLER (2010): "Telling the Truth May Not Pay Off: An Empirical Study of Centralized University Admissions in Germany," *The B.E. Journal of Economic Analysis and Policy*, 10(1).

CALSAMIGLIA, C., G. HAERINGER, AND F. KLIJN (2010): "Constrained School Choice: An Experimental Study," *American Economic Review*, 100(4), 1860–74.

CARVALHO, J. R., AND T. MAGNAC (2009): "College Choice and Entry Exams," *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.

CRAWFORD, V. P., AND N. IRIBERRI (2007): "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75(6), 1721–1770.

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

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.

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

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.

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.

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. (2010): "Are Boys Left Behind? The Evolution of the Gender Achievement Gap in Beijing's Middle Schools," *Economics of Education Review*, 29(3), 383–399.

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

——— (2011): "The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools," *Journal of Human Resources*, 46(1), 123–153.

MCFADDEN, D., AND K. TRAIN (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15(5), 447–470.

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

SCHMEIDLER, D. (1973): "Equilibrium Points of Nonatomic Games," *Journal of Statistical Physics*, 7(4), 295–300.

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

TRAIN, K. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, second edition edn.

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 Lemma 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 $\overline{q}_{s^*}$, then the number of students unassigned is $I - \sum_s q_s + \overline{q}_{s^*}$. Since $I \leq \sum_s q_s$, $I - \sum_s q_s + \overline{q}_{s^*} < \overline{q}_{s^*}$. 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. Since $I \leq \sum_s q_s$, at that point, the number of available seats is at least the number of remaining students, and 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), 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' and $K' < K$. In any realization of the game (any lottery number), 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 for sure if submitting C'. Moreover, there are cases that s is available in Round K' but not in Round K. 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 1. To show the existence of a symmetric equilibrium, I make use of the results in Schmeidler (1973) by reconstructing the Bayesian game into a game of complete information with a nonatomic continuum of players. Theorem 1 in Schmeidler (1973) establishes the existence.

Within the current incomplete information setting, each player i is facing $(I-1)$ players without knowing their types. Given the distribution of $\{u_{-i,s}\}_{s=1}^{S}$ being common knowledge, it is equivalent to say that i is playing against a continuum of players each of whom is of type $\{u_{j,s}\}_{s=1}^{S}$, $j \neq i$.

More formally, the game of school choice can be re-written into a game of complete information where the set of players T is R^S endowed with a measure μ such that for any measurable set $\hat{T} \subset T$,

$$\mu\left(\hat{T}\right) \equiv (I-1)\int 1\left(\{u_s\}_{s=1}^{S} \in \hat{T}\right) dG\left(\mathbf{X}, \mathbf{Z}\right) dF_{\Sigma}\left(\boldsymbol{\varepsilon}\right),$$

where each player is indexed by $\{u_s\}_{s=1}^{S} \in R^S$.

With some abuse of notation, now define a strategy σ as a measurable function from T to $\Delta\left(\mathcal{C}\right)$. The payoff to player $\{u_s\}_{s=1}^{S}$ is

$$V\left[\sigma\left(\{u_s\}_{s=1}^{S}\right), \sigma\right] = \sum_{s=1}^{S} A_s\left[\sigma\left(\{u_s\}_{s=1}^{S}\right), \sigma\right] \max\left(u_s, 0\right),$$

where, with $C^n = \left(C_1^n, ..., C_m^n, ..., C_{(I-1)}^n\right)$,

$$\begin{aligned} & A_s\left[\sigma\left(\{u_s\}_{s=1}^{S}\right), \sigma\right] \\ = & \sum_{n=1}^{L^{(I-1)}} \left\{\int_{\{\hat{u}_s\}_{s=1}^{S} \in T} \Pr\left[C_m^n \text{ is played under } \sigma\left(\{\hat{u}_s\}_{s=1}^{S}\right)\right] d\mu\right\} a_s\left(\sigma\left(\{u_s\}_{s=1}^{S}\right), C^n\right). \end{aligned}$$

It can be verified that the above notations are equivalent to the original ones given the independent types across parents, and $V\left[\sigma\left(\{u_s\}_{s=1}^{S}\right), \sigma\right]$ is continuous in σ.

To apply Schmeidler's theorem, one need to show that $\forall C, C' \in \mathcal{C}$ the following set is measurable,

$$\left\{\{u_s\}_{s=1}^{S} \in R^S | V\left[C, \sigma\right] > V\left[C', \sigma\right]\right\},$$

where $V\left[C, \sigma\right] > V\left[C', \sigma\right]$ is equivalent to

$$\begin{aligned} & \sum_{s=1}^{S} \left\{A_s\left[C, \sigma\right] - A_s\left[C', \sigma\right]\right\} \max\left(u_s, 0\right) \\ = & \sum_{n=1}^{L^{(I-1)}} \left\{ \begin{array}{c} \int_{\{\hat{u}_s\}_{s=1}^{S} \in T} \Pr\left[C_m^n \text{ is played under } \sigma\left(\{\hat{u}_s\}_{s=1}^{S}\right)\right] d\mu \\ \times\left[a_s\left(C, C^n\right) - a_s\left(C', C^n\right)\right] \max\left(u_s, 0\right) \end{array} \right\} \\ > & \ 0, \end{aligned}$$

which is linear in $\{u_s\}_{s=1}^{S}$. The above set is therefore measurable. By Schmeidler's Theorem 1, an equilibrium always exist.

(i) From i's perspective, for any $j \neq i$, $\Pr\left(u_{j,s} > 0 > u_{j,s'}, \text{ given s } \& \ \forall s' \neq s\right) > 0$ given the continuous distribution assumption on ε_j. In this case, since j only has one acceptable school (s), given any equilibrium beliefs, the best response for j is rank s at top. Therefore, Pr (s is top ranked by all $j \neq i$) > 0 which implies that $A_s(C, \sigma_{-i}) < 1$ for all C such that s is top ranked.

Now suppose that $A_s\left(C, \sigma^*_{-i}\right) = 1$ for some C such that s is not top ranked, there must be $A_s\left(C', \sigma^*_{-i}\right) = 1$ such that s is top ranked in C' (from Lemma 1). Therefore, $A_s\left(C, \sigma^*_{-i}\right) < 1$ for all s and for all $C \neq (0, ..., 0)$.

Similarly, conditional on being rejected by previous choices, the probability of being accepted by s is less than one, unless s is the Sth (the last) choice. Suppose that $C = \left(c^1, ..., c^K, s, c^{K+2}, ..., c^S\right)$, where $1 \leq K \leq (S-2)$ and $c^k \neq 0$, $\forall k = 1, ..., K$. Then there is a strictly positive probability that (i) $(q_{c^1} + 1)$ students' preferences are such that c^1 is the only acceptable school and (ii) q_{c^k} students' preferences are such that c^k is the only acceptable school, for $k \in \{2, ..., K, s\}$. This is true because $\sum_{k=1}^{K} q_{c^k} + q_s + 1 \leq I$ which is implied by $\sum_{s=1}^{S} q_s \geq I$ and $\sum_{s=1}^{S} q_s - q_{s'} < I$, for any s'.

For those students, their best response is to rank their only acceptable school first, and therefore, there is a positive probability that a student who submits C is rejected by s conditional on she is rejected by $c^1, ..., c^K$ as well.

Since $\sum_{s=1}^{S} A_s\left(C, \sigma^*_{-i}\right) = 1$, the above two results implies that $A_s\left(C, \sigma^*_{-i}\right) > 0$ for all s and $C \neq (0, 0, 0, 0)$.

(ii) Note that in equilibrium i's value function is $V_i\left(\sigma^*_i, \sigma^*_{-i}\right) = \sum_{s=1}^{S} A_s\left(\sigma^*_i, \sigma^*_{-i}\right) \max(u_{i,s}, 0)$. From above, $A_s\left(C, \sigma^*_{-i}\right) \in (0, 1)$ for all s and for all $C \neq (0, ..., 0)$. If $C^*_i \neq C^{**}_i \neq (0, ..., 0)$ are played with positive probability in σ^*_i given $\{u_{i,s}\}_{s=1}^{S}$ or $(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i)$, then $V_i\left(C^*_i, \sigma^*_{-i}\right) = V_i\left(C^{**}_i, \sigma^*_{-i}\right)$, or,

$$\sum_{s=1}^{S} \left[A_s\left(C^*_i, \sigma^*_{-i}\right) - A_s\left(C^{**}_i, \sigma^*_{-i}\right)\right] \max(u_{i,s}, 0) = 0. \tag{5}$$

Since $C^*_i \neq C^{**}_i$, there is at least 2 schools such that $A_s\left(C^*_i, \sigma^*_{-i}\right) \neq A_s\left(C^{**}_i, \sigma^*_{-i}\right)$, as a result of part (iv) in Proposition 2.

Since i has at most one unacceptable school and $u_{i,s} \neq u_{i,s'}$ for all $s \neq s'$ with probability one, equation (5) holds *ex ante* with probability zero. This proves that i plays mixed strategies with probability zero.

(iii) The argument in (ii) implies that if i plays a mixed strategy, i has at least two unacceptable schools with probability one.

Suppose that $C^*_i = \left(c^1_i, ..., c^S_i\right)$ is played with positive probability in σ^*_i, and that $\exists k \in \{1, ..., (S-2)\}$, $u_{i,c^k_i} < 0$ and $u_{i,c^{k+1}_i} > 0$. By applying results in Lemma 1 and (i), one may show that C^*_i must be strictly dominated by excluding the unacceptable school from the list. Therefore, any C^*_i must exclude or include the unacceptable schools at the bottom.
■

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 most preferred 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 most preferred 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, while she will be matched with the most preferred school when submitted C' instead. Thus, C is strictly dominated by 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 most preferred school.

(ii) Since including an otherwise omitted school always weakly increases the probability of being accepted by that school (Lemma 2), adding the acceptable school after the last choice of a partial list, always weakly improves the expected utility. If there are multiple acceptable schools are omitted, adding them to the list from the best to the worse will also weakly improves the expected utility.

(iii) Suppose the submitted list of i is $C = \left\{c^1, ..., c^S\right\}$ such that $c^K = \widehat{s}$ which is worst school, $1 \leq K < S$, such that $u_{i,\widehat{s}} = \min_{t=1,...,S}\left\{u_{i,t}\right\}$ and $\exists t \in \left\{K+1, ..., S\right\}$ s.t., $u_{i,c^t} > 0$. Consider an alternative list, $C' = \left\{c^1, ...c^{K-1}, \check{s}, c^{K+1}, ..., c^S\right\}$, where $u_{i,\check{s}} = \max_{t=1,...,S}\left\{u_{i,t}\right\} > 0$, i.e., replace the worst school with the best one.

Given any realization of the game, if the student is accepted by any school of $c^1, ...c^{K-1}$ when submitting C, she will be still accepted by that school when submitting C' instead. By Lemma 2, $A_{i,\widehat{s}}\left(C, \sigma^*_{-i}\right) \geq A_{i,\widehat{s}}\left(C', \sigma^*_{-i}\right)$, and the decrease in the probability is distributed to $\check{s}, c^{K+1}, ..., c^S$ and $\widehat{s}$ as well. Since $u_{i,\widehat{s}} = \min_{t=1,...,S}\left\{u_{i,t}\right\}$ and $u_{i,c^t_i} > 0$, C' weakly improves i's expected utility. Similar arguments can be made if $u_{i,\widehat{s}} < 0$.

(iv) The strict increase can be seen easily in terms of conditional probabilities. $A_{i,s}\left(C, \sigma^*_{-i}\right)$ equals:

$$\begin{aligned}
&\Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma^*_{-i}) \\
&* \Pr(i \text{ is accepted by } s \text{ given } C|\sigma^*_{-i}; i \text{ is rejected by schools ranked above } s)
\end{aligned}$$

If $A_{i,s}\left(C, \sigma^*_{-i}\right) \in (0,1), \forall s$ and $\forall C \neq (0, ..., 0)$, then the above two terms are both in $(0,1)$ unless s is ranked as Sth after all other schools.

Suppose that s is ranked as kth choice in C and ranked as k'th in C', where $1 \leq k' < k \leq S$ and the 1st to ($k'-1$)th choices are the same in both lists. We want to show that $A_{i,s}\left(C', \sigma^*_{-i}\right) > A_{i,s}\left(C, \sigma^*_{-i}\right)$.

From Lemma 2, $A_{i,s}\left(C', \sigma^*_{-i}\right) \geq A_{i,s}\left(C, \sigma^*_{-i}\right)$, if $A_{i,s}\left(C', \sigma^*_{-i}\right) = A_{i,s}\left(C, \sigma^*_{-i}\right)$, then

$$\begin{aligned}
&\Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma^*_{-i}) \\
&* \Pr(i \text{ is accepted by } s \text{ given } C|\sigma^*_{-i}; i \text{ is rejected by schools ranked above } s) \\
= \; &\Pr(i \text{ is rejected by schools ranked above } s \text{ in } C'|\sigma^*_{-i}) \\
&* \Pr(i \text{ is accepted by } s \text{ given } C'|\sigma^*_{-i}; i \text{ is rejected by schools ranked above } s).
\end{aligned}$$

We would also have

$$\begin{aligned} & \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma^*_{-i}) \\ < \;& \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C'|\sigma^*_{-i}), \end{aligned}$$

because the probability of being rejected by schools ranked between $(k'+1)$th to $(k-1)$th in C is positive. Otherwise, it would imply $A_{i,\hat{s}}\left(\hat{C}, \sigma_{-i}\right) = 0$ for some $\hat{C}$ and $\hat{s}$.

Together, the above equality and inequality imply that

$$\begin{aligned} & \Pr(i \text{ is accepted by } s \text{ given } C|\sigma^*_{-i}; i \text{ is rejected by schools ranked above } s) \\ > \;& \Pr(i \text{ is accepted by } s \text{ given } C'|\sigma^*_{-i}; i \text{ is rejected by schools ranked above } s) \end{aligned}$$

which is impossible because s ranked higher in C' than it is in C.

This proves that $A_{i,s}\left(C', \sigma^*_{-i}\right) > A_{i,s}\left(C, \sigma^*_{-i}\right)$. Similarly, the same is true if s is otherwise omitted from the list.

Since now the increase is always strict, it is straightforward to construct the strict dominance for (ii) and (iii). ■

Proof of Lemma 3.

For parent i, with equilibrium beliefs, $B_i\left(C, \sigma^*_{-i}\right) \equiv \left(A_{i,1}\left(C, \sigma^*_{-i}\right), ..., A_{i,S}\left(C, \sigma^*_{-i}\right)\right) \in [0,1)^S$, where $A_{i,1}\left(C, \sigma^*_{-i}\right) \in [0,1)$ if C has School 1 as 3rd or 4th choice and $A_{i,s}\left(C, \sigma^*_{-i}\right) \in (0,1)$ otherwise.

If School 1 is the worst or an unacceptable school for i, the lemma satisfies trivially given the results in Proposition 2.

Suppose School 1 is not the worst, and, without loss of generality, assume that $u_{i,2} > u_{i,3} > u_{i,4}$. Since there are at least three acceptable schools, one may transform i's utilities into $\hat{u}_{i,s} \equiv \max\left(u_{i,s}, 0\right) - u_{i,3}$. Therefore, $\hat{u}_{i,2} > 0$, $\hat{u}_{i,3} = 0$, $\hat{u}_{i,4} < 0$, and $\hat{u}_{i,4} < \hat{u}_{i,1} = u_{i,1} - u_{i,3}$. Therefore, parent i's problem is equivalent to:

$$\max_C \sum_{s=1}^{S} A_{i,s}\left(C, \sigma^*_{-i}\right) \hat{u}_{i,s}.$$

Denote further that $A_{i,1}\left(C^k, \sigma^*_{-i}\right) \hat{u}_{i,1} + V^k\left(C^k\right)$ as the value function when School 1 is fixed as kth choice and C^k is optimally chosen to maximize the expected utility.

The necessary and sufficient conditions for School 1 not being ranked as top 2 choices are, for $k = 1, 2$,

$$\begin{aligned} A_{i,1}\left(C^3, \sigma^*_{-i}\right) \hat{u}_{i,1} + V^3\left(C^3\right) &> A_{i,1}\left(C^k, \sigma^*_{-i}\right) \hat{u}_{i,1} + V^k\left(C^k\right), \\ A_{i,1}\left(C^4, \sigma^*_{-i}\right) \hat{u}_{i,1} + V^4\left(C^3\right) &> A_{i,1}\left(C^k, \sigma^*_{-i}\right) \hat{u}_{i,1} + V^k\left(C^k\right); \end{aligned}$$

or equivalently,

$$\hat{u}_{i,1} < \min \left\{ \begin{array}{c} \frac{V^3(C^3)-V^1(C^1)}{A_{i,1}(C^1,\sigma^*_{-i})-A_{i,1}(C^3,\sigma^*_{-i})}, \frac{V^3(C^3)-V^2(C^2)}{A_{i,1}(C^2,\sigma^*_{-i})-A_{i,1}(C^3,\sigma^*_{-i})}, \\ \frac{V^4(C^4)-V^1(C^1)}{A_{i,1}(C^1,\sigma^*_{-i})-A_{i,1}(C^4,\sigma^*_{-i})}, \frac{V^4(C^4)-V^2(C^2)}{A_{i,1}(C^2,\sigma^*_{-i})-A_{i,1}(C^4,\sigma^*_{-i})} \end{array} \right\} \equiv \bar{u}_i.$$

where all the denominators are positive by Lemma 2 and Proposition 1.

It needs to be shown that the minimum operation always returns a non-negative value, or $V^k\left(C^k\right) - V^{k'}\left(C^{k'}\right) \geq 0$ for $k = 3, 4$ and $k' = 1, 2$. Recall that $A_{i,1}\left(C^k, \sigma^*_{-i}\right)\hat{u}_{i,1} + V^k\left(C^k\right)$ is defined as the value function when kth choice is School 1. Therefore, $V^k\left(C^k\right)$ is the part of the expected utility which is from schools other than School 1 while the kth choice is taken. Under Assumption ZERO-PROB, by Part (iv) of Proposition 2**,** moving $s \neq 1$ towards the top strictly increases the probability of being accepted by s. Therefore, the arguments similar to those in the proof of Proposition 2's Part (iii) can be used to verify that $V^k\left(C^k\right) - V^{k'}\left(C^{k'}\right) \geq 0$ for $k = 3, 4$ and $k' = 1, 2$.

This proves that when $(u_{i,1} - u_{i,s'}) < \bar{u}_i$, i does not rank School 1 as top 2, and $\bar{u}_i$ is non-negative and is a function of i's beliefs and preferences but not of $u_{i,1}$. ■

A.2 Data Cleaning and Likelihood Estimation

A.2.1 Data Cleaning and Imputation

The data set that I use in this paper is a subsample of the data set that has been used in Lai, Sadoulet, and de Janvry (2009), Lai (2010), and Lai, Sadoulet, and de Janvry (2011).

Students' submitted lists are the most important variable. There are two types of "technical errors" among the lists as defined in Lai, Sadoulet, and de Janvry (2009): (i) repeated choice of a school and (ii) applying to schools not accessible to the neighborhood with the assignment system. Among the 914 students, there are 6 cases of type (i) error and 3 of type (ii) error.

The first 6 students submitted lists as follows: $(2,3,3,3)$, $(2,3,1,2)$, $(2,1,3,2)$, $(1,1,0,0)$–two cases, and $(1,2,2,3)$. I replace these lists by $(2,3,0,0)$, $(2,3,1,4)$, $(2,1,3,4)$, $(1,0,0,0)$, and $(1,2,3,4)$ respectively. The replacement for the first 4 lists is straightforward, as they are payoff-equivalent in any realization of the game. Replacing $(1,2,2,3)$ by $(1,2,3,4)$ is because this student shows a preference of School 3 over School 4. The results do not change in a few cases that I have experimented when $(1,2,2,3)$ is replaced by $(1,2,0,0)$.

The second 3 cases are those who have submitted $(2,3,1,s')$–2 cases and $(2,1,s'',4)$, where $s' \neq s'' \notin \{0,1,2,3,4\}$. I replace the first list by $(2,3,1,4)$, as they are always payoff equivalent. $(2,1,s'',4)$ is either replaced by $(2,1,0,0)$ or $(2,1,3,4)$. $(2,1,0,0)$ is payoff equivalent in the observed play of the game. I also consider $(2,1,3,4)$ as an alternative because the code for School 3 in the application is 15, while the code for s'' is 25. Therefore, it is likely that $(2,1,s'',4)$ is submitted or recorded as a typo. I present the results when $(2,1,s'',4)$ is replaced by $(2,1,3,4)$.

The main explanatory variables are $Distance_{i,s}$, Own_Score_i, $Parent\_Income_i$, $Parent\_Edu_i$, and $Award_i$.

$Distance_{i,s}$ measures the walking distance between i's home address and school s, and both addresses are from 1999. I use the Chinese version of Google Maps, http://ditu.google.cn/, to get the walking distance. Students' home addresses are from the administrative data, and there are 4 students missing home address. Their distances are assigned at the medians.

Own_Score_i is the sum of student i's scores of Chinese and math in grade 6 which is the final year of elementary school. They scores are from administrative data, but there are 125 missing values. To impute, I follow the 3 steps: (i) I regress these test scores on their test scores from the two semesters of grade 7 controlling middle school and elementary school fixed effects, then I do the out-of-sample prediction. (ii) I run similar regressions but with $Parent\_Edu_i$ as main regressor and then do out-of sample predictions. (iii) The remaining missing values are replaced by the median.

$Parent\_Income_i$ is the sum of father's and mother's income. There are 108 missing values in father's income and 100 in mother's. Some of the missing values in $Parent\_Income_i$ are replaced by the households disposable income plus the average difference between $Parent\_Income_i$ and the disposable income. I regress their own income on different combinations of their own and their spouse's education, political affiliations, and ages, the disability status (of either of them), and their spouse's income, and then do out-of-sample prediction to further impute $Parent\_Income_i$.

$Parent\_Edu_i$ is the average years of schools of parents. There are 49 missing values. I regress $Parent\_Edu_i$ on different subsets of the variables, $Parent\_Income_i$, father's and mother's political affiliations, father's and mother's job stabilities, and the disability status (of either of them). Then I do out-of-sample predictions to impute $Parent\_Edu_i$.

$Award_i$ are calculated from 6 questions in the 2002 survey. These questions ask students if they have received any awards at district level or above in 6 different categories during the six years of elementary study – all-round excellence, excellence in specific subjects, in science and technology contests, in arts and sports, in student leadership, and others. For the responses to each question, it takes one of the values of 0, 1, or missing (which is treated as 0).

A.2.2 Optimization in the Maximum Likelihood Estimation

The objective function, the negative of likelihood function in each case, is minimized mainly using Chris Sims Matlab "csminwel" algorithm which is a quasi-Newton method with BFGS update of the estimated inverse Hessian.[25] Based on my experiments, it is much faster than other Matlab packages.

As discussed in details in Appendix A.3, I use a logit-smoothed accept-reject simulator to get the choice probabilities. Different values of scale factor, λ, which determines the degree of smoothing are experimented: $\lambda = 0.05, 0.025, 0.01, 0.005$. For a given case, i.e., a given likelihood function, the following procedure is followed to get the estimates.

(i) For any given value of λ, to avoid local optimum problem, I use Sims' algorithm to minimize the objective function four times each of which has randomly chosen starting values. Among the four sets, I find the set of parameters which minimizes the objective function.

(ii) For a given λ, I find the set of estimates minimizing the objective function among the 3 sets which are estimated given a different value of λ. Using this set of parameter estimates as starting values, I minimize the objective function one more time for each given λ. The algorithm used for $\lambda = 0.01$ is the "simulannealbnd" algorithm which is a simulated annealing algorithm canned in Matlab. For other values of λ, Sims' algorithm is used again.

(iii) For each λ, I now have 5 sets of estimates from which I choose the one minimizing the objective function as the final estimates.

In the paper, I only report results from $\lambda = 0.01$, and results are similar for different values of λ. All results are available upon request.

[25]The package is available at: http://sims.princeton.edu/yftp/optimize/

A.2.3 Simulated Maximum Likelihood with Equilibrium Constraints

For the estimation in the case of Bayesian Nash equilibrium, the realized play of the game is used to calculate the equilibrium beliefs. It is also tempting to consider an alternative approach, a (simulated) maximum likelihood with equilibrium constraints as follows:

$$\max_{\boldsymbol{\theta},B} \sum_{i=1}^{I} \ln \left[\Pr \left(C_i | \mathbf{X}_i, \mathbf{Z}_i, B; \boldsymbol{\theta} \right) \right], s.t., \overline{B} \left(\cdot, \sigma_{-i} | B; \boldsymbol{\theta} \right) = B, \tag{6}$$

where the constraint restricts the beliefs to be a fixed point in equilibrium. $\overline{B}$ is defined as the "beliefs" implied by B, i.e., $\overline{B} \left(C, \sigma_{-i} | B; \boldsymbol{\theta} \right) = \left[A_1 \left(C, \sigma_{-i} | B; \boldsymbol{\theta} \right), ..., A_S \left(C, \sigma_{-i} | B; \boldsymbol{\theta} \right) \right]$, and $A_s \left(C, \sigma_{-i} | B; \boldsymbol{\theta} \right) = \sum_{n=1}^{L^{(I-1)}} \Pr \left(C_{-i}^n \text{ played under } \sigma_{-i} \text{ given } B \text{ and } \boldsymbol{\theta} \right) a_s \left(C, C_{-i}^n \right)$. $\overline{B}$ is also formulated by simulation which is described in Appendix A.6.

The problem with this approach is that given $\boldsymbol{\theta}$, there might be multiple equilibria. In other words, the fixed point, $\overline{B} \left(\cdot, \sigma_{-i} | B; \boldsymbol{\theta} \right) = B$, may not be unique for given $\boldsymbol{\theta}$, although in many numerical computation procedures there might be a unique solution. I therefore do not consider this approach in the paper. Instead, the method is used to solve the Bayesian Nash equilibrium in counterfactual analyses.

A.3 Heterogeneous information on Students' Characteristics

This appendix considers that the case when some parents know more than others about other students/parents' characteristics. Therefore, $(\boldsymbol{X}_i, \boldsymbol{Z}_i)$ is no longer private information.

Proposition 3 *Consider the following scenario:*

(i) Every parent has the same ability to process information;

(ii) Parent i also knows the realization of $\mathcal{X}^i \equiv \{\bar{\boldsymbol{X}}_{i_1}, ..., \bar{\boldsymbol{X}}_{i_F}\}$, where $\bar{\boldsymbol{X}}_i \equiv (\boldsymbol{X}_i, \boldsymbol{Z}_i)$, F is fixed and $\mathcal{X}^i$ may be different across parents.

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

Proof of Proposition 3.

A student's decision is to choose one of the L possible lists. Fix the order of all the lists, and let $\boldsymbol{d}_i = (d_{i,1}, ..., d_{i,L})'$, and $d_{i,l} = 1$ if the lth list is chosen by student i and $d_{i,l} = 0$ otherwise. Thus, $\sum_{l=1}^{L} d_{i,l} = 1$.

Without loss of generality, consider student 1's decision and suppose $\mathcal{X}^1 = \{\bar{\boldsymbol{X}}_2, ..., \bar{\boldsymbol{X}}_{F+1}\}$. Her perceived probability of other students' choices is a function of her information set – $\overline{\boldsymbol{X}}_1$, $\mathcal{X}^1$ and the distributions of $\overline{\boldsymbol{X}}_i$ and $\boldsymbol{\varepsilon}_i$, $i > 1$.

Let $(\pi_{i,1}, ..., \pi_{i,L})$ be student 1's belief about the probability that each list is being chosen by student i. Given the continuous distribution of $\boldsymbol{\varepsilon}$, $E(d_{i,l}) = \pi_{i,l} \in (0,1)$ and $\sum_{l=1}^{L} \pi_{i,l} = 1$ for all l and i.

For $i = 2, ..., F+1$, the beliefs are a function of the realization of $\mathcal{X}^1$,

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

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

$$(\pi_{i,1}, ..., \pi_{i,L}) \equiv (\overline{\pi}_1, ..., \overline{\pi}_L),$$

which is not a function of $\mathcal{X}^1$.

Consider a vector of random variables, $\boldsymbol{N}^I \equiv \sum_{i=2}^{I} \boldsymbol{d}_i = (N_1^I, N_2^I, ..., N_L^I)' \in \mathbb{N}^L$, which are the numbers of students submitting each list, i.e.,

$$N_l^I = \sum_{i=2}^{I} d_{i,l}, \sum_{l=1}^{L} N_l^I = I - 1, N_l^I \geq 0 \text{ and } N_l^I \leq I - 1.$$

In any realization of the game, $\boldsymbol{N}^I$ is a sufficient statistics to calculate the probability of being accepted by each school for Student 1, given the anonymity of the mechanism. Therefore, in the following I focus on the distribution of $\boldsymbol{N}^I$. Two definitions are also introduced:

$$\boldsymbol{\mu}^I \equiv \frac{1}{I-1}\left(\sum_{i=2}^{F+1}\pi_{i,1}+(I-F-1)\,\overline{\pi}_1,...,\sum_{i=2}^{F+1}\pi_{i,L}+(I-F-1)\,\overline{\pi}_L\right),$$

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

Consider the number of parents grows, i.e., $I \to \infty$,

$$\lim_{I\to\infty}\boldsymbol{\mu}^I = (\overline{\pi}_1,...,\overline{\pi}_L) \equiv \boldsymbol{\mu},$$

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

where Q is a finite, positive definite matrix, since it is the covariance matrix for d_i, for $i > (F+1)$.

To use the Multivariate Lindeberg–Feller Central Limit Theorem (see for example, Greene (1999), page 117), the following conditions are checked and are satisfied:

$$\lim_{I\to\infty}\left[(I-1)\,Q^I\right]^{-1} Var(\boldsymbol{d}_i) = \lim_{I\to\infty}\left(\sum_{j=2}^{I} Var(\boldsymbol{d}_j)\right)^{-1} Var(\boldsymbol{d}_i) = 0, \forall i = 2,...,I.$$

Therefore,

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

Moreover, $\lim_{I\to\infty}\sqrt{I-1}\left(\boldsymbol{\mu}^I-\boldsymbol{\mu}\right)=0$, and thus,

$$\sqrt{I-1}\left(\frac{\boldsymbol{N}^I}{I-1}-\boldsymbol{\mu}\right) \xrightarrow{d} N(\boldsymbol{0},Q)\,, as\ I \to \infty. \tag{7}$$

Similarly, when $\bar{\boldsymbol{X}}_i$ is private information, with $\widetilde{\boldsymbol{N}}^I$ as the counterpart of $\boldsymbol{N}^I$, by the Multivariate Lindberg–Levy Central Limit Theorem, one can show that

$$\sqrt{I-1}\left(\frac{\widetilde{\boldsymbol{N}}^I}{I-1}-\boldsymbol{\mu}\right)\xrightarrow{d}N\left(\mathbf{0},Q\right),as\ I\rightarrow\infty. \tag{8}$$

One need to prove that the sequences of random variables, $\sqrt{I-1}\left(\frac{\boldsymbol{N}^I}{I-1}-\boldsymbol{\mu}\right)$ and $\sqrt{I-1}\left(\frac{\widetilde{\boldsymbol{N}}^I}{I-1}-\boldsymbol{\mu}\right)$, would lead to Student 1 having the same beliefs when I grows. Namely, given $\mathbf{n}^I$ as an any realization of $\boldsymbol{N}^I$ and $\widetilde{\boldsymbol{N}}^I$,

$$\lim_{I\rightarrow\infty}\left[\Pr\left(\boldsymbol{N}^I=\mathbf{n}^I\right)-\Pr\left(\widetilde{\boldsymbol{N}}^I=\mathbf{n}^I\right)\right]=0, \tag{9}$$

which is true because of the convergence in (7) and (8), and because

$$\begin{aligned}
& \lim_{I\rightarrow\infty}\Pr\left(\boldsymbol{N}^I=n^I\right)\\
= \;& \lim_{I\rightarrow\infty}\Pr\left[\sqrt{I-1}\left(\frac{\boldsymbol{N}^I}{I-1}-\boldsymbol{\mu}\right)=\sqrt{I-1}\left(\frac{\mathbf{n}^I}{I-1}-\boldsymbol{\mu}\right)\right]\\
= \;& \lim_{I\rightarrow\infty}\Pr\left[\sqrt{I-1}\left(\frac{\boldsymbol{N}^I}{I-1}-\boldsymbol{\mu}\right)\in Ball\left(\sqrt{I-1}\left(\frac{\mathbf{n}^I}{I-1}-\boldsymbol{\mu}\right),\frac{1}{2\sqrt{I-1}}\right)\right]\\
= \;& \Phi_Q\left[\sqrt{I-1}\left(\frac{\boldsymbol{N}^I}{I-1}-\boldsymbol{\mu}\right)\in Ball\left(\sqrt{I-1}\left(\frac{\mathbf{n}^I}{I-1}-\boldsymbol{\mu}\right),\frac{1}{2\sqrt{I-1}}\right)\right]\\
= \;& \lim_{I\rightarrow\infty}\Pr\left(\widetilde{\boldsymbol{N}}^I=\mathbf{n}^I\right),
\end{aligned}$$

where $Ball\left(\sqrt{I-1}\left(\frac{\mathbf{n}^I}{I-1}-\boldsymbol{\mu}\right),\frac{1}{2\sqrt{I-1}}\right)$ is an open ball centered at $\sqrt{I-1}\left(\frac{\mathbf{n}^I}{I-1}-\boldsymbol{\mu}\right)$ with a radius of $\frac{1}{2\sqrt{I-1}}$, and Φ_Q is the distribution function for $N\left(0,Q\right)$. The second-to-last equation comes from the definition of convergence in distribution (see for example Bhattacharya (1977) in a multidimensional setting).

By definition, given the information $\mathcal{X}^1$, the beliefs of Student 1 are, $\forall s$,

$$A_{1,s}\left(C,\sigma_{-i},\mathcal{X}^1\right)=\sum_{n=1}^{L^{(I-1)}}\Pr\left(C_{-i}^n\text{ played under }\sigma_{-i}|\mathcal{X}^1\right)a_s\left(C,C_{-i}^n\right),$$

which can be re-written as:

$$A_{1,s}\left(C,\sigma_{-i},\mathcal{X}^1\right)=\sum_{\forall\mathbf{n}^I}\Pr\left(\boldsymbol{N}^I=\mathbf{n}^I\text{ under }\sigma_{-i}|\mathcal{X}^1\right)\bar{a}_s\left(C,\mathbf{n}^I\right),$$

where $\bar{a}_s\left(C, \mathbf{n}^I\right)$ is the probability that Student 1 is accepted by s while others' submitted lists are such that $\mathbf{n}^I$ is realized. By the result in (9), as $I \to \infty$,

$$\left[A_{1,s}\left(C, \sigma_{-i}, \mathcal{X}^1\right) - A_{1,s}\left(C, \sigma_{-i}\right)\right] \to 0,$$

where $A_{1,s}\left(C, \sigma_{-i}\right)$ is the one when $\bar{\boldsymbol{X}}_i$ is private information.

Since this can be proved this for any other student, the beliefs converge: $B_i\left(C\right) \to B\left(C\right), \forall i, \forall C \in \mathcal{C}$. ■

Corollary *Under the same conditions at in Proposition 3, and that $\bar{X}_i$ is now common knowledge, the beliefs converge to a common belief, $B_i\left(C\right) \to B\left(C\right)$, $\forall i$, $\forall C \in \mathcal{C}$.*

In this corollary, the difference between any two students, i and j, is their information about their opponents, the realizations of $\bar{\boldsymbol{X}}_{-i}$ and $\bar{\boldsymbol{X}}_{-j}$. However, the difference between $\bar{\boldsymbol{X}}_{-i}$ and $\bar{\boldsymbol{X}}_{-j}$ is very limited, since $\bar{\boldsymbol{X}}_{-i} = \left(\bar{\boldsymbol{X}}_{-i,j}, \bar{\boldsymbol{X}}_j\right)$ and $\bar{\boldsymbol{X}}_{-j} = \left(\bar{\boldsymbol{X}}_{-i,j}, \bar{\boldsymbol{X}}_i\right)$ where $\left(-i, j\right)$ denotes the students other than i and j. By the same argument in Proposition 3, the beliefs converge.

A.4 Assumption on Non-Participants

One of the maintained assumptions is as follows:

AM.4. A parent does not participate, or submits $(0, ..., 0)$, if and only if no school is acceptable.

It should be less of a concern that if a parent finds no school acceptable, she does not participate. I discuss why the reverse might not be true in reality and consider alternative ways to model the decision of non-participation.

In the data, as Table 2 shows, there are 181 non-participants. Among them, only 71 are enrolled in a school other than the four inside schools, and the other 110 (60.77%) are enrolled in one of the four schools. This implies that 60.77% of the non-participants may find at least one school acceptable. One possible explanation is that they take the outside option without making extra payments, as they might have succeeded in the entrance exam or have earned city-level awards. Indeed, the non-participants have higher test scores and hold more awards, as Table 4 shows. If these parents are asked to make a payment, the amount might be low since they are more likely to be well connected, given their higher income and educational attainment, or the marginal disutility of making such a payment is low for these parents. This implies that they choose the outside option even when there is a school s such that $u_{i,s} = 0$ or $u_{i,s}$ is close to zero. Therefore, the bias due to assumption AM.4 might be small.

Besides, a parent may still choose not to participate in the assignment mechanism even when she finds some schools acceptable, for the following reasons:

First, parents may be uncertainty averse or ambiguity averse. In the game, they understand that the outcome of the mechanism is uncertain, and they do not know the probabilities of each event. Uncertainty aversion thus makes parents to choose a certain outcome, the outside option in this case, even when there is an acceptable school.

Second, although they have subjective probabilities of each event, they might not use expected utility theory to make decisions, but use the prospect theory or other alternatives.

Third, in particular, the value of outside option might increase over time. For example, the lump-sum payment to the accepting school may decrease (increase) after the mechanism if there are fewer (more) people who would like to pay. Besides, after everyone gets the assignment from the mechanism, parents may have a better assessment on the peer quality at each school, and thus they may change their preferences.

Given the considerations above, it might be more realistic to consider the following assumption:

AM.4'. A parent does not participate or submits $(0, ..., 0)$ if and only if the expected utility from participating is lower than a threshold, $\underline{u}$.

Unfortunately, to calculate the expected utility, one has to specify the beliefs. Therefore, this is not feasible in the case where beliefs are allowed to be heterogeneous.

Another alternative is as follows:

AM.4". A parent does not participate or submits $(0, ..., 0)$ if and only if $u_{i,s} < \underline{u}$ for all s.

This assumption would be sufficient to identify the model. However, $\underline{u}$ has no reasonable interpretation, as it does not correspond to the cost of participation or the potential decrease in the value of outside option. I thus do not consider this either in the paper.

A.5 Characterization of Choice Probabilities

A.5.1 General Case with All-Positive Beliefs

Given $S = 4$, I group the lists into 15 groups, g_n, $n = 1, ..., 15$: There are one group with $(0,0,0,0)$, four groups with only one school being ranked, six groups with two schools being ranked, and four groups with all schools are ranked. The main criteria of grouping are the number of schools included in the list and also the identities of them. The choice probabilities now should be interpreted as the probabilities of choosing a group, g_n, or choosing any list within that group, $C_i \in g_n$.

Also recall that $m_{K,l}$, $l \geq K$, is the probability that a l-school list is submitted while only K schools are acceptable.

Characterization of Choice Probabilities When $S = 4$, the conditional probability of i choosing a group g_i, $\Pr(C_i \in g_n | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta})$, is

(i) if $C_i \in g_1 = \{(0,0,0,0)\}$, $\Pr(u_{i,s} < 0, \text{ for all } s | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta})$;

(ii) if $C_i \in g_n = \{(c^1, 0, 0, 0)\}$ ($n = 2, ..., 5$ given the identity of c^1),

$$m_{1,1} * \Pr(u_{i,c^1} > 0 > u_{i,s}, \text{ for } s \neq c^1 | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta})$$

(iii) if $C_i \in g_n = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\}$ ($n = 6, ..., 11$ given the identities of c^1 and c^2),

$$\begin{aligned} & m_{2,2} * \Pr(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{ for } s \neq c^1, c^2 | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ & + m_{1,2} * \left[\begin{array}{c} \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ + \Pr(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \end{array} \right] \end{aligned}$$

(iv) if $C_i \in g_n = \{\text{full lists s.t. the 4th is always } c^4\}$, ($n = 12, ..., 15$ given the identity of c^4),

$$\begin{aligned} & \Pr\left(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right) \\ & + m_{1,4} * \left[\begin{array}{c} \Pr(u_{i,c^1} > 0 > \max(u_{i,c^2}, u_{i,c^3}) > \min(u_{i,c^2}, u_{i,c^3}) > u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ + \Pr(u_{i,c^2} > 0 > \max(u_{i,c^1}, u_{i,c^3}) > \min(u_{i,c^1}, u_{i,c^3}) > u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ + \Pr(u_{i,c^3} > 0 > \max(u_{i,c^1}, u_{i,c^2}) > \min(u_{i,c^1}, u_{i,c^2}) u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \end{array} \right] \\ & + m_{2,4} * \left[\begin{array}{c} \Pr(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ + \Pr(u_{i,c^1}, u_{i,c^3} > 0 > u_{i,c^2} > u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \\ + \Pr(u_{i,c^2}, u_{i,c^3} > 0 > u_{i,c^1} > u_{i,c^4} | \boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \end{array} \right] \end{aligned}$$

A.5.2 General Case with Some Possibly Zero Elements in Beliefs

Putting together Proposition 2, Assumptions UNACCEPTABLES, MIXING, A-ZERO-PROB, and A-ZERO-PROB-MIXING, I re-group the lists and characterize the choice probabilities as follows:

Characterization of Choice Probabilities When $S = 4$, the conditional probability of i choosing a group $\bar{g}_n$, $n = 1, ..., 18$, $\Pr\left(C_i \in \bar{g}_n | \boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)$, is

(i) if $C_i \in \bar{g}_1 = \{(0,0,0,0)\}$, $\Pr\left(u_{i,s} < 0, \text{ for all } s|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)$;

(ii) if $C_i \in \bar{g}_n = \{(c^1,0,0,0)\}$ ($n = 2, ..., 5$ given the identity of c^1),

$$m_{1,1} * \Pr\left(u_{i,c^1} > 0 > u_{i,s}, \text{ for } s \neq c^1|\boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right)$$

(iii) if $C_i \in \bar{g}_n = \{(c^1,c^2,0,0), (c^2,c^1,0,0)\}$, s.t., c^1 or $c^2 = 1$, ($n = 6,7,8$ given the identities of c^1, c^2),

$$\begin{aligned}&m_{2,2} * \Pr\left(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\&+m_{1,2} * \left[\begin{array}{c}\Pr\left(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right)\\+\Pr\left(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right)\end{array}\right]\end{aligned}$$

(iv) if $C_i \in \bar{g}_n = \{(c^1,c^2,0,0), (c^2,c^1,0,0)\}$ s.t., $c^1, c^2 \neq 1$, ($n = 9, 10, 11$ given the identities of c^1, c^2),

$$\begin{aligned}&m_{2,2} * \Pr\left(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\&+m_{1,2} * \left[\begin{array}{c}\Pr\left(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right)\\+\Pr\left(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,s}, \text{ for } s \neq c^1, c^2|\boldsymbol{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}\right)\end{array}\right]\\&+\left[\Pr\left(u_{i,1}, u_{i,c^1}, u_{i,c^2} > 0; u_{i,1} - \min\left(u_{i,c^1}, u_{i,c^2}\right) < \bar{u}_i; u_{i,4} < 0; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\right]\\&* \Pr\left(\begin{array}{c}\text{Omitting School 1 and the worst when only the worst is unacceptable}\\\text{and School 1 is acceptable and is optimally ranked outside top 2.}\end{array}\right)\\&+\left[\Pr\left(\min\{u_{i,s}\}_{s=1}^4 > 0; u_{i,1} - \min\left(u_{i,c^1}, u_{i,c^2}\right) < \bar{u}_i; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\right]\\&* \Pr\left(\begin{array}{c}\text{Omitting School 1 and the worst when even the worst is acceptable}\\\text{and School 1 is not the worst and is optimally ranked outside top 2.}\end{array}\right)\end{aligned}$$

(v) if $C_i \in \bar{g}_n = \{$full list s.t. School 1 (denoted as c^1) is 1st or 2nd and the 4th is always $c^4\}$, ($n = 12, 13, 14$ given the identities of c^4),

$$\begin{aligned}&\Pr\left(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^1} - \min\left(u_{i,c^2}, u_{i,c^3}\right) > \bar{u}_i; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\&+m_{1,4} * \left[\begin{array}{c}\Pr\left(u_{i,c^1} > 0; u_{i,c^4} < \min\left(u_{i,c^2}, u_{i,c^3}\right) < \max\left(u_{i,c^2}, u_{i,c^3}\right) < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\+\Pr\left(u_{i,c^2} > 0; u_{i,c^4} < u_{i,c^3} < u_{i,c^1} < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\+\Pr\left(u_{i,c^3} > 0; u_{i,c^4} < u_{i,c^2} < u_{i,c^1} < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\end{array}\right]\\&+m_{2,4} * \left[\begin{array}{c}\Pr\left(u_{i,c^1}, u_{i,c^2} > 0; u_{i,c^4} < u_{i,c^3} < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\+\Pr\left(u_{i,c^1}, u_{i,c^3} > 0; u_{i,c^4} < u_{i,c^2} < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\end{array}\right]\end{aligned}$$

(vi) if $C_i \in \bar{g}_n = \{$full list C s.t. School 1 (denoted as c^3) is 3rd and the 4th is always $c^4\}$, ($n = 15, 16, 17$ given the identities of c^1 and c^2),

$$\begin{aligned}&\Pr\left(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^3} - \min\left(u_{i,c^1}, u_{i,c^2}\right) < \bar{u}_i; u_{i,c^4} < 0|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\&* \Pr\left(\begin{array}{c}\text{Including School 1 and the worst when only the worst is unacceptable}\\\text{and School 1 is acceptable and is optimally ranked outside top 2.}\end{array}\right)\\&+\Pr\left(u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 > 0; u_{i,c^3} - \min\left(u_{i,c^1}, u_{i,c^2}\right) < \bar{u}_i; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\&* \Pr\left(\begin{array}{c}\text{Ranking School 1 third when the worst is acceptable and}\\\text{School 1 is not the worst and optimally ranked outside top 2.}\end{array}\right)\\&+m_{1,4} * \left[\begin{array}{c}\Pr\left(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\\+\Pr\left(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\end{array}\right]\\&+m_{2,4} * \left[\Pr\left(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)\right]\end{aligned}$$

(vii) if $C_i \in \bar{g}_{18} = \{$full list s.t. School 1 (denoted as c^4) is ranked 4th$\}$,

$$
\begin{aligned}
&\Pr\left(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^4} = \min\left\{u_{i,s}\right\}_{s=1}^{4} |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
&+ \left[\begin{array}{c}
\Pr\left(u_{i,c^3} = \min\left\{u_{i,s}\right\}_{s=1}^{4} > 0; u_{i,c^4} - \min\left(u_{i,c^1}, u_{i,c^2}\right) < \bar{u}_i; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
+\Pr\left(u_{i,c^2} = \min\left\{u_{i,s}\right\}_{s=1}^{4} > 0; u_{i,c^4} - \min\left(u_{i,c^1}, u_{i,c^3}\right) < \bar{u}_i; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
+\Pr\left(u_{i,c^1} = \min\left\{u_{i,s}\right\}_{s=1}^{4} > 0; u_{i,c^4} - \min\left(u_{i,c^2}, u_{i,c^3}\right) < \bar{u}_i; |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)
\end{array}\right] \\
&* \Pr\left(\begin{array}{c}
\text{Ranking School 1 fourth when the worst is acceptable and} \\
\text{School 1 is not the worst and is optimally ranked outside top 2.}
\end{array}\right) \\
&+m_{1,4} * \left[\begin{array}{c}
\Pr\left(u_{i,c^1} > 0; u_{i,c^4} < \min\left(u_{i,c^2}, u_{i,c^3}\right) < \max\left(u_{i,c^2}, u_{i,c^3}\right) < 0 |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
+\Pr\left(u_{i,c^2} > 0; u_{i,c^4} < \min\left(u_{i,c^1}, u_{i,c^3}\right) < \max\left(u_{i,c^1}, u_{i,c^3}\right) < 0 |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
+\Pr\left(u_{i,c^3} > 0; u_{i,c^4} < \min\left(u_{i,c^1}, u_{i,c^2}\right) < \max\left(u_{i,c^1}, u_{i,c^2}\right) < 0 |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)
\end{array}\right] \\
&+m_{2,4} * \left[\begin{array}{c}
\Pr\left(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4} |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
+\Pr\left(u_{i,c^1}, u_{i,c^3} > 0 > u_{i,c^2} > u_{i,c^4} |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right) \\
\Pr\left(u_{i,c^2}, u_{i,c^3} > 0 > u_{i,c^1} > u_{i,c^4} |\boldsymbol{X}_i, \mathbf{Z}_i; \bar{\boldsymbol{\theta}}\right)
\end{array}\right]
\end{aligned}
$$

A.5.3 Everyone Is Truth-Telling

Now suppose everyone reports their true preference ranking. To characterize the choice probability, one also need to use Assumptions UNACCEPTABLES and MIXING because of the unacceptable schools. In the following, grouping is not needed.

Characterization of Choice Probabilities Given that researchers observe $(\boldsymbol{X}_i, \mathbf{Z}_i)$, the equilibrium beliefs B^*, and with $\boldsymbol{\theta}$ denoting the unknown parameters, in equilibrium, the conditional probability of i choosing C_i in equilibrium, $\Pr(C_i|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$, is:

(i) if $C_i = (0,0,0,0)$, $\Pr(u_{i,s} < 0$, for all $s|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$;

(ii) if $C_i = (c^1, 0, 0, 0)$,

$m_{1,1} * \Pr(u_{i,c^1} > 0 > u_{i,s}$, for $s \neq c^1|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

(iii) if $C_i = (c^1, c^2, 0, 0)$,

$m_{2,2} * \Pr(u_{i,c^1} > u_{i,c^2} > 0 > u_{i,s}$, for $s \neq c^1 \neq c^2|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

$+m_{1,2} * \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}$, for $s \neq c^1 \neq c^2\ |\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

(iv) if $C_i = (c^1, c^2, c^3, c^4)$,

$\Pr(u_{i,c^1} > u_{i,c^2} > u_{i,c^3} > 0 > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

$+m_{1,4} * \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

$+m_{2,4} * \Pr(u_{i,c^1} > u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4}|\boldsymbol{X}_i, \mathbf{Z}_i, B^*; \boldsymbol{\theta})$

A.6 The Logit-Smoothed Reject-Accept Simulator and Solving the Equilibrium

This appendix describes the logit-smoothed reject-accept simulator, how to simulate the choice probabilities, and how to find equilibrium beliefs when the parameters in the utility function and parents' strategies are given.

The simulator is implemented in the following steps, as described in Chapter 5 of Train (2009).

1. Draw a value of the 4-dimensional vector of errors, $\boldsymbol{\varepsilon}_i^r = L_i^r \boldsymbol{\eta}$, as follows:
   - Draw 4 values from a standard normal density using a random number generator. Stack these values into a vector, and label the vector $\boldsymbol{\eta}^r$. In the paper, randomized Halton sequences are used to reduced variance of the simulator, as suggested in Chapter 9 of Train (2009)
   - Calculate $\boldsymbol{\varepsilon}_i^r = L_i^r \boldsymbol{\eta}$, where L_i is the Cholesky factor of Σ.
2. Repeat Step 1 for $r = 2, ..., 300$, and calculate $u_{i,s}^r$ given $(\boldsymbol{X}_i, \mathbf{Z}_i)$ and parameters.
3. To calculate logit formula for corresponding events and/or choice probabilities, for example, $u_{i,4} = \min \{u_{i,s}\}_{s=1}^4$, I define:

$$S^r = \frac{1}{1+\exp\left[\left(u_{i,4}^r - u_{i,1}^r\right)/\lambda\right]} \frac{1}{1+\exp\left[\left(u_{i,4}^r - u_{i,2}^r\right)/\lambda\right]} \frac{1}{1+\exp\left[\left(u_{i,4}^r - u_{i,3}^r\right)/\lambda\right]},$$

   where $\lambda > 0$ is a scale factor and I experiment it with $\lambda = 0.05, 0.025, 0.01, 0.005$. Those presented in the paper is from $\lambda = 0.01$.
4. The simulated probability of the corresponding event is then $\widetilde{\Pr} = \frac{1}{300}\sum_{r=1}^{300} S^r$.

It is easy to verify that $\widetilde{\Pr}$ is strictly positive and twice differentiable.

With the help of this simulator, the following procedure can be used to solve the Bayesian Nash equilibrium in the Monte Carlo experiment. The basic idea is illustrated in the following mapping:

Given the common strategy $\sigma_i(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\varepsilon}_i; B)$, start with beliefs $B(C), \forall C \in \mathcal{C}$

$\Downarrow$

Possible Profiles: $[C, C_{-1}^n]$	Prob(accepted by each school $\vert C_{-1}^n$)	Prob(C_{-1}^n Chosen)
$[C, C_{-1}^1]$	$a_s(C, C_{-1}^1)$	$p(C_{-1}^1 \vert B)$
$[C, C_{-1}^2]$	$a_s(C, C_{-1}^2)$	$p(C_{-1}^2 \vert B)$
...	...	...
$[C, C_{-1}^{L(I-1)}]$	$a_s(C, C_{-1}^{L(I-1)})$	$p(C_{-1}^{L(I-1)} \vert B)$

$\Downarrow$

$A_s(C, B) = \sum_{n=1}^{L(I-1)} p(C_{-1}^n \vert B)\, a_s(C, C_{-1}^n), \forall C \in \mathcal{C}, \forall s = 1, ..., S$

Implied Probabilities: $\overline{B}(C, B) = (A_1(C, B), ..., A_S(C, B))$

Mapping from Beliefs to the Implied Probabilities for Student 1

Since everyone has the same beliefs, 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 NC (= 20,000) profiles of choice lists, $\left\{C_i = \left(c_i^1, ..., c_i^S\right)\right\}_{i=2}^I$, Given each profile, $\{C_i\}_{i=2}^I$, student 1 tries all (S!) full choice lists. Combine them together, I create $S! \times NC$ profiles of $\{C_i\}_{i=1}^I$. Among 20,000 profiles,

   - a quarter of them are drawn from the distribution of observed lists plus 9 "imputed" observation each of which is one of the 9 lists which are not observed in the data;
   - another quarter are random draws from the L (=41) possible lists;
   - another quarter are such that a half of the students are randomly fixed at one of the L (=41) possible lists, while the other half take a random draw from the L lists;
   - the last quarter are random draws from a distribution of true preference orders which are predicted using the estimates from the $Zero\_Belief$ case.

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

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

3. Repeat Step 2 with different lottery number draws, $rs = 2, ..., 1000$, and calculate

the probabilities of Student 1 being admitted by every s respectively.

$$\begin{aligned}
& \widetilde{\Pr}\left(\text{Student 1 assigned to } s|C_1,\{C_i\}_{i=2}^{I}\right) \\
= & \frac{1}{1000}\sum_{rs=1}^{1000}\mathbf{1}^{rs}\left(\text{Student 1 assigned to } s|C_1,\{C_i\}_{i=2}^{I}\right), s=1,...S.
\end{aligned}$$

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.

   The above 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 $r = 1,...,300$ draws of $\{\boldsymbol{\eta}_i^r\}_{i=2}^{I}$. Given the candidate belief, B, the simulated probability of student i a list C_k is

$$\widetilde{P}\left(C_k|\boldsymbol{X}_i,\{\boldsymbol{z}_s\}_{s=1}^{S};\boldsymbol{\theta}\right)=\frac{1}{300}\sum_{r=1}^{300}\mathbf{1}^{r}\left(C_i|\boldsymbol{X}_i,\boldsymbol{Z}_i;\boldsymbol{\theta};B\right), k=1,...,L,$$

   where $\mathbf{1}^{r}\left(C_i|\boldsymbol{X}_i,\boldsymbol{Z}_i;\boldsymbol{\theta};B\right)$ is an indicator function of C_i being choosing as a best response given B. Note that $\mathbf{1}^{r}\left(C_i|\boldsymbol{X}_i,\{\boldsymbol{z}_s\}_{s=1}^{S};\boldsymbol{\theta}\right)$ may be weighted by the corresponding mixing probabilities.

6. Calculate the average choice probability for the L choice lists:

$$\widetilde{P}_k=\frac{1}{914}\sum_{i=1}^{914}\widetilde{P}\left(C_i|\boldsymbol{X}_i,\{\boldsymbol{z}_s\}_{s=1}^{S};\boldsymbol{\theta}\right), k=1,...,L.$$

7. Calculate the probability of the profiles $\left\{\left\{C_i^{(t)}\right\}_{i=2}^{I}\right\}_{t=1}^{NC}$ 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

$$\widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^{I}\text{ realized}\right)=\frac{1}{K}\prod_{i=2}^{I}\left[\prod_{k=1}^{L}\left(\widetilde{P}_k\right)^{\mathbf{1}\left(C_i^{(t)}=C_k\right)}\right],$$

where K is a normalization term,

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

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

$$\begin{aligned} & \widetilde{\Pr}\left(\text{Student 1 assigned to } s \text{ when submitting } C\right) \\ = \quad & \frac{1}{NC}\sum_{t=1}^{NC} \widetilde{\Pr}\left(\text{Student 1 assigned to } s|C, \left\{C_i^{(t)}\right\}_{i=2}^{I}\right) \times \widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^{I} \text{ realized}\right). \end{aligned}$$

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

9. The equilibrium belief is a fixed point: $\bar{B}(\cdot, B) = B$.

Note that the above Steps 6 and 7 can be replaced by the following one step:

6' Calculate the probability of the profiles $\left\{\left\{C_i^{(t)}\right\}_{i=2}^{I}\right\}_{t=1}^{NC}$ 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

$$\widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^{I} \text{ realized}\right) = \frac{1}{K}\prod_{i=2}^{I} \widetilde{P}\left(C_i^{(t)}|\boldsymbol{X}_i, \{\boldsymbol{z}_s\}_{s=1}^{S}; \boldsymbol{\theta}\right),$$

where K is a normalization term,

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

The issue with this step in practice is that many $\widetilde{\Pr}\left(\left\{C_i^{(t)}\right\}_{i=2}^{I} realized\right)$ are very likely to be zero, since each individual might have a very low probability choosing a given list. Replacing Step 6' with Steps 6 and 7 solves this problem while introducing some simulation error. To be more precise, This two procedures converge to random draws from two multinomial normal distributions with the same mean but different variances (difference is bounded), as can be shown using the same arguments in Appendix A.3.

A.7 Additional Tables

Table A-1: Attention on Different Aspects of Uncertainties and School Quality

	Quota	Prob(Bad School)	Others' App.	School Quality	
	Full List[a]	Full List[a]	Full List[a]	Full Sample[a]	Full List[a]
	(1)	(2)	(3)	(4)	(5)
Mean(Dep V)	4.232	4.468	2.814	4.151	4.136
Std Dev(Dep V)	0.863	0.740	1.213	0.460	0.467
$Parent\_Edu_i$	0.016	-0.000	-0.056*	-0.001	-0.004
	(0.018)	(0.016)	(0.030)	(0.008)	(0.010)
$Parent\_Inc_i$	-0.084**	-0.003	-0.052	0.043**	0.052**
	(0.042)	(0.036)	(0.070)	(0.021)	(0.023)
Own_Score_i	0.053	1.037**	-1.003	0.224	-0.261
	(0.609)	(0.518)	(1.011)	(0.252)	(0.339)
$Awards_i$	0.025	0.048	0.022	0.008	0.002
	(0.038)	(0.033)	(0.064)	(0.016)	(0.021)
$Girl_i$	-0.020	-0.008	0.077	0.026	0.028
	(0.068)	(0.058)	(0.112)	(0.030)	(0.038)
$Attn\_Others_i$	-0.029	0.009		0.063***	0.069***
	(0.029)	(0.024)		(0.013)	(0.016)
$Attn\_Q_i$	0.529***	0.379***	0.610***		
	(0.082)	(0.071)	(0.139)		
Attn on Quota		0.339***	-0.080	0.145***	0.161***
		(0.037)	(0.079)	(0.024)	(0.030)
Prob(Bad School)	0.465***		0.033	0.156***	0.164***
	(0.051)		(0.093)	(0.020)	(0.025)
Constant	0.255	-3.95	6.824	1.198	3.544**
	(3.15)	(2.681)	(5.220)	(1.297)	(1.745)
Observations	457	457	457	676	457
R-squared	0.339	0.344	0.075	0.301	0.299

Results are from OLS regressions, and other variables include fixed effects for elementary schools.
a. The full sample includes every parent whose relevant variables are not missing, and the subsample with full lists are those who submit a full list.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-2: Deviation from Best-Responding and Truth-Telling Prediction: Partial Lists

Rank-Ordered Lists	Observed Data Percent[a]	Deviation from the Prediction of Best Responding: Mean[b] (10^{-2})	Best Responding: Std. Dev.	Truth-Telling: Mean[b] (10^{-2})	Truth-Telling: Std. Dev.
$(1,4,0,0)$	0.00%	-0.06***	0.001	-0.06***	0.001
$(1,3,0,0)$	0.77%	0.07	0.087	0.07	0.087
$(1,2,0,0)$	1.09%	-0.28	0.104	-0.28	0.104
$(2,1,0,0)$	0.66%	0.54**	0.081	0.54**	0.081
$(2,3,0,0)$	2.95%	2.63***	0.169	2.82***	0.169
$(2,4,0,0)$	0.22%	0.15	0.047	0.20	0.047
$(3,1,0,0)$	0.11%	-0.23**	0.033	-0.23**	0.033
$(3,4,0,0)$	0.22%	0.10	0.047	0.15	0.047
$(3,3,0,0)$	0.11%	-1.13***	0.034	-0.76***	0.033
$(4,1,0,0)$	0.00%	0.00***	0.000	0.00***	0.000
$(4,3,0,0)$	0.22%	0.21	0.047	0.22	0.047
$(4,2,0,0)$	0.00%	-0.01***	0.000	0.00***	0.000
$(2,0,0,0)$	0.11%	0.11	0.033	0.11	0.033
$(3,0,0,0)$	0.66%	0.01	0.081	0.01	0.081
$(1,0,0,0)$	0.22%	-0.17	0.047	-0.17	0.047
$(4,0,0,0)$	0.11%	0.03	0.033	0.03	0.033
$(0,0,0,0)$	19.80%	-0.71	0.381	-0.71	0.381

A t-test is run for the null hypothesis that each mean equals zero. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

a. It is the percentage out of the total of 914 students.

b. The means should be interpreted as percentages points, i.e., -0.06 means -0.06 percentage points.

Table A-3: Who Is Truth-Telling More Often: Regression Analysis of Deviations from Truth-Telling

	Most Under-Used List: (1,3,2,4)			Most Over-Used List: (2,3,1,4)			Most Likely True Preference: (1,2,3,4)		
	Deviation from Truth_Telling Mean: -0.177; Std Dev: 0.223			Deviation from Truth_Telling Mean: 0.199; Std Dev: 0.408			Deviation from Truth_Telling Mean: -0.113; Std Dev: 0.356		
$Parent\_Inc_i$	0.003	0.003	-0.002	-0.013	0.031	0.029	0.034*	-0.019	0.001
	(0.007)	(0.009)	(0.010)	(0.021)	(0.029)	(0.023)	(0.018)	(0.022)	(0.022)
$Parent\_Edu_i$	0.008**	0.004	0.003	-0.006	0.008	0.005	0.009	-0.012*	-0.005
	(0.004)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)	(0.006)	(0.007)	(0.007)
Own_Score_i	0.307***	0.452***	0.372***	-0.345	0.014	-0.125	0.373***	0.405**	0.681***
	(0.082)	(0.098)	(0.093)	(0.225)	(0.300)	(0.246)	(0.140)	(0.206)	(0.204)
$Awards_i$	0.015*	0.013	0.008	-0.032***	0.003	-0.009	0.014	-0.025	0.001
	(0.009)	(0.014)	(0.014)	(0.012)	(0.018)	(0.014)	(0.013)	(0.021)	(0.019)
$Girl_i$	-0.002	-0.009	-0.009	0.023	0.040	0.039	-0.004	-0.007	-0.006
	(0.014)	(0.016)	(0.016)	(0.027)	(0.028)	(0.028)	(0.023)	(0.025)	(0.025)
$P_i^{TT=BR}$		-0.366**			-0.932			1.680***	
		(0.158)			(0.607)			(0.275)	
$Gain_i$			0.867*			4.550*			-5.716***
			(0.512)			(2.496)			(0.980)
$Attn\_U_i$		-0.006	-0.006		0.039**	0.039**		-0.011	-0.011
		(0.012)	(0.012)		(0.017)	(0.017)		(0.019)	(0.019)
$Attn\_Q_i$		-0.002	-0.001		-0.028	-0.027		0.052*	0.051*
		(0.017)	(0.017)		(0.035)	(0.035)		(0.029)	(0.029)
$Attn\_Others_i$		-0.008	-0.008		0.018	0.018		-0.019*	-0.019*
		(0.006)	(0.006)		(0.012)	(0.012)		(0.011)	(0.011)
$P_{i,k}^{BR}$		-0.535**	-0.414		-0.031	-0.524		-1.652***	-1.424***
		(0.258)	(0.257)		(1.812)	(1.805)		(0.519)	(0.525)
Obs.	914	810	810	914	810	810	914	810	810
R-Squared	0.065	0.083	0.081	0.053	0.067	0.069	0.059	0.107	0.099

Definitions of $P_i^{TT=BR}$ and $Gain_i$ are in Table 11. $P_{i,k}^{TT}$ is the probability of the list (in the dependent variable) being a best response. Elementary school fixed effects are included. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-4: Who Strategizes Better: Regression Analysis of Sophistication Measures

	Most Likely True Preference Order: (1,2,3,4)				
	Dependent V.: Deviation from Best-Responding Prediction Mean: 0.0324; Std. Dev: 0.346				
$Parent\_Inc_i$	0.005	-0.000	0.006	-0.006	0.000
	(0.017)	(0.021)	(0.020)	(0.022)	(0.021)
$Parent\_Edu_i$	-0.001	-0.007	-0.007	-0.012*	-0.012*
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Own_Score_i	0.266*	0.494**	0.537***	0.632***	0.678***
	(0.136)	(0.219)	(0.172)	(0.231)	(0.182)
$Awards_i$	-0.011	-0.008	-0.005	-0.020	-0.017
	(0.013)	(0.019)	(0.015)	(0.021)	(0.016)
$Girl_i$	-0.004	-0.004	-0.005	-0.009	-0.009
	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)
$p_i^{TT=BR}$		-0.184		-0.106	
		(0.357)		(0.388)	
$Gain_i$			1.822		1.421
			(1.177)		(1.302)
$Attn\_U_i$				-0.012	-0.011
				(0.019)	(0.019)
$Attn\_Q_i$				0.053*	0.053*
				(0.028)	(0.028)
$Attn\_Others_i$				-0.018*	-0.018*
				(0.011)	(0.011)
$P_{i,k}^{TT}$		-1.062***	-1.260***	-1.136***	-1.310***
		(0.376)	(0.360)	(0.403)	(0.386)
Obs.	914	914	914	810	810
R-Squared	0.029	0.042	0.044	0.057	0.057

Definitions of $p_i^{TT=BR}$ and $Gain_i$ can be found in Table 11. $P_{i,k}^{TT}$ is the probability that the list (in the dependent variable) is the true preference ranking All regressions include a constant and elementary school fixed effects.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.