

Gaming the Boston School Choice Mechanism in Beijing

YINGHUA HE *

September 16, 2017

(First Version: November, 2009)

Abstract

The Boston mechanism has been criticized for its poor incentive and welfare performance compared with the deferred-acceptance mechanism (DA). Using school choice data from Beijing where the Boston mechanism without school priorities is adopted, I investigate parents' behavior and allow for possible mistakes. Evidence shows that parents are overcautious because they play "safe" strategies too often. There is no evidence that wealthier/more-educated parents are more adept at strategizing. If others behave as indicated in the data, an average naïve parent who always reports her true preferences experiences a utility loss in switching from the Boston to the DA mechanism (equivalent to random serial dictatorship in this setting), corresponding to an 8% increase in the distance from home to school or substituting a 13% chance at the best school with an equal chance at the second-best school. She has a 27% (55%) chance of being better (worse) off. If parents are instead either sophisticated (they always play a best response against others) or naïve, the results are mixed: under DA, naïve parents enjoy a utility gain on average when less than 80% of the population is naïve, while still about 42% are worse off and only 39% are better off. Sophisticated parents always lose more.

KEYWORDS: *Boston Immediate-Acceptance Mechanism, Gale-Shapley Deferred-Acceptance Mechanism, School Choice, Overcautious, Strategy-Proofness, Moment Inequalities, Maximin Preferences*

*Rice University and Toulouse School of Economics. Email address: yinghua.he@gmail.com. This paper was previously circulated under the title "Gaming the School Choice Mechanism." I am deeply indebted to Fang Lai for generous sharing with me the data. 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, Atila Abdulkadiroğlu, Jushan Bai, Steve Berry, Ivan Canay, Yeon-Koo Che, Pierre-André Chiappori, Navin Kartik, Wojciech Kopczuk, Dennis Kristensen, Michel Le Breton, Greg Lewis, Sanxi Li, Hong Luo, Thierry Magnac, Serena Ng, Brendan O'Flaherty, Parag Pathak, Michael Riordan, Jonah Rockoff, Johannes Schmieder, Xiaoxia Shi, Herdis Steingrimsdottir, Priscilla K. Yen, and participants at various seminars and conferences. This research has received financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement No. 295298, as well as from Columbia's Wueller and Vickrey research prizes and the Program for Economic Research.

1 Introduction

Because monetary transfers are usually precluded in the placement of students in public schools, a centralized mechanism for school assignment 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 (a.k.a. the immediate-acceptance mechanism), one of the most popular mechanisms in practice. It was used by the Boston Public Schools (BPS) from 1989 to 2005.¹ The main criticism of the mechanism is that it encourages parents to “game the system.” Namely, parents may have incentives to misreport their preferences when submitting rank-order lists of schools (Abdulkadiroglu and Sonmez 2003). On the other side, schools employ a strict priority ordering of students, usually using lotteries as tie-breakers. Each school begins with students who rank it first and assigns seats in order of students’ priority at that school. Then, each school that *still has available seats* considers *unassigned* 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 obtaining 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, may affect parents’/students’ welfare. Experimental and empirical evidence in the literature suggests that parents strategize at heterogenous levels (e.g., Abdulkadiroglu, Pathak, Roth, and Sonmez (2006), Chen and Sonmez (2006), and Lai, Sadoulet, and de Janvry (2009)). Moreover, Pathak and Sonmez (2008) consider two types: sincere (or naïve) parents, who always reveal their true preferences, and sophisticated ones always best responding. They theoretically show that the mechanism may give an advantage to sophisticated parents.

These results were instrumental in the BPS’s decision (Abdulkadiroglu, Pathak, Roth, and Sonmez 2005). The Boston School Committee voted in 2005 to replace the Boston mechanism with the student-proposing deferred-acceptance mechanism (henceforth, DA; Gale and Shapley (1962)), which is strategy-proof: reporting true preferences is a weakly dominant strategy (Dubins and Freedman 1981, Roth 1982).

One of the key arguments for the reform is that the Boston mechanism might penalize less-sophisticated parents, while DA offers them protection. 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*

¹There are many school districts that still use the mechanism, e.g., Cambridge, MA; Charlotte-Mecklenburg, NC; St. Petersburg, FL; Minneapolis, MN; and Providence, RI. It was the prevalent college admissions mechanism in China in the 1980s and 1990s and remains in practice in some provinces (Chen and Kesten 2017).

²In real life, this is known to some parents. For instance, Boston’s West Zone Parents Group, recommended two types of strategies 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.*”

strategize or do not strategize well.” More importantly, policy makers are worried that poorer 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 systematic evidence relating parents’ sophistication level to family background,³ and there are mixed, mainly theoretical and experimental, results on naïve parents’ welfare. Additionally, a recent strand of literature provides results in favor of the Boston mechanism (e.g., Featherstone and Niederle (2016), Miralles (2008)). In a setting without priorities, Abdulkadiroglu, Che, and Yasuda (2011) show that some naïve parents can even be better off under the mechanism, although this result is not robust to the introduction of priority groups (Trojan 2012). 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 naïve** and (ii) **whether the Boston mechanism harms naïve parents relative to DA.**

In the data, 914 students apply to four middle schools under a version of the Boston mechanism whereby 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.⁴ At the time of application, the lottery is unknown, and parents’ preferences are private information. Parents maximize expected utility by selecting a rank-order list of schools under uncertainties from two sources: other parents’ behavior and the lottery.

The data contain parents’ submitted lists of schools and family backgrounds 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 type I extreme value distributed errors as in conditional logit models. More importantly, the paper explicitly considers potentially heterogeneous sophistication beyond the dichotomy between naïve and sophisticated parents in the literature, which unfortunately implies that parents’ preferences and beliefs are not jointly identified. I therefore rely on identifying conditions that are independent of parents’ sophistication level or beliefs.

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

³The main difficulty is that researchers do not observe parents’ sophistication and true preferences. Taking an advantage of a web-based school choice program, Dur, Hammond, and Morrill (Forthcoming) assume students who log in the application website once as naïve and those who log in more than once as sophisticated. They correlate student sophistication with student race/ethnicity, gender, and English proficiency, although information on family background such as parental income and education is not available.

⁴In 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, DA is not *ex ante* efficient because it satisfies the other two properties.

⁵Hastings, 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 in others. I also estimate the model under the truth-telling assumption, and this assumption is rejected when tested against the model with strategic behavior.

of their preference and sophistication. A parent is sophisticated if she assesses correctly; her subjective beliefs—the perceived admission probability at each school when submitting a given list—match what is implied by the true distribution. This is rather demanding because it requires the parent to have perfect information about how others make mistakes. It is therefore important to allow for less sophisticated parents with inaccurate beliefs and even naïve ones who disregard uncertainty and always report truthfully.

While being potentially incorrect, beliefs satisfy the properties imposed by the mechanism, e.g., moving a school upward in a list (weakly) increases the probability of admission at that school. Such properties lead to a set of dominated strategies, for instance, ranking an unacceptable school first. As a minimum requirement of parents' rationality, I assume that these dominated strategies are never played in equilibrium.

For estimation, one has to consider multiple equilibrium, in particular, **multiple best responses** for each parent. This issue occurs because (i) there are unacceptable schools that are worse than an outside option and (ii) some elements in parents' beliefs might be zero or degenerate. One may arbitrarily rank unacceptable schools after acceptable ones and obtain the same payoff; it is also payoff-irrelevant to rank or to exclude a school with zero admission probability. I show that the multiplicity exists in the data and provide solutions.

Four cases, with sequential reductions in the number of assumptions, are considered and estimated: (i) truth-telling (everyone is naïve and always reports true preferences), (ii) Bayesian Nash equilibrium (everyone is sophisticated and plays a best response against others), (iii) heterogeneous sophistication with non-degenerate beliefs (some may make mistakes, but all have non-degenerate beliefs), and (iv) heterogeneous sophistication with degenerate beliefs (some may make mistakes and have degenerate beliefs). The last case is shown to fit the data significantly better than the other three.

The estimation relies on characterizing choice probabilities of rank-order lists of schools submitted by parents. The first two cases, truth-telling and Bayesian Nash equilibrium, deliver choice probability of each list as a function of data and unknown parameters. Since the model is both coherent and complete as in Tamer (2003), maximum likelihood estimation can be applied. However, in the last two cases with heterogeneous sophistication, one can only characterize choice probabilities for some lists and derive probability bounds for others. The model thus becomes incomplete (Tamer 2003). As a remedy, I group some lists together as a single outcome similar to Bresnahan and Reiss (1990), Bresnahan and Reiss (1991), and Berry (1992), resulting in a limited-information maximum likelihood estimator. To utilize the information in the bounds, I also use a method of moment equalities and inequalities à la Andrews and Shi (2013). In Monte Carlo simulations, the estimation approaches are shown to work as predicted.

The evidence from Beijing rejects the hypothesis that all parents are naïve (or sophisticated). The results, both model-based and reduced-form, imply that parents understand the rules well, but they are **overcautious**,

as they avoid top-ranking the best school more often than their best responses would prescribe.⁶ A good outside option, which is positively correlated with parents' income and student achievements, offsets overcautiousness slightly because their true preference order is more likely to be a best response. There is no evidence that high-income and more-educated parents are more sophisticated.

From linked survey data, I have 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 effects of replacing the Boston mechanism with the DA mechanism, I simulate outcomes under both mechanisms while assuming that preferences do not change across mechanisms. When schools rank all students in a single lottery, DA is equivalent to the random serial dictatorship under which students sequentially choose among the remaining schools in the order determined by the lottery. Although parents can have heterogeneous beliefs/sophistications in the estimation, I evaluate the welfare of two types, naïve and sophisticated, as a benchmark in the counterfactual. When other parents behave as in the data and therefore may have heterogeneous sophistication, an average parent, either naïve or sophisticated, suffers a significant utility loss under DA. For the naïve type, the average utility loss is 0.22, equivalent to an 8% increase in the distance from home to school (or substituting a 13% chance at the best school with an equal chance at the second best); and only 27% of them would be better off, while 55% being worse off. Sophisticated parents suffer more from such a reform: The average utility loss is -0.99, equivalent to a 40% increase in the distance or substituting a 60% chance at the best school with an equal chance at the second best school; 66% of them are worse off under DA, while only 15% are better off. Students from high-income families, or with more-educated parents or higher grades are less likely to be worse off.

In the second counterfactual, every parent is either sophisticated or naïve. Switching from the Boston to the DA mechanism has mixed effects. When there are not as many naïve parents (<80%), DA helps naïve ones while hurting sophisticated parents. However, when at least 80% of parents are naïve, no type benefits from DA. Furthermore, in terms of individual welfare change, 39% naïve parents may be better off under DA, while 42% are actually worse off. The effect of DA is more negative on sophisticated parents: Only 20% are better off, while 61% are worse off.

The different results from the two counterfactual exercises highlight the significance of parents' overcautiousness, and more generally, the importance of allowing additional behavioral types of parents beyond being naïve or sophisticated, which is another justification for allowing parents to have heterogeneous so-

⁶This 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.

phistication and heterogeneous beliefs in the preference estimation.

Overcautiousness may imply that parents are *maximin*, i.e., they maximize their expected payoff in the worst-case scenario. It is feasible to investigate this behavior hypothesis because the preference estimates based on dominated strategies are still valid when parents are maximin. Indeed, maximin predictions fit the data better on some dimensions, e.g., avoiding top-ranking the top two popular schools. However, maximin parents would have been more overcautious in that 20% (or 183) parents should have top-ranked the third-most popular school instead of ranking the two most-popular schools.

Recently, empirical studies with school choice data have surged. Several papers study the same or similar questions on the Boston mechanism with a structural approach while restricting the behavioral types of parents (Agarwal and Somaini 2016, Calsamiglia, Fu, and Güell 2016, Hwang 2016), with survey data on student preferences (De Haan, Gautier, Oosterbeek, and Van der Klaauw 2015), with survey data on individual beliefs (Kapoor, Neilson, and Zimmerman 2016), or with data on student behavior in a dynamic implementation of the mechanism (Dur, Hammond, and Morrill Forthcoming). The literature using school choice data from DA is also growing, e.g., Abdulkadiroglu, Agarwal, and Pathak (2015), Abdulkadiroglu, Angrist, Narita, and Pathak (2015), and Fack, Grenet, and He (2015).

Multiple Equilibria versus Multiple Best Responses. This paper encounters a special kind of multiple equilibria—multiple best responses—which occurs when one has several best responses given others’ strategies. In general, it implies multiple equilibria, but the reverse is not true. The literature on games focuses on multiple equilibria, while a multiplicity of best responses is more relevant in school choice games.

School choice, or centralized student-school matching, differs from usual matching games in the literature because researchers observe only matching outcomes but not agent actions in the latter, largely because those markets are often decentralized and agent actions are unobserved, and therefore, the empirical analysis relies on outcomes. For example, in the marriage market, researchers observe who matches with whom but never know to whom a husband or a wife has proposed (Choo and Siow 2006); the same is true in a range of matching games (Fox 2009). Chiappori and Salanié (2016) provide a recent survey. Without data on agent actions, it is difficult if not impossible to consider multiple best responses.

Common games in the literature usually have multiplicity in equilibrium but not necessarily in best responses. For example, in an entry model with multiple equilibria (Bresnahan and Reiss 1991), in a market profitable for only one firm to enter, entry by any of the firms is equilibrium. However, given other firms’ actions, each firm has a unique best response. In contrast, this uniqueness in best response disappears in school choice: given others’ strategies/actions, it is payoff-irrelevant if one bottom-ranks or omits her unacceptable schools and those with zero admission probability.

The recent student-school matching literature usually assumes away this multiplicity. The papers on the Boston mechanism exclude outside option and/or degenerate beliefs or impose a selection rule among best responses (Agarwal and Somaini 2016, Calsamiglia, Fu, and Güell 2016, Hwang 2016).⁷ A similar approach that assumes away the multiplicity of best responses is taken in Arcidiacono (2005).

Other Related Literature. There is a growing empirical literature on assignment problems, e.g., Budish and Cantillon (2012) on course allocations, Braun, Dwenger, and Kubler (2010) on university admissions in Germany, and Carvalho, Magnac, and Xiong (2014) on university admissions 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. Another related strand of literature is an estimation of simultaneous games of incomplete information. Most studies need consistent beliefs to derive moment conditions or choice probabilities, e.g., Seim (2006), Bajari, Hong, Krainer, and Nekipelov (2010), Aradillas-Lopez (2010), and Aradillas-Lopez (2012).

For estimation and inference, I use the results from the recent development in moment inequalities, in particular, Andrews and Shi (2013). There is a growing literature on moment inequalities and set identification, on which Tamer (2010) provides a review. These approaches have also been increasingly adopted in empirical studies, e.g., on strategic behavior in voting (Kawai and Watanabe 2013).

Organization of the Paper. Section 2 describes the two school choice mechanisms and the data from Beijing. Section 3 models school choice under the Boston mechanism as a Bayesian game. Restrictions on parents' behavior are derived, and I characterize choice probabilities and propose estimation methods. Section 4 presents reduced-form results and supporting evidence for the assumptions, and section 5 shows the model estimation. In particular, I present the correlation between sophistication and family background in both sections. Section 6 presents the counterfactual analyses for replacing the Boston with DA. The paper concludes with section 7.

2 Two Mechanisms, Background, and Data

The **DA** mechanism, works as follows:

- (i) Schools announce enrollment quotas, and students submit rank-ordered lists of schools.
- (ii) Each school forms a strict priority ordering of students with rules that are determined by state or local laws. In the Beijing data, it only depends on a lottery, while in the Boston public schools, it is determined

⁷This multiplicity in best responses may also occur in school choice under DA. Realizing this potential issue, Fack, Grenet, and He (2015) propose a method focusing on matching outcomes instead of strategies.

by sibling enrollment, distance to schools, and a lottery.

(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., a single lottery, to rank all students, DA is equivalent to the **serial dictatorship** (Abdulkadiroglu and Sonmez 1998); following their priority order determined by the lottery, students sequentially choose their favorite among the schools that still have available seats.

Similarly, the **Boston mechanism**, also known as the **Immediate-Acceptance mechanism**, asks students to submit rank-order lists, uses pre-defined 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 are no seats left or no such students left.

Generally, in

Round k . The k th 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 their k th choice in order of their priority at that school until either there are no seats left or no such students left.

The process terminates after any round k when every student is assigned a seat at a school or if the only students who remain unassigned listed no more than k choices. Unassigned students are then matched with available seats randomly. In each round, the assignment is final under the Boston mechanism but not under the DA mechanism.

2.1 The Boston Mechanism in Beijing

I study school choice in a neighborhood of Beijing's Eastern City District in 1999. The whole district, as a county-level administrative region, is divided into multiple neighborhoods each of which has access to a pre-determined set of middle schools with certain capacity quotas. The middle schools and quotas are

designed to give every student access to schools of good and medium quality.⁸ The neighborhood in question consists of 914 students who can 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. Although there are 28 public middle schools in the district in total, due to the school choice design, the neighborhood has access to only four of them when going through the mechanism, while some schools may enroll students from multiple neighborhoods (e.g., School 1 in the data). A detailed description of the education system can be found in Lai, Sadoulet, and de Janvry (2009).

The neighborhood adopted a version of the Boston mechanism in which schools' ranking over students was determined solely 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.⁹

Students' outside option was mainly to attend the 28 public schools **without going through the mechanism**, including the four to which they could apply through the mechanism. There are three ways to do so: (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.¹⁰ (ii) Besides the quota announced, some best-performing 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. It should be emphasized that a school seat that is given to the mechanism cannot be rescinded, while there are a pre-determined number of seats to be allocated outside the mechanism. Therefore, students who are admitted by a school in the mechanism are not affected by students who get access to the same school outside the mechanism.

Other possible outside options were not very relevant at that time. Specifically, private schools were not well developed in 1999. Besides, transferring out of the district were possible only when there was a formal relocation of parents or an even higher payment made to the out-of-district accepting school.

⁸This is similar to the Home-Based plan used by the Boston Public Schools since 2014. Depending on a student's home, she has a choice of at least six schools. Among them, at least four must be of high or medium quality, as determined by standardized test scores (Seelye 2013).

⁹The same mechanism was used in all of Beijing's neighborhoods in 1999, including those in other districts.

¹⁰Such payments, or "*ze xiao fei*" (literally, "fees for choosing a school"), may depend on the student's ability and parents' connections. Unfortunately, information on these payments is not publicly available. Since 2008, the education authority of Beijing has regulated that such fees for high schools cannot be more than 30,000 yuan (Source: <http://www.bjesr.cn/mbjy/2012-05-22/1131.html>). This figure is slightly above the average annual 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 gain acceptance into a high-quality elementary school in 2011 (Source: http://blog.sina.com.cn/s/blog_6ce3959f0102dr2x.html).

2.2 Data

The data in this study come from two sources: submitted lists, elementary school enrollment, grade six test scores, and home addresses in 1999 are provided in administrative data, while all other information comes from a district-level survey in early 2002. Unfortunately, there is no information on the lottery and students' initial assignments in 1999. It is worth emphasizing that for the given neighborhood, the data include every potential participant in the school choice game, although some of them choose not to participate.

Chinese middle schools offer 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 educational attainment and income, as well as retrospective information on the preparedness for making school choice decisions in 1999.

2.2.1 Heterogeneity among the Middle Schools

The four schools that the students could apply to are highly differentiated on two dimensions: enrollment quota and quality. Table 1 shows that School 1 has the smallest quota, 63 seats, although 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. School 1 also has the best quality measured by the performance of its graduating class in the high school entrance exam in 1999. The exam is city-wide and high stakes and therefore is a factor weighed heavily by parents 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.

Table 1: Middle Schools: Quota and Quality

Schools	Quota	<i>School_Score_s</i> : 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		

Notes: 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 scores among all 28 public schools in the district. Note that the 914 students are not allowed to apply to the other 24 schools through the mechanism, although they may be admitted by any of the 28 schools through other channels.

2.2.2 Student Characteristics and Behavior: A First Look

Using elementary school enrollment, I identify 914 qualified applicants in this neighborhood in 1999.¹¹ The distribution of their submitted lists in 1999 and middle school enrollment in 2002 are shown in Table 2; 181 (20%) of the students did not participate in the centralized mechanism and took their outside option directly,¹² while 110 of the them were still enrolled in one of the four schools in 2002. The majority submitted a full list with three or four schools; merely 7% submitted a partial list that ranks one or two schools. Overall, in 2002, only 92 (10%) of the students were enrolled in a school other than the four schools. The best in the district, School 1, enrolled 147 of the 914 students, more than double its quota, while 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

Notes: a. “Other” means one of the other 24 public middle schools in the district. Due to the design of the school choice program, these 24 schools accept students from other neighborhoods through the same mechanism while being available to students in the data as outside options.

Since neither students’ initial placement nor the realized lottery is observed in the data, enrollment information is not used in the estimation. Nonetheless, Table 2 indicates that not everyone enrolls in the school prescribed by the mechanism, highlighting the importance of “outside option” – the ways outside the centralized mechanism to enroll in a school. Such “non-compliance” is not unique to Beijing and is also observed in places such as Denver (Abdulkadiroglu, Angrist, Narita, and Pathak 2015) and New York City (Abdulkadiroglu, Agarwal, and Pathak 2015), although the exact channels may differ across the cities.

I then look at students’ family background ($Parent_Inc_i$, $Parent_Edu_i$), ability (Own_Score_i , $Awards_i$),

¹¹The 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) total enrollment quota is usually set to be larger than the number of students intentionally; (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, this impact is plausibly negligible. I therefore focus on the 914 observed students.

¹²This non-trivial fraction does not necessarily imply that the school choice in Beijing is fundamentally different from others due its outside option because other studies usually do not observe non-participants in school choice. In the current paper, the information on non-participants helps preference estimation, as shown later in section 3.

gender ($Girl_i$), and distance to each school ($Distance_{i,s}$). Table 3 presents their definitions and sources.

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

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 for the full sample and 3 subsamples – non-participants, participants with partial lists, and those submitting full lists. Non-participants have richer and more educated parents; they also have higher test scores and have earned more awards, which 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

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

Notes: Standard deviations are in parentheses. a. The mean here is that of all 4 distances.

3 Model: School Choice as a Bayesian Game

In this section, school choice under the Boston mechanism is formalized as a Bayesian game. To simplify presentation, the formulation below is specific to the empirical study with four schools, although the results can be readily extended to more general settings with more schools, as in an earlier version (He 2012).

The game consists of

- (i) a set of students/parents, $\{i\}_{i=1}^I$;
- (ii) a set of 4 schools, $\{s\}_{s=1}^4$, and an outside option denoted as School 0;
- (iii) a capacity vector, $\{q_s\}_{s=1}^4$; $\sum_{s=1}^4 q_s \geq I$, $\sum_{s=1}^4 q_s - q_{s'} < I$, and $q_{s'} > 0$, $\forall s'$;
- (iv) students' rank-order lists, $\{C_i\}_{i=1}^I$, where $C_i = (c_i^1, \dots, c_i^4)$, $c_i^k \in \{s\}_{s=0}^4$, $\forall k = 1, \dots, 4$;
- (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. $\sum_{s=1}^4 q_s \geq I$; no group of three schools has enough seats to accommodate all students, $\sum_{s=1}^4 q_s - q_{s'} < I$, $\forall s'$. Parents or students submit their choice lists, $C_i = (c_i^1, \dots, c_i^4)$ where $c_i^k (= 0, 1, \dots, 4)$ is the k th choice. C_i is a full list if it ranks all 4 schools (i.e., $c_i^k \neq 0$ for all k); otherwise, it is a partial list. They may submit partial lists or submit $(0, \dots, 0)$. In the latter case, the student is a non-participant and is not considered in the mechanism.

After collecting $\{C_i\}_{i=1}^I$, the mechanism assigns each student a random number that determines her priority at all schools. All students thus 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 previously. Having learned their assignments, students can choose an outside option if they are not satisfied.

Below, I use “student” and “parent” interchangeably. After the model set-up, I present the benchmark case where everyone is (equally) sophisticated with a common prior. The model is then extended to allow parents to have heterogeneous sophistication. The definition of sophistication is formalized in due course.

3.1 Set-Up

The utility of student i attending school s is defined as

$$u_{i,s} = \alpha_s + \mathbf{X}_i \beta_X + \mathbf{Z}_{i,s} \beta_Z + \varepsilon_{i,s}, s \in \{1, \dots, 4\}; \text{ and } u_{i,0} = \varepsilon_{i,0};$$

where α_s is school s 's fixed effect; $\mathbf{X}_i \in \mathbb{R}^{K_1}$ are i 's characteristics, such as his or her test score, parents' income, and parents' education, etc.; $\mathbf{Z}_{i,s} \in \mathbb{R}^{K_2}$ are students' school-specific attributes, e.g., 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 $\varepsilon_i \equiv \{\varepsilon_{i,s}\}_{s=0}^4$. The utility

functions are normalized such that $\alpha_0 = 0$ and $\mathbf{X}_i\beta_X$ measures the effect of \mathbf{X}_i on the differences between the outside and all inside options. If a school is worse than the outside option, $u_{i,s} < u_{i,0}$, 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 have full information on their own preferences, $\{u_{i,s}\}_{s=0}^4$, as well as the functional form of $u_{i,s}$ and its parameters.

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

AM.3. $\varepsilon_i \perp (\mathbf{X}_i, \mathbf{Z}_i)$, and $\varepsilon_{i,s}$ is i.i.d. over i and s with type I extreme value (Gumbel) distribution, with C.D.F. $F(\varepsilon)$. ε_i is private information of i , and its distribution is common knowledge.

AM.4. A parent does not participate, i.e., submitting $(0, \dots, 0)$, if and only if no school is acceptable.

The assumption that $(\mathbf{X}_i, \mathbf{Z}_i)$ is private information is made to simplify notations. When I is large, similar results hold if $(\mathbf{X}_i, \mathbf{Z}_i)$ is common knowledge or if i knows a fixed number of others' $(\mathbf{X}_j, \mathbf{Z}_j)$. Appendix D discusses this possibility in detail. AM.3 rules out correlations between any $\varepsilon_{i,s}$ and $\varepsilon_{i,s'}$ and imposes independence among $\varepsilon_{i,s}$ and $(\mathbf{X}_i, \mathbf{Z}_i)$. While AM.3 can be relaxed (as in He 2012), it greatly reduces computational burden.¹³ AM.4 requires that the outside option does not change after parents observe the matching outcome, and it also rules out parents' possible uncertainty aversion.

3.2 A Road Map for Theoretical Analysis

Before advancing to an equilibrium analysis of parents' behavior, I provide a road map (Figure 1). In a traditional game with incomplete information, Bayesian Nash equilibrium in which everyone has a common prior is usually considered. The following analysis begins with this, as it serves as a benchmark and is used in the counterfactual analysis. Equivalently, the assumption is that parents have the same information set and the same ability to process information; therefore, everyone plays a best response against others.

However, the common-prior approach assumes away all possible mistakes, while there are concerns about heterogeneous sophistication among parents, e.g., sophisticated and naïve (Pathak and Sonmez 2008). I further generalize the model to allow for all possible mistakes while imposing a minimal rationality requirement that parents do not play dominated strategies. Parents' information and sophistication can be summarized by beliefs, i.e., parents' perceived admission probabilities at school, given a list of schools. An inevitable question then arises: should certain "extreme beliefs" be allowed, e.g., whether some parents actually consider it impossible to be accepted by some schools under some circumstances?

¹³AM.3 rules out the omitted variables that are correlated with the observables, $(\mathbf{X}_i, \mathbf{Z}_i)$. This requires, for example, distance from home to school to be independent of ε . If students have moved closer to their preferred schools before the school choice game, distances are negatively correlated with the taste shocks.

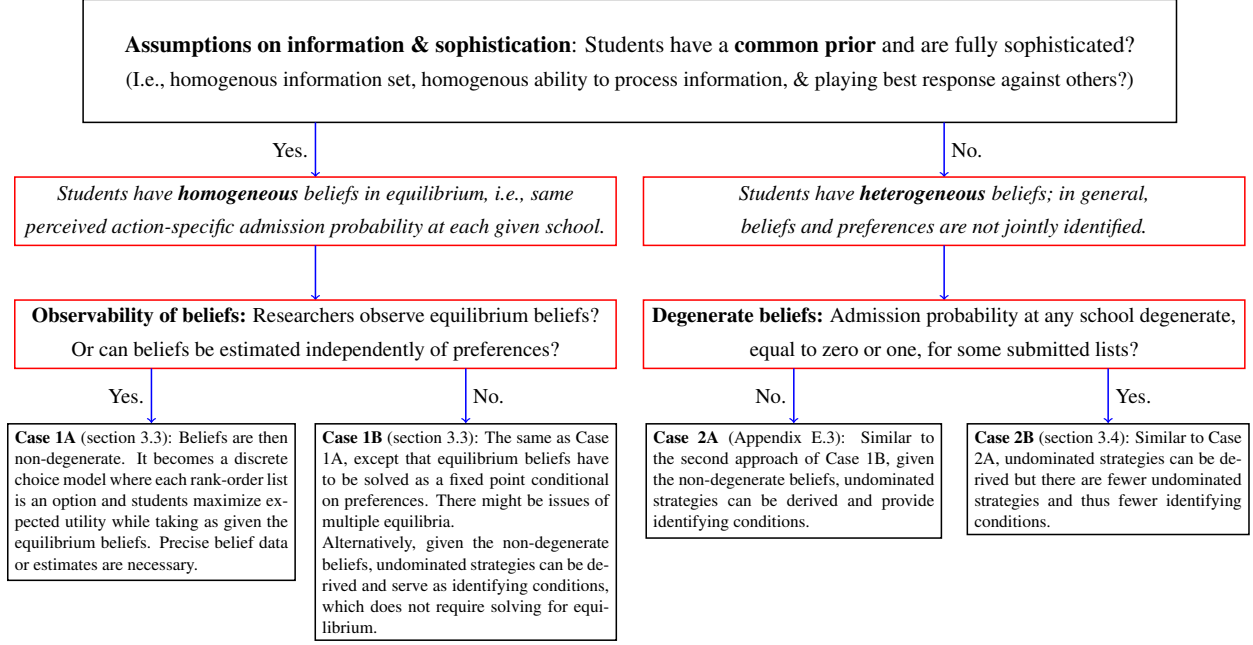


Figure 1: Modeling Assumptions and Estimation: A Road-map.

These considerations lead to four subcases considered in the paper (Figure 1):

- Case 1: Parents have a common prior, and depending on researcher’s information, two subcases arise: (1A) Researchers have data on beliefs or can independently estimate beliefs, and (1B) no information on beliefs is available, or beliefs cannot be identified separately from preferences.
- Case 2: Parents have heterogeneous beliefs, which may be due to heterogeneity in either information or information-processing ability. There are again two subcases: (2A) Parents have non-degenerate beliefs, discussed in Appendix E.3), and (2B) parents may have degenerate beliefs (in section 3.4).

In the following, I consider Case 1 and then Case 2B. Other cases are discussed in Appendix E.

3.3 Benchmark: Homogeneous Sophistication

Suppose that everyone is (equally) sophisticated and is endowed with a common prior. Therefore, they have the same information and correctly use it in the same way. The purpose of these restrictions is to provide a benchmark for analyzing parents’ behavior and a theoretical foundation for counterfactual analysis.

3.3.1 Strategy, Payoff, and Decision Making

A strategy $\sigma_i(\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i)$, possibly in mixed strategies, is a mapping from i ’s “type” space to the set of all probability distributions over possible lists: $\mathbb{R}^{K_1+4K_2+5} \rightarrow \Delta(\mathcal{C})$. The total number of pure strategies or possible lists in \mathcal{C} is finite, $L \equiv 41$, including 24 full lists, 12 two-school lists, 4 one-school list, and non-

participation $(0, 0, 0, 0)$.¹⁴ Each element in \mathcal{C} is a rank-order list of k distinct schools, for $k \in \{0, 1, 2, 4\}$.

The payoff to i can be characterized in two steps by sequentially assuming that (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 $\sigma_i = C$, the expected payoff to i is

$$V_i(C, C_{-i}) \equiv \sum_{s=1}^4 [a_s(C, C_{-i}) \max(u_{i,s}, u_{i,0})],$$

where only $\max(u_{i,s}, u_{i,0})$ matters because parents may take the outside option whenever the assignment is unacceptable; and $a_s(C, C_{-i})$ is the admission probability of student i at s given (C, C_{-i}) , which is determined mechanically by the algorithm and the random lottery.

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

(i) *A seat is guaranteed if participating: $\forall C \neq (0, \dots, 0)$, $\sum_{s=1}^4 a_s(C, C_{-i}) = 1$;*

(ii) *Given two lists that agree on the top K choices, if school s is ranked as K^{th} choice, the probability of being accepted by s is the same when submitting either of the two lists: $a_s(C, C_{-i}) = a_s(\bar{C}, C_{-i})$, $\forall C, \bar{C} \in \mathcal{C}$, s.t., $c^K = \bar{c}^K = s$, $c^k = \bar{c}^k$, $\forall k < K \leq 4$.*

(iii) *Moving a school up (or including an otherwise omitted one) in the list weakly increases the probability of being accepted by that school: $a_s(\bar{C}, C_{-i}) \geq a_s(C, C_{-i})$, $\forall C, \bar{C} \in \mathcal{C}$, s.t., $c^K = \bar{c}^{K'} = s$, $K' < K \leq 4$, and $c^k = \bar{c}^k$, $\forall k < K'$.*

(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; these properties can be easily verified, given the mechanism.¹⁵

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 thus be defined in the same way as $V_i(C, C_{-i})$. Now, instead, suppose that

¹⁴Here, I consider only lists that are not always payoff equivalent. For example, the lists with 3 schools are excluded, as one may always find a payoff-equivalent full list for any three-school list. Lists in which one school appears multiple times are excluded either because of the same reason or because they are strictly dominated. Other obviously dominated lists are not considered either, e.g., 0 is ranked first and followed by other schools. This step is innocuous, as students never submit these lists when maximizing expected utility. In other words, the choice probability of such a list is zero in equilibrium.

¹⁵In a more general school choice problem where schools have a few priority groups, e.g., in Boston (Abdulkadiroglu and Sonmez 2003), parts (i)–(iii) of Lemma 1 are still satisfied. However, part (iv) should be modified to take into account that students with higher priorities can be accepted by s before student i is considered, under the assumption that everyone knows the number of students in each priority group at each school.

(σ_i, σ_{-i}) is given. i 's expected payoff is defined as

$$\begin{aligned} V_i(\sigma_i, \sigma_{-i}) &\equiv \sum_{n=1}^{L^{(I-1)}} \{ \Pr(C_{-i}^n \text{ played under } \sigma_{-i}) V_i(\sigma_i, C_{-i}^n) \} \\ &= \sum_{s=1}^4 \sum_{n=1}^{L^{(I-1)}} [\Pr(C_{-i}^n \text{ played under } \sigma_{-i}) a_s(\sigma_i, C_{-i}^n)] \max(u_{i,s}, u_{i,0}), \end{aligned}$$

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

$$\int \int \Pr(C_{-i}^n \text{ played under } \sigma_{-i}(\mathbf{X}_{-i}, \mathbf{Z}_{-i}, \varepsilon_{-i})) dG(\mathbf{X}_{-i}, \mathbf{Z}_{-i}) dF(\varepsilon_{-i}).$$

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

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

The expected payoff is thus simplified as $V_i(\sigma_i, \sigma_{-i}) = \sum_{s=1}^4 A_s(\sigma_i, \sigma_{-i}) \max(u_{i,s}, u_{i,0})$.

Furthermore, denote $B(\sigma_i, \sigma_{-i}) \equiv (A_1(\sigma_i, \sigma_{-i}), \dots, A_4(\sigma_i, \sigma_{-i})) : \Delta(\mathcal{C}) \rightarrow [0, 1]^4$ as i 's beliefs. By definition, $A_s(\sigma_i, \sigma_{-i})$ is a probability-weighted sum of $a_s(C, C_{-i})$, for all C and C_{-i} . Therefore, it can be verified that the properties of $a_s(C, C_{-i})$ in Lemma 1 still hold for $A_s(\sigma_i, \sigma_{-i})$.

With the beliefs, a parent is sophisticated if her subjective beliefs are the same as objective ones.

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

Given her beliefs, parent i chooses the best response to maximize her expected utility:

$$\sigma_i(\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i) \in \arg \max_{\hat{\sigma}_i \in \Delta(\mathcal{C})} \sum_{s=1}^4 A_s(\hat{\sigma}_i, \sigma_{-i}) \max(u_{i,s}, u_{i,0}). \quad (1)$$

It should be emphasized that i 's **best response need not be unique**: (i) the operator $\max(u_{i,s}, u_{i,0})$ creates multiple payoff-equivalent actions if some schools are unacceptable; and (ii) additional payoff-equivalent actions arise if $A_s(\sigma_i, \sigma_{-i})$ is degenerate, i.e., equal to zero, for some s . Multiplicity in best responses implies multiple equilibria and thus creates challenges for empirical analysis because choice probabilities of actions can no longer be characterized.

3.3.2 Bayesian Nash Equilibrium

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

$$\sigma^*(\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i) \in \arg \max_{\sigma \in \Delta(\mathcal{C})} \sum_{s=1}^4 A_s(\sigma, \sigma_{-i}^*) \max(u_{i,s}, u_{i,0}), \text{ given } (\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i), \forall i;$$

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

The existence and a characterization of such an equilibrium are 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 $A_s(C, \sigma_{-i}^*) \in (0, 1) \forall s, \forall C \neq (0, \dots, 0)$;
- (ii) if at most one school is unacceptable, almost surely i plays a pure strategy, i.e., i has a unique best response.
- (iii) if i plays mixed strategies, almost surely she has at least two unacceptable schools; furthermore, she plays with a positive probability only the lists in which the unacceptable schools are either excluded or included after the acceptable ones.
- (iv) given how everyone chooses mixed strategies, i.e., how to include unacceptables in the list, everyone has a unique best response with a probability one.

The paper has data on one game play and therefore guarantees that there is only one equilibrium in the data-generation process. Under the parametric assumption and with the identifying conditions to be derived later, the parameters are shown to be identified. The multiplicity of equilibria nonetheless matters for counterfactual analyses where a single equilibrium must be selected. I defer this discussion until Section 6.

Proposition 1 clarifies that the multiplicity of equilibrium is now due only to **multiple best responses** or the presence of multiple unacceptable schools because in any equilibrium, it is payoff-equivalent for parents to arbitrarily rank unacceptable schools when there are multiple schools, as long as all unacceptable schools are ranked after the acceptable ones. To rule out some of this arbitrariness, following the result in part (iv) above, I make the following assumption:

Assumption UNACCEPTABLES If an expected-utility-maximizing parent includes some or all of her unacceptable schools in her list, they are ranked according to the true preference order among themselves after the acceptable schools. Moreover, the excluded unacceptable schools are always less preferable than those included.

It is not implausible that parents follow this strategy in real life for reasons such as “playing safe.” Although the above assumption does not prescribe which unacceptable schools should be included in the list, it implies restrictions on actions played in mixed strategies.

Lemma 2 (i) If i has two unacceptable schools, say $u_{i,1} > u_{i,2} > u_{i,0} > u_{i,3} > u_{i,4}$, i only mixes over two lists: $(c_i^1, c_i^2, 0, 0)$ and $(c_i^1, c_i^2, 3, 4)$, where $c_i^1, c_i^2 \in \{1, 2\}$, whenever both are best responses;

(ii) If i only has three unacceptable schools, say $u_{i,1} > u_{i,0} > u_{i,2} > u_{i,3} > u_{i,4}$, i mixes over three lists: $(1, 0, 0, 0)$, $(1, 2, 0, 0)$, and $(1, 2, 3, 4)$.

Because this lemma is a straightforward implication of Assumption UNACCEPTABLES, its proof is omitted. Note that for any three-school list, one may find a full list that is always payoff-equivalent, e.g., $(1, 2, 3, 0)$ and $(1, 2, 3, 4)$. I therefore always pool these schools together and use the latter to denote both. The following assumption further clarifies how parents mix the actions in equilibrium mixed strategies:

Assumption MIXING Suppose i has k schools acceptable, $k \in \{1, 2\}$, and let $m_i^{k,l}$ denote the probability that an l -school list is submitted, where $l \in \{1, 2, 4\}$ if $k = 1$ and $l \in \{2, 4\}$ if $k = 2$. Furthermore, $\sum_l m_i^{k,l} = 1$; $m_i^{k,l} = m^{k,l}(\mathbf{X}_i) + \eta_i^{k,l}$, $E(\eta_i^{k,l}) = 0$ and $\eta_i^{k,l} \perp (\varepsilon_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{B})$ for all l and k .

Later on, $m^{k,l}(\mathbf{X}_i)$ is assumed to be a linear function of (a sub-vector of) \mathbf{X}_i . The idiosyncratic mixing probabilities, $\eta_i^{k,l}$, are independent of $(\varepsilon_i, \mathbf{X}_i, \mathbf{Z}_i, B)$, where B should be interpreted as i 's beliefs, and there is no further restriction on the distribution of $\eta_i^{k,l}$. Under this assumption, any two parents with the same \mathbf{X}_i may adopt different probability distributions over possible payoff-equivalent lists, as long as all those lists are best responses, while on average, their mixing probabilities are $m^{k,l}(\mathbf{X}_i)$.

An restriction of Assumption MIXING is that $m_i^{k,l}$ depends only on the number of acceptable schools but not on their identities or their preference order. Although this can be relaxed, there would be too many parameters to estimate. For example, allowing $m^{k,l}(\mathbf{X}_i)$ to depend on identities of acceptable schools would result in 14 functions of mixing probabilities, while under Assumption MIXING, there are only 3.

Remark 1 To simplify notations, in the following, I focus on a special case where $\eta_i^{k,l} = 0$, for all k and l and therefore $m_i^{k,l} = m^{k,l}(\mathbf{X}_i)$ for all i . Allowing $\eta_i^{k,l} \neq 0$ makes two differences: (1) the choice

probabilities that are to be characterized should be interpreted as expected choice probabilities; and (2) the maximum likelihood estimation should be called pseudo- or quasi-maximum likelihood or replaced by the generalized method of moments. A more detailed discussion is available in Appendix H.

(Approximate) Equilibrium under the Two Assumptions: Existence. One may be concerned with the equilibrium existence under Assumptions UNACCEPTABLES and MIXING, especially counterfactual analyses require solving such an equilibrium. Proposition 1 establishes the equilibrium existence without the two assumptions, and not extending the proof, one may consider that the proposition shows the existence of an “approximate” equilibrium under the two assumptions.

Suppose that equilibrium does not exist under the two assumptions and that B^* is an equilibrium belief system as in Proposition 1. One can argue that it is an approximate equilibrium when everyone best responds to B^* under the restrictions. First and foremost, the two assumptions do not require individuals to play sub-optimal strategies, as they only restrict how they choose among payoff-equivalent strategies. Therefore, the two assumptions only matter in the sense that they change the admission probabilities of others. Second, the two assumptions do not affect the admission probability at the first choice given any strategy because they are all about the unacceptable schools ranked after acceptable ones by participating parents. Given the mechanism, the assumptions then only have a second-order effect on the admission probabilities at the second or later choices. Third, given the high popularity of Schools 1 and 2 (re-confirmed later with estimation results), both of them are filled in the first round, and thus, the two assumptions have no effect on the admission probabilities of these two schools, no matter how they are ranked. Fourth, because there are more seats in total than the number of applicants, there must be a school (School 4, as shown later) whose capacity constraint is not binding after the four rounds. Therefore, admission probabilities at this school are not affected by the two assumption, either. Lastly, the above argument left us with the admission probabilities at School 3 to be affected by the two assumptions. However, this effect is bound to be small. According to estimation results to be shown later, only 4.2% of students find School 3 unacceptable conditional on having at least one acceptable school. Even if they all include School 3 on their lists, there is a high probability that they are accepted by their earlier choices.

Estimation. With Proposition 1 and Assumptions UNACCEPTABLES and MIXING, one can fully characterize the choice probability of each list being played in equilibrium if the equilibrium beliefs, B^* , are observed. To estimate B^* , one possibility is to use the observed game play. That is, one can treat each observed player as a random draw from the population, and the empirical action distribution approximates the true distribution of equilibrium actions when the number of players is large. By bootstrapping from the

observed individual strategies, one may approximate $A_s(\sigma^*, \sigma_{-i}^*)$ and thus B^* (denoted as \hat{B}^*).

The characterization of choice probabilities is provided in Appendix E based on which the simulated maximum likelihood estimation can be applied. However, the approach crucially relies on the accuracy of \hat{B}^* . Additionally, simulation is necessary in the maximum likelihood estimation, introducing additional errors. The results from Monte Carlo simulations with the true B^* and the approximated equilibrium beliefs in Appendix H confirm these concerns.

On the other hand, the structure of equilibrium and its properties in Proposition 1 contain rich identifying information that is independent of B^* . I defer the exploration until the next subsection, where I discuss the estimation with heterogeneous sophistication. The alternative approach is shown to perform even better than the simulated maximum likelihood with the true equilibrium beliefs B^* in Monte Carlo simulations (Appendix H).

3.3.3 Assumptions and Flexible Heterogeneous Sophistication: An Evaluation

It is worth empirically evaluating the assumptions, UNACCEPTABLES and MIXING, which are introduced due to multiple best responses but may seem ad hoc and restrictive. I also emphasize the need to allow for flexible heterogeneous sophistication, which is considered in the next section.

Assumptions UNACCEPTABLES and MIXING. To formalize the discussion, let us supplement i 's choice list C_i by inserting her outside option, School 0, before the highest-ranked unacceptable school in C_i . That is, $\bar{C}_i = (\bar{c}_i^1, \dots, \bar{c}_i^4, \bar{c}_i^5)$ such that $\bar{c}_i^{k_i} = 0$ for $k_i = \max_{k=1, \dots, 5} \{k \text{ s.t. } u_{i, c_i^k} > u_{i, 0} \text{ for all } t < k\}$, $\bar{c}_i^k = c_i^k$ for all $k < k_i$, and $\bar{c}_i^k = c_i^{(k-1)}$ for all $k > k_i$. For example, for a parent ranking (1, 2, 3, 4) with Schools 3 and 4 as unacceptable, \bar{C} is (1, 2, 0, 3, 4). When being rejected by earlier choices, i will “enroll” in her outside option for sure, and therefore, every school ranked after an outside option is payoff-irrelevant.

Assumption UNACCEPTABLES entails two behavioral patterns for parents with multiple unacceptable schools: (a) among the payoff-equivalent lists that differ in the inclusion of unacceptable schools, parents sometimes include unacceptable schools. That is, in \bar{C}_i with $k_i < 5$, i ranks some schools after outside option, $\bar{c}_i^k \neq 0$ for $k > k_i$; (b) when ranking unacceptable schools, they rank unacceptable schools truthfully after acceptable ones. In other words, in \bar{C}_i , choices lower than k_i are ranked truthfully among themselves.

The first pattern is consistent with the fact that parents sometimes do not enroll in the assigned school in the data (Table 2) and in other places, e.g., Denver and New York City. The assumption is rejected if (almost) everyone goes to the assigned school, e.g., in Paris (Fack, Grenet, and He 2015). This is exactly the issue of multiple best responses: a parent may rank unacceptable schools because the marginal cost and benefit of adding more choices are both zero.

The second pattern includes two aspects. An unacceptable school, if ranked, is ranked below all acceptables, which is implied by the utility-maximization assumption. Besides, unacceptable schools are ranked truthfully among themselves, which is a behavioral assumption for equilibrium selection. This phenomenon is difficult to evaluate with observational data, although it is not unimaginable that parents do so in reality.

Therefore, I also use an external dataset to evaluate the assumptions. In the lab experiment of Cal-samiglia, Haeringer, and Klijn (2010), students play a school choice game under three mechanisms. Everyone has a “district school,” which approximates an outside option in list \bar{C}_i : if it is ranked first under the Boston, i is for sure accepted by district school; if rejected by earlier choices, she is accepted by her district school whenever it is ranked under DA, which creates multiple payoff-equivalent lists.

More importantly, one can rank up to three choices, in contrast to other lab experiments, where students are required to rank a given number of choices. The detailed analysis in Appendix B demonstrates that

- (i) The majority (95%) of the students rank additional schools after “outside option” (district school) so that they exhaust the three choices.
- (ii) Among the students in part (i), the majority (82%) rank their second and third choices truthfully, conditional on a district school being ranked first.

These results provide empirical justifications for Assumption UNACCEPTABLES and thus lead me to use Assumption MIXING to model how parents rank unacceptable schools. As described earlier, it restricts that the mixing probabilities do not depend on preferences or identities of acceptable schools. Indeed, this phenomenon is also supported by the experimental data, as Appendix B shows that

- (i) Preferences do not affect students’ probability of ranking additional schools after the “outside option”.
- (ii) Students do not rank the next-preferred school after district school more often when the utility difference between the two is smaller.

These results provide empirical support to Assumptions UNACCEPTABLES and MIXING.

Necessity of Allowing Flexible Heterogeneous Sophistication. The above analysis focuses on parents playing a Bayesian Nash equilibrium with a common prior and thus rules out all possible mistakes.

This assumption, however, is usually too strong. In particular, there is evidence suggesting that students underestimate admission probabilities or are **overcautious**, when playing the game under the Boston mechanism. This evidence is consistent with the “small-school bias” in the experimental literature (Chen and Sonmez 2006), wherein students avoid schools with small quotas. More directly, students in Amsterdam under-predict admission probabilities, and 5.5% do not apply to their most-preferred schools, even though they would have been admitted (De Haan, Gautier, Oosterbeek, and Van der Klaauw 2015). Besides, the education literature documents abundant evidence showing that students, especially those with disadvan-

tagged backgrounds, do not apply to better schools or universities due to information frictions (Hoxby and Turner 2013).

3.4 General Case: Heterogeneous Sophistication

I now relax the sophistication assumption and allow parents to make mistakes forming their beliefs. Mistakes, which are associated with heterogeneous beliefs, can be caused by heterogeneity in information and/or in the ability to process information. I derive the structure on the beliefs and the set of dominated strategies.

3.4.1 Equilibrium, Sophistication, and Dominated Strategies

To highlight the heterogeneity in beliefs, denote i 's belief as $B_i(C, \sigma_{-i}) \equiv \{A_{i,s}(C, \sigma_{-i})\}_{s=1}^4 \in [0, 1]^4$, $\forall C$, where $A_{i,s}(C, \sigma_{-i}) \equiv \sum_{n=1}^{L^{(I-1)}} \Pr_i(C_{-i}^n \text{ played under } \sigma_{-i}) a_s(C, C_{-i}^n)$ and the probability measure $\Pr_i(\cdot)$ is i 's subjective assessment of an event's likelihood.

Given the above notation, the extent to which parents can make mistakes is not unrestricted. They may be wrong when assessing others' behavior and thus $\Pr_i(\cdot)$ is individual specific; however, they have full information on the rules of the game and thus correctly consider $a_s(C, C_{-i}^n)$.

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 $A_{i,s}(\sigma_i, \sigma_{-i})$, and the proof of the following lemma is omitted.

Lemma 3 *Given σ_{-i} , $A_{i,s}(\sigma_i, \sigma_{-i})$ has the same properties as $a_s(C, C_{-i})$ in Lemma 1.*

i 's strategy can now be rewritten as an explicit correspondence of beliefs, $\sigma_i[\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i; B_i(\cdot, \sigma_{-i})]$ and again, I focus on symmetric equilibrium.

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

$$\sigma^{**}[\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i; B_i(\cdot, \sigma_{-i}^{**})] \in \arg \max_{\sigma \in \Delta(C)} \sum_{s=1}^4 A_{i,s}(\sigma, \sigma_{-i}^{**}) \max(u_{i,s}, u_{i,0}), \forall i.$$

Here, the only requirement is that everyone is a subjective-expected-utility maximizer; there is no fixed-point restriction on subjective beliefs; and therefore, the existence of such an equilibrium is guaranteed. This definition also provides a measure of sophistication in terms of predicting the game play.

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

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

where $\Pr(C_{-i}^n \text{ played under } \sigma_{-i}^{**} \text{ given } B_{-i})$ is the objective (correct) probability of C_{-i}^n 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 the knowledge of their beliefs in addition to that of the preference distribution. Being sophisticated demands extremely rich information and thus may be difficult to achieve, and therefore, it is important to consider other possible beliefs. Unfortunately, this phenomenon leads to a non-identification of beliefs.

Proposition 2 *Given the choice lists submitted by parents with heterogeneous sophistication, the beliefs, which are individual-specific, are not identified even when parents' true cardinal preferences are observed.*

The proof of the proposition is not provided because of its obviousness, but its basic intuition is as follows. Since beliefs are individual specific, there is only one observation from a given parent to estimate her beliefs. Moreover, beliefs are high-dimensional: there are 24 unique elements for each parent when there are four schools. From Definition 3, one can derive 23 inequalities for a given parent because the submitted list is optimal among all 24 full lists, which is far from sufficient to solve for the 24 elements. If parents' cardinal preferences are not observed, there is no hope to jointly identify preferences and beliefs.

It is therefore necessary to derive conditions that do not depend on beliefs. Given the properties of Lemma 3, a set of dominated strategies, which are belief-independent, can be derived.

Proposition 3 *Suppose i has at least one acceptable school. Given belief $B_i(\cdot, \sigma_{-i}^{**})$,*

- (i) top-ranking an unacceptable or the least-preferred school is **strictly** dominated;*
- (ii) ranking an unacceptable or the least-preferred before an acceptable school is **weakly** dominated;*
- (iii) excluding an acceptable school from the list is **weakly** dominated;*
- (iv) if beliefs are non-degenerate, i.e., $A_{i,s'}(C, \sigma_{-i}^{**}) \in (0, 1)$, for all s' and $C \neq (0, \dots, 0)$, moving school s upward in the list **strictly** increases the probability of being accepted by s , and the dominance in (ii) and (iii) becomes **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. Additionally, 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 the better school. If a school is unacceptable, putting it at the bottom or omitting it also increases the likelihood of getting into better schools. These results also apply to a more general setting with pre-determined priorities (Abdulkadiroglu and Sonmez 2003).

According to Proposition 3, the truth-telling strategy is not dominated, nor is any Bayesian Nash equilibrium with homogeneous sophistication (σ^*). Moreover, as shown in Proposition 1, in any σ^* , beliefs are non-degenerate, and therefore, strict dominance can thus be obtained (part iv). These results provide identifying conditions that are formalized in Appendix E.

3.4.2 Degenerate Beliefs

Unfortunately, the weak-dominance results in parts (ii) and (iii) of Proposition 3 still lead to potential multiple best responses, creating issues for empirical study. Moreover, when parents are allowed to make mistakes, one cannot rule out that parents have degenerate beliefs, i.e., some zero-admission probabilities. Certainly, this possibility depends on the context in question, and I therefore tailor the modeling to what is observed.

In the data, School 1 has the smallest quota, 63 seats, and is also the “best” school with the highest average test score of graduating students (Table 1). Table 5 further presents how parents rank the four schools. In total, 228 parents top-rank School 1, 360% of its capacity, and thus, the only possibility to be accepted by it is ranking it first. Despite its high quality, 11% of the participating parents ranked School 1 fourth or omitted it. In contrast, only 3% and 9% did so with Schools 2 and 3, the two lower-quality schools.

A possible rationalization is that certain parents expect a zero-admission probability for School 1 when ranking it low and therefore decide to omit it or rank it fourth. A later result from survey data, Table 8 in section 4.2, also confirms this conjecture: Among 579 parents who claim School 1 as their best school, 61 (11%) of them rank it fourth or omit it, which can be explained only by degenerate beliefs. In contrast, for Schools 2 and 3, such degenerate beliefs are less of a concern: they all have a larger quota; fewer parents rank them fourth or omit it (Table 5); and only two parents rank the claimed best school fourth or omit it (Table 8).

Table 5: How Each School is Ranked

School	Quota	First Choice		Second Choice		Third Choice		Fourth Choice		Omitted by Participants	
		Freq.	Pct.	Freq.	Pct.	Freq.	Pct.	Freq.	Pct.	Freq.	Pct.
1	63	228	25%	157	17%	242	26%	68	7%	38	4%
2	227	431	47%	206	23%	64	7%	11	1%	21	2%
3	310	66	7%	333	36%	252	28%	56	6%	26	3%
4	360	8	1%	27	3%	107	12%	530	58%	61	7%

Notes: The percentages represent the fraction of the total 914 students. There are 181 (20%) non-participants.

These patterns in the data lead to the following assumption:

Assumption ZERO-PROB Some parents may expect that the admission probabilities at School 1 are zero

if the school is not top ranked, while other belief elements are always non-degenerate, (in $(0, 1)$) for all parents.

As a consequence of Assumption ZERO-PROB, School 1 might be ranked as fourth choice or omitted by some parents even when it is neither unacceptable nor the least preferred.

Lemma 4 *Under Assumptions UNACCEPTABLES and ZERO-PROB, suppose i 's beliefs are degenerate, i.e., some elements are equal to zero according to Assumption ZERO-PROB. If her preferences and beliefs are such that the list $(c^1, c^2, 1, c^4)$ is a best response, $u_{i,c^4} = \min \{u_{i,s}\}_{s=1}^4$, and $u_{i,s} > u_{i,0}$ for $s \neq c^4$, then the pure strategies that can be played with a positive probability in equilibrium are*

- (i) $(c^1, c^2, 1, c^4)$, $(c^1, c^2, c^4, 1)$, and $(c^1, c^2, 0, 0)$, if $u_{i,c^4} > u_{i,0}$; or
- (ii) $(c^1, c^2, 1, c^4)$ and $(c^1, c^2, 0, 0)$ but never $(c^1, c^2, c^4, 1)$, if $u_{i,c^4} < u_{i,0}$.

The proof of the above lemma is straightforward and thus omitted, while a sketch is as follows. First, note that $(c^1, c^2, 1, c^4)$ and $(c^1, 1, c^2, c^4)$ are not payoff-equivalent even if the probability of being accepted by School 1 is zero because ranking School 1 second wastes a round in the mechanism and decreases the chance of being accepted by c_2 . Second, if i 's beliefs are degenerate as in Assumption ZERO-PROB, $(c^1, c^2, c^4, 1)$, $(c^1, c^2, 0, 0)$, and $(c^1, c^2, 1, c^4)$ are always payoff-equivalent. However, Assumption UNACCEPTABLES specifies that i never ranks an unacceptable school before an acceptable one, and therefore, $(c^1, c^2, c^4, 1)$ is never played if $u_{i,c^4} < u_{i,0}$.

As evident in Lemma 4, the only consequence of having degenerate beliefs is that the school in question is omitted or moved to fourth place, which justifies Assumption ZERO-PROB when it rules out zero-admission probability at the second-most popular school, School 2. At most, the number of parents who can be possibly affected by the zero probabilities at School 2 is 33 (3%) (those ranking it fourth or omit it, Table 5). Besides, none of the 33 claims School 2 is the best school when responding to a survey question (Table 8).

Lemma 4 does not describe how parents mix the payoff-equivalent strategies. It specifies i 's possible actions when $(c^1, c^2, 1, c^4)$ is her best response, while whether $(c^1, c^2, 1, c^4)$ is a best response depends on her preferences and subjective beliefs. Allowing heterogeneous sophistication, however, would make the paper internally inconsistent if further assumptions on beliefs were introduced. Keeping these considerations in mind, I characterize choice probabilities for estimation based on Proposition 3 and Lemma 4.

Grouping of Lists. Under the above assumptions, the model is incomplete because of its inability to predict a unique distribution of lists conditional on preferences (Tamer 2003). I therefore categorize the lists into 9 groups, g_n , $n = 1, \dots, 9$, which “restores” completeness.

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) 3 groups of two-school lists ranking School 1; and (c) all other lists. These groups are shown in Table 6 and are mutually exclusive and collectively exhaustive.

Table 6: The 9 Groups of Lists When Beliefs are Degenerate

Groups of non-participation or one-school lists	Groups of two-school lists and other lists
Group 1: $g_1 = \{(0, 0, 0, 0)\}$	Group 6: $g_6 = \left\{ \begin{array}{l} (1, 2, 0, 0) \\ (2, 1, 0, 0) \end{array} \right\}$
Group 2: $g_2 = \{(1, 0, 0, 0)\}$	Group 7: $g_7 = \left\{ \begin{array}{l} (1, 3, 0, 0) \\ (3, 1, 0, 0) \end{array} \right\}$
Group 3: $g_3 = \{(2, 0, 0, 0)\}$	Group 8: $g_8 = \left\{ \begin{array}{l} (1, 4, 0, 0) \\ (4, 1, 0, 0) \end{array} \right\}$
Group 4: $g_4 = \{(3, 0, 0, 0)\}$	
Group 5: $g_5 = \{(4, 0, 0, 0)\}$	Group 9: $g_9 = \left\{ \begin{array}{l} \text{All lists that are not in any} \\ \text{of the other eight groups} \end{array} \right\}$

For type-(a) groups, since they are independent of beliefs, the choice probabilities are similar to traditional discrete choice models, although the mixing probabilities have to be considered. For the 3 groups of type (b), the grouping is based only on which school, together with School 1, is included in the list but not on their ranking. For example, $(s, 1, 0, 0)$ and $(1, s, 0, 0)$ are in the same group, but not $(s, s', 0, 0)$, given $s \neq s' \neq 1$. The choice probabilities for these groups have two sources: either the two included schools are the only acceptable schools or only one of them is acceptable while the other is unacceptable but better than the two excluded schools. The contributions of both sources are weighted by mixing probabilities. The type-(c) group is the residual group.

Characterization of Choice Probabilities The choice probability of a group, g_n , should be interpreted as choosing any list within that group, $C_i \in g_n$. Recall that $m^{k,l}(\mathbf{X}_i)$, $l \geq k$, is the (expected) probability that an l -school list is submitted while only k schools are acceptable, conditional on \mathbf{X}_i . The conditional probability of i choosing a list in group g_n , $\Pr(C_i \in g_n | \mathbf{X}_i, \mathbf{Z}_i; \theta)$, equals

- (i) $\Pr(u_{i,s} < u_{i,0}, \text{ for all } s | \mathbf{X}_i, \mathbf{Z}_i; \theta)$, if $C_i \in g_1 = \{(0, 0, 0, 0)\}$;
- (ii) $m^{1,1}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1 | \mathbf{X}_i, \mathbf{Z}_i; \theta)$, if $C_i \in g_n = \{(c^1, 0, 0, 0)\}$ for $n = 2, \dots, 5$, given the identity of c^1 ;

(iii) if $C_i \in g_n = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\}$ for $n = 6, \dots, 8$, given $c^1 = 1$ or $c^2 = 1$,

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

(iv) the residual probability, if $C_i \in g_9$.

With the functional form assumptions on the utility shocks and the mixing probabilities (to be specified in Section 5), all nine choice probabilities mentioned above can be re-written as functions of the observables $(\mathbf{X}_i, \mathbf{Z}_i)$ and unknown parameters θ . Three points should be highlighted here: (i) none of the choice probabilities depends on beliefs; (ii) after grouping, the model is “complete” in the sense that it implies a unique distribution of groups given a distribution of preferences; and (iii) identification means that the model implies a unique distribution of preferences given a distribution of groups.

Identification and Maximum Likelihood Estimation The model is now equivalent to a discrete choice with 9 options. The choice probabilities imply the likelihood:

$$L_{DB}(\theta) \equiv \sum_{i=1}^I \sum_{n=1}^9 [1_{(C_i \in g_n)} \log [\Pr(C_i \in g_n | \mathbf{X}_i, \mathbf{Z}_i; \theta)]] ,$$

where the subscript *DB* stands for Degenerate Beliefs. Intuitively, identification is similar to discrete choice models and requires that $\Pr(\cdot | \cdot, \cdot; \theta) \neq \Pr(\cdot | \cdot, \cdot; \theta_0)$ if θ is not equal to the true value θ_0 . In the current context, this amounts to the identification in a conditional logit model even with the mixing probabilities, while parameters in the utility function are normalized. Additional discussion is available in Appendix E.3.

As stated previously, this modeling and estimation approach can be applied to data from a Bayesian Nash equilibrium with a common prior. Moreover, the non-degenerate beliefs in a Bayesian Nash equilibrium provide more identifying conditions, as does an equilibrium with non-degenerate heterogeneous beliefs. The detailed characterization of choice probability of list groups are provided in Appendix E.3, and Appendix H presents Monte Carlo results showing that this approach performs even better than a maximum likelihood approach, even when the true equilibrium beliefs are observed.

When lists are grouped together, some identifying information is necessarily dropped, which may cause an efficiency loss in the estimation. I therefore derive bounds on certain choice probabilities and include them in the estimation, even though the model cannot point-predict them.

Bounds on Choice Probabilities. The details are provided in Appendix F, while a brief summary is as follows. I start with the 41 individual lists and then discuss various groupings of lists. Among the individual lists, the model predicts the choice probabilities of five lists (Table 6) and implies bounds on the other 36.

For example, for a two-school list, $(c^1, c^2, 0, 0)$, the lower bound of its choice probability is

$$\underline{\Pr}(C_i = (c^1, c^2, 0, 0) | \mathbf{X}_i, \mathbf{Z}_i; \theta) \equiv m^{1,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta).$$

The model accommodates that some parents submit a two-school list even though only c^1 is acceptable. Moreover, in this case, the preference ranking of all four schools and the outside option is partially determined, which constitutes a lower bound because these parents must submit $(c^1, c^2, 0, 0)$ with probability $m^{1,2}(\mathbf{X}_i)$ and there can be other parents, with multiple acceptable schools, submitting $(c^1, c^2, 0, 0)$.

If c^1 or $c^2 = 1$, i.e., the most popular school is included in the two-school list, the upper bound of the choice probability, denoted by $\overline{\Pr}(C_i = (c^1, c^2, 0, 0) | \mathbf{X}_i, \mathbf{Z}_i; \theta)$, equals:

$$\underline{\Pr}(C_i = (c^1, c^2, 0, 0) | \mathbf{X}_i, \mathbf{Z}_i; \theta) + m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta).$$

The extra term is from the parents who only have c^1 and c^2 acceptable. As some such parents may submit $(c^2, c^1, 0, 0)$ instead of $(c^1, c^2, 0, 0)$ and no one else submits this list, it becomes an upper bound.

If $c^1, c^2 \neq 1$, the upper bound probability of choosing $(c^1, c^2, 0, 0)$ is even bigger because of the degenerate beliefs (Assumption ZERO-PROB). The additional term is

$$\Pr[u_{i,c^1}, u_{i,c^2}, u_{i,1} > \max(u_{i,0}, \min_s(u_{i,s})) | \mathbf{X}_i, \mathbf{Z}_i; \theta],$$

because one may omit Schools 1 and c^4 when $(c^1, c^2, 1, c^4)$ is her best response (Lemma 4).

For a full list, (c^1, c^2, c^3, c^4) , the lower bound is

$$\underline{\Pr}(C_i = (c^1, c^2, c^3, c^4) | \mathbf{X}_i, \mathbf{Z}_i; \theta) \equiv m^{1,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta),$$

which is similar to the lower bound for two-school lists. If $c^4 \neq 1$, the upper bound is

$$\begin{aligned} & \underline{\Pr}(C_i = (c^1, c^2, c^3, c^4) | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ & + m^{2,4}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ & + \Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{aligned}$$

Intuitively, one may submit (c^1, c^2, c^3, c^4) when c^4 is the worst and others are acceptable (the last term);

some parents may also submit a full list even if they only have two schools acceptable (the second term).

When $c^4 = 1$, the upper bound is larger and has an extra term:

$$\Pr \left(u_{i,c^3} = \min \{u_{i,s}\}_{s=1}^4 > u_{i,0} | \mathbf{X}_i, \mathbf{Z}_i; \theta \right).$$

Namely, due to Assumption ZERO-PROB and Lemma 4, one may play $(c^1, c^2, c^3, 1)$ when $(c^1, c^2, 1, c^3)$ is also a best response.

Grouping lists together results in more bounds of choice probabilities. More precisely, there are 3 groups whose choice probabilities can be exactly characterized (Table 6) and 19 groups with bounds only.

For example, grouping $(c^1, c^2, 0, 0)$ and $(c^2, c^1, 0, 0)$ together when c^1 or $c^2 = 1$ leads to the choice probabilities characterized above. However, if $c^1, c^2 \neq 1$, only non-binding bounds can be derived based on Proposition 3 and Lemma 4. In this case, the lower bound is

$$\begin{aligned} & \underline{\Pr} \left(C_i \in \{ (c^1, c^2, 0, 0), (c^2, c^1, 0, 0) \} | \mathbf{X}_i, \mathbf{Z}_i; \theta \right) \\ \equiv & m^{1,2}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,s}, s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right] \\ & + m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta); \end{aligned}$$

and the upper bound is

$$\underline{\Pr} \left(C_i \in \left\{ \begin{array}{l} (c^1, c^2, 0, 0), \\ (c^2, c^1, 0, 0) \end{array} \right\} | \mathbf{X}_i, \mathbf{Z}_i; \theta \right) + \Pr \left(\begin{array}{l} u_{i,c^1}, u_{i,c^2}, u_{i,1} > u_{i,0}, \\ u_{i,c^1}, u_{i,c^2}, u_{i,1} > \min_s (u_{i,s}) \end{array} \middle| \mathbf{X}_i, \mathbf{Z}_i; \theta \right),$$

where the last term exists because parents may omit School 1 when the beliefs are degenerate. For brevity, I discuss the construction of other bounds in Appendix F.

Moment Conditions: Equalities and Inequalities. The exact choice probabilities and the bounds on them lead to moment conditions. In total, there are 9 (conditional) moment equalities and 110 (conditional) moment inequalities, and equalities and inequalities are collectively denoted as (in)equalities:

$$\begin{aligned} E[h_j(C_i, \mathbf{X}_i, \mathbf{Z}_i; \theta) | \mathbf{X}_i, \mathbf{Z}_i; \theta] & \equiv E \left[1_{(C_i \in g'_j)} - \Pr(C_i \in g'_j | \mathbf{X}_i, \mathbf{Z}_i; \theta) | \mathbf{X}_i, \mathbf{Z}_i; \theta \right] = 0, j = 1, \dots, 9; \\ E[h_{(9+j)}(C_i, \mathbf{X}_i, \mathbf{Z}_i; \theta) | \mathbf{X}_i, \mathbf{Z}_i; \theta] & \equiv E \left[1_{(C_i \in g''_j)} - \underline{\Pr}(C_i \in g''_j | \mathbf{X}_i, \mathbf{Z}_i; \theta) | \mathbf{X}_i, \mathbf{Z}_i; \theta \right] \geq 0, j = 1, \dots, 55; \\ E[h_{(64+j)}(C_i, \mathbf{X}_i, \mathbf{Z}_i; \theta) | \mathbf{X}_i, \mathbf{Z}_i; \theta] & \equiv E \left[\overline{\Pr}(C_i \in g''_j | \mathbf{X}_i, \mathbf{Z}_i; \theta) - 1_{(C_i \in g''_j)} | \mathbf{X}_i, \mathbf{Z}_i; \theta \right] \geq 0, j = 1, \dots, 55; \end{aligned}$$

where g' and g'' correspond to the groupings for the equalities and inequalities, respectively.

Binding Moment Inequalities. In the data, several lists are not played by anyone, and most are in Group 9, the residual group.¹⁶ Thus, there are 6 binding lower bounds on choice probability and 6 moment equalities; therefore, a version of the semiparametric maximum likelihood in Tamer (2003) can be adopted. Namely, given a list in Group 9 that is not played by anyone, the (semiparametric) estimate of the choice probability of this list is its theoretical lower bound, as it has to be the maximum of the lower bound and zero. Therefore, the choice probability of the rest of the lists in Group 9 can be re-characterized by subtracting this estimated probability. Tamer (2003) (Theorem 3) shows that the maximum likelihood estimator using these refined choice probabilities gains efficiency.

Estimation with Moment (In)equalities. The over-identifying information in the moment inequalities can lead to a reduction of the asymptotic mean-squared estimation error (Moon and Schorfheide 2009). To obtain consistent point estimates with moment (in)equalities, I follow the approach of Andrews and Shi (2013), which is valid for both point and partial identifications. The objective function is a test statistic, $T_I(\theta)$, of the Cramer-von Mises type with the modified method of moments (or sum function).

A few extra notations are needed to derive $T_I(\theta)$. First, each conditional moment condition, h_j , is transformed into a vector of unconditional moments, $(m_{j,1}, \dots, m_{j,K})$, such that for each k

$$E[m_{j,k}(\theta)] = E(h_j f_{k,j}(\mathbf{X}_i, \mathbf{Z}_i)) \begin{cases} = 0, & \text{for } j = 1, \dots, 9; \\ \geq 0, & \text{for } j = 10, \dots, 119. \end{cases}$$

where $f_{k,j}(\mathbf{X}_i, \mathbf{Z}_i)$ is the k th instrument for the moment condition h_j constructed from exogenous variables $(\mathbf{X}_i, \mathbf{Z}_i)$; and for $j = 10, \dots, 119$, $f_{k,j}(\mathbf{X}_i, \mathbf{Z}_i) \geq 0$, so that the moment inequalities still hold. Among the numerous ways to construct the instruments, I use the “crude” ones. That is,

$$f_{k,j}(\mathbf{X}_i, \mathbf{Z}_i) = \begin{cases} (\mathbf{X}_i, \mathbf{Z}_i)_k, & j = 1, \dots, 9; \\ [(\mathbf{X}_i, \mathbf{Z}_i)_k]^2, & j = 10, \dots, 119. \end{cases}$$

where $(\mathbf{X}_i, \mathbf{Z}_i)_k$ denotes the k th element in the vector $(\mathbf{X}_i, \mathbf{Z}_i)$. With the unconditional moments, the test statistic is constructed as follows:

$$T_I(\theta) = \sum_{j=1}^9 \sum_{k=1}^K \left[\frac{\bar{m}_{j,k}(\theta)}{\hat{\sigma}_{j,k}(\theta)} \right]^2 + \sum_{j=10}^{119} \sum_{k=1}^K \left[\frac{\bar{m}_{j,k}(\theta)}{\hat{\sigma}_{j,k}(\theta)} \right]_-^2 \quad (2)$$

¹⁶In contrast, in the Monte Carlo simulations, each of the 41 individual lists is observed being chosen.

where $\bar{m}_{j,k}(\theta)$ and $\hat{\sigma}_{j,k}(\theta)$ are the sample mean and standard deviation of $m_{j,k}(\theta)$, respectively; and the operator $[\]_-$ is such that $[a]_- = \min\{0, a\}$. The point estimate is a solution to the following problem:

$$\hat{\theta}_{DB}^{Ineq} = \arg \min_{\theta} T_I(\theta).$$

Moreover, as the model is point identified with the moment equalities (of the choice probabilities), one can use maximum likelihood estimation, which results in the point estimation denoted by $\hat{\theta}_{DB}^{Eq}$.

There is a large and growing body of literature on moment (in)equalities, e.g., Chernozhukov, Hong, and Tamer (2007) (and an application in Ciliberto and Tamer (2009)), Andrews and Soares (2010), and Andrews and Shi (2013), and a survey is available in Tamer (2010). This paper follows the approach in Bugni, Canay, and Shi (2016) to construct the *marginal confidence intervals* for each coordinate of θ . For a given coordinate θ_k , the authors provide a test for the null hypothesis $H_0 : \theta_k = \gamma$, for any given $\gamma \in \mathbb{R}$. The confidence interval for the true value of θ_k is the convex hull of all γ 's for which H_0 is not rejected.¹⁷

3.5 Other Cases and Relationship among Them

To reiterate, I also consider the cases with homogeneous sophistication (or common prior) and heterogeneous sophistication with non-degenerate beliefs. The relationship among all the cases is such that

$$\begin{aligned} \left(\begin{array}{c} \text{Bayesian Nash Equ.} \\ \text{w/ common prior} \end{array} \right) &\subset \left(\begin{array}{c} \text{Non-degenerate} \\ \text{heterogeneous beliefs} \end{array} \right) \subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ equalities only} \end{array} \right) \\ \left(\begin{array}{c} \text{Bayesian Nash Equ.} \\ \text{w/ common prior} \end{array} \right) &\subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ (in)equalities} \end{array} \right) \subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ equalities only} \end{array} \right) \end{aligned}$$

The Bayesian Nash equilibrium under homogeneous sophistication is the most restrictive one and thus is nested in all other cases, where sophistication may be heterogeneous. The case where all elements in everyone's belief are non-degenerate is nested in the case where some elements are allowed to be zero/degenerate but only equalities are considered. It is certainly more restrictive when both equalities and inequalities are considered. However, the case with (in)equalities still nests the Bayesian Nash equilibrium because all the bounds on choice probabilities are satisfied in Bayesian Nash equilibrium. On the other hand, there is no nesting relationship between non-degenerate beliefs and degenerate Beliefs with (in)equalities.

The literature has also focused on naïve parents who always report true preferences under the Boston mechanism. I therefore consider that everyone reports truthfully and follows the aforementioned mixing probabilities to include unacceptable schools in choice list. Appendix E provides the characterization of

¹⁷Discussion is also available in an application of the approach, Fack, Grenet, and He (2015).

choice probabilities and discuss estimation approach. The truth-telling case is nested in the two cases with heterogeneous sophistication:

$$\begin{aligned} \text{Truth-telling} &\subset \left(\begin{array}{c} \text{Non-degenerate} \\ \text{heterogeneous beliefs} \end{array} \right) \subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ equalities only} \end{array} \right) \\ \text{Truth-telling} &\subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ (in)equalities} \end{array} \right) \subset \left(\begin{array}{c} \text{Degenerate heterogeneous} \\ \text{beliefs w/ equalities only} \end{array} \right) \end{aligned}$$

There is, however, no clear nesting structure between truth-telling and Bayesian Nash equilibrium.

The nesting structures imply that the two approaches with degenerate beliefs always result in consistent estimates. In contrast, estimates from truth-telling or non-degenerating beliefs are only consistent when the identifying assumptions are satisfied. Monte Carlo simulations in Appendix H confirm these predictions.

3.6 Post-Estimation Analysis: Sophistication and Incentives

Ideally, we would like to measure which parents are more strategic and sophisticated in the game at an individual level; however, since the preference estimates can indicate only the distribution of preferences conditional on $(\mathbf{X}_i, \mathbf{Z}_i)$, sophistication can only be measured conditional on $(\mathbf{X}_i, \mathbf{Z}_i)$.

With either heterogeneous or homogeneous sophistication, a parent is defined as sophisticated if her beliefs coincide with the objective equilibrium beliefs. Therefore, the distance between one's beliefs and the equilibrium ones and that between one's strategy and her best response provide a measure of sophistication. Given the large number of players, the empirical beliefs, \hat{B} , as discussed in Section 3.3.2, provide an approximation of the equilibrium beliefs. In particular, with heterogeneity in sophistication, it is impossible to solve for the equilibrium because the joint distribution of preference and sophistication is unknown and not estimated. The measures will be calculated by replacing equilibrium beliefs B^* with \hat{B} , and θ with $\hat{\theta}$.

3.6.1 Probability of Observing a Given Action

Under the assumption that everyone with $(\mathbf{X}_i, \mathbf{Z}_i)$ plays a best response, the model can predict the choice probability for each list, $P_{i,k}^{BR} \equiv \Pr(C_i = C^k | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$, $k = 1, \dots, 41$. Given that 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,

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

One may test the hypothesis that everyone plays a best response by running a regression for each k :

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

where $\mathbf{W}_{i,k}$ is a vector of variables other than \mathbf{X}_i and \mathbf{Z}_i . Under the null, $(\delta_0, \delta_X, \delta_Z, \delta_W) = 0$.

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

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

Similarly, one can regress $(d_{i,k} - P_{i,k}^{TT})$ on $(\mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_{i,k})$ and test the truth-telling hypothesis under which no coefficient should be significantly different from zero.

3.6.2 Incentives to Be Strategic

In real life, it is not implausible that finding a best response is costly. 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:

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

Under the assumption that the cost of finding a best response is lower if truth-telling itself is a best response, a high $P_i^{TT=BR}$ means a lower decision cost for i to play a best response.

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

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

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.¹⁸ If $Gain_i$ is higher, i has a greater incentive to find her best response. $P_i^{TT=BR}$ and $Gain_i$ are later included in regression (3) to test if parents' behavior is affected by incentives.

4 Reduced-Form Results and Supporting Evidence for the Assumptions

Before reporting the model estimates, I present reduced-form results from the data. The results, which are independent of the model, are shown to be consistent with the model assumptions and predictions.

¹⁸ 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 : $V_i^{BR} = \int_{\varepsilon_i} \max_{\sigma \in \Delta(C)} \sum_{s=1}^S A_s(\sigma, \sigma_{-i}^*) \max(u_{i,s}, 0) dF(\varepsilon_i)$; V_i^{TT} is similarly defined.

4.1 Understanding the Rules of the Game

An important assumption is that parents understand the rules of the game, and therefore, their beliefs follow the structure specified in Lemma 3. 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 7 shows the summary statistics, although **the self-reported beliefs are not used anywhere in estimation.**¹⁹ The empirical beliefs, calculated from the submitted lists, and the self-reported beliefs share the same pattern, although they do not exactly match. Also consistent with Lemma 3, parents on average expect that moving a school up in the list increases the probability of being accepted by that school.

Table 7: Empirical Beliefs and Parents' Self-Reported Beliefs

School	Ranked as 1st Choice				Ranked as 2nd Choice			
	Empirical Beliefs ^a	Survey Responses ^b			Empirical Beliefs ^a	Survey Responses ^b		
		Mean	Std. Dev	# Obs ^c		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

Notes: a. Calculated from the actual submitted lists. Each entry shows the probability of 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?" These responses are not used in any estimation. c. The 1st and 2nd choices are self-reported and thus are not necessarily the submitted ones.

4.2 Undominated Strategies, Truth-Telling, and Degenerate Beliefs

Lemma 3 leads to the dominated strategies in Proposition 3, and parents should not play these strategies in equilibrium. Table 8 shows the distribution of parents' first choice: 24.9% top rank School 1, while 47.0% top rank School 2. Another survey question asks, "Among those to which you could apply, which school was the best?" It is not necessarily about the most-preferred school, but the correlation between the "best" and the most-preferred schools is probably positive. Among the 699 valid responses, 82.8% claim School 1 as the best. Comparing the submitted first choices with the most recognized school, the difference is significant, which is inconsistent with the truth-telling hypothesis.

If everyone understands the rules, the first-choice school should never be the least-preferred school (Proposition 3), which is consistent with the data in Table 8: Only 8 parents top-rank School 4, while even fewer people claim it as the best school.

Proposition 3 also predicts that the last-choice school (or the omitted school, conditional on participating) should either be an unacceptable/the least-preferred or a school with zero admission probability. The

¹⁹Since the questions are asked after the assignment, the results may be affected by the fact that the student is accepted or rejected by that school. For this reason, I use their self-reported top two choices, which are not necessarily their submitted top two choices.

Table 8: 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	(24.9%)	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-Participant Other ^c		181	(19.8%)	33	(4.7%)					
Total	960	914	(100%)	699	(100%)					

Notes: a. Responses to a survey question: "Among those to which you could apply, which school was the best?" These responses are not used in any estimation. 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, which may be due to misreporting.

last five columns in Table 8 confirm this prediction by showing how parents rank the claimed best school. For Schools 2-4, only two parents rank the claimed best school fourth or omit it. Therefore, as discussed in section 3.4.2, degenerate beliefs are less of a concern for these schools.

Although 579 parents claim School 1 as the best, 36 (6.2%) parents rank it fourth, while another 25 (4.3%) participants exclude School 1 altogether. Since this school has the smallest quota, only those who top-rank it have a chance of getting in, and even then, the admission probability is merely 26.7%. It is thus highly plausible that a parent might expect that there is no chance of getting into School 1 when ranking it low, in line with the discussion in section 3.4.2.

4.3 Attention to Uncertainty

Additionally, there are several survey questions on parents' perceptions of the importance of 12 different factors of school choice. Parents rate these factors on a scale of 1-5, with 5 being very important. Three of the factors are related to the game's uncertainty: (i) admission quota and the possibility of being accepted; (ii) the probability of being assigned to a bad school; and (iii) consideration of other parents' applications. Since the third factor may also be correlated with school quality because other parents' applications reveal their school preferences, I create $Attn_U_i$ (attention to uncertainty) as the average of responses to the first two factors and use the third for $Attn_Others_i$ (attention to others' applications). The nine other factors deal with school quality: teachers' quality, peer quality, etc. I define $Attn_Q_i$ (attention to quality) as the average of the responses to these questions.

A sophisticated parent cares about uncertainty, which implies a positive correlation between sophistication and $Attn_U_i$. While the correlation between $Attn_U_i$ and parents' performance in the game is investigated in section 5.2, I first explore how family background is correlated with $Attn_U_i$.

Table 9 presents the regression results of $Attn_U_i$ on family background and student characteristics.

Table 9: Attention to 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)	4.34	4.36	4.36	4.35	4.19
Std Dev(Dep V)	0.74	0.72	0.71	0.70	0.90
<i>Parent_Edu_i</i>	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	-0.02 (0.05)
<i>Parent_Inc_i</i>	-0.04 (0.04)	-0.05 (0.03)	-0.06* (0.03)	-0.07* (0.04)	0.55** (0.23)
<i>Own_Score_i</i>	0.23 (0.41)	0.29 (0.40)	0.31 (0.40)	0.94* (0.51)	1.74 (3.57)
<i>Awards_i</i>	0.02 (0.03)	0.04 (0.03)	0.04 (0.03)	0.06* (0.03)	0.05 (0.08)
<i>Girl_i</i>	-0.05 (0.05)	-0.04 (0.05)	-0.03 (0.05)	-0.02 (0.06)	-0.25 (0.21)
<i>Attn_Others_i</i>	-0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.02)	0.07 (0.09)
<i>Attn_Q_i</i>	0.82*** (0.06)	0.82*** (0.06)	0.80*** (0.06)	0.75*** (0.06)	0.76*** (0.24)
Constant	-0.01 (2.14)	-0.17 (2.09)	-0.14 (2.08)	-3.23 (2.65)	-12.60 (18.43)
Observations	676	605	597	457	71
R-squared	0.27	0.29	0.28	0.28	0.36

Notes: a. The full sample includes every parent for whom relevant variables are not missing. The participants include those who submit a list that is not (0,0,0,0). The subsample (≥ 2 schools) includes participants whose submitted lists have at least 2 schools. The subsample with a full list includes those who submit a full list. Elementary school fixed effects are included. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Column (1) shows that family background has no significant correlation with $Attn.U_i$ in the full sample. I exclude non-participants (column 2) and those submitting partial lists (columns 3 and 4). The negative coefficient on income becomes significant at the 10% level and larger in magnitude, particularly in the subsample of those submitting full lists. As a comparison, this coefficient is significantly positive (at the 5% level) for non-participants (column 5). This negative correlation between $Attn.U_i$ and parents' income is robust if $Attn.U_i$ is broken down into attention to admission quota and attention to the probability of getting into bad schools (Table J15 in Appendix J).²⁰

5 Model Estimation and Test Results

This section presents the estimates from four cases: (i) degenerate beliefs with moment equalities allowing heterogeneous sophistication; (ii) degenerate beliefs with (in)equalities; (iii) non-degenerate and possibly heterogeneous beliefs, which include Bayesian Nash equilibrium and truth-telling as special cases; and (iv)

²⁰In Table J15, the same regression is run for $Attn.Others_i$: 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. Contrary to those from the $Attn.U_i$ regressions, parents' income is significantly positively correlated with $Attn.Q_i$.

truth-telling. Throughout, the utility function is specified as

$$\begin{aligned}
u_{i,s} = & \alpha_s + \beta_{X,1}Own_Score_i + \beta_{X,2}Parent_Inc_i + \beta_{X,3}Parent_Edu_i + \beta_{X,4}Awards_i + \beta_{X,5}Girl_i \\
& + \beta_{Z,1}Distance_{i,s} + \beta_{Z,2}Own_Score_i \times School_Score_s + \beta_{Z,3}Parent_Inc_i \times School_Score_s \\
& + \beta_{Z,4}Parent_Edu_i \times School_Score_s + \beta_{Z,5}Awards_i \times School_Score_s + \varepsilon_{i,s},
\end{aligned} \tag{4}$$

where α_s is middle school fixed effects; $School_Score_s$ is the (log) average test score of graduating students from 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 all schools, captures the observed heterogeneity of the outside option. Finally, $(\varepsilon_{i,1}, \dots, \varepsilon_{i,4})$ are *i.i.d.* type I extreme values.

To guarantee them in $[0, 1]$, the average mixing probabilities (Assumption MIXING) are as follows:²¹

$$\begin{aligned}
m^{1,2}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{1,2} + \gamma_2^{1,2}Parent_Inc_i + \gamma_3^{1,2}Parent_Edu_i \right] \right), \\
m^{1,4}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{1,4} + \gamma_2^{1,4}Parent_Inc_i + \gamma_3^{1,4}Parent_Edu_i \right] \right) (1 - m^{1,2}(\mathbf{X}_i)), \\
m^{2,4}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{2,4} + \gamma_2^{2,4}Parent_Inc_i + \gamma_3^{2,4}Parent_Edu_i \right] \right).
\end{aligned} \tag{5}$$

where \arctan is the inverse trigonometric function, inverse tangent.

5.1 Estimation Results

Table 10 presents the coefficients of the main variables in the utility function. Standard errors for maximum likelihood estimates are from the outer product of gradients; the 95% marginal confidence intervals in the case with (in)equalities (column 2) are based on Bugni, Canay, and Shi (2016).²²

The results show that parents dislike distance, while the way in which parents value the $School_Score_s$ varies. The mixing probabilities evaluated at three combinations of parents' income and education (10th, 50th, and 90th percentiles of both variables) are also reported. In all estimations except one, wealthier and more-educated parents are less likely to include unacceptable schools in their lists, while the majority tend to submit full lists.

One may be interested in the marginal effects of variables in the interaction terms. For example, as used later in the main analyses, Table 11 shows these results for the case of degenerate beliefs with (in)equalities. Both Own_Score_i and $Parent_Inc_i$ decrease the utility of every school, because they lead to better outside options, while $Parent_Edu_i$, $Awards_i$, and $Girl_i$ have a positive effect. Each variable has impacts of a

²¹Note that $m^{1,1}(\mathbf{X}_i) = 1 - m^{1,2}(\mathbf{X}_i) - m^{1,4}(\mathbf{X}_i)$ and $m^{2,2}(\mathbf{X}_i) = 1 - m^{2,4}(\mathbf{X}_i)$.

²²The confidence intervals are wide, partly because the method tends to construct conservative intervals.

Table 10: Preferences for Schools: Model Estimation Results from Different Cases

	Degenerate Beliefs		Non-Degenerate Beliefs (3)	Truth-Telling (4)
	Equalities Only (1)	Equalities & Inequalities ^a (2)		
<i>Distance_{i,s}</i>	-0.27*** (0.06)	-2.95 [-8.64, 0.05]	-0.24*** (0.07)	-0.26*** (0.06)
<i>Own_Score_i × School_Score_s</i>	-1.46 (14.53)	79.59 [-6.87, 160.98]	49.91*** (6.90)	52.00*** (6.39)
<i>Parent_Inc_i × School_Score_s</i>	28.86*** (0.02)	0.62 [-1.27, 32.51]	1.03 (1.06)	1.40 (0.90)
<i>Parent_Edu_i × School_Score_s</i>	-2.20*** (0.02)	2.08 [-3.46, 13.70]	0.39 (0.37)	0.44 (0.29)
<i>Awards_i × School_Score_s</i>	-1.15*** (0.17)	-2.21 [-7.14, 4.33]	2.22** (1.00)	3.02*** (0.71)
<i>Girl_i × School_Score_s</i>	6.01*** (0.02)	3.55 [-1.24, 15.86]	0.15 (1.24)	-0.56 (1.01)
<i>Own_Score_i</i>	4.04 (91.25)	-1259.00 [-1992.14, 83.30]	-314.12*** (43.53)	-326.15*** (40.11)
<i>Parent_Inc_i</i>	-181.85*** (0.14)	-10.41 [-192.14, -3.30]	-6.87 (6.64)	-9.22 (5.64)
<i>Parent_Edu_i</i>	13.72*** (0.09)	-12.60 [-21.90, 15.82]	-2.53 (2.34)	-2.91 (1.81)
<i>Awards_i</i>	6.91*** (1.06)	17.32 [-10.90, 23.55]	-14.21** (6.25)	-19.23*** (4.43)
<i>Girl_i</i>	-37.62*** (0.13)	-17.34 [-40.07, -4.51]	-0.76 (7.80)	3.67 (6.35)
Mixing Probabilities				
$m^{1,2}$: at p10 of income & education	0.19	0.23	0.05	21.08
$m^{1,2}$: at median income & education	0.13	0.17	0.05	25.09
$m^{1,2}$: at p90 of income & education	0.11	0.09	0.05	30.62
$m^{1,4}$: at p10 of income & education	93.40	99.57	95.18	74.75
$m^{1,4}$: at median income & education	94.87	1.84	94.91	70.36
$m^{1,4}$: at p90 of income & education	95.51	0.27	94.51	64.33
$m^{2,4}$: at p10 of income & education	93.37	99.99	85.37	99.88
$m^{2,4}$: at median income & education	88.58	0.06	74.45	99.82
$m^{2,4}$: at p90 of income & education	63.70	0.01	47.36	99.51

Notes: Middle school fixed effects are included in all cases, as specified in equation (4). $m^{k,l}$ and $l \geq k$ denote the average mixing probability that an l -school list is submitted when only k schools are acceptable. The function is parameterized in equation (5). The table reports the function evaluated at the 10th, 50th, and 90th percentiles of income and education. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

a. Estimation with both equalities and inequalities are based on Andrews and Shi (2013); the 95% marginal confidence intervals in brackets are calculated following the approach in Bugni, Canay, and Shi (2016).

similar size on each school. However, a higher own score, higher parental income and education, and being a girl make the student value school scores slightly more and thus prefer School 1 more.

Goodness of Fit Comparison It is not obvious how the four cases differ in Table 10, so I compare them in terms of the goodness of fit. Details are collected in Appendix I, where I also discuss the sources of biases. For each school, I calculate the predicted average utility and the probability of being most preferred. These predictions are then compared with claimed best school in the survey and actual enrollment in 2002; the

Table 11: Marginal Effects of Individual Characteristics

School	Own_Score_i	$Parent_Inc_i$	$Parent_Edu_i$	$Awards_i$	$Girl_i$
1	-7.55	-0.06	0.59	3.32	5.11
2	-7.61	-0.07	0.45	3.47	4.87
3	-7.63	-0.07	0.39	3.53	4.77
4	-7.69	-0.07	0.23	3.71	4.49

Notes: The calculation is based on the estimates from degenerate beliefs with (in)equalities. For $Parent_Inc_i$ and Own_Score_i , the table reports the change in each school's utility as a result of a 1% increase in the variable. For $Parent_Edu_i$, $Awards_i$ and $Girl_i$, it reports the change in utility when the variable increases by one unit or from 0 to 1.

results are in favor of the degenerate beliefs with (in)equalities. Moreover, the estimates for this case fit the moment equalities well. Therefore, this set of estimates is used for the following analysis.

5.2 Sophistication and Incentives to Be Strategic

This section investigates who strategizes better and how parents respond to incentives. Measures of sophistication and incentives are constructed using estimated preferences and empirical equilibrium beliefs. Equivalently, the observed individual actions are assumed to be random draws from the distribution of equilibrium actions, and therefore, the empirical beliefs, constructed by bootstrapping from the empirical distribution of actions, provide an approximation.²³

5.2.1 Deviations from Best Response and Truth-Telling: Overcautiousness

As section 3.6.1 shows, deviations from playing the best response on average should be zero if everyone plays best responses, so are deviations from truth-telling if everyone is truth-telling. For the 24 full lists, Table 12 presents summary statistics for how the observed behaviors deviate from best responses and truth-telling. For each list, I run a t-test for the null hypothesis that the mean of each deviation equals zero independently, and only in 1 case (out of 48) is the null not rejected at the 5% level.

Table 12 highlights the importance of distinguishing between best response and truth telling, as they lead to contrasting predictions. For example, (1, 2, 3, 4) is the most common true preference order and played by 13.89% of the parents. The truth-telling assumption predicts that 24.86% should choose that list, whereas only 9.66% do so under the best response hypothesis. A large discrepancy is also found for the

²³To calculate the empirical beliefs, \hat{B}^* , the distribution of equilibrium strategies is approximated by the 914 observations augmented by the 9 lists that are not played by anyone. Additionally, 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 with 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 process for the 5,000 samples, I calculate $\hat{B}^* = \sum_{n=1}^{5000} \hat{B}^n / 5000$.

Table 12: Deviation from Best Response and Truth-Telling Predictions: Full Lists

List	Observed Fraction ^a (In Percentage Points) (1)	Deviation (in Percentage Points) from the Prediction of		
		Best Responding ^b (2)	Truth-Telling ^b (3)	Maximin Preferences ^{b,d} (4)
(1,3,2,4)	5.14	-11.40	-8.09	0.20
(1,3,4,2)	0.77	-0.21	-2.22	-0.06
(1,2,3,4)	13.89	4.24	-10.96	10.86
(1,2,4,3)	2.41	2.02	-0.25	0.29
(1,4,2,3)	0.22	0.19	-0.93	-0.32
(1,4,3,2)	0.00	-6.03	-1.83	-1.84
(1,?,?,?) ^c	22.43	-11.20	-24.29	9.13
(2,3,4,1)	4.70	4.69	4.53	4.50
(2,3,1,4)	21.33	12.18	19.62	-5.32
(2,4,3,1)	1.09	0.40	0.97	0.97
(2,4,1,3)	0.77	0.76	0.56	-0.67
(2,1,4,3)	2.74	2.29	1.74	2.24
(2,1,3,4)	12.58	-7.46	1.96	9.22
(2,?,?,?) ^c	43.22	12.86	29.38	10.94
(3,4,2,1)	0.22	-0.31	-0.30	-0.43
(3,4,1,2)	0.22	-0.32	-1.54	-2.69
(3,2,4,1)	0.88	0.79	0.59	0.61
(3,2,1,4)	4.16	3.01	2.31	3.97
(3,1,2,4)	0.88	-3.69	-4.81	-19.27
(3,1,4,2)	0.22	-1.31	-2.37	-3.07
(3,?,?,?) ^c	6.56	-1.83	-6.12	-20.88
(4,3,2,1)	0.55	0.55	-0.02	0.14
(4,3,1,2)	0.00	0.00	-2.26	-2.84
(4,2,3,1)	0.00	-1.43	-0.20	-0.21
(4,2,1,3)	0.00	0.01	-0.22	-1.06
(4,1,2,3)	0.00	-0.02	-0.48	-0.14
(4,1,3,2)	0.00	-5.51	-1.69	-1.45
(4,?,?,?) ^c	0.55	-6.42	-4.87	-5.56

Notes: a. This reports the percentage out of the total of 914 students who submit the given list. b. This is the average individual deviation. A positive number implies the observed fraction is higher than the model prediction. All deviations are different from zero at 5% level except the list (2, 4, 1, 3) under the assumption of best responses. c. (c^1 , ?, ?, ?) denotes all the full lists that rank c^1 as first choice. d. Maximin preferences are introduced and discussed later in section 6.4

most under-used list, (2, 3, 1, 4).

These results provide additional evidence that parents are neither all-best-responding nor all-truth-telling in the game. Instead, parents are **overcautious** in that they low-rank School 1 and top-rank School 2 too often. In the data, 22.43% top-rank School 1 in a full list, while the model predicts that 33.62% should have optimally done so given other parents' behavior. For School 2, 43.22% rank it first in a full list, whereas the model predicts only 30.35% should have done so had they best responded. Top ranking School 2 and avoiding School 1 is certainly *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 too many parents choose the “safe” strategy, the overcautiousness leads to a coordination failure.

I repeat this exercise for the partial lists, including $(0, \dots, 0)$, and details are in appendix Table J16.

5.2.2 Incentives to Be Strategic

Table 13 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 $P_i^{TT=BR}$) but also have a slightly greater incentive to move away from always truth-telling (higher $Gain_i$). The correlation between the two measures is -0.42.

The probability that truth-telling is a best response, $P_i^{TT=BR}$, is relatively high at 0.54 (standard deviation 0.30). In both regressions, $Parent_Inc_i$, $Parent_Edu_i$, and Own_Score_i are positively correlated with $P_i^{TT=BR}$, whether controlling for $Gain_i$ or not.

$Gain_i$ is the utility gain when changing from always truth-telling to always best responding, measured as a fraction of the latter. The mean gain is 0.03 with a standard deviation of 0.05. Whether controlling $P_i^{TT=BR}$ or not, $Gain_i$ is positively correlated with $Parent_Edu_i$ and Own_Score_i .

Table 13: Determinants of Incentive to Be Strategic: Regression Analysis

	$P_i^{TT=BR}$		$Gain_i$	
Mean(Dep V)	0.54		0.03	
Std Dev(Dep V)	(0.30)		(0.05)	
$Gain_i$		-2.88*** (0.13)		
$P_i^{TT=BR}$				-0.13*** (0.01)
$Parent_Inc_i$	0.03** (0.01)	0.02** (0.01)	-0.00 (0.00)	0.00 (0.00)
$Parent_Edu_i$	0.01* (0.00)	0.01*** (0.00)	0.00** (0.00)	0.00*** (0.00)
Own_Score_i	2.26*** (0.18)	2.49*** (0.20)	0.08*** (0.02)	0.38*** (0.05)
$Awards_i$	-0.00 (0.01)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
$Girl_i$	-0.01 (0.02)	-0.00 (0.01)	0.00 (0.00)	0.00 (0.00)
Obs.	914	914	914	914
R^2	0.36	0.60	0.05	0.41

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

5.2.3 Who Strategizes Better?

To investigate which parents strategize better, Table 14 reports how family background affects deviations from best responses. I focus on two lists, the most under-used list $(1, 3, 2, 4)$ and the most over-used $(2, 3, 1, 4)$, relative to the best response prediction. Neither of them is dominated *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 – 12 percentage points more often than best-responding parents would, and the list, (1, 3, 2, 4), is under-used by 11 percentage points.

Table 14: Who Strategizes Better: Regression Analysis of Deviations from Best Responding

	Most Under-Used List: (1,3,2,4)				Most Over-Used List: (2,3,1,4)			
	Depend. V: Deviation from Best Responding Mean: -0.11; Std Dev: 0.32				Depend. V: Deviation from Best Responding Mean: 0.12; Std Dev: 0.42			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Parent_Inc_i</i>	0.02 (0.01)	0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
<i>Parent_Edu_i</i>	-0.01 (0.01)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)
<i>Own_Score_i</i>	-0.42*** (0.11)	-0.52*** (0.12)	-0.68*** (0.17)	-0.64*** (0.18)	-0.84*** (0.25)	-0.73*** (0.25)	-0.86*** (0.33)	-0.90** (0.37)
<i>Awards_i</i>	0.02* (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	-0.04** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)
<i>Girl_i</i>	-0.05** (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.04 (0.03)
$P_i^{TT=BR}$			0.12** (0.05)	0.12** (0.05)			0.10 (0.07)	0.12 (0.08)
<i>Gain_i</i>			-1.27*** (0.18)	-1.33*** (0.20)			-1.34*** (0.33)	-1.30*** (0.38)
<i>Attn_U_i</i>				-0.00 (0.01)				0.04** (0.02)
<i>Attn_Q_i</i>				-0.02 (0.02)				-0.02 (0.04)
<i>Attn_Others_i</i>				-0.00 (0.01)				0.01 (0.01)
$P_{i,k}^{TT}$		-1.32*** (0.07)	-1.22*** (0.06)	-1.24*** (0.07)		0.87 (0.72)	0.79 (0.71)	0.53 (0.77)
Obs.	914	914	914	810	914	914	914	810
R-Squared	0.06	0.33	0.40	0.41	0.04	0.05	0.09	0.09

Notes: The regressions follow equation 3 with the outcome variable, $d_{i,k} - P_{i,k}^{BR}$, defined for the list in question. The probability that truth-telling is the best response, $P_i^{TT=BR}$, and the gain of switching from truth-telling to playing the best response, $Gain_i$, are defined in Table 13. $P_{i,k}^{TT}$ is the probability that the list (in the dependent variable) is in the true preference order. Elementary school fixed effects are included. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In the left half of Table 14, following equation 3, I regress the deviations on 4 sets of regressors, with or without controlling incentive measures and/or attention measures (columns 1-4). For the most under-used list, higher Own_Score_i robustly leads to more under-utilization of the strategy (1, 3, 2, 4). When including incentive measures (columns 2-4), the effects of $Parent_Inc_i$, $Parent_Edu_i$, Own_Score_i , $Awards_i$, and $Girls_i$ are close to zero. Additionally, $P_i^{TT=BR}$ reduces the deviation, and $Gain_i$ increases the deviation; the attention measures have no significant effect.

The above results imply that family background does not offset underutilization of the strategy. However, the lower cost of finding a best response (higher $P_i^{TT=BR}$) reduces this deviation, and parents' income and education are positively correlated with $P_i^{TT=BR}$ (Table 13). Moreover, the potential gain of switching

from always truth-telling to always playing the best responds, $Gain_i$, increases the deviation, which implies the difficulty of finding a best response. It is unlikely that wealthier and more-educated parents are more sophisticated: they do not respond to incentives more optimally.²⁴

The last 4 columns in Table 14 show the results for the most over-used list (2, 3, 1, 4). In all regressions, only Own_Score_i and $Awards_i$ among the individual characteristics have a significant (negative) effect. For the two incentive measures, only $Gain_i$ is significant. Surprisingly, attention to 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), the results are presented in appendix Table J18.

To summarize, these results show that, on average, parents are overcautious because they avoid top-ranking School 1 more often than they should. There is no evidence that wealthier and more-educated parents are more sophisticated, but the game is easier for them to play because their true preference ranking is more likely to coincide with their best response. Poorer parents pay more attention to uncertainty, but it does not mitigate—and sometimes exacerbates—the overcautiousness.

6 Counterfactual Analyses

When analyzing the welfare effects of replacing the Boston mechanism by DA, I follow the literature and focus on two types of parents: naïve and sophisticated. Although evidence shows that more types exist in the data, these two focal types provide a benchmark. Given the equilibrium beliefs, the welfare of a sophisticated type is the upper bound of achievable payoffs, while that of a naïve parent is likely to be a reasonable lower bound because one may choose to be truth-telling when failing to find a best response.

I consider two experiments: (i) all other parents behave as in the data, and (ii) parents are either naïve or sophisticated. The assumption for the first experiment is the same as the above analyses on strategic behavior, and I use the bootstrapped empirical equilibrium beliefs as an approximation of the equilibrium.

6.1 Simulating Outcomes under the Two Mechanisms

In simulations, everyone reports truthfully under DA. The equilibrium admission probabilities are obtained by drawing a large number of profiles of preferences rankings, simulating the outcomes, and weighting them by the probability of observing each profile.

²⁴When $Gain_i$ is interacted with family background and included in regression (3), the coefficients of the interaction terms are $Gain_i * Parent_Inc_i$: 0.06, s.e. = 0.31; $Gain_i * Parent_Edu_i$: -0.18, s.e. = 0.08; which imply that more educated parents are more cautious such that they under-play the list more when $Gain_i$ is higher.

Under the Boston mechanism in the second experiment, equilibrium beliefs are solved as a fixed point. Eleven cases, each with 0, 10%, ..., or 100% naïve parents, are considered. Naïve parents in each case are randomly chosen, and the remaining are sophisticated. A naïve 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 G. When all parents are naïve, the equilibrium probabilities of being assigned to each school are calculated similar to that in DA. Throughout, the mixing probabilities are held constant, as estimated from the data. Therefore, due to part (iv) of Proposition 1, each parent has a unique best response conditional on equilibrium beliefs.

After solving an equilibrium, I simulate parents' actions and calculate welfare. For each parent, 1,000 profiles of preferences are constructed using random draws of utility shocks, and each of them (with each draw) plays two types of games – DA and the Boston; in the latter, parents play it as sophisticated and then as naïve. Each parent's average welfare (averaged over the 1000 draws/plays) is studied.

6.2 Other Parents Behaving as in the Data

The first experiment considers the equilibrium in the data, so the empirical beliefs, which are constructed as in footnote 23, are taken as the equilibrium beliefs. Table 15 reports the effects of changing from the Boston mechanism to the DA mechanism in Beijing, while other parents behave as they do in the data under the Boston mechanism.

Switching to DA hurts both naïve and sophisticated parents on average (columns 1 and 2). For naïve parents, the average utility loss is 0.22, equivalent to an 8% increase in the distance from home to school (or replacing a 13% chance at the best school with that at the second best). Sophisticated parents suffer more from such a reform: The average utility loss is -0.99, equivalent to a 40% increase in the distance or substituting a 60% chance at the best school with an equal chance at the second best.

Inter-personal welfare comparison is implicitly assumed when calculating average welfare, so in columns (3)-(7), I consider individual welfare change. On average, 19% of the parents, whether naïve or sophisticated, achieve the same welfare in either case, as they do not participate (column 5). Among naïve parents, 27% are better off, and surprisingly, 55% are worse off. Based on regressions (columns 3 and 6), naïve students from high-income families are more likely to be unaffected by the reform because of better outside options. Moreover, students with better grades are much less likely to be worse off.

Among sophisticated parents, 66% are worse off under DA, while 15% are better off (columns 4 and 7). Students from high-income families or those with more-educated parents or higher grades are less likely to be worse off. The main reason is again their better outside options.

Table 15: Welfare Effects of Replacing the Boston Mechanism with DA: Regression Analyses Given the Observed Equilibrium

	Mean Utility Diff ^a		Prob(Better off)		Prob(Indiff.) ^b	Prob(Worse off)	
	Naïve (1)	Sophist. (2)	Naïve (3)	Sophist. (4)	Naïve/Sophist. (5)	Naïve (6)	Sophist. (7)
mean(Dep V)	-0.22	-0.99	0.27	0.15	0.19	0.55	0.66
S.D(Dep V)	1.06	1.92	0.30	0.23	0.37	0.36	0.38
<i>Parent_Inc_i</i>	0.02 (0.03)	0.19*** (0.04)	-0.03** (0.01)	-0.01 (0.01)	0.05** (0.02)	-0.02 (0.01)	-0.04** (0.02)
<i>Parent_Edu_i</i>	0.02 (0.01)	-0.03** (0.01)	0.00 (0.01)	0.00 (0.00)	0.01** (0.01)	-0.02*** (0.01)	-0.02*** (0.01)
<i>Own_Score_i</i>	10.95*** (1.05)	27.32*** (0.66)	0.53*** (0.16)	0.80*** (0.10)	2.14*** (0.26)	-2.68*** (0.32)	-2.94*** (0.29)
<i>Awards_i</i>	-0.06** (0.03)	-0.21*** (0.02)	-0.00 (0.01)	-0.00 (0.01)	0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
<i>Girl_i</i>	0.01 (0.05)	-0.13*** (0.04)	0.02 (0.02)	0.01 (0.01)	-0.04* (0.02)	0.02 (0.02)	0.03 (0.02)
Observations	914	914	914	914	914	914	914
R-squared	0.55	0.91	0.10	0.17	0.24	0.38	0.40

Notes: Each column regresses the dependent variable on the listed explanatory variables and the fixed effects of elementary schools. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Under DA, everyone is truth-telling. Under the Boston mechanism, naïve parents are truth-telling, and sophisticated ones play a best response, given the empirical equilibrium beliefs. a. The mean utility difference is defined as the average expected utility obtained under the DA minus that obtained from the Boston mechanism. b. The probability of being indifferent between the Boston and DA mechanisms is the same for both naïve and sophisticated parents because these are parents who do not participate at all under either mechanism.

6.3 Equilibrium with Naïve and Sophisticated Parents Only

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

Figure 2 reports the welfare effects when the Boston mechanism is replaced by the DA mechanism.²⁵ The first sub-figure reports the average expected utility normalized under the DA mechanism.²⁶ When the fraction of naïve parents is low (< 80%), DA benefits naïve parents and hurts sophisticated ones on average: the average loss for sophisticated parents is small, 0.1 (equivalent to 3% increase in the distance to school or substituting a 6% chance at the best school with an equal chance at the second best), while the average gain for naïve parents is rather big, 0.46 (equivalent to 17% decrease in the distance or substituting a 28% chance at the second best school with an equal chance at the best). However, when the fraction of naïve parents is 80% or larger, naïve parents are also hurt by DA, while the loss for sophisticated parents is three times as

²⁵In all three subfigures, there is a big discontinuity when the fraction of naïve parents is 40%. Because this discontinuity may be due to simulation errors, I do not single it out when discussing the results, while not omitting the result, either.

²⁶Given that schools do not rank students *ex ante*, the school choice problem is equivalent to the assignment problem to which Bogomolnaia and Moulin (2001) propose a mechanism called the Probabilistic Serial. The results of the DA presented here are an approximation of the Probabilistic Serial, as Che and Kojima (2010) show the two mechanisms are asymptotically equivalent. I thank a referee for bringing this point to my attention.

large as before.

The overall average welfare is a weighted sum of the two groups' welfare. The DA mechanism performs better when there are 20-70% naïve parents; otherwise, the Boston mechanism achieves higher overall average welfare.

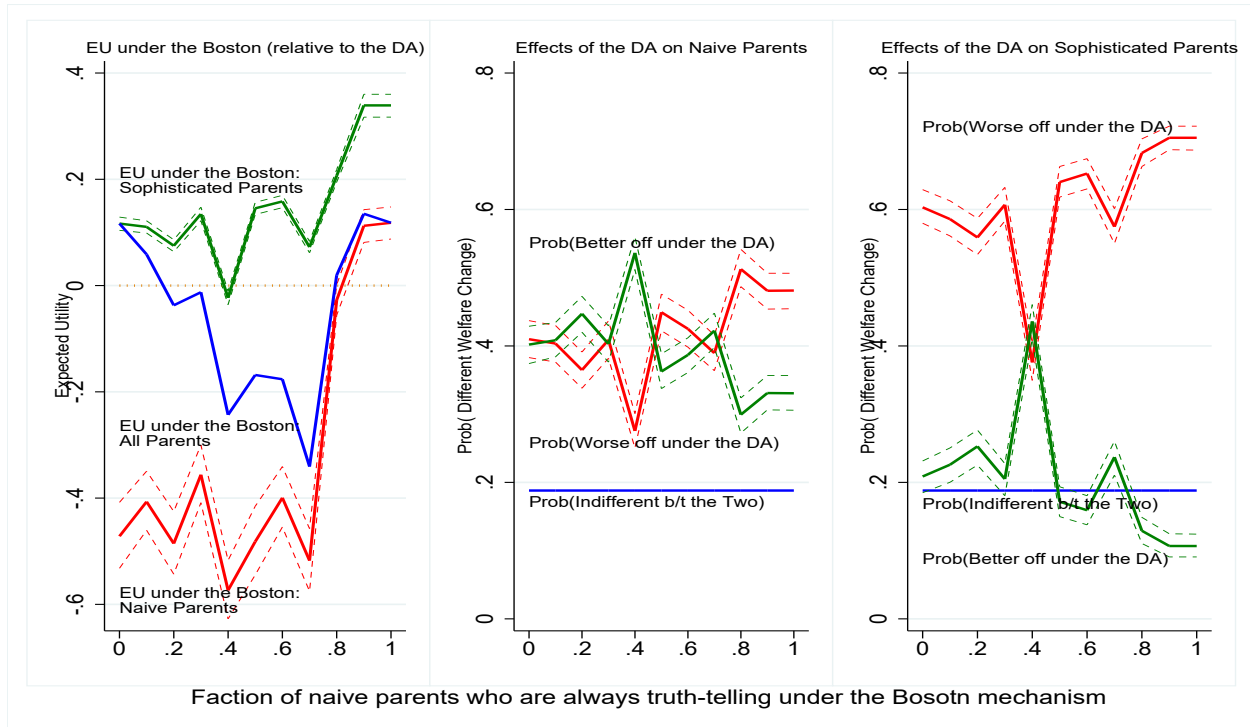


Figure 2: Welfare Effects of Replacing the Boston with DA: Naïve and Sophisticated Parents

Notes: In all simulations, parents play the DA and the Boston mechanisms. Everyone is truth-telling under DA; under the Boston mechanism, naïve parents are always truth-telling, and sophisticated parents always play the best response. Equilibrium is solved, with fractions of naïve parents ranging from 0% to 100%. Given each equilibrium, all parents behave sophisticatedly and then naïvely under the Boston mechanism. Welfare effects in each case are calculated for everyone. The average expected utility under DA is normalized to zero. The dashed lines depict the 95% confidence intervals for the corresponding measures; they are the 2.5th and 97.5th percentiles among the welfare measures evaluated at the point estimates across the 1000 simulation samples.

Furthermore, the middle and right subfigures show the probabilities of welfare changes for naïve and sophisticated parents, respectively. Approximately 19% of the parents are always indifferent between the mechanisms because they do not participate. Among naïve parents, approximately 39% of them are better off under DA when averaging over all possible fractions of naïve parents, while 42% are worse off. The reason that so many naïve parents are hurt by DA is that the Boston mechanism provides a coordination opportunity; thus, sophisticated parents avoid popular schools and leave a greater opportunity for naïve ones. For sophisticated parents, DA is more harmful: on average, only 20% are better off, while 61% are worse off. Both types of parents are more likely to be worse off when there are more naïve parents.

The different results from the two counterfactual exercises highlight the significance of parents' over-

cautiousness, and, more generally, the importance of allowing flexible behavioral types of parents besides being naïve or sophisticated. Because DA delivers the same outcome in both experiments, one can compare the welfare between the two counterfactuals. Given overcautiousness, the Boston mechanism helps 55% of naïve parents (Table 15): Others’ overcautiousness gives a naïve parent greater odds at the best school. In contrast, in the case with two types of parents, the fraction of naïve parents who are better off under the Boston mechanism is between 28-51% and 42% on average (Figure 2).

6.4 Overcautiousness and Maximin Preferences

Overcautiousness can be an indicator of maximin preferences (Gilboa and Schmeidler 1989). When a maximin parent i has little information about others’ behaviors, i considers a set of beliefs to be possible. Being uncertainty averse, i takes into account the minimal expected utility over all possible beliefs and chooses a strategy to maximize that. One possible belief system can be that the majority of parents are truth-telling and thus top-rank School 1, which leads to a low expected utility for i had i top-ranked School 1. With maximin preferences, i tends to top rank a different school, e.g., School 2.

Below, I first compare the observed actions to those predicted by maximin preferences and show that it is unlikely that all parents are as uncertainty averse as maximin, although it fits the data better than the best-responding and truth-telling assumptions in some aspects. I then measure the welfare effects of replacing the Boston mechanism by the DA mechanism when everyone has maximin preferences, which shows DA harms more than half of them.

More formally, the maximin strategy of i , σ^{mm} , maximizes her lowest possible payoff:

$$\sigma^{mm}(\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i) \in \arg \max_{\sigma \in \Delta(\mathcal{C})} \left\{ \min_{C_{-i} \in \mathcal{C}^{I-1}} \sum_{s=1}^4 [a_s(\sigma, C_{-i}) \max(u_{i,s}, u_{i,0})] \right\},$$

where the minimization over C_{-i} amounts to choosing the worst-case scenario for i , given her own strategy σ . Moreover, truth-telling is still a dominant strategy under DA for maximin parents.

It is straightforward to simulate σ^{mm} , but \mathcal{C}^{I-1} contains too many possible profiles. The simulation draws 68,000 profiles, based on which the “maximin” operation is performed. Combining σ^{mm} with the 1000 simulated preference profiles, I calculate the outcomes by running the Boston with lotteries.

The outcomes lead to column (4) of Table 12, which reports the difference between observed actions and maximin predictions. Several interesting patterns emerge: for the lists top-ranking School 1 or 2, i.e., (1, ?, ?, ?) or (2, ?, ?, ?), parents choose it more often than the maximin prediction, although the predictions are more similar than those from best-responding or truth-telling. Moreover, the maximin prediction top-

ranks School 3 significantly more often than observed (by 21 percentage points). Indeed, with maximin preferences, School 3, which is the third most popular school with a large capacity (310 seats), seems to be an even safer choice. These results, however, imply that not all parents have maximin preferences.

Table 16: Welfare Effects of Replacing the Boston Mechanism with DA on Maximin Students

<i>Panel A: Summary Statistics</i>				
	All Parents	Under DA, parents who are		
		Better off	Indifferent	Worse off
Fraction (Percentage Points)	100	18.93	23.59	57.48
Average Difference in Utility	-0.22	0.19	0.00	-0.44
<i>Panel B: Regression Analysis</i>				
Outcome Variable:	Average Utility Difference	Pr(Better Off)	Pr(Indifferent)	Pr(Worse Off)
<i>Parent_Inc_i</i>	0.01 (0.01)	-0.01 (0.01)	0.04 ** (0.02)	-0.03 ** (0.01)
<i>Parent_Edu_i</i>	0.00 (0.00)	0.00 (0.00)	0.01 * (0.01)	-0.01 ** (0.01)
<i>Own_Score_i</i>	0.01 (0.10)	-0.45 *** (0.08)	1.69 *** (0.27)	-1.24 *** (0.21)
<i>Awards_i</i>	0.02 ** (0.01)	-0.01 ** (0.00)	0.02 (0.01)	-0.01 (0.01)
<i>Girl_i</i>	-0.01 (0.01)	0.02 ** (0.01)	-0.03 (0.02)	0.02 (0.02)
Observations	914	914	914	914
R-squared	0.13	0.17	0.19	0.16

Notes: Each column reports a regression of the outcome variable on the listed explanatory variables and the fixed effects of elementary schools. The welfare measures are calculated from 1000 simulated samples.

I then investigate the welfare effects of switching from the Boston mechanism to the DA mechanism when everyone is maximin. Parents under DA report true ordinal preferences and play σ^{mm} under the Boston mechanism. Panel A of Table 16 shows that DA decreases average welfare by 0.22, equivalent to increasing distance by 8% or moving a student's 13% chance at the best school to the second best. Furthermore, the majority, 57%, are substantially worse off, while only 19% of them are slightly better off.

Panel B of Table 16 studies who is more likely to be worse off: those who have less-educated or less-affluent parents (the last regression). Similar to the previous results, given that they have better outside options for their children, more educated and/or wealthier parents tend to have only one acceptable school and thus rank them truthfully, even when they have maximin preferences. They have a better chance at the top school, when others are maximin and thus avoid the school.

7 Discussion and Concluding Remarks

This paper uses data from Beijing on school choice under the Boston (Immediate-Acceptance) mechanism to answer two questions: (i) whether poorer and/or less-educated parents are more likely to be naïve, and

(ii) whether the Boston mechanism harms naïve parents more than DA does.

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, and parents may make mistakes when strategizing. I then derive a set of dominated strategies as identifying conditions and obtain preference estimates that are robust to a wide range of parents' strategic mistakes.

Based on the estimated preferences, the analysis of observed parent behavior rejects the hypotheses that everyone is of the same type: naïve or sophisticated. Parents are revealed to be overcautious in the sense that they avoid top-ranking the most popular school more often than prescribed by best responding. A better outside option offsets the overcautiousness to some degree because it makes the true preference order more likely to be a best response. Although wealthier and more-educated parents tend to have a better outside option, there is no evidence that they are more sophisticated. On the contrary, poorer parents pay more attention to uncertainty in the game, which indicates that they strive to find a best response. However, paying more attention to uncertainty does not help and sometimes even worsens their overcautiousness.

Given parents' observed behavior, especially their overcautiousness, when replacing the Boston mechanism by DA, both naïve and sophisticated parents suffer a utility loss. For naïve parents, the loss is equivalent to an 8% increase in the distance to the school or substituting a 13% chance at the best school with an equal chance at the second best. The magnitude of this effect is more than 4 times as large for sophisticated parents. Additionally, only 27% of the naïve parents are better off under DA, while 55% are actually worse off. The negative effects decrease with parents' income and education because of outside options.

If every parent is either sophisticated or naïve and no one is overcautious, switching from the Boston mechanism to DA has mixed effects. In terms of average expected utility, when there are not as many naïve parents (<80%), DA helps naïve parents while hurting sophisticated parents slightly; but when at least 80% of parents are naïve, the naïve ones do not benefit from DA, either. Furthermore, in terms of individual welfare change, 39% of naïve parents are better off under DA, while 42% are actually worse off. The effect of DA is more negative for sophisticated parents: only 20% are better off, while 61% are worse off.

The difference between the above two counterfactual analyses highlights the heterogeneity of parents' sophistication in the data. When one considers a two-type model, the overcautiousness is assumed away, while the data suggest that there are more types other than being naïve or sophisticated.

7.1 The Boston Mechanism in Practice?

A few caveats are in order regarding the paper's external validity. Schools in Beijing have no pre-determined priorities for students; Abdulkadiroglu, Che, and Yasuda (2011) show that Bayesian Nash equilibrium under the Boston mechanism can dominate that under DA with single tie-breaker (or, equivalently, Random Serial Dictatorship). However, this result is not robust to the introduction of priority groups (Troyan 2012).²⁷

The game in Beijing looks “simpler” relative to the usual applications (Abdulkadiroglu and Sonmez 2003) because it only has four schools. On the other hand, the conflict of interests among parents is high because most prefer the same schools, which may encourage more “gaming”. Because the most popular school has the smallest capacity, naïve parents may have an unusual advantage in this setting: By top-ranking the most popular school, naïve parents can effectively “deter” sophisticated parents from applying to that school. Additionally, the importance of an outside option is also pronounced in Beijing, as it significantly reduces the difficulty of playing the game for some parents but not others – when a parent has only one acceptable school in the system, the strategy is to top-rank that school regardless of uncertainties.

More importantly, the above welfare evaluation does not consider the “cost of strategizing,” which can be substantial under a non-strategy-proof mechanism. The cost may include many aspects, but Chen and He (2017) capture it partially through information acquisition cost. Under the Boston mechanism, students often find it beneficial to learn others' preferences and strategies and are thus willing to acquire this costly information, whereas under DA, this incentive is absent. Results from a lab experiment show that the potential welfare gain of the Boston mechanism decreases and sometimes disappears due to this costly information acquisition. In other words, the cost of strategizing offsets the potential gain from improved match quality.

These results and concerns can be a hint of considering alternative cardinal mechanisms, e.g., the pseudo-market mechanism (Hylland and Zeckhauser 1979, He, Miralles, Pycia, and Yan 2015). On one hand, it further improves upon the efficiency of the Boston mechanism by explicitly using cardinal preferences to optimally assign students to schools. On the other hand, it has better incentive properties: students have no incentives to misreport under the pseudo-market mechanism when the market is large. The various practical issues associated with such mechanisms, however, call for future research efforts.

²⁷Troyan (2012) partially restores the dominance results by introducing a new welfare criterion.

References

- ABDULKADIROGLU, A., N. AGARWAL, AND P. A. PATHAK (2015): “The Welfare Effects of Coordinated Assignment: Evidence from the NYC HS Match,” *National Bureau of Economic Research Working Paper Series*, No. 21046.
- ABDULKADIROGLU, A., J. D. ANGRIST, Y. NARITA, AND P. A. PATHAK (2015): “Research design meets market design: Using centralized assignment for impact evaluation,” Discussion paper, National Bureau of Economic Research.
- ABDULKADIROGLU, A., Y.-K. CHE, AND Y. YASUDA (2011): “Resolving Conflicting Preferences in School Choice: the Boston Mechanism Reconsidered,” *American Economic Review*, 101(1), 399–410.
- ABDULKADIROGLU, A., P. A. PATHAK, A. E. ROTH, AND T. SONMEZ (2005): “The Boston Public School Match,” *The American Economic Review, Papers and Proceedings*, 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.
- AGARWAL, N., AND P. SOMAINI (2016): “Demand Analysis using Strategic Reports: An application to a school choice mechanism,” Discussion paper.
- ANDREWS, D. W. K., AND X. SHI (2013): “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81(2), 609–666.
- ANDREWS, D. W. K., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78(1), 119–157.
- ARADILLAS-LOPEZ, A. (2010): “Semiparametric Estimation of a Simultaneous Game with Incomplete Information,” *Journal of Econometrics*, 157(2), 409–431.
- (2012): “Pairwise Difference Estimation of Incomplete Information Games,” *Journal of Econometrics*, 168(1), 120–140.
- ARCIDIACONO, P. (2005): “Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?,” *Econometrica*, 73(5), 1477–1524.
- 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.
- BERRY, S. T. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4), 889–917.
- BHATTACHARYA, R. N. (1977): “Refinements of the Multidimensional Central Limit Theorem and Applications,” *The Annals of Probability*, 5(1), 1–27.
- BOGOMOLNAIA, A., AND H. MOULIN (2001): “A New Solution to the Random Assignment Problem,” *Journal of Economic Theory*, 100(2), 295–328.

- 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).
- BRESNAHAN, T., AND P. REISS (1991): “Entry and Competition in Concentrated Markets,” *Journal of Political Economy*, 99, 977–1009.
- BRESNAHAN, T. F., AND P. C. REISS (1990): “Entry in Monopoly Markets,” *Review of Economic Studies*, 57(4), 531–553.
- BUDISH, E., AND E. CANTILLON (2012): “Strategic Behavior in Multi-Unit Assignment Problems: Theory and Evidence from Course Allocation,” *American Economic Review*, 102(5), 2237–71.
- BUGNI, F. A., I. A. CANAY, AND X. SHI (2016): “Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models,” *Quantitative Economics*.
- CALSAMIGLIA, C., C. FU, AND M. GÜELL (2016): “Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives,” .
- CALSAMIGLIA, C., G. HAERINGER, AND F. KLIJN (2010): “Constrained School Choice: An Experimental Study,” *American Economic Review*, 100(4), 1860–74.
- CARVALHO, J. R., T. MAGNAC, AND Q. XIONG (2014): “College Choice Allocation Mechanisms: Structural Estimates and Counterfactuals,” *Mimeo*.
- CHE, Y.-K., AND F. KOJIMA (2010): “Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms,” *Econometrica*, 78(5), 1625–1672.
- CHEN, Y., AND Y. HE (2017): “Information Acquisition and Provision in School Choice: An Experimental Study,” *mimeo*.
- CHEN, Y., AND O. KESTEN (2017): “Chinese college admissions and school choice reforms: A theoretical analysis,” *Journal of Political Economy*, 125(1), 99–139.
- CHEN, Y., AND T. SONMEZ (2006): “School Choice: An Experimental Study,” *Journal of Economic Theory*, 127(1), 202–231.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75(5), 1243–1284.
- 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.
- CHIAPPORI, P.-A., AND B. SALANIÉ (2016): “The Econometrics of Matching Models,” *Journal of Economic Literature*.
- CHOO, E., AND A. SIOW (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114(1).
- CILIBERTO, F., AND E. TAMER (2009): “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, 77(6), 1791–1828.
- DE HAAN, M., P. A. GAUTIER, H. OOSTERBEEK, AND B. VAN DER KLAUW (2015): “The performance of school assignment mechanisms in practice,” .
- DUBINS, L. E., AND D. A. FREEDMAN (1981): “Machiavelli and the Gale-Shapley Algorithm,” *American Mathematical Monthly*, 88(7), 485–494.

- DUR, U., R. G. HAMMOND, AND T. MORRILL (Forthcoming): “Identifying the harm of manipulable school-choice mechanisms,” Discussion paper.
- FACK, G., J. GRENET, AND Y. HE (2015): “Estimating Preferences in School Choice Mechanisms: Theoretical Foundation and Empirical Approaches,” *Manuscript, Toulouse School of Economics*.
- FEATHERSTONE, C. R., AND M. NIEDERLE (2016): “Boston versus deferred acceptance in an interim setting: An experimental investigation,” *Games and Economic Behavior*, 100, 353–375.
- FOX, J. T. (2009): “Matching models: Empirics,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, Basingstoke.
- GALE, D. E., AND L. S. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 69(1), 9–15.
- GILBOA, I., AND D. SCHMEIDLER (1989): “Maxmin expected utility with non-unique prior,” *Journal of Mathematical Economics*, 18(2), 141 – 153.
- GREENE, W. H. (1999): *Econometric Analysis*. Prentice-Hall, 4th edition edn.
- HASTINGS, J., T. KANE, AND D. STAIGER (2008): “Heterogeneous Preferences and the Efficacy of Public School Choice,” *Mimeo*, Yale University.
- HE, Y. (2012): “Gaming the Boston School Choice Mechanism in Beijing,” *Toulouse School of Economics Working Papers*, 12-345, Toulouse School of Economics.
- HE, Y., A. MIRALLES, M. PYCIA, AND J. YAN (2015): “A Pseudo-Market Approach to Allocation Problems with Priorities,” *Mimeo*.
- 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.
- HOBY, C., AND S. TURNER (2013): “Expanding college opportunities for high-achieving, low income students,” *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014).
- HWANG, S. I. M. (2016): “A Robust Redesign of High School Match,” .
- HYLLAND, A., AND R. ZECKHAUSER (1979): “The Efficient Allocation of Individuals to Positions,” *Journal of Political Economy*, 87(2), 293–314.
- KAPOR, A., C. NEILSON, AND S. ZIMMERMAN (2016): “Heterogeneous Beliefs and School Choice,” .
- KAWAI, K., AND Y. WATANABE (2013): “Inferring Strategic Voting,” *The American Economic Review*, 103(2), 624–662.
- 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.

- MIRALLES, A. (2008): “School Choice: The Case for the Boston Mechanism,” *Mimeo*.
- MOON, H. R., AND F. SCHORFHEIDE (2009): “Estimation with overidentifying inequality moment conditions,” *Journal of Econometrics*, 153(2), 136–154.
- 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.
- SEELYE, K. Q. (2013): “Boston Schools Drop Last Remnant of Forced Busing,” *The New York Times*, March 14, 2013, <http://www.nytimes.com/2013/03/15/education/boston-schools-adopt-new-placement-plan-for-students.html>.
- SEIM, K. (2006): “An Empirical Model of Firm Entry with Endogenous Product-type Choices,” *Rand Journal of Economics*, 37(3), 619–640.
- TAMER, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70(1), 147–165.
- (2010): “Partial Identification in Econometrics,” *Annual Review of Economics*, 2, 167–195.
- TRAIN, K. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, second edition edn.
- TROYAN, P. (2012): “Comparing school choice mechanisms by interim and ex-ante welfare,” *Games and Economic Behavior*, 75(2), 936–947.
- ZHOU, L. (1990): “On A Conjecture by Gale about One-Sided Matching Problems,” *Journal of Economic Theory*, 52(1), 123–135.

Online Appendix for
Gaming the Boston School Choice Mechanism in Beijing

Yinghua He

September 16, 2017

List of Appendices

A. Proofs	A.2
B. Evaluation of the Assumptions: Evidence from Experimental Data	A.6
C. Data Cleaning and Imputation	A.12
D. Heterogeneous Information on Students' Characteristics	A.13
E. Choice Probabilities and Estimation: Additional Cases	A.16
F. Degenerate Beliefs: Moment Equalities and Inequalities	A.20
G. The Logit-Smoothed Reject-Accept Simulator and Equilibrium Solver	A.24
H. Monte Carlo Simulations	A.28
I. Goodness of Fit and Comparison across Cases	A.33
J. Additional Tables	A.35

A 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 \bar{q}_{s^*} , then the number of students unassigned is $I - \sum_s q_s + \bar{q}_{s^*}$. Since $I \leq \sum_s q_s$, $I - \sum_s q_s + \bar{q}_{s^*} < \bar{q}_{s^*}$. 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 leftover school.

(ii) Suppose C and \bar{C} have the same top 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 must be assigned to that school when submitting \bar{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 \bar{C} . This means she has the same probability to be assigned to any given school among the first K choices when she submits C or \bar{C} .

(iii) Suppose C and \bar{C} have the same first $K' - 1$ choices. School s is listed as K th choice in C , but as K' th choice in \bar{C} and $K' < K$. In any realization of the game (any lottery number), if the student is rejected by s when submitting \bar{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 \bar{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 $\hat{T} \subset T$,

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

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 $\Delta(\mathcal{C})$. The payoff to player $\{u_s\}_{s=1}^S$ is

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

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

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

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,

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

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

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

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 $j \neq i$, $\Pr(u_{j,s} > 0 > u_{j,s'}, \text{ given } s \ \& \ \forall s' \neq s) > 0$ 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, $\Pr(s \text{ is top ranked by all } j \neq i) > 0$ which implies that $A_s(C, \sigma_{-i}) < 1$ for all C such that s is top ranked.

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

Similarly, conditional on being rejected by previous choices, the probability of being accepted by s is less than one, unless s is the S th (the last) choice. Suppose that $C = (c^1, \dots, c^K, s, c^{K+2}, \dots, c^S)$, where $1 \leq K \leq (S-2)$ and $c^k \neq 0, \forall k = 1, \dots, K$. 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 $k \in \{2, \dots, K, s\}$. 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, \dots, c^K as well.

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

(ii) Note that in equilibrium i 's value function is $V_i(\sigma_i^*, \sigma_{-i}^*) = \sum_{s=1}^S A_s(\sigma_i^*, \sigma_{-i}^*) \max(u_{i,s}, u_{i,0})$. From above, $A_s(C, \sigma_{-i}^*) \in (0, 1)$ for all s and for all $C \neq (0, \dots, 0)$. If $C_i^* \neq C_i^{**} \neq (0, \dots, 0)$ are played

with positive probability in σ_i^* given $\{u_{i,s}\}_{s=1}^S$ or $(\mathbf{X}_i, \mathbf{Z}_i, \varepsilon_i)$, then $V_i(C_i^*, \sigma_{-i}^*) = V_i(C_i^{**}, \sigma_{-i}^*)$, or,

$$\sum_{s=1}^S [A_s(C_i^*, \sigma_{-i}^*) - A_s(C_i^{**}, \sigma_{-i}^*)] \max(u_{i,s}, u_{i,0}) = 0. \quad (6)$$

Since $C_i^* \neq C_i^{**}$, there is at least 2 schools such that $A_s(C_i^*, \sigma_{-i}^*) \neq A_s(C_i^{**}, \sigma_{-i}^*)$, as a result of part (iv) in Proposition 3.

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 (6) holds *ex ante* with probability zero. This proves that i plays pure 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 $C_i^* = (c_i^1, \dots, c_i^S)$ is played with positive probability in σ_i^* , and that $\exists k \in \{1, \dots, (S-2)\}$, $u_{i,c_i^k} < 0$ and $u_{i,c_i^{k+1}} > 0$. 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 the unacceptable schools or include them at the bottom.

(iv) From (ii) and (iii), i has multiple best responses in equilibrium only when she has multiple unacceptable schools. Therefore, when one fixes how everyone chooses a mixed strategy, there is a unique equilibrium strategy for everyone with probability one. ■

Proof of Proposition 3.

(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 3), 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 $C = \{c^1, \dots, c^S\}$ such that $c^K = \hat{s}$ which is worst school, $1 \leq K < S$, such that $u_{i,\hat{s}} = \min_{t=1,\dots,S} \{u_{i,t}\}$ and $\exists t \in \{K+1, \dots, S\}$ s.t., $u_{i,c^t} > u_{i,0}$. Consider an alternative list, $C' = \{c^1, \dots, c^{K-1}, \check{s}, c^{K+1}, \dots, c^S\}$, where $u_{i,\check{s}} = \max_{t=1,\dots,S} \{u_{i,t}\} > u_{i,0}$, 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, \dots, c^{K-1} when submitting C , she will be still accepted by that school when submitting C' instead. By Lemma 3, $A_{i,\hat{s}}(C, \sigma_{-i}^*) \geq A_{i,\hat{s}}(C', \sigma_{-i}^*)$, and the decrease in the probability is distributed to $\check{s}, c^{K+1}, \dots, c^S$ and \hat{s} as well. Since $u_{i,\hat{s}} = \min_{t=1,\dots,S} \{u_{i,t}\}$ and $u_{i,c^t} > u_{i,0}$, C' weakly improves i 's expected utility. Similar arguments can be made if $u_{i,\hat{s}} < u_{i,0}$.

(iv) The strict increase can be seen easily in terms of conditional probabilities. $A_{i,s}(C, \sigma_{-i}^*)$ 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 $A_{i,s}(C, \sigma_{-i}^*) \in (0, 1)$, $\forall s$ and $\forall C \neq (0, \dots, 0)$, then the above two terms are both in $(0, 1)$ unless s is

ranked as S th after all other schools.

Suppose that s is ranked as k th 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 $A_{i,s}(C', \sigma_{-i}^*) > A_{i,s}(C, \sigma_{-i}^*)$.

From Lemma 3, $A_{i,s}(C', \sigma_{-i}^*) \geq A_{i,s}(C, \sigma_{-i}^*)$, if $A_{i,s}(C', \sigma_{-i}^*) = A_{i,s}(C, \sigma_{-i}^*)$, 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 $A_{i,\hat{s}}(\hat{C}, \sigma_{-i}) = 0$ 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 $A_{i,s}(C', \sigma_{-i}^*) > A_{i,s}(C, \sigma_{-i}^*)$. 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). ■

B Evaluation of the Assumptions: Evidence from Experimental Data

This appendix uses data from the lab experiment in Calsamiglia, Haeringer, and Klijn (2010) to evaluate the key assumptions.

Specifically, I use the data of the school choice experiment in which every student can only rank up to three choices. Under each of the three mechanisms (the Boston mechanism, DA and the Top-Trading-Cycles mechanism (TTC)),²⁸ students are explicitly instructed that:

“Please write down a ranking of up to 3 schools.”

That is, students are allowed to rank fewer than 3 choices, which is in contrast to other lab experiments that require students to rank a given number of schools. Both DA and TTC are strategy-proof when students are allowed to rank all schools. In the experiment, however, students can only rank three out of the total of seven schools, which may lead to strategic behaviors.

In each session, there are 36 students competing for 36 seats across seven schools. Moreover, each student has a “district school” at which she always has the highest priority. Among all out-of-district students, schools use a single lottery to rank them. When submitting the rank-order list, students play a one-shot game of incomplete information: Everyone knows her own preferences and district status at each school but does not know others’ preferences or the lottery determining schools’ priority ranking over students.

Each student is guaranteed admission at her district school in any of the following three scenarios:

- (i) Under the Boston mechanism, the district school is top ranked.²⁹
- (ii) Under DA, the district school is ranked anywhere, and the student is not admitted by any school ranked above the district school.
- (iii) Under TTC, the district school is ranked anywhere, and the student is not admitted by any school ranked above the district school.

Therefore, in any of the scenarios, the admission probability is degenerate, and the student will not be accepted by any school that is ranked below the district school, if there is any. This leads to a set of payoff-equivalent lists for the student. In this sense, district school amounts to outside option in the extended choice list \bar{C}_i , as defined in Section 3.3.3.

The sample of students who are in any of the above three scenarios is defined as the **MIXING sample**, as they are the ones who can mix over multiple payoff-equivalent strategies.³⁰ In the following, I focus on the MIXING sample and provide evidence showing that:

- (i) The majority of students in the MIXING Sample tend to rank up to three schools, even when it does not matter for their payoffs (Section B.1);
- (ii) When they rank more schools, the choices ranked after the district school are very often ranked according to their true preference order (Section B.2);

²⁸Under TTC, each student is asked to submit a choice list, i.e., a rank-order list of schools. Given how schools rank students and the submitted lists, the final assignment is determined in the following procedure:

Round k , $k \geq 1$: Every student that has not been assigned a school in the previous round applies to the highest ranked available school in her choice list that has not rejected her yet (if there is no such school then the student “applies” to herself). Each school with vacant seats “points” to the student with highest priority among the students that have not been assigned a seat yet. Together with the above described procedure for the students, it creates a cycle or cycles of students and schools. If a student is in a cycle she is assigned a seat at the school she applies to (or to herself if she is in a self-cycle). If a school is in a cycle then its number of vacant seats is decreased by one. If a school has no longer vacant seats, then it is no longer available and students that applied to it are rejected.

TTC terminates when every student has been assigned to a school or to herself.

²⁹Under the Boston, admission to the district school is not guaranteed if it is ranked second or third even when the student is rejected by higher choices. It is because the district school may have enough students ranking it as first choice.

³⁰In the data, there are some students ranking the same school twice. I delete the lower ranked one and consider that these students do not rank up to three schools.

- (iii) The mixing probability, or the probability of ranking more schools after the district school, is not significantly correlated with student preferences (Section B.3).

B.1 Choice among Payoff-Equivalent Strategies

Focusing on the students in the MIXING sample, Table B1 shows how many schools students rank after district school. Panel A presents the distribution of students rank zero, one, or two schools after district school, conditional on how district school is ranked (first or second choice); Panel B shows the same distribution among all students in the MIXING sample. Under any of the three mechanisms, students tend to rank three choices: Only 3.41% of the students choose not to rank any school after district school and 1.14% leaves blank one place in their list, while the majority (95.45%) rank three choices. Across mechanisms, students under the Boston mechanism tend to rank fewer choices after district school, although 92.54% still rank three choices.

Table B1: Ranking Behavior with Degenerate Beliefs among Students in the MIXING Sample

	Mechanism			All 3 Mechanisms
	Boston	DA	TTC	
<i>Panel A: By the Rank of District School</i>				
District school ranked first				
No school ranked after district school (column percentage)	3 4.48	2 3.33	0 0	5 2.84
One school ranked after district school (column percentage)	2 2.99	0 0.00	0 0.00	2 1.14
Two schools ranked after district school (column percentage)	62 92.54	31 51.67	29 59.18	122 69.32
District school ranked second				
No schools ranked after district school (column percentage)	-- --	1 1.67	0 0	1 0.57
One school ranked after district school (column percentage)	-- --	26 43.33	20 40.82	46 26.14
<i>Panel B: All Students in the MIXING Sample</i>				
District school ranked first or second				
No school ranked after district school (column percentage)	3 4.48	3 5	0 0	6 3.41
≥ one school ranked after district school (column percentage)	64 95.54	57 95	49 100	170 96.59
Total	67	60	49	176

Notes: Data source: Calsamiglia, Haeringer, and Klijn (2010).

Each student has a “district school” at which the student is guaranteed admission under either of the following three scenarios:

- (i) Under the Boston mechanism, the district school is top ranked.
- (ii) Under DA, the district school is ranked anywhere, and the student is not admitted by any school above the district school.
- (iii) Under TTC, the district school is ranked anywhere, and the student is not admitted by any school above the district school.

In any of the scenarios, the admission probability is degenerate. The percentages are out of the total number of students under each mechanism.

Table B1 highlights the importance of the multiple best response and thus multiple equilibria. Moreover, it shows that empirically students tend to rank more schools, even though it does not affect their payoff. In this setting, this is probably because the cost of ranking more schools is zero.

B.2 Assumption UNACCEPTABLE

In the context of the experiment, Assumption UNACCEPTABLE translates into that students rank them truthfully among the choices after district school. To test this, I look at the second and third choices of the

students who rank district school as their first choice. I then further compare this with the second and third choices of those who do not rank district school as first choice.

Table B2: Truth-telling among the second and third choices: Conditional on the first choice

Truth-Telling	First Choice = District School ^a				First Choice \neq District School ^b			
	Mechanism			All 3 Mechanisms	Mechanism			All 3 Mechanisms
	Boston	DA	TTC		Boston	DA	TTC	
No	9	8	5	22	28	0	3	31
(column percentage)	14.52	25.81	17.24	18.03	51.85	0	25	41.33
Yes	53	23	24	100	26	9	9	44
(column percentage)	85.48	74.19	82.76	81.97	48.15	100	75	58.67
Total	62	31	29	122	54	9	12	75

Notes: This table reports whether students rank their second and third choices truthfully. The sample only includes students who have ranked three choices and whose second and third choices are not district school.

^aThis part of the analysis only focuses on students whose first choice is a district school. Therefore, it does not affect their payoffs no matter how they rank the second and third choices.

^bThis part of the analysis only focuses on students whose first choice is not district school, which implies that district school is not included in the list at all.

When district school is ranked first under any of the three mechanisms, 82% of the students rank their second and third choice according to their preferences; in contrast, only 59% do so when the first choice is not district school. Under the Boston, the comparison is 85% versus 48%.

Table B2 therefore shows that Assumption UNACCEPTABLE is plausible: When ranking schools that do not affect their payoffs, students tend to rank them truthfully.

B.3 Assumption MIXING

Assumption MIXING in the main text requires that the mixing probabilities are independent of preferences.

Table B3 investigates who rank more schools after district school. With the MIXING sample, Columns (1) and (2) report results from a linear probability model in which the binary outcome variable equals to one if the student ranks three choices with district school as the first or second choice. It shows that the probability of ranking more schools are not affected by the student's preference of her district school nor how district school is ranked. Column (2) shows the same results.

As a robustness check, Columns (3) and (4) use a probit model to take into account the discrete nature of the outcome variable, although similar results are obtained.

In the main text, the mixing probability, or the decision to rank an unacceptable school is independent of the cardinal utility associated with that school. To see if the data from the experiment is consistent with this assumption, I investigate what affects students' decision to rank the next preferred school after district school. For example, for a student in the MIXING sample whose district school is her k th preferred school ($k < 7$), the analysis is on whether she ranks her $(k + 1)$ th preferred school after her district school ($= 1$ if yes; $= 0$ otherwise). In particular, I am interested in the utility difference between the district school and the $(k + 1)$ th school.

The results from a linear probability model and a probit model are presented in Table B4, where the sample excludes those who only rank more preferred schools than the district school after the district school. Overall, there is no evidence showing that a student would rank the next preferred school more often when it is more preferable in term of cardinal preferences. In fact, columns (4) and (8) show that students do the opposite: When the next preferred school is much worse relative to the district school, the student ranks it after the district school more often.

Additional results are presented in Table B5, where I use the full MIXING sample, including those who only rank more preferred schools than the district school after the district school. Again, there is no evidence

Table B3: Regression Analysis: Who Rank Three Choice while District School is 1st or 2nd Choice

	Linear Probability		Probit Model ^a	
	(1)	(2)	(3)	(4)
District school ranked as first choice	-0.03 (0.04)	0.01 (0.04)	-0.41 (0.47)	0.14 (0.66)
<u>In true preferences, district school ranked as:</u>				
2nd	0.03 (0.07)	0.00 (0.07)	0.18 (0.58)	-0.16 (0.73)
3rd	0.05 (0.07)	0.04 (0.07)	0.47 (0.68)	0.60 (0.84)
4th	0.02 (0.07)	0.01 (0.07)	0.11 (0.63)	0.08 (0.80)
5th	0.08 (0.12)	0.04 (0.12)		
6th	0.07 (0.11)	0.06 (0.11)		
7th	0.08 (0.22)	0.19 (0.22)		
<u>Mechanism</u>				
DA		0.05 (0.04)		0.26 (0.50)
TTC		0.10** (0.04)		
<u>Other Controls</u>				
Experiment at Universitat Pompeu Fabra		0.11*** (0.03)		1.29** (0.54)
Preferences randomly generated		0.09** (0.04)		1.03* (0.54)
Constant	0.95*** (0.07)	0.80*** (0.09)	1.79** (0.70)	0.52 (1.03)
<i>N</i>	176	176	166	119
<i>R</i> ²	0.01	0.10		

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This table reports estimates from a linear probability model (columns (1) and (2)) and a probit model (columns (3) and (4)), in which the outcome variable is whether the student rank three choices (= 1 if yes; = 0 otherwise). Only students in the MIXING sample is included. That is, everyone in the sample ranks her district school as first or second choice under DA or TTC or rank it as first choice under the Boston.

^a Some observations are omitted because some variables (District school ranked 5th, 6th, or 7th, and TTC) perfectly predict the outcome variable.

^b Definitions of variables:

“DA” = 1 if the mechanism is DA; = 0 otherwise (i.e., it is either TTC or the Boston mechanism).

“TTC” = 1 if the mechanism is TTC; = 0 otherwise (i.e., it is either DA or the Boston mechanism).

“Experiment at Universitat Pompeu Fabra” = 1 if the experiment is conducted at that university; = 0 if it is at Universitat Autònoma de Barcelona.

“Preferences Randomly Generated” = 1 if the students preferences are randomly generated (i.e., independent of district-school status); = 0 if the preferences are created in the following way: the ordinal preference is determined by three factors: a schools quality, its proximity, and a random factor; based on the ordinal preferences, the payoff to be received by each participant when assigned to a given school are determined.

showing that a student would rank the next preferred school more often when it is more preferable in term of cardinal preferences.

Table B4: Ranking the Next-Preferred School after District School

	Linear Probability Model				Probit Model ^a			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Utility diff. b/t school district school & the next preferred school	0.02 (0.05)	0.02 (0.05)	0.08 (0.05)	0.11* (0.06)	0.10 (0.19)	0.11 (0.19)	0.34 (0.26)	0.90* (0.50)
District school ranked as first choice		-0.02 (0.08)	0.02 (0.08)	0.07 (0.09)		-0.08 (0.32)	0.12 (0.35)	0.41 (0.41)
<u>In true preferences, district school ranked as^b</u>								
2nd			0.17 (0.14)	0.18 (0.14)			0.75 (0.63)	1.24 (0.81)
3rd			-0.26 (0.16)	-0.24 (0.16)			-0.61 (0.65)	-0.10 (0.82)
4th			-0.24 (0.19)	-0.23 (0.20)			-0.57 (0.73)	-0.26 (0.88)
6th			0.19 (0.39)	0.23 (0.40)				
<u>Mechanism</u>								
DA				0.02 (0.10)				0.12 (0.42)
TTC				0.11 (0.09)				0.53 (0.43)
<u>Other Controls</u>								
Experiment at Universitat Pompeu Fabra				0.17** (0.08)				0.82** (0.37)
Random Preferences				0.12 (0.09)				0.85 (0.53)
Constant	0.77*** (0.09)	0.79*** (0.11)	0.56*** (0.21)	0.31 (0.25)	0.72** (0.35)	0.77* (0.41)	-0.16 (0.97)	-2.60 (1.80)
<i>N</i>	114	114	114	114	114	114	113	113
<i>R</i> ²	0.00	0.00	0.15	0.20				

Notes: This table reports estimates from a linear probability model (columns (1)-(4)) and a probit model (columns (5)-(8)), in which the outcome variable is whether the next preferred school after district school is ranked as second or third choice (= 1 if yes; = 0 otherwise). The sample includes students in the MIXING sample whose second and third choices are not their district schools, excludes those who rank more preferred schools than district school after their district school, and further deletes those whose district school is the least preferred school.

a. Some observations are omitted because some variables (District school ranked 5th and 6th) perfectly predict the outcome variable.

b. "District school ranked 5th in true preferences" is omitted due to collinearity.

Definitions of the variables:

"Utility diff. b/t school district school & the next preferred school" = "utility of district school" – "utility of the next preferred school after district school".

Other definitions are available in the notes of Table B3.

Table B5: Ranking the next preferred school after district school

	Linear Probability Model				Probit Model ^a			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Utility diff. b/t school district school & the next preferred school	-0.06 (0.03)	-0.06* (0.04)	0.01 (0.03)	0.01 (0.03)	-0.15 (0.09)	-0.16* (0.09)	0.06 (0.12)	0.06 (0.12)
District school ranked as first choice		0.05 (0.09)	0.02 (0.06)	0.03 (0.07)		0.13 (0.22)	0.12 (0.29)	0.17 (0.34)
<u>In true preferences, district school ranked as</u>								
2nd			0.07 (0.11)	0.07 (0.11)			0.33 (0.50)	0.28 (0.53)
3rd			-0.66*** (0.11)	-0.65*** (0.12)			-1.89*** (0.50)	-1.97*** (0.53)
4th			-0.72*** (0.12)	-0.72*** (0.12)			-2.22*** (0.55)	-2.30*** (0.59)
5th			-0.82*** (0.20)	-0.82*** (0.20)				
6th			-0.63*** (0.18)	-0.61*** (0.19)			-1.76** (0.78)	-1.72** (0.81)
<u>Mechanism</u>								
DA				-0.03 (0.07)				-0.18 (0.34)
TTC				0.08 (0.07)				0.38 (0.36)
<u>Other Controls</u>								
Experiment at Universitat Pompeu Fabra				0.07 (0.06)				0.32 (0.27)
Random Preferences				0.03 (0.06)				0.11 (0.29)
Constant	0.64*** (0.08)	0.61*** (0.09)	0.77*** (0.13)	0.70*** (0.16)	0.37* (0.20)	0.29 (0.24)	0.68 (0.60)	0.46 (0.72)
<i>N</i>	175	175	175	175	175	175	171	171
<i>R</i> ²	0.02	0.02	0.56	0.58				

Notes: This table reports estimates from a linear probability model (columns (1)-(4)) and a probit model (columns (5)-(8)), in which the outcome variable is whether the next preferred school after district school is ranked as second or third choice (= 1 if yes; = 0 otherwise). The sample includes students in the MIXING sample whose second and third choices are not their district schools and whose district school is not the least-preferred school.

a. Some observations are omitted because some variables (District school ranked 5th) perfectly predict the outcome variable. Definitions of the variables can be found in Tables B3 and B4.

C 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 , $Parent_Inc_i$, $Parent_Edu_i$, and $Awards_i$.

$Distance_{i,s}$ 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.

$Parent_Inc_i$ 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 $Parent_Inc_i$ are replaced by the households disposable income plus the average difference between $Parent_Inc_i$ 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 $Parent_Inc_i$.

$Parent_Edu_i$ is the average years of schools of parents. There are 49 missing values. I regress $Parent_Edu_i$ on different subsets of the variables, $Parent_Inc_i$, 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 $Parent_Edu_i$.

$Awards_i$ 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).

D Heterogeneous Information on Students' Characteristics

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

Proposition 4 Consider the following scenario:

- (i) Every parent has the same ability to process information;
- (ii) Parent i also knows the realization of $\mathcal{X}^i \equiv \{\bar{\mathbf{X}}_{i_1}, \dots, \bar{\mathbf{X}}_{i_F}\}$, where $\bar{\mathbf{X}}_i \equiv (\mathbf{X}_i, \mathbf{Z}_i)$, 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, $B_i(C) \rightarrow B(C)$, $\forall i, \forall C \in \mathcal{C}$.

Proof of Proposition 4. A student's decision is to choose one of the L possible lists. Fix the order of all the lists, and let $\mathbf{d}_i = (d_{i,1}, \dots, d_{i,L})'$, and $d_{i,l} = 1$ if the l th 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 $\mathcal{X}^1 = \{\bar{\mathbf{X}}_2, \dots, \bar{\mathbf{X}}_{F+1}\}$. Her perceived probability of other students' choices is a function of her information set $-\bar{\mathbf{X}}_1, \mathcal{X}^1$ and the distributions of $\bar{\mathbf{X}}_i$ and $\varepsilon_i, i > 1$.

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

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

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

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

$$(\pi_{i,1}, \dots, \pi_{i,L}) \equiv (\bar{\pi}_1, \dots, \bar{\pi}_L),$$

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

Consider a vector of random variables, $\mathbf{N}^I \equiv \sum_{i=2}^I \mathbf{d}_i = (N_1^I, N_2^I, \dots, N_L^I)' \in \mathbb{N}^L$, 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, \mathbf{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 \mathbf{N}^I . Two definitions are also introduced:

$$\begin{aligned} \boldsymbol{\mu}^I &\equiv \frac{1}{I-1} \left(\sum_{i=2}^{F+1} \pi_{i,1} + (I-F-1)\bar{\pi}_1, \dots, \sum_{i=2}^{F+1} \pi_{i,L} + (I-F-1)\bar{\pi}_L \right), \\ \mathbf{Q}^I &\equiv \frac{1}{I-1} \begin{bmatrix} \sum_{i=2}^{F+1} \pi_{i,1}(1-\pi_{i,1}) & \dots & \sum_{i=2}^{F+1} \pi_{i,1}\pi_{i,L} \\ + (I-F-1)\bar{\pi}_1(1-\bar{\pi}_1) & \dots & + (I-F-1)\bar{\pi}_1\bar{\pi}_L \\ \sum_{i=2}^{F+1} \pi_{i,1}\pi_{i,2} & \dots & \sum_{i=2}^{F+1} \pi_{i,2}\pi_{i,L} \\ + (I-F-1)\bar{\pi}_1\bar{\pi}_2 & \dots & + (I-F-1)\bar{\pi}_2\bar{\pi}_L \\ \dots & \dots & \dots \\ \sum_{i=2}^{F+1} \pi_{i,1}\pi_{i,L} & \dots & \sum_{i=2}^{F+1} \pi_{i,L}(1-\pi_{i,L}) \\ + (I-F-1)\bar{\pi}_1\bar{\pi}_L & \dots & + (I-F-1)\bar{\pi}_L(1-\bar{\pi}_L) \end{bmatrix}, \end{aligned}$$

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

$$\lim_{I \rightarrow \infty} \boldsymbol{\mu}^I = (\bar{\pi}_1, \dots, \bar{\pi}_L) \equiv \boldsymbol{\mu},$$

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

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:

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

Therefore,

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

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

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

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

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

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

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

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

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

where $\text{Ball} \left(\sqrt{I-1} \left(\frac{\mathbf{n}^I}{I-1} - \boldsymbol{\mu} \right), \frac{1}{2\sqrt{I-1}} \right)$ 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 $N(0, Q)$. 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, $\forall s$,

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

which can be re-written as:

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

where $\bar{a}_s(C, \mathbf{n}^I)$ 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 $I \rightarrow \infty$,

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

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

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

■

Corollary *Under the same conditions as in Proposition 4, and that $\bar{\mathbf{X}}_i$ is now common knowledge, the beliefs converge to a common belief, $B_i(C) \rightarrow B(C), \forall i, \forall C \in \mathcal{C}$.*

In this corollary, the difference between any two students, i and j , is their information about their opponents, the realizations of $\bar{\mathbf{X}}_{-i}$ and $\bar{\mathbf{X}}_{-j}$. However, the difference between $\bar{\mathbf{X}}_{-i}$ and $\bar{\mathbf{X}}_{-j}$ is very limited, since $\bar{\mathbf{X}}_{-i} = (\bar{\mathbf{X}}_{-i,j}, \bar{\mathbf{X}}_j)$ and $\bar{\mathbf{X}}_{-j} = (\bar{\mathbf{X}}_{-i,j}, \bar{\mathbf{X}}_i)$ where $(-i, j)$ denotes the students other than i and j . By the same argument in Proposition 4, the beliefs converge.

E Choice Probabilities and Estimation: Additional Cases

As outlined in Figure 1, this appendix presents the characterization of choice probabilities under the Bayesian Nash equilibrium and then the truth-telling assumption and discuss the estimation approach. Besides, I also discuss the case with heterogeneous but non-degenerate beliefs.

E.1 Bayesian Nash Equilibrium

To characterize choice probabilities of each list, assume first the equilibrium beliefs B^* are known.

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

- (i) if $C = (0, 0, 0, 0)$, $\Pr(u_{i,s} < u_{i,0}, \text{ for all } s | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$;
- (ii) if $C = (c^1, 0, 0, 0)$, $m^{1,1}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1 | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$;
- (iii) if $C = (c^1, c^2, 0, 0)$,
 $m^{2,2}(\mathbf{X}_i) * \Pr(C_i \text{ is a best response; } u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$
 $+ m^{1,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1 \neq c^2 | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$;
- (iv) if $C = (c^1, c^2, c^3, c^4)$,
 $\Pr(C \text{ is a best response; } u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0} | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$
 $+ m^{1,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$
 $+ m^{2,4}(\mathbf{X}_i) * \Pr(C \text{ is a best response; } u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i, B^*; \theta)$.

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 $m^{1,1}(\mathbf{X}_i)$ times the probability that only one school is acceptable, because parents may submit a two-school or a full list ($m^{1,2}(\mathbf{X}_i), m^{1,4}(\mathbf{X}_i) \geq 0$). Note that Assumption MIXING is used here. Namely $m_i^{K,l} \equiv m^{K,l}(\mathbf{X}_i) + \eta_i^{K,l}$, $E(\eta_i^{K,l}) = 0$, and $\eta_i^{K,l} \perp (\varepsilon_i, \mathbf{X}_i, \mathbf{Z}_i, B)$.

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,³¹ and thus the probability of optimally ranking two acceptable schools is weighted by the mixing probability, $m^{2,2}(\mathbf{X}_i)$. 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.

Parts (iii) and (iv) require the probability of the given list being a best response to be calculated. Since the payoff from a given list involves a weighted sum of five i.i.d. type I extreme values, the choice probability does not follow a logistic form any more. They are therefore simulated by the smoothed logit-smoothed accept-reject simulator (Chapter 5, Train (2009)) which is described in Appendix G.

Estimation Since the equilibrium beliefs, B^* , are unknown, one possibility is to use the with empirical beliefs, \hat{B} , which is bootstrapped from the observed actions, as an approximation. The model is then estimated by the method of simulated maximum likelihood.

It is also tempting to consider an alternative approach, a method of simulated maximum likelihood with equilibrium constraints as follows:

$$\max_{\theta, B} L(\theta), \text{ s.t., } \bar{B}(\cdot, \sigma_{-i} | B; \theta) = B, \quad (10)$$

³¹It is also possible to submit a three-school list, but it is equivalent to submitting a four-school/full list.

where $L(\theta)$ is the likelihood function, and the constraint restricts the beliefs to be a fixed point in equilibrium. \bar{B} is defined as the “beliefs” implied by B , i.e.,

$$\bar{B}(C, \sigma_{-i}|B; \theta) = [A_1(C, \sigma_{-i}|B; \theta), \dots, A_S(C, \sigma_{-i}|B; \theta)],$$

and

$$A_s(C, \sigma_{-i}|B; \theta) = \sum_{n=1}^{L^{(I-1)}} \Pr(C_{-i}^n \text{ played under } \sigma_{-i} \text{ given } B \text{ and } \theta) a_s(C, C_{-i}^n).$$

\bar{B} is also formulated by simulation which is described in Appendix G.

The main issue with this approach is the computational burden. For any given θ , one has to solve for the equilibrium, which may not be unique. Due to this computational difficulty, I do not consider this approach in the paper. Instead, the fixed point, $\bar{B}(\cdot, \sigma_{-i}|B; \theta) = B$, is only used to solve the Bayesian Nash equilibrium in counterfactual analyses.

Monte Carlo Results The simulated maximum likelihood estimation is tested in Monte Carlo. Appendix H shows that the approach with empirical beliefs \hat{B} does not work well, although it works fine with the true equilibrium beliefs. As discussed in the main text, the case with non-degenerate beliefs is an alternative way to estimate the model under the Bayesian Nash equilibrium, and the simulation results show that it even dominates the approach with the true equilibrium beliefs.

E.2 Everyone Is Truth-Telling

Suppose that everyone reports her true preference ranking and also follows Assumptions UNACCEPTABLES and MIXING when there are unacceptable schools.

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

- (i) if $C_i = (0, 0, 0, 0)$, $\Pr(u_{i,s} < u_{i,0}, \text{ for all } s|\mathbf{X}_i, \mathbf{Z}_i; \theta)$;
- (ii) if $C_i = (c^1, 0, 0, 0)$, $m^{1,1}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1|\mathbf{X}_i, \mathbf{Z}_i; \theta)$;
- (iii) if $C_i = (c^1, c^2, 0, 0)$,
 $m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,c^2} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1 \neq c^2|\mathbf{X}_i, \mathbf{Z}_i; \theta)$
 $+ m^{1,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1 \neq c^2|\mathbf{X}_i, \mathbf{Z}_i; \theta)$
- (iv) if $C_i = (c^1, c^2, c^3, c^4)$,
 $\Pr(u_{i,c^1} > u_{i,c^2} > u_{i,c^3} > u_{i,0}, u_{i,c^4} = \min(u_{i,s})|\mathbf{X}_i, \mathbf{Z}_i; \theta)$
 $+ m^{2,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4}|\mathbf{X}_i, \mathbf{Z}_i; \theta)$
 $+ m^{1,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4}|\mathbf{X}_i, \mathbf{Z}_i; \theta)$

Estimation Given the assumptions on the mixing probabilities and the utility shocks, the choice probabilities can be fully parameterized as functions of unknown parameters θ . Therefore the maximum likelihood estimation can be applied.

E.3 General Case: Heterogeneous Sophistication

I now relax the sophistication assumption and allow parents to make mistakes when forming their beliefs. This appendix supplements section 3.4 by considering non-degenerate beliefs.

The key result is Proposition 3 on the dominated strategies. With non-degenerate beliefs, the dominance becomes strict (part iv of Proposition 3), which is in contrast to degenerate beliefs. Moreover, as shown in Proposition 1, in homogenous-sophistication Bayesian Nash equilibrium (σ^*), beliefs are non-degenerate, and strict dominance can also be obtained in σ^* .

E.3.1 Estimation with Non-Degenerate Beliefs

In the following, let us assume that beliefs are non-degenerate and that parents are rational in the sense that they do not play dominated strategies. Notice that under these assumptions, the model cannot predict a unique distribution of rank-order lists given preferences and thus is incomplete (Tamer (2003)). However, this will no longer be the case after grouping certain lists together.

Given $S = 4$, I assign the lists into 15 groups, g_n , $n = 1, \dots, 15$. 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 of two-school lists; and (c) 4 groups of full lists. They are shown in Table E6, and the groups are both mutually exclusive and exhaustive.

Table E6: The 15 Groups of Lists When Beliefs are Non-Degenerate

Groups of non-participation or one-school lists	Groups of two-school lists	Groups of four-school lists
Group 1: $\{(0, 0, 0, 0)\}$	Group 6: $\left\{ \begin{array}{l} (1, 2, 0, 0) \\ (2, 1, 0, 0) \end{array} \right\}$	Group 12: $\left\{ \begin{array}{l} 6 \text{ full lists} \\ \text{s.t. } (c^1, c^2, c^3, 1) \end{array} \right\}$
Group 2: $\{(1, 0, 0, 0)\}$	Group 7: $\left\{ \begin{array}{l} (1, 3, 0, 0) \\ (3, 1, 0, 0) \end{array} \right\}$	Group 13: $\left\{ \begin{array}{l} 6 \text{ full lists} \\ \text{s.t. } (c^1, c^2, c^3, 2) \end{array} \right\}$
Group 3: $\{(2, 0, 0, 0)\}$	Group 8: $\left\{ \begin{array}{l} (1, 4, 0, 0) \\ (4, 1, 0, 0) \end{array} \right\}$	Group 14: $\left\{ \begin{array}{l} 6 \text{ full lists} \\ \text{s.t. } (c^1, c^2, c^3, 3) \end{array} \right\}$
Group 4: $\{(3, 0, 0, 0)\}$	Group 9: $\left\{ \begin{array}{l} (1, 4, 0, 0) \\ (4, 1, 0, 0) \end{array} \right\}$	Group 15: $\left\{ \begin{array}{l} 6 \text{ full lists} \\ \text{s.t. } (c^1, c^2, c^3, 4) \end{array} \right\}$
Group 5: $\{(4, 0, 0, 0)\}$	Group 10: $\left\{ \begin{array}{l} (2, 3, 0, 0) \\ (3, 2, 0, 0) \end{array} \right\}$	
	Group 11: $\left\{ \begin{array}{l} (2, 4, 0, 0) \\ (4, 2, 0, 0) \end{array} \right\}$	

For type-(a) groups, since they are independent of beliefs, the choice probabilities are similar to traditional discrete choice models, although the mixing probabilities have to be considered. 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 them is acceptable, and the other is unacceptable but better than the two excluded schools. 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 distinct. The choice probabilities for these groups come from three sources: (i) the top three schools are all acceptable; (ii) only two of the top three are acceptable; and (iii) only one of the top three is acceptable. Again, the contributions of last two sources are weighted by the mixing probabilities.

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 $m^{k,l}(\mathbf{X}_i)$, $l \geq k$, is the (expected) probability that an l -school list is submitted while only k schools are acceptable, conditional on \mathbf{X}_i .

Characterization of Choice Probabilities When there are four schools, the conditional probability of i choosing some list C_i in group g_n , $\Pr(C_i \in g_n | \mathbf{X}_i, \mathbf{Z}_i; \theta)$, is

- (i) if $C_i \in g_1 = \{(0, 0, 0, 0)\}$, $\Pr(u_{i,s} < u_{i,0}, \text{ for all } s | \mathbf{X}_i, \mathbf{Z}_i; \theta)$;
- (ii) if $C_i \in g_n = \{(c^1, 0, 0, 0)\}$ ($n = 2, \dots, 5$ given the identity of c^1),

$$\begin{aligned}
 & m^{1,1}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\
 \text{(iii) if } C_i \in g_n & = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\} \text{ (} n = 6, \dots, 11 \text{ given the identities of } c^1 \text{ and } c^2\text{),} \\
 & m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\
 & + m^{1,2}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,s}, \text{ for } s \neq c^1, c^2 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right] \\
 \text{(iv) if } C_i \in g_n & = \{\text{full lists s.t. the 4th is always } c^4\}, \text{ (} n = 12, \dots, 15 \text{ given the identity of } c^4\text{),} \\
 & \Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\
 & + m^{1,4}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > \max(u_{i,c^2}, u_{i,c^3}) > \min(u_{i,c^2}, u_{i,c^3}) > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2} > u_{i,0} > \max(u_{i,c^1}, u_{i,c^3}) > \min(u_{i,c^1}, u_{i,c^3}) > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^3} > u_{i,0} > \max(u_{i,c^1}, u_{i,c^2}) > \min(u_{i,c^1}, u_{i,c^2}) > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right] \\
 & + m^{2,4}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^1}, u_{i,c^3} > u_{i,0} > u_{i,c^2} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2}, u_{i,c^3} > u_{i,0} > u_{i,c^1} > u_{i,c^4} | \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right]
 \end{aligned}$$

With the functional-form assumptions on the utility shocks and on the mixing probabilities (to be specified in Section 5), all the above 15 choice probabilities can be re-written as functions of the observables $(\mathbf{X}_i, \mathbf{Z}_i)$ and unknown parameters θ . Three points should be highlighted here: (i) none of the choice probabilities depends on beliefs; (ii) after grouping, the model is complete, as it implies a unique distribution of groups given a distribution of preferences; and (iii) identification amounts to that the model implies a unique distribution of preferences given a distribution of groups.

Identification and Estimation After the grouping, the model is essentially a discrete choice with 15 options. With the choice probabilities, the likelihood function can be written as:

$$L_{NDB}(\theta) \equiv \sum_{i=1}^I \sum_{n=1}^{15} [1_{(C_i \in g_n)} \log [\Pr(C_i \in g_n | \mathbf{X}_i, \mathbf{Z}_i; \theta)]] ,$$

where the subscript *NDB* stands for Non-Degenerate Beliefs. I denote the estimator as $\hat{\theta}_{NDB}$.

Intuitively, identification is similar to discrete choice models with five options. It requires that $\Pr(\cdot | \cdot, \cdot; \theta) \neq \Pr(\cdot | \cdot, \cdot; \theta_0)$ if θ is not equal to the true value θ_0 . In the current context, this is essentially the same as the identification assumption for a conditional logit model even with the mixing probabilities, while parameters in the utility function are already normalized. It can be best illustrated by the full-list groups (without considering the mixing probability for the moment): Observing a group being chosen, researchers learn which school is the least preferred by the student, whereas in a conditional logit model the most preferred is revealed. In either case, the choice probability can be readily specified. Distinct from traditional discrete models are the mixing probabilities, which may present a complication. The probability of choosing a full list also has to take into the ranking of the outside option. That is, one can submit a full list when she has one, two, or three acceptable schools. Moreover, the empirical probability of choose a one-school list identifies the probability that the listed school being the only acceptable one (weighted by the mixing probability), and the two-school lists play a similar role. To separately identify preferences and mixing probabilities, an “exclusion” condition is crucial: The choice probability of the group/list $(0, 0, 0, 0)$ is independent of the mixing probabilities.

As previously discussed, this modeling and estimation approach can be applied to data that are generated either from a Bayesian Nash equilibrium or an equilibrium with heterogeneous but non-degenerate beliefs. In the case of Bayesian Nash equilibrium, Appendix H presents Monte Carlo results showing that this approach performs even better than a maximum likelihood approach where the true equilibrium beliefs are observed.

F Degenerate Beliefs: Moment Equalities and Inequalities

Now heterogeneous sophistication is allowed, and some of the elements in beliefs may be zero for some parents. Putting together Proposition 3, Assumptions UNACCEPTABLES, MIXING, and ZERO-PROB, I derive bounds of the choice probabilities of individual lists and of groups of lists.

With the 41 individual lists, choice probability bounds lead to 5 moment equalities and 72 ($=36*2$) moment inequalities. Grouping lists together results in more probability bounds and therefore more moment conditions: 3 moment equalities and 38 ($=19*2$) moment inequalities. The characterization of the upper and lower bounds is detailed in the following three tables, Tables F7, F8, and F9.

Table F7: Bounds of Choice Probabilities: Individual Lists

Lists	# Moment (In)equ.	Lower or Upper Bounds
Partial lists without grouping		
(0,0,0,0)	1	$\Pr(u_{i,s} < u_{i,0}, \text{ for all } s \mathbf{X}_i, \mathbf{Z}_i; \text{mathbf}\theta)$
$(c^1, 0, 0, 0)$	4	Lower=Upper $m^{1,1}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,s}, s \neq c^1 \mathbf{X}_i, \mathbf{Z}_i; \theta)$
$(c^1, c^2, 0, 0)$	6×2	Lower $m^{1,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta)$
s.t. c^1 or $c^2 = 1$		Upper $\text{lower bound} + m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta)$
$(c^1, c^2, 0, 0)$	6×2	Lower $m^{1,2}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta)$
s.t. $c^1, c^2 \neq 1$		Upper $\text{lower bound} + m^{2,2}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta)$ $+ \Pr(u_{i,c^1}, u_{i,c^2}, u_{i,1} > u_{i,0} \ \& \ u_{i,c^1}, u_{i,c^2}, u_{i,1} > \min_s(u_{i,s}) \mathbf{X}_i, \mathbf{Z}_i; \theta)$
Full lists without grouping		
(c^1, c^2, c^3, c^4)	18×2	Lower $m^{1,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta)$
s.t. $c^4 \neq 1$		Upper $\text{lower bound} + m^{2,4}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta)$ $+ \Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 \mathbf{X}_i, \mathbf{Z}_i; \theta)$
(c^1, c^2, c^3, c^4)	6×2	Lower $m^{1,4}(\mathbf{X}_i) * \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta)$
$c^4 = 1$		Upper $\text{lower bound} + m^{2,4}(\mathbf{X}_i) * \Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta)$ $+ \Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 \mathbf{X}_i, \mathbf{Z}_i; \theta)$ $+ \Pr(u_{i,c^3} = \min\{u_{i,s}\}_{s=1}^4 > u_{i,0} \mathbf{X}_i, \mathbf{Z}_i; \theta)$

Table F8: Bounds of Choice Probabilities: Groups of Lists, Part I

Group of Lists	# Moment (In)equ.	Lower or Upper Bounds
Partial Lists with grouping		
$(c^1, c^2, 0, 0)$ & $(c^2, c^1, 0, 0)$ c^1 or $c^2 = 1$	3	Lower=Upper $m^{1,2}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right]$
	3×2	Lower $m^{1,2}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ + \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right]$
$(c^1, c^2, 0, 0)$ & $(c^2, c^1, 0, 0)$ $c^1, c^2 \neq 1$		Upper lower bound + $\Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,s}, s \neq c^1, c^2 \mathbf{X}_i, \mathbf{Z}_i; \theta)$ lower bound + $\Pr(u_{i,c^1}, u_{i,c^2} > u_{i,1} > u_{i,0} \text{ \& } u_{i,c^1}, u_{i,c^2}, u_{i,1} > \min_s(u_{i,s}) \mathbf{X}_i, \mathbf{Z}_i; \theta)$
	Full lists with grouping	
(c^1, c^2, c^3, c^4) & (c^2, c^1, c^3, c^4) s.t. $c^4 \neq 1$	9×2	Lower $m^{1,4}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right]$
		Upper lower bound + $\Pr(u_{i,c^1}, u_{i,c^2} > u_{i,0} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta)$ lower bound + $\Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min_{s=1}^4 \{u_{i,s}\} \mathbf{X}_i, \mathbf{Z}_i; \theta)$
(c^1, c^2, c^3, c^4) & (c^2, c^1, c^3, c^4) s.t. $c^4 = 1$	3×2	Lower $m^{1,4}(\mathbf{X}_i) * \left[\begin{array}{l} \Pr(u_{i,c^1} > u_{i,0} > u_{i,c^2} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta) \\ \Pr(u_{i,c^2} > u_{i,0} > u_{i,c^1} > u_{i,c^3} > u_{i,c^4} \mathbf{X}_i, \mathbf{Z}_i; \theta) \end{array} \right]$
		Upper lower bound + $\Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > u_{i,0}; u_{i,c^4} = \min_{s=1}^4 \{u_{i,s}\} \mathbf{X}_i, \mathbf{Z}_i; \theta)$ + $\Pr(u_{i,c^3} = \min_{s=1}^4 \{u_{i,s}\} > u_{i,0} \mathbf{X}_i, \mathbf{Z}_i; \theta)$

Table F9: Bounds of Choice Probabilities: Groups of Lists, Part II

Group of Lists	# Moment (In)equ.	Lower or Upper Bounds
Full Lists with grouping		
Full lists s.t. the 4th is c^4 & $c^4 \neq 1$	3 × 2	Lower
		Upper
Full lists s.t. the 4th is School 1	1 × 2	Lower
		Upper

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

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

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

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

G The Logit-Smoothed Reject-Accept Simulator and Equilibrium Solver

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, similar to the steps in Chapter 5 of Train (2009).

- (i) Calculate the choice probabilities for $(0, 0, 0, 0)$ and one-school lists, which are independent of beliefs. For example, the choice probability for $(c, 0, 0, 0)$, $c \neq 0$, is:

$$m^{1,1}(\mathbf{X}_i) \frac{\exp(v_{i,c})}{1 + \sum_{s=1}^4 \exp(v_{i,s})} \frac{1}{1 + \sum_{s \neq c} \exp(v_{i,s})},$$

where $v_{i,s}$ is the deterministic part of the utility of i at school s and $m^{1,1}(\mathbf{X}_i)$ is the average probability of submitting a one-school list when i has one school acceptable.

- (ii) Draw a value of the 5-dimensional vector of errors, η^r , which is i.i.d. uniform $U(0, 1)$, and calculate $\varepsilon_i^r = -\ln(-\ln(\eta^r))$.
- (iii) Calculate $\hat{u}_{i,s}^r = v_{i,s} + \varepsilon_{i,s}^r - \varepsilon_{i,0}^r = u_{i,s}^r - u_{i,0}^r$ given $(\mathbf{X}_i, \mathbf{Z}_i)$ and parameters.
- (iv) Given the beliefs and thus the admission probabilities, calculate the expected utility of submitting a two-school or full list, $V_{i,(c^1,c^2,0,0)}^r$ and $V_{i,(c^1,c^2,c^3,c^4)}^r$ for all c^1, c^2, c^3 and c^4 .
- (v) To calculate the choice probabilities of full or two-school lists, for example, full list (c^1, c^2, c^3, c^4) , the probability of choosing (c^1, c^2, c^3, c^4) when only c^1 is acceptable is defined as follows (independent of the simulated errors):

$$P_{i,1}^r = m^{1,4}(\mathbf{X}_i) \frac{\exp(v_{i,c^1})}{1 + \sum_{s=1}^4 \exp(v_{i,s})} \frac{1}{1 + \sum_{s \neq c^1} \exp(v_{i,s})} \frac{\exp(v_{i,c^2})}{\sum_{s \neq c^1} \exp(v_{i,s})} \frac{\exp(v_{i,c^3})}{\sum_{s \neq c^1, c^2} \exp(v_{i,s})};$$

the probability of choosing it when only c^1 and c^2 are acceptable:

$$P_{i,2}^r = m^{2,4}(\mathbf{X}_i) \left[\frac{\frac{1}{1 + \exp(-\hat{u}_{i,c^1}^r/\lambda)} \frac{1}{1 + \exp(-\hat{u}_{i,c^2}^r/\lambda)} \frac{1}{1 + \exp(\hat{u}_{i,c^3}^r/\lambda)} \frac{1}{1 + \exp(-\hat{u}_{i,c^4}^r/\lambda)}}{1 + \exp[(\hat{u}_{i,c^4}^r - \hat{u}_{i,c^3}^r)/\lambda]} \frac{1}{1 + \exp[(V_{i,(c^2,c^1,0,0)}^r - V_{i,(c^1,c^2,0,0)}^r)/\lambda]} \right];$$

and the probability when at least c^1, c^2 , and c^3 are acceptable:

$$P_{i,3}^r = \left[\frac{\frac{1}{1 + \exp(-\hat{u}_{i,c^1}^r/\lambda)} \frac{1}{1 + \exp(-\hat{u}_{i,c^2}^r/\lambda)} \frac{1}{1 + \exp(-\hat{u}_{i,c^3}^r/\lambda)}}{1 + \exp[(\hat{u}_{i,c^4}^r - \hat{u}_{i,c^1}^r)/\lambda]} \frac{1}{1 + \exp[(\hat{u}_{i,c^4}^r - \hat{u}_{i,c^2}^r)/\lambda]} \frac{1}{1 + \exp[(\hat{u}_{i,c^4}^r - \hat{u}_{i,c^3}^r)/\lambda]} \right] \frac{1}{\prod_{\forall (c^{1'}, c^{2'}, c^{3'}, c^4) \neq (c^1, c^2, c^3, c^4)} \frac{1}{1 + \exp[(V_{i,(c^{1'}, c^{2'}, c^{3'}, c^4)}^r - V_{i,(c^1, c^2, c^3, c^4)}^r)/\lambda]}},$$

where the last term is the probability that (c^1, c^2, c^3, c^4) delivers the highest expected utility among those ranking c^4 at the bottom. In all expressions, $\lambda > 0$ is a scale factor and I experiment it with $\lambda = 0.01, 0.005, 0.001$. Results presented in the paper are from $\lambda = 0.001$. Then for this simulation, r , the choice probability of choosing (c^1, c^2, c^3, c^4) is:

$$P_i^r = P_{i,1}^r + P_{i,2}^r + P_{i,3}^r.$$

- (vi) Repeat steps 2 to 5 for $r = 1, \dots, 300$, and the simulated probability of the corresponding event is then $\widetilde{\text{Pr}}_i = \frac{1}{300} \sum_{r=1}^{300} P_i^r$.

It is easy to verify that $\widetilde{\text{Pr}}$ is strictly positive and twice differentiable with respect to beliefs and parameters in utility function.

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

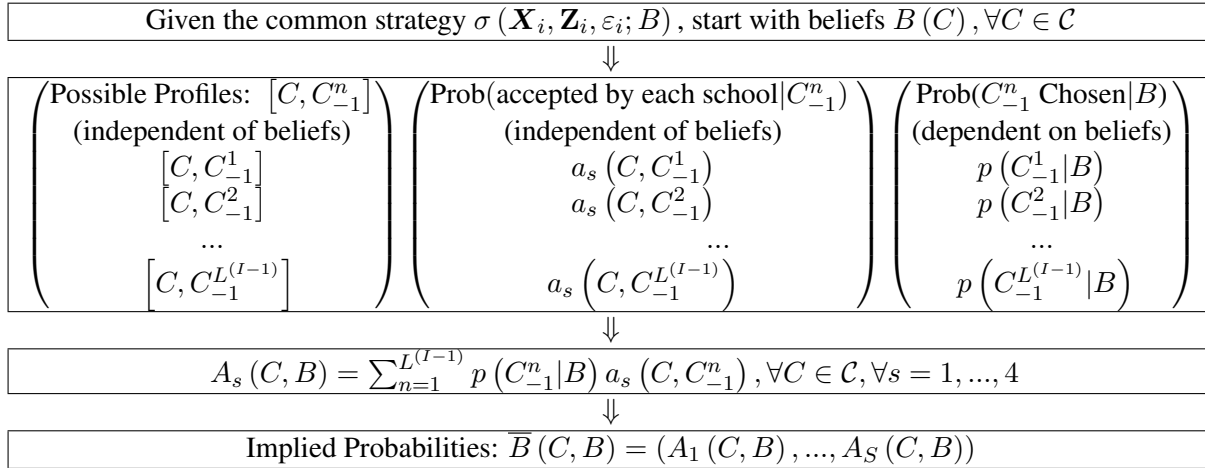


Figure H1. Mapping from Beliefs to the Implied Probabilities for Student 1

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:

- (i) Draw NC ($\geq 10,000$) profiles of choice lists, $\{C_i = (c_i^1, \dots, c_i^S)\}_{i=2}^I$. Given each profile, $\{C_i\}_{i=2}^I$, student 1 tries all ($S! = 24$) full choice lists. Combine them together, I create $S! \times NC$ profiles of $\{C_i\}_{i=1}^I$. For the profiles, there are various ways to draw them to make the belief non-degenerate, for example,
- a quarter of them are random draws from students' true preference orders.
 - another quarter are random draws from the 24 possible full lists;
 - the third quarter are such that at least q_s choosing a list ranking s ($= 1, 2, 3, 4$) as top choice, and others randomly choose a possible list;
 - the last quarter of profiles are such that at least q_s choosing a full list ranking s ($= 1, 2, 3, 4$) as top choice, at least $q_{s'}$ choose full lists ranking $s' \neq s$ as first choice, and others randomly choose a possible list.

After solving the equilibrium beliefs, one can draw another set of profiles under the assumption that everyone chooses a best response against the solved "equilibrium." This two-stage approach is adopted in the Monte Carlo simulation. All profiles from the two steps are pooled together to solve the equilibrium.

- (ii) Given each profile of lists, $\{C_1, \{C_i\}_{i=2}^I\}$, create a set of random lottery numbers, $r_s = 1$, and then run the admission process to see which school admits student 1, i.e., get the values for the following

indicator functions:

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

- (iii) Repeat Step 2 with different lottery number draws, $rs = 2, \dots, 1000$, and calculate the probabilities of Student 1 being admitted by every s respectively.

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

- (iv) Repeat Steps 2 and 3 for all $S! = 24$ 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, and thus it is run only simulated once.

- (v) Simulate the choice probability of each list by the logit-smoothed accept-reject simulator. First, simulate $r = 1, \dots, 300$ draws of $\{\eta_i^r\}_{i=2}^I$. Given the candidate belief, B , the simulated probability of student i choosing a list C_k is:

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

where $P_i^r (C_i | \mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}; B)$ is probability of C_i being choosing as a best response given B that is from the logit-smoothed accept-reject simulator described above.

- (vi) Calculate the average choice probability for the L (41) choice lists:

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

- (vii) 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 $\left\{ C_i^{(t)} \right\}_{i=2}^I = (C_2^{(t)}, C_3^{(t)}, \dots, C_I^{(t)})$, then

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

where K is a normalization term,

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

(viii) Calculate the implied probability of Student 1 being admitted by school s as follows, $\forall s = 1, \dots, S$:

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

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

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

Note that the above Steps (vi) and (vii) can be replaced by the following one step:

(vi') Calculate the probability of the profiles $\left\{ \{C_i^{(t)}\}_{i=2}^I \right\}_{t=1}^{NC}$ simulated in Step 1 being realized, i.e., if

$$\{C_i^{(t)}\}_{i=2}^I = (C_2^{(t)}, C_3^{(t)}, \dots, C_I^{(t)}), \text{ then}$$

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

where K is a normalization term,

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

The issue with this step in practice is that many $\widetilde{\Pr} \left(\{C_i^{(t)}\}_{i=2}^I \text{ realized} \right)$ are very likely to be zero or very close to it, since each individual might have a very low probability choosing a given list. Replacing Step (vi') with Steps (vi) and (vii) solves this problem. To be more precise, Step (vi') assumes observed characteristics $(\mathbf{X}_i, \mathbf{Z}_i)$ are common knowledge, while Steps (vi) and (vii) assumes they are private information while the joint distribution is common knowledge. These two approaches converge when the number of students grows (by using similar arguments as in Appendix D).

H Monte Carlo Simulations

This section presents Monte Carlo results that show how the estimation approaches perform under various data generating processes.

H.1 Model

In the following Monte Carlo exercises, students have access to four schools and play the game under the Boston mechanism as specified in the main text. Student i 's utility function at school $s \in \{1, 2, 3, 4\}$ is simplified as:

$$u_{i,s} = \alpha + \alpha_s + \beta \text{Distance}_{i,s} + \varepsilon_{i,s},$$

where α_s is the middle school fixed effect in addition to the common effect α ; and $\text{Distance}_{i,s}$ is the distance from i 's home to school s (in natural logarithm). The utility of the outside option is normalized: $u_{i,0} = \varepsilon_{i,0}$. ($\varepsilon_{i,0}, \varepsilon_{i,1}, \dots, \varepsilon_{i,4}$) are i.i.d. type I extreme value across students.

Recall that Assumption MIXING specifies i 's mixing probabilities as $m_i^{k,l} \equiv m^{k,l}(\mathbf{X}_i) + \eta_i^{k,l}$, where $m_i^{k,l}$ denotes the probability that an l -school list is submitted by i while only k schools are acceptable, $l \in \{k, k+1, \dots, (S-2), S\}$. Besides, $E(\eta_i^{k,l}) = 0$, $\eta_i^{k,l} \perp (\varepsilon_i, \mathbf{X}_i, \mathbf{Z}_i, B)$, for all k, l ; and $\sum_{l=k}^{S-2} m_i^{k,l} + m_i^{k,S} = 1$. The conditional expectations of mixing probabilities, $m^{1,2}(\mathbf{X}_i)$, $m^{1,4}(\mathbf{X}_i)$, and $m^{2,4}(\mathbf{X}_i)$ are assumed to have the following functional form:

$$\begin{aligned} m^{1,2}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{1,2} + \gamma_2^{1,2} \text{Parent_Inc}_i \right] \right), \\ m^{1,4}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{1,4} + \gamma_2^{1,4} \text{Parent_Inc}_i \right] \right) (1 - m^{1,2}(\mathbf{X}_i)), \\ m^{2,4}(\mathbf{X}_i) &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan \left[\gamma_1^{2,4} + \gamma_2^{2,4} \text{Parent_Inc}_i \right] \right). \end{aligned}$$

where Parent_Inc_i is annual income of i 's parents (in natural logarithm); \arctan is the inverse trigonometric function arctangent, which guarantees that all the above terms lie in $(0, 1)$.

In order to simplify notations, in the main text I focus on a special case where $\eta_i^{k,l} = 0$, for all k and l and therefore $m_i^{K,l} = m^{K,l}(\mathbf{X}_i)$ for all i . In the simulation, I re-introduce $\eta_i^{k,l} \neq 0$. There are several ways to introduce $\eta_i^{k,l}$, and experiments show that the results are not sensitive to the distribution of $\eta_i^{k,l}$. One example is as follows:

$$\begin{aligned} \eta_i^{1,1} &= \min \left(\hat{\eta}_i^{1,1}, \underline{m}^{1,1}, \frac{\underline{m}^{1,4}}{2} \right), \text{ where } \hat{\eta}_i^{1,1} \sim N \left(0, \left(\frac{1}{2} \min \left(\underline{m}^{1,1}, \frac{\underline{m}^{1,4}}{2} \right) \right)^2 \right); \\ \eta_i^{1,2} &= \min \left(\hat{\eta}_i^{1,2}, \underline{m}^{1,2}, \frac{\underline{m}^{1,4}}{2} \right), \text{ where } \hat{\eta}_i^{1,2} \sim N \left(0, \left(\frac{1}{2} \min \left(\underline{m}^{1,2}, \frac{\underline{m}^{1,4}}{2} \right) \right)^2 \right); \\ \eta_i^{2,2} &= \min \left(\hat{\eta}_i^{2,2}, \underline{m}^{2,2}, \underline{m}^{2,4} \right), \text{ where } \hat{\eta}_i^{2,2} \sim N \left(0, \left(\frac{1}{2} \min \left(\underline{m}^{2,2}, \underline{m}^{2,4} \right) \right)^2 \right); \end{aligned}$$

where $\underline{m}^{k,l} = \min_i m^{k,l}(\mathbf{X}_i)$. Because $\sum_{l=k}^{S-2} m_i^{k,l} + m_i^{k,S} = 1$, the above three expressions also determine $\eta_i^{1,4}$ and $\eta_i^{2,4}$.

One may have noticed that the choice probabilities characterized in the paper should be interpreted as expected choice probabilities. Besides, the maximum likelihood estimation should be called pseudo- or quasi-maximum likelihood, as the true likelihood function contains some error terms related to $\eta_i^{k,l}$. For the estimations based on maximum likelihood, one may use the generalized method of moments (GMM) given

the presence of $\eta_i^{k,l}$, but the performance of the maximum likelihood is shown to be good enough. More importantly, the estimation used for most of the analyses, the case of degenerate beliefs with (in)equalities, is robust to the presence of $\eta_i^{k,l}$.

As estimation codes are inevitably “local” in the sense that their performance varies with primitives, I try to make the Monte Carlo exercises as close as possible to the real data. Each school has the same quota as in the main text. Both $Distance_{i,s}$ and $Parent_Inc_i$ are constructed with real data plus some random noise, while parameters are selected such that there are about 20% students who has no acceptable school as in the real data. In the end, the mean utilities of the four school are $(0.89, 0.11, -0.74, -1.12)$. To reduce computational burden, in all the estimations with simulated data, I only focus on four parameters, $(\alpha, \beta, \gamma_1^{2,4}, \gamma_2^{2,4})$, whereas all others are assumed to be known constants.

H.2 Data Generating Processes

Corresponding to the cases discussed in the main text, data are simulated under three data generating processes (DGP’s):

- (i) Truth-Telling: Each student reports her true preferences.
- (ii) Bayesian Nash Equilibrium: Facing the equilibrium beliefs, which are necessarily non-degenerate, each student submits a rank-order list that maximizing her expected utility.
- (iii) Degenerated Beliefs: Each student is endowed with individual-specific (subjective) beliefs with respect to which she maximizes expected utility by choosing a rank-order list. Besides, some elements in the belief system are degenerate. Specifically, for some students and any list (c^1, c^2, c^3, c^4) such that $c^3 = 1$ or $c^4 = 1$ (i.e. the most popular school being ranked 3rd or 4th), the (subjective) probability of being accepted by school 1 is zero. In the simulation, there are 817 such students. All other elements in the belief system are in $(0, 1)$ for everyone.

In all cases, students may include unacceptable schools in their lists following Assumptions UNACCEPTABLES and MIXING with mixing probabilities specified above.

To simulate choice data in the first scenario (truth-telling) is straightforward. Every student first truthfully rank the schools up to the least-preferred acceptable school, and then according to the mixing probabilities, she draws a lottery to decide if she includes unacceptable schools or not. If unacceptable schools are included, they are ranked after all acceptable schools and are ranked truthfully among all unacceptables (Assumption UNACCEPTABLES).

In the case of Bayesian Nash equilibrium, I first solve the equilibrium beliefs following the steps in Appendix G. Given the equilibrium beliefs, each student chooses a rank-order list, which only includes acceptable schools, to maximize her expected utility, and then she follows the same rules to decide if she includes unacceptable schools.

In the last scenario (degenerate beliefs), for every student, I generate a belief system that satisfies the structure specified in the main text, and some of them have degenerate beliefs. Students then choose a rank-order list to maximize expected utility with respect to their individual beliefs. The decision of including unacceptable schools is the same as in other cases. However, for students whose beliefs are degenerate, another mixing decision is made according to Lemma 4 but with an equal probability choosing any of the pure strategies. Note that these mixing probabilities are not estimated.

H.3 Estimation Approaches

After the data are simulated, there are four different estimation approaches to be taken, as discussed in the main text. Recall that the key to estimation is to specify the choice probabilities, which leads to either maximum likelihood estimation (MLE) or the Andrews-Shi approach to moment equalities and inequalities.

- (i) Truth-Telling: The choice probabilities for individual lists can be fully specified (see Appendix E for detailed derivations) and thus an MLE approach can be taken.
- (ii) Bayesian Nash Equilibrium: The leading approach under this assumption is to use MLE with choice probabilities for grouped lists that are derived under non-degenerate beliefs as in subsection E.3.1. I also present results showing that the MLE approach using either true equilibrium beliefs or approximated ones.
- (iii) Degenerated Beliefs with Equalities Only: I present results from an MLE approach that only uses moment equalities.
- (iv) Degenerated Beliefs with Both Equalities and Inequalities: The Andrews-Shi approach is used to obtain point estimates.

Clearly some approaches deliver consistent estimates under specific DGP's, while others always do. Table H10 summarizes the consistency of each estimation approach. Note that both estimation approaches with degenerate beliefs are consistent under all DPG's, while the other two approaches, Truth-Telling and Bayesian Nash equilibrium (non-degenerate beliefs), are only consistent under their corresponding DGP's respectively. In the Monte Carlo exercises below, data sets are generated from the three DGP's, and the four estimation approaches are applied to each data set.

Table H10: Monte Carlo Simulations: (In)Consistency of Performed Estimation Approaches

True Data Generating Process (DGP)	Estimation Approach			
	Truth-Telling	Non-Degenerate Beliefs	Degenerate Beliefs:	
			Equalities Only	Inequalities
Truth-Telling	Yes	Yes	Yes	Yes
Bayesian Nash	No	Yes	Yes	Yes
Degenerated Beliefs	No	No	Yes	Yes

H.4 Results

Results from 150 Monte Carlo samples are shown in Table H11, and they confirm the predictions. The truth-telling approach only delivers consistent results when the assumption is the same as the DGP; the non-degenerate-belief estimates are consistent when the DGP is either truth-telling or Bayesian Nash equilibrium; and the other two non-degenerate cases are always consistent. Besides, preference parameters (α, β) are more precisely estimated than the mixing probabilities, but for welfare analyses, the mixing probabilities only have second-order effects as discussed in section 3.3.2.

One may have also noticed that between the two cases with degenerate beliefs, the estimates from (in)equalities are less precise (high root-mean-square error, or RMSE), despite the theoretical advantages. One of the major reasons is the choice of instrumental variables. In the simulations, they are the “crude” ones, $(Distance_{i,1}, Distance_{i,2}, Distance_{i,3}, Parent_Inc_i)$ for moment equalities and their squared terms for the inequalities. This certainly creates some inefficiency, whereas the MLE of the case with equalities utilizes information efficiently. One may also wonder if an alternative choice of instruments, such as the hypercubes as proposed by Andrews and Shi. Results show no improvement while the computational burden is greatly increased.

Another reason for the inferior performance of the Andrews-Shi approach is the choice of tuning parameters. Specifically, the objective function, equation (2), may encounter close-to-zero denominators, $\hat{\sigma}_{j,k}(\theta)$, for some values of θ , and thus they are replaced by $\max(\hat{\sigma}_{j,k}(\theta), \underline{\sigma})$, where $\underline{\sigma}$ is a small number.

Table H11: Monte Carlo Simulations: Results from 150 Repetitions

	α			β			$m^{2,4}$ at $Parent_Inc_i = 0$			$m^{2,4}$ at $p75\ Parent_Inc_i$		
	mean	median	RMSE	mean	median	RMSE	mean	median	RMSE	mean	median	RMSE
<i>DGP</i>	<i>Estimation: Truth-Telling (consistent only if DGP = Truth-Telling)</i>											
Truth-Telling	-0.00	0.00	0.14	-1.00	-1.00	0.04	0.23	0.21	0.09	0.05	0.05	0.02
Bayesian Nash	0.02	0.02	0.10	-0.73	-0.73	0.257	0.19	0.18	0.10	0.04	0.04	0.02
Degenerate Beliefs	-0.32	-0.31	0.33	-0.67	-0.67	0.33	0.25	0.25	0.06	0.08	0.08	0.03
<i>DGP</i>	<i>Estimation: Non-Degenerate Beliefs (consistent only if DGP= Truth-Telling or Bayesian Nash)</i>											
Truth-Telling	-0.00	-0.00	0.14	-1.00	-0.99	0.05	0.23	0.21	0.09	0.05	0.05	0.02
Bayesian Nash	0.00	0.00	0.13	-1.01	-1.01	0.05	0.23	0.22	0.09	0.05	0.04	0.02
Degenerate Beliefs	-0.27	-0.27	0.29	-0.58	-0.58	0.42	0.24	0.24	0.06	0.08	0.07	0.03
<i>DGP</i>	<i>Estimation: Degenerate Beliefs – Equalities (always consistent)</i>											
Truth-Telling	0.00	-0.01	0.16	-1.00	-1.01	0.07	0.24	0.23	0.11	0.05	0.04	0.02
Bayesian Nash	0.00	0.00	0.14	-1.02	-1.01	0.08	0.23	0.21	0.12	0.04	0.04	0.02
Degenerate Beliefs	0.00	0.00	0.16	-1.01	-1.00	0.07	0.24	0.23	0.12	0.05	0.04	0.02
<i>DGP</i>	<i>Estimation: Degenerate Beliefs – Inequalities (always consistent)</i>											
Truth-Telling	0.00	-0.02	0.22	-0.95	-0.86	0.23	0.32	0.20	0.36	0.05	0.00	0.07
Bayesian Nash	0.01	-0.01	0.24	-1.03	-0.97	0.24	0.31	0.16	0.35	0.05	0.00	0.07
Degenerated Beliefs	-0.00	-0.01	0.22	-1.02	-0.93	0.31	0.28	0.22	0.30	0.05	0.03	0.07

Notes: RMSE = root-mean-square error; for consistent estimates, RMSE is close to standard deviation. Mixing probabilities are calculated based on the estimates at zero (or mean/median) income and its 75th percentile (p75).

Based on the method in Andrews and Shi, one can also test if a given vector of parameters is in the 95% confidence set. Results cannot reject the hypothesis that each set of the estimates, consistent or inconsistent, is in the confidence set. This indicates the confidence set may not be as informative as one would like.

H.5 Bayesian Nash Equilibrium

Estimation under Bayesian Nash equilibrium is discussed in subsection 3.3.2 where several approaches are proposed. Below, I compare the performance of each approach in the Monte Carlo simulations (Table H12).

A good benchmark is the maximum likelihood estimation assuming that the true equilibrium beliefs B^* are observed. With the characterization of choice probabilities of each list detailed in Appendix E, I use simulation to form the likelihood function. This is necessary as the choice probabilities depend on a mixture of 5 type I extreme values. The results show that it performs well with (α, β) – the preference parameters, but not very well with mixing probabilities. More importantly, its performance is not better than those from the MLE with non-degenerate beliefs which is also much easier computationally. This justifies that it is preferable to apply the latter when the data is generated from a Bayesian Nash equilibrium with unknown beliefs.

The other candidate approach is to approximate the equilibrium beliefs by bootstrapping from a single realization of the game. This approach treats each observed player as a random draw from the population, and the empirical action distribution is an approximate of the true distribution of equilibrium actions when there are a large number of players. By bootstrapping from the observed individual strategies, one may approximate B^* by \hat{B}^* . Unfortunately, this approximate does not work well in the simulation. Sometimes the estimated \hat{B}^* is not quite close to B^* (the biggest root-mean-squared error relative to B^* that elements in \hat{B}^* can have is about 0.49). Certainly, this error is partly due to the computational error in solving the equilibrium. This leads to inconsistent estimates as shown in the bottom lines of Table H12.

Table H12: Monte Carlo Simulations for Approaches to Estimation of Bayesian Nash Equilibrium

α			β			$m^{2,4}: Parent_Inc_i = 0$			$m^{2,4}: p75\ Parent_Inc_i$		
true value: 0			true value: -1			true value: 0.25			true value: 0.05		
mean	median	RMSE	mean	median	RMSE	mean	median	RMSE	mean	median	RMSE
<u>Estimation: Non-Degenerate Beliefs</u>											
0.00	0.00	0.13	-1.01	-1.01	0.05	0.23	0.22	0.09	0.05	0.04	0.02
<u>Estimation: True Equilibrium Beliefs</u>											
-0.03	-0.08	0.32	-0.96	-0.95	0.09	0.40	0.37	0.24	0.06	0.05	0.05
<u>Estimation: Bootstrapped Empirical Equilibrium Beliefs</u>											
-0.14	-0.18	0.70	-0.73	-0.73	0.31	0.46	0.44	0.29	0.15	0.07	0.23

Notes: Data are generated by a Bayesian Nash equilibrium. RMSE = root-mean-square error; for consistent estimates, RMSE is close to standard deviation. Mixing probabilities are calculated based on the estimates at zero (or mean/median) income and its 75th percentile (p75).

I Goodness of Fit and Comparison across Cases

This section provide discussion on goodness of fit of the four cases in Table 10.

I start with the predicted preferences. The left half of Table I13 reports each school’s average utility level from four cases relative to School 4. The average utility of School 4 is normalized to zero and considered as a benchmark because it is the same school for everyone, while the outside option may be different schools. To calculate these numbers, the deterministic part of the utility is calculated for each student, and the table shows the average over all students. Two cases with degenerate beliefs predict an order (1, 2, 3, 4) in terms of average utility, while the other two predict School 2 as the most valued school on average.

The right half of the table further shows the predicted probability of each school being the most preferred, again calculated for each student and averaged over all of them. The pattern is similar to the left half of the table. Also notice that all cases predict a similar probability of outside option being the most preferred, despite that the average utility of outside option varies substantially across cases.

I then compare these results with the claimed best school in the survey. The distribution of responses is available in a previous table (Table 8), and School 1 is considered as the best school by 83% of students. Furthermore, it confirms the most likely preference ranking as (1, 2, 3, 4). These results are closer to the predictions from the degenerate-belief case with (in)equalities.

One may be suspicious about the quality of data from the survey. Indeed, if parents answer this question with the objectively best school instead of their most preferred ones, the survey would over-estimate the popularity of School 1. I then look at students’ actual enrollment recorded in 2002, three years after school choice. The numbers are available in Table 2, and the ratio of the actual enrollment to the quota announced in 1999 is 233% for School 1, and 99%, 92%, and 46% for Schools 2, 3, and 4, respectively. These actual enrollment numbers clearly show the popularity of School 1 and a ranking of popularity as (1, 2, 3, 4).

Table I13: Goodness of Fit across Cases: Predicted Preferences

School	Average Utility ^b					Prob(Being Most Preferred) ^c				
	1	2	3	4	outside	1	2	3	4	outside
Degenerate Belief: (In)equalities ^a	4.32	2.67	2.34	0	-33.57	49%	14%	13%	5%	19%
Degenerate Belief: Equalities ^a	4.90	4.67	4.49	0	4.38	33%	23%	23%	1%	20%
Non-Degenerate Beliefs ^a	2.10	2.77	2.11	0	2.04	19%	37%	20%	3%	20%
Truth-Telling	2.28	3.25	2.11	0	2.33	17%	45%	15%	2%	20%
Claimed Best School in survey ^d						83%	8%	4%	0%	5%

Notes: a. Degenerate beliefs are such that some parents may have zero/degenerate elements in the belief system as in Assumption ZERO-PROB. b. Average utility is the deterministic part of the utility of each school averaged over all students. The average utility of School 4 is normalized to zero. c. This takes into account the probability that the outside option is the most preferred. d. Claimed best school is from the responses to a survey question, “Among those to which you could apply, which school was the best?” Each probability is the fraction of parents claiming that school as the best among all respondents. The distribution of responses is available in Table 8.

I.1 Sources of Biases

The above results show that the degenerate-belief case with (in)equalities dominates others in terms of model fit. The following analyses will be based on this set of estimates. Before moving on, it might be useful to discuss where the differences in model fit come from.

The over-estimation of School 2’s value in the truth-telling case is obvious. There are 431 (47%) students ranking it top and only 8 (1%) ranking it fourth, and the truth-telling assumption takes this ranking information literally. Another bias, the under-estimation of School 1’s value, is due to the same reason: There are only 228 (25%) students ranking it top and more importantly 41 ranking it bottom.

In the case with non-degenerate beliefs, the model do not take each ranking as literal as the truth-telling

assumption. However, it assumes that a bottom-ranked school is the least preferred. The 41 students who rank School 1 as fourth choice lead to a low estimate of School 1’s value, while the mere 8 students ranking School 2 as bottom choice result in a high valuation of School 2. As a comparison, there are 45 (5%) and 464 (51%) ranking Schools 3 and 4 as fourth choice, respectively.

This under-estimation of School 1’s value is gone in the degenerate-belief cases, as it is assumed that ranking School 1 at bottom does not mean it is the least preferred (Lemma 4) as a consequence of the degenerate beliefs (Assumption ZERO-PROB).

It is intriguing that the difference between the two degenerate-belief cases are non-negligible, although Monte Carlo simulations show the opposite (Appendix H). Certainly it implies that the over-identifying information in the moment inequalities is rich. The deeper reason is the mixing probabilities that make parents often submit full lists even when they only have one or two acceptable schools. When only the moment equalities are considered, all the full lists are grouped together (Table 6), and thus most of the identifying information is from partial lists. However, there are only 7% choosing partial lists and 20% not participating, which implies that 73% of parents are considered as choosing the same option while they certainly do not. The “outcome variable,” their choice among the 9 groups (Table 6), does not provide much variation, and the estimates become imprecise. In contrast, the case with (in)equalities uses information within each group in addition to the choice among the 9 options, which constitutes (over-)identifying conditions.

In the Monte Carlo simulations, both approaches work well, as the probability of submitting full lists is reasonable – $m^{1,4}$ on average is 47% and $m^{2,4}$ is 45% (Appendix H). There are a significant number of students choosing each of the 9 options, and thus the data provides enough variation in the outcome variable.

I.2 Estimates from the Degenerate Beliefs with (in)equalities

I now focus on the estimates from the inequalities and start with additional evidence on the goodness of fit. Although the model cannot predict the choice probability of each individual list, it can still predict choice probabilities of a subset of rank-order lists. Panel A in Table I14 includes the ratio of correct predictions for the lists in the moment equalities. If the predicted probability for a given list or a given group of lists is above 0.5, the student is predicted to choose that list/group. Because the groups of lists are mutually exclusive, although not exhaustive, a predicted probability greater than 0.5 means that the group in question is the most likely one. The fractions of correct prediction are at least 68%, and all others are above 90%.

Table I14: Goodness of Fit of Degenerate Beliefs with (In)equalities: Fraction of Correct Predictions

<i>Panel A: Moment Equalities^a</i>							
Groups of Two-School Lists			One-School Lists				Non-Participating
(4, 1, 0, 0)	(3, 1, 0, 0)	(2, 1, 0, 0)	(4, 0, 0, 0)	(3, 0, 0, 0)	(2, 0, 0, 0)	(1, 0, 0, 0)	(0, 0, 0, 0)
(1, 4, 0, 0)	(1, 3, 0, 0)	(1, 2, 0, 0)					
98%	92%	91%	100%	98%	98%	94%	68%
<i>Panel B: Binding Lower Bounds^b</i>							
Groups of Four-School Lists		Single Lists					
(4, 1, 3, 2)	(1, 4, 3, 2)	(4, 2, 0, 0)	(4, 3, 1, 2)	(4, 2, 1, 3)	(4, 1, 2, 3)	(4, 2, 3, 1)	
93%		99%	100%	99%	100%	94%	

Notes: For every student, the predicted probability of choosing a given list or a given group of lists is calculated by using the estimates from the degenerate beliefs with (in)equalities. a. For the lists that are in the moment equalities, if the predicted probability is above 0.5, it is predicted that the student chooses that list or that group of lists. b. For the lists that are in the binding lower bounds (no one chooses these lists in the data), it is predicted that the student chooses that list or that group of lists, if the predicted probability is above 1/41.

Moreover, there are six binding lower bounds of choice probability, because no one chooses these lists and thus the lower bounds have to be as close to zero as possible. For these groups, a student is predicted to choose a given list/group if the predicted probability is above 1/41 (i.e., when it cannot be the least likely option). Panel B of Table I14 shows that the prediction is again highly accurate.

J Additional Tables

Table J15: Attention on Different Aspects of Uncertainties and School Quality

	Quota Full List ^a (1)	Prob(Bad School) Full List ^a (2)	Others' App. Full List ^a (3)	School Quality Full Sample ^a Full List ^a (4) (5)	
Mean(Dep V)	4.23	4.47	2.81	4.15	4.14
Std Dev(Dep V)	0.86	0.74	1.21	0.46	0.47
<i>Parent_Edu_i</i>	0.02 (0.02)	-0.00 (0.02)	-0.06* (0.03)	-0.00 (0.01)	-0.00 (0.01)
<i>Parent_Inc_i</i>	-0.08** (0.04)	-0.00 (0.04)	-0.05 (0.07)	0.04** (0.02)	0.05** (0.02)
<i>Own_Score_i</i>	0.05 (0.61)	1.04** (0.52)	-1.00 (1.01)	0.22 (0.25)	-0.26 (0.34)
<i>Awards_i</i>	0.03 (0.04)	0.05 (0.03)	0.02 (0.06)	0.01 (0.02)	0.00 (0.02)
<i>Girl_i</i>	-0.02 (0.07)	-0.00 (0.06)	0.08 (0.11)	0.03 (0.03)	0.03 (0.04)
<i>Attn_Others_i</i>	-0.03 (0.03)	0.01 (0.02)		0.06*** (0.01)	0.07*** (0.02)
<i>Attn_Q_i</i>	0.53*** (0.08)	0.38*** (0.07)	0.61*** (0.14)		
Attn on Quota		0.34*** (0.04)	-0.08 (0.08)	0.15*** (0.02)	0.16*** (0.03)
Prob(Bad School)	0.47*** (0.05)		0.03 (0.09)	0.16*** (0.02)	0.16*** (0.03)
Constant	0.26 (3.15)	-3.95 (2.68)	6.82 (5.22)	1.20 (1.30)	3.54** (1.75)
Observations	457	457	457	676	457
R-squared	0.34	0.34	0.08	0.30	0.30

Notes: Results are from OLS regressions, and other variables include fixed effects for elementary schools. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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.

Table J16: Deviation from Best-Responding and Truth-Telling Prediction: Partial Lists

List	Observed Fraction ^a (in percentage points)	Deviation (in percentage points) from the Prediction of		
		Best Responding ^b	Truth-Telling ^b	Maximin Preferences ^b
(4,3,0,0)	0.22	0.21	0.21	0.20
(4,2,0,0)	0.00	0.00	0.00	0.00
(4,1,0,0)	0.00	-0.01	-0.02	-0.02
(3,4,0,0)	0.22	0.21	0.21	0.20
(3,2,0,0)	0.11	0.10	0.09	0.09
(3,1,0,0)	0.11	-0.01	-0.08	-0.14
(2,3,0,0)	2.95	2.95	2.95	2.94
(2,4,0,0)	0.22	0.22	0.22	0.22
(2,1,0,0)	0.66	0.41	0.43	0.26
(1,3,0,0)	0.77	0.53	0.40	0.43
(1,2,0,0)	1.09	0.68	0.21	0.28
(1,4,0,0)	0.00	-0.02	-0.03	-0.03
(4,0,0,0)	0.11	0.08	0.08	0.11
(3,0,0,0)	0.22	0.17	0.17	0.20
(2,0,0,0)	0.11	0.06	0.06	0.11
(1,0,0,0)	0.66	0.04	0.04	0.52
(0,0,0,0)	19.80	0.98	0.98	0.99

Notes: a. This reports the percentage out of the total of 914 students who submit the given list.

b. This is the average individual deviation. A positive number implies the observed fraction is higher than the model prediction.

Table J17: 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: -0.08; Std Dev: 0.27			Deviation from Truth.Telling Mean: 0.20; Std Dev: 0.41			Deviation from Truth.Telling Mean: -0.11; Std Dev: 0.44		
<i>Parent_Inc_i</i>	0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.04* (0.02)	0.02 (0.02)	0.02 (0.02)
<i>Parent_Edu_i</i>	-0.00 (0.00)	0.01 (0.00)	0.01 (0.00)	-0.00 (0.01)	-0.00 (0.01)	-0.00 (0.01)	-0.01 (0.01)	-0.01* (0.01)	-0.02** (0.01)
<i>Own_Score_i</i>	0.41*** (0.12)	0.84*** (0.19)	0.95*** (0.22)	-0.25 (0.22)	0.04 (0.30)	0.14 (0.34)	0.58*** (0.19)	-0.99*** (0.19)	-0.90*** (0.21)
<i>Awards_i</i>	0.02** (0.01)	0.01 (0.01)	0.01 (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	0.03* (0.02)	0.03* (0.01)	0.02 (0.02)
<i>Girl_i</i>	-0.03* (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.03 (0.03)	0.02 (0.03)	0.04 (0.03)	-0.02 (0.03)	-0.02 (0.03)	-0.03 (0.03)
<i>P_i^{TT=BR}</i>		-0.00 (0.05)	-0.01 (0.05)		-0.15** (0.07)	-0.15* (0.08)		0.59*** (0.07)	0.60*** (0.08)
<i>Gain_i</i>		0.54*** (0.16)	0.60*** (0.18)		-0.56 (0.34)	-0.53 (0.37)		-0.47 (0.31)	-0.55 (0.36)
<i>Attn_U_i</i>			-0.01 (0.01)			0.04** (0.02)			-0.01 (0.02)
<i>Attn_Q_i</i>			0.01 (0.02)			-0.03 (0.03)			0.05 (0.03)
<i>Attn_Others_i</i>			-0.01* (0.01)			0.02 (0.01)			-0.02 (0.01)
<i>P_{i,k}^{BR}</i>		-0.62*** (0.04)	-0.66*** (0.04)		-0.15** (0.07)	-0.15* (0.08)		-0.47*** (0.13)	-0.50*** (0.14)
Obs.	914	914	810	914	914	810	914	914	810
R-Squared	0.07	0.25	0.27	0.05	0.06	0.07	0.12	0.28	0.29

Notes: Definitions of $P_i^{TT=BR}$ and $Gain_i$ are in Table 13. $P_{i,k}^{TT}$ 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 J18: 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: 0.04; Std Dev.: 0.38			
<i>Parent_Inc_i</i>	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)
<i>Parent_Edu_i</i>	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)
<i>Own_Score_i</i>	1.11*** (0.18)	1.10*** (0.17)	0.86*** (0.23)	1.02*** (0.25)
<i>Awards_i</i>	0.02 (0.01)	0.01 (0.01)	0.01 (0.02)	-0.00 (0.02)
<i>Girl_i</i>	0.00 (0.02)	0.01 (0.02)	0.01 (0.02)	0.00 (0.03)
$p_i^{TT=BR}$			0.08 (0.08)	0.08 (0.09)
<i>Gain_i</i>			0.70*** (0.26)	0.75** (0.30)
<i>Attn_U_i</i>				-0.01 (0.02)
<i>Attn_Q_i</i>				0.06* (0.03)
<i>Attn_Others_i</i>				-0.02* (0.01)
$P_{i,k}^{TT}$		-0.26*** (0.06)	-0.28*** (0.08)	-0.31*** (0.08)
Obs.	914	914	914	810
R^2	0.10	0.12	0.13	0.15

Notes: Definitions of $p_i^{TT=BR}$ and $Gain_i$ can be found in Table 13. $P_{i,k}^{TT}$ is the probability that the list (in the dependent variable) is the true preference ranking. All regressions include elementary school fixed effects.