Matching Mechanisms and Justified Envy: Stability on the Intensive Margin

Le Kang
Wei Ha
Yang Song*
Sen Zhou

July 2022

Abstract

We use millions of student records in Chinese college admissions and construct four intensive-margin stability measures, including the number of blocking students, pairs, and triplets, and a novel measure that captures the *degree* of justified envy (JE). Utilizing the staggered timing of provincial reform, we show that the parallel mechanism outperforms the Immediate Acceptance mechanism in all JE measures. The four measures lead to different conclusions on the matching outcomes' distance to stability and the improvement from mechanism reform. We also document a distributional pattern, where middle-ranking students suffer higher levels of JE and enjoy greater improvement from mechanism change.

JEL Codes: C78, I23, I28.

*Le Kang, Institute of Education, Nanjing University; China Institute for Educational Finance Research, Peking University. Sen Zhou, China Institute for Educational Finance Research, Peking University. Wei Ha, Institute of Economics of Education, Peking University. Corresponding author: Yang Song, Department of Economics, Colgate University. 13 Oak Dr, Hamilton, NY 13346. ysong@colgate.edu.

1 Introduction

Well-functioning matching mechanisms facilitate fair and efficient student assignment in school districts and centralized college admissions. Prior literature has focused on three properties of matching mechanisms: stability, efficiency, and strategy-proofness. Among them, stability has a clear link to fairness. In the common school choice and college admissions contexts,[1] an unstable matching is often due to unfairness.[2] For example, many would consider it unfair when your preferred school ranks you higher than a student assigned to them, yet you are not assigned to this school. At the same time, you and your preferred school both have an incentive to deviate from the assignment, thus creating instability.

With a long history of exam-based civil service candidate selection systems, Chinese citizens have a particularly strong distaste for unfair assignment outcomes. News stories about high-achieving students failed by the matching process and assigned to a low-ranked college stir up social anxiety and dissatisfaction with the assignment system. This kind of news stories played a critical role in the Chinese government's reform of college admissions in the past few decades. In Chinese college admissions, students with justified envy (JE),[3] who are faced with other students scoring lower but placing higher, are more likely to repeat a year and retake the college entrance exam (Kang et al., 2020). This is costly to both individual students and society, considering the valuable time of youth and the discounted value of a delayed entry into the labor market. Unfair and unstable assignments could also have efficiency implications if student ability and college quality are complementary in the talent and knowledge production process.[4]

Most of the previous theoretical discussions related to stability have been based on a binary concept. However, an unstable matching where only one student has JE is fairer than an unstable matching where half of the students share in the unjust feeling. The binary stability measure does not capture such a difference, nor the heterogeneous impacts of instability on different students. Even though a few recent theory studies have started to compare justified envy by set-inclusion or counts of blocking students, pairs, or triplets (Abdulkadiroğlu et al., 2020; Bonkoungou and Nesterov, 2020; Dogan and Ehlers, 2021, 2022; Kwon and Shorrer, 2020), the understanding of the relationship between these measures and the heterogeneous impacts of JE on different students is lacking. Further, the existing intensive-margin measures of blocking students, pairs, and triplets are all count measures, but a case where one student ranks one percentile below you yet gets into a slightly better school, feels fairer than the case where one student ranks 50 percentiles below you yet gets into a much better school. The count measures do not capture the varying *degree* of

[1] See Romm, Roth and Shorrer (2020) for discussions on the contexts in which justified envy-free and stability do not coincide.

[2] Instability could also be due to wastefulness, i.e. unassigned slots preferred by a student over their assignment.

[3] A more detailed definition is given in the next section.

[4] Dillon and Smith (2020) find evidence for complementarity between student ability and college quality in generating higher earnings using NLSY79&97. Bleemer (2022) find that the top percent policy in elite California universities benefited the marginal minority students more than it hurt the displaced slightly higher-performing non-minority students, although this net positive effect not only comes from the interaction between student ability and college quality, but also from the interaction between racial discrimination and the sheepskin effect.

unfairness: how much lower the envied student is ranked and how much higher you would value the better school.

Our paper addresses this gap by using empirical data to construct existing and new intensive-margin fairness measures, comparing them under different mechanisms, and exploring the distributional patterns of JE over student performance. Section 2 describes the four main JE measures we consider and their relationship with one another. In addition to the existing measures, we create a novel measure, tridimensional envy, which incorporates how much lower the envied student is ranked, how much higher the school is ranked, and how many students one envies. We use an illustrative example to show that these four measures do not have a monotonic relationship with one another. When one measure increases, another measure may or may not increase. We then give a brief theoretical background of the two main mechanisms we consider: the Immediate Acceptance (IA, or Boston) mechanism and the parallel mechanism–a hybrid of IA and the Deferred Acceptance mechanism.

This study is set in the context of Chinese college admissions. The setting is unique in three ways. First, it's arguably the largest matching market in the world, where millions of students compete for limited seats at hundreds of universities every year. Second, the college entrance exams and matching assignments are centralized at the provincial level. In the past two decades, many provinces changed their matching mechanism from the IA mechanism to the parallel mechanism. The temporal and geographical variations in the reform provide a unique opportunity for estimating the causal effects of such mechanisms. Lastly, not knowing student preferences is often an obstacle in empirical research in matching. To construct JE, we need to make an assumption on student preferences.[5] The Chinese higher education system provides a context where students share overall homogeneous preferences over Tier I universities. We further relax the homogeneous preference assumption by allowing students to have heterogeneous preferences within a rank group (e.g., Peking and Tsinghua in China, or Harvard and Yale in the U.S.) and only assume homogeneous preferences over groups of universities. The rank groups are categorized based on the admission cutoff scores, i.e. the equilibrium "price" of seats at universities. Section 3 gives more details on the setting.

We use millions of individual-level data to construct JE measures for 398 matching games at the province-track-year level between 2005 and 2011.[6] Section 4 describes data and summary statistics. All of the matching outcomes are unstable, with at least some justified envy, highlighting the insufficiency of looking at stability as a binary measure. Comparing these intensive-margin JE measures yields intriguing observations. The numbers of blocking students and pairs suggest that the realized matching outcomes in Chinese college admissions are much closer to the least stable matching than to the stable matching, offering a pessimistic picture of how the IA and parallel mechanisms perform. If we assign a stability score from 0 to 100, with 100 being stable, the matching outcomes would get failing grades of 14 for the number of blocking students and 31 for

[5]Note that assuming homogeneous preference implies that there is no way to improve Pareto efficiency, limiting our ability to compare efficiency across mechanisms.

[6]Admissions are separate for STEM and Non-STEM tracks.

the number of blocking pairs. However, when we look at the number of blocking triplets and tridimensional envy, the matchings are quite close to the stable matching, scoring at 92 for the number of blocking triplets and 95 for tridimensional envy.

Using a generalized difference-in-differences framework described in Section 5, we find that the parallel mechanism reduces all JE measures and brings the matching outcome closer to a stable matching on the intensive margin. The improvement from IA to parallel is statistically significant but small for the blocking students, 7.1 percentage points or 8.2% of the sample mean. The improvement in the number of blocking pairs is around 26% of the sample mean. But the relative magnitudes of improvement from reform are large (55-60% of sample mean) for the blocking triplets and tridimensional envy, which incorporates more information from the matching outcome. We further investigate how the magnitudes of effects change over time, as well as the relationship between improvement in JE and the number of parallel choices in the first choice band.[7] Contrary to what one would expect from a related theoretical result (Chen and Kesten, 2017), the magnitude of effects does not always increase with the number of parallel choices. The effect sizes for six parallel choices are smaller than those for five parallel choices, especially for blocking students and pairs. Section 6 presents these results and a variety of robustness checks.

Finally, we find an inverted-U shape relationship between student priority order (CEE scores) and their JE levels. The highest-ranked and lowest-ranked students have lower levels of JE under both IA and parallel mechanisms than those ranked in the middle. They also benefit more from the change in mechanism in the reduction of JE measures. Although a low level of JE is perhaps not surprising for the lower-ranked students given that one needs to have a higher priority order to justifiably envy others, the comparison between highest-ranked and mid-ranked students is interesting. We find evidence supporting that congestion in mid-ranked universities may explain the inverted-U shape we observe. In particular, the preferences for top universities are clearer, while students may be indifferent towards a group of mid-ranked universities. The limited number of choices in the parallel mechanism and the lack of coordination on which mid-ranked universities to include in choice lists lead to larger variations in the cutoff scores among mid-ranked universities. These are discussed in Section 7, followed by discussions for future research in Section 8.

This paper adds to the matching literature in school choice, pioneered by Abdulkadiroğlu and Sönmez (2003). While the majority of this literature has been theoretical and experimental, we have seen a growing body of empirical studies.[8] Although experimental studies help test theoretical predictions and provide valuable insights, analyses using real-world data are useful for determining external validity (Levitt and List, 2007) both in terms of higher stakes and larger game sizes. This study provides empirical estimates to generate useful comparisons with experimental findings

[7] Choice bands can be roughly understood as "rounds" in the matching process. There are multiple choice bands, or rounds, in the parallel mechanism: assignment within each choice band is temporary while assignment across choice bands is final. In this context, seats in Tier I colleges are basically all finalized after the first choice band is considered.

[8] See Abdulkadiroğlu, Agarwal and Pathak (2017); Agarwal and Somaini (2018); Calsamglia, Fu and Güell (2020); Kapor, Neilson and Zimmerman (2018); Chen, Jiang and Kesten (2020); Terrier, Pathak and Ren (2021); Akbarpour et al. (2022) for a few examples.

(Chen and Kesten, 2019). The percentage of students with JE in Chinese college admissions is over 80%, much higher than the approximate 10% in the four-school experimental setting and the approximate 20% in the six-school setting. One student with a very low score and an unfairly high-ranked college assignment could lead to a majority of students having justified envy. Chen and Kesten (2019) find in the experimental setting that the parallel mechanism reduces the number of blocking pairs by 50% compared with IA in the four-school environment, and by around 18% in the six-school environment. We find a 28% reduction in the real world setting with over a hundred schools.

More importantly, we are the first to use large-scale empirical data to compare multiple fairness measures, including a novel measure accounting for degree of justified envy, and study their distributional patterns. Until recently, most theoretical studies have focused on a binary concept of stability. A few recent theoretical studies on justified envy have focused on comparisons based on set inclusion or counts of blocking students, pairs, or triplets (Abdulkadiroğlu et al., 2020; Bonkoungou and Nesterov, 2020; Dogan and Ehlers, 2021, 2022; Kwon and Shorrer, 2020).[9] Chen, Jiang and Kesten (2020) and Abdulkadiroğlu et al. (2020) are two empirical studies that have measured and compared justified envy on the intensive margin. Abdulkadiroğlu et al. (2020) use data from New Orleans and Boston and show that TTC admits significantly less JE than random serial dictatorship. While DA underperforms TTC in efficiency (measured by whether the first choice is assigned) by around 1%, it significantly outperformed TTC in fairness (measured by the number of blocking pairs). Chen, Jiang and Kesten (2020) use data from one Chinese county and find that the percent of blocking students is lower under the parallel mechanism than the IA mechanism.

Our results suggest that focusing on one measure could yield misleading judgments on the performance of a matching mechanism or the change resulting from adopting a new mechanism. According to the number of blocking students and pairs, both IA and parallel mechanisms perform poorly, and the improvement from parallel reform is small. According to the number of blocking triplets and tri-dimensional envy measures, however, both mechanisms are very close to the stable matching, and the relative magnitude of improvement from parallel reform is large. These observations highlight the differences between these measures and call for more theoretical analyses on blocking triplets and tri-dimensional envy (or other measures for degree of envy), as they receive less attention thus far yet capture more information on how stable a matching is. The observations on the nonlinear relationship between parallel choice band and effect size and the distribution of JE measures across student performance also warrant further investigation.

We also add to the understanding of Chinese college admissions and the parallel mechanism. Chen and Kesten (2017) characterize the parallel mechanism within a system of matching mechanisms as in between IA and DA, and they show that the parallel mechanism is more stable than

[9] Abdulkadiroğlu et al. (2020); Kwon and Shorrer (2020); Dogan and Ehlers (2022) study the justified envy minimal mechanism among efficient mechanisms in one-to-one matching markets. Dogan and Ehlers (2021) show the impossibility to reduce the number of blocking pairs or the set of blocking students while achieving Pareto efficiency improvement over the DA mechanism. Bonkoungou and Nesterov (2020) show that the IA and parallel mechanisms are not comparable in terms of counts of blocking students. We give more descriptions at the end of Section 2.

the IA mechanism.[10] In addition to experimental studies (Li, Wang and Zhong, 2016; Lien, Zheng and Zhong, 2016, 2017; Chen and Kesten, 2019), a few empirical studies have compared matching mechanisms in the Chinese college admission context. However, they face limitations in their outcome measures Ha, Kang and Song (2020) or use data from only one county or university Wu and Zhong (2014); Chen, Jiang and Kesten (2020).[11] This study leverages administrative data on the entire universe of students admitted to Tier I universities in China, allowing us to have a more granular understanding of fairness under two mechanisms.

2 Justified Envy and Theoretical Background

In this section, we define some theoretical concepts and the justified envy measures, describe the matching mechanisms, and summarize related theoretical background.

2.1 Justified envy measures

The theoretical literature on matching focuses on three main properties: stability, strategy-proofness, and efficiency. The justified envy measures are closely connected to stability. We will first set up the problem and introduce some notations and definitions.

A college admissions problem includes N students, denoted by $I = i_1, i_2, ...i_n, ...i_N$, each to be assigned to a school in $S = s_1, s_2, ...s_j, ...s_J \cup \emptyset$, where $J \geq 2$ and $\emptyset$ denotes the null assignment, or a student's outside option. Each university has a limited number of seats, or capacity q_j. Denote the preference order for student i by P_i and the preference order for university s by $\succ_s$. A matching $m : I \to S$ is a list of assignments that assign each student to a school, including the null assignment, and allows no university to admit more than its capacity.

Student i ***justifiably envies*** student j for school s if i would rather be assigned to school s, where some student j, who has lower priority than i, is assigned. In this case, student i is a ***blocking student***; student i and and school s are a ***blocking pair***; and student i, j and school s are a ***blocking triplet***. These three are among the main outcomes in our empirical analysis. A matching is ***stable*** if it is non-wasteful and free of justified envy. To be ***non-wasteful***, a matching outcome either has no unfilled seats or has no student preferring a school with any remaining seat to their own assignment. Stability is a binary measure and does not distinguish between a matching with 1% of blocking students and one with 50%.

Consider a matching problem with three universities and six students. Schools share homogeneous preferences: $s_1 P s_2 P s_3$, and students share homogeneous preferences: $i_1 \succ i_2 ... \succ i_6$. It is easy to see that the unique stable outcome, also the fairest outcome, would be when i_1 and i_2 are assigned to s_1, i_3 and i_4 assigned to s_2, and i_5 and i_6 assigned to s_3 - a perfectly assortative

[10] "More stable" is defined such that given any matching game, if IA produces a stable matching then so does parallel; but not the other way around.

[11] Ha, Kang and Song (2020) use university-level college admissions data and show that changing from IA to a parallel mechanism likely improved stability and increased stratification. Wu and Zhong (2014) use data from one elite university and find that pre-exam submission combined with the IA mechanism produces better ex-ante fairness than post-exam submission combined with the parallel mechanism.

matching. Any other assignment would exhibit justified envy, some more than others. The least fair matching game (hereafter worst matching) would be to reverse the assignment, assigning the lowest-ranked students to the most preferred university, and so on. In this worst matching, the number of blocking students (BS) is 4, the number of blocking pairs (BP) is 6, and the number of blocking triplets (BT) is 12. To illustrate, consider the matching in Figure 1, where i_1 is assigned to s_1, i_2 to s_3,i_3 to s_1, i_4 to s_2, i_5 to s_3, and i_6 to s_2. The arrows point from the blocked students to the student(s) they justifiably envy, and the shaded areas capture the areas with justified envy. Each origin of these arrows consists of a blocking student (BS=2), each arrow consists of a blocking triplet (BT=4), and the number of unique combinations of blocking student and blocking university gives us the number of blocking pairs (BP=3).

Note that all three measures are counts, and none capture the extent of justified envy. Consider two cases: in one case, another student ranks one percentile below you yet got into a slightly better school than you, and in the second case another student ranks 50 percentiles below you yet got into a much better school than you. Both will yield the same count measures, yet the latter induces a higher disutility from unfairness and creates a higher level of threat to matching stability.

Therefore, we create a novel measure that takes into account the degree to which a student justifiably envies the other. For any student i, denote by T_i the rank of student i's assigned university and r_i as student i's rank. Consider student i as the focal student. i justifiably envies j if $r_i < r_j$ and $T_i > T_j$. Denote by D_{ij}^T the degree of JE from i to any student j in the dimension of university admitted. D_{ij}^T equals zero if i does not justifiably envy j and equals $T_i - T_j$ otherwise (note that $T_j < T_i$ is a necessary condition for i to envy j). Similarly, the degree of JE in the dimension of student rankings, D_{ij}^r, equals zero if i does not justifiably envy j and equals $r_j - r_i$ otherwise (note that $r_i < r_j$ is the other necessary condition for i to justifiably envy j). Finally, we construct a measure for degree of envy: $D_{ij}^d = \sqrt{(D_{ij}^T)^2 + (D_{ij}^r)^2}$, which is the diagonal distance, or the length of the justified envy vector between i and j. Note that this is two-dimensional, accounting for differences in student rank and university rank. For each student i, the ***tridimensional envy***(TE) equals $\sum_j D_{ij}^d$, which sums the degree of JE over all the students i envies. This is equivalent to summing the lengths of all justified envy vectors initiating from i.

These four JE measures have a non-monotonic relationship with one another. By nonmonotonic relationship, we mean that if JE measure A is larger in matching 1 than in matching 2, this does not necessarily imply that JE measure B is also larger in matching 1. One can easily construct examples where JE measures have opposing comparisons between two matching outcomes. For example, consider matching outcome U (student 3, 4 assigned to school 1; student 5, 6 assigned to school 2; student 1, 2, assigned to school 3) and matching outcome V (student 1, 3 assigned to school 1; student 2, 6 assigned to school 2; student 4 and 5 assigned to school 3) under the setting of our illustrative matching game. Matching U has fewer blocking students but more blocking pairs, and much more blocking triplets and tridimensional envy than Matching V.[12]

[12]We provide an exhaustive list of all possible matching outcomes in a matching problem with six students and three school tiers here: `https://github.com/grasiasong/matching`.

To compare the justified envy measures across matching games with varying numbers of schools and students, standardized measures are needed. This is because these measures grow non-linearly with the game size, so controlling for the game size in our empirical analysis would not be sufficient. To standardize the JE measures, we first construct the worst matching outcome given a matching problem: assigning the lowest-ranked students to the most preferred school up to its total quota, then the next lowest-ranked students to the second-best university up to its total quota, and so on. This worst possible matching outcome gives us the largest JE measures for that matching game size. We then divide a given JE measure by the corresponding JE measure of the worst possible matching to obtain a standardized JE measure. In the illustrative example, the standardized blocking student measure equals actual BT (4) divided by the worst game's BT (12), which equals one-third. All the standardized measures range from zero to one, with zero being the stable matching and one being the worst matching.

2.2 Matching mechanisms

Two matching mechanisms are of central relevance to this paper, the Immediate Acceptance (IA) mechanism and the parallel mechanism, which is a hybrid between the IA and the Deferred Acceptance (DA) mechanisms. The IA mechanism operates as follows:

Step 1. Each school s considers only the students who list it as their first choice. Up to q_s students with the highest preference orders by school s are assigned to school s.

Step k, $k \geq 2$. School s with q_s^k remaining seats consider the students who are unassigned at the beginning of step k and list school s as their kth choice. The q_s^k students with the highest preference orders by school s are assigned to school s.

The algorithm continues until all student lists are considered or get assigned to a school. The IA mechanism has been criticized for its lack of strategy-proofness and stability, conditions which are achieved by the DA mechanism (Gale and Shapley, 1962). The student-proposing DA works as follows:

Step 1. For each school s, up to q_s applicants who listed school s as their top choice and have the highest preference orders by school s are *tentatively* assigned to school s. The remaining applicants are rejected.

Step k, $k \geq 2$. Each student rejected from a school at step $k-1$ is considered by the kth school listed. For each school s, up to q_s students who have the highest preference orders by school s among *both* the new applicants *and* those tentatively held from the earlier step are tentatively assigned to school s. The remaining applicants are rejected.

The algorithm continues until all student lists are considered. The key difference between IA and DA is that under IA, assignment after every choice is final; under DA, assignment after every choice is temporary. The parallel mechanism is in-between the IA and DA mechanisms. Students list several parallel universities within each choice band. Within each choice band, the parallel

mechanism operates like DA; across choice bands, it operates like IA. Here we describe how a parallel mechanism with two parallel choices in each choice band operates, as an example:

Step 1. For each school s, up to q_s applicants who listed school s as their top choice and have the highest preference orders by school s are tentatively assigned to school s. The remaining applicants are rejected.

Step 2. Each student rejected from a school at step 1 is considered at their second choice school. For each school s, up to q_s students who have the highest preference orders by school s among *both* the new applicants *and* those tentatively on hold from step 1 are assigned to school s. The remaining applicants are rejected. All assignments are final after all students' two parallel choices in the first choice band are considered.

Step 3. Step 1 and Step 2 are repeated using the two parallel choices in the second choice band for students who have not yet been assigned to any school. Seats assigned during the previous choice band consideration are no longer available. The algorithm continues until no student proposal is rejected.

As discussed by Chen and Kesten (2017), these three matching mechanisms can be grouped into a family of application-rejection mechanisms, $\phi^e, e \in \{1, 2, ..., \infty\}$. When $e = 1$, ϕ^e is equivalent to the IA mechanism. When $2 \leq e < \infty$, ϕ^e is equivalent to the parallel mechanism. When $e = \infty$, ϕ^e is equivalent to the DA mechanism. Chen and Kesten (2017) show that as e, the number of parallel choices in each choice band, increases by an integer of $k > 1$, the mechanism becomes less manipulable and more stable.[13] "More stable" is defined such that given any matching game, if ϕ^e produces a stable matching then so does ϕ^{2e}; but not the other way around.

Although most prior theoretical studies focus on stability as the main fairness measure, several recent theoretical studies have focused on justified envy. Abdulkadiroğlu et al. (2020) show that JE can be minimized by the Top Trading Cycle (TTC) mechanism among the Pareto efficient and strategy-proof mechanisms. They show that TTC admits fewer blocking pairs on average when every possible priority profile is considered or when participantsâ priorities are drawn uniform randomly. In their empirical analysis using data from New Orleans and Boston, they show that although TTC outperforms random serial dictatorship, DA only underperforms in efficiency (measured by whether the first choice is assigned) by around 1% while it significantly outperformed in fairness (measured by the number of blocking pairs). In a related paper, Dogan and Ehlers (2022) show that the top trading cycles (TTC) mechanism is minimally unstable among efficient and strategy-proof mechanisms in one-to-one matching, not only in terms of set inclusion but also in some cardinal comparisons. Kwon and Shorrer (2020) show that efficiency-adjusted deferred acceptance (EADA) mechanism is justified-envy minimal, defined in regards to set inclusion, in the class of efficient mechanisms, while TTC is not.

Dogan and Ehlers (2021) look at potential efficient and minimally unstable Pareto improvements over the deferred acceptance (DA) mechanism. In particular, they focus on comparing the number

[13] Chen and Kesten (2017) also hypothesize a more general result where the manipulability and stability improve monotonically with the number of choices in the parallel choice band and provide some simulation analysis to support the hypothesis.

of blocking pairs and the set of blocking students and show that there is no Pareto improvement over the DA mechanism that is minimally unstable among efficient assignments. Bonkoungou and Nesterov (2020) give examples where a parallel mechanism generates more or fewer blocking students than the IA mechanism to show that the IA and parallel mechanisms are not comparable in terms of counts of blocking students. This paper does not attempt to compare mechanisms theoretically. Instead, we focus on the empirical characterization of the average justified envy and the distribution of justified envy, as well as changes therein induced by the mechanism change, with the hope of bringing unanswered questions to the table.

3 Chinese College Admissions

The college admissions system in China is centralized at the provincial level. Universities allocate quotas to each province, and the provincial education bureau assigns these seats to students based on their College Entrance Exams (CEEs) scores and submitted college preferences. Students choose either a STEM or non-STEM track at the end of 10th grade, then study different subjects in the last two years of high school and take the CEEs at the end of 12th grade. All students study and take the CEEs in Chinese, mathematics, and English. In addition to these three subjects, students in the STEM track study and get tested on physics, chemistry, and biology, while students in the non-STEM track do so for politics, history, and geography.[14] Due to the subject differences, the CEEs and college admission processes are separate by track. In sum, students are competing internally with others in the same track-province-year.

3.1 Matching reform in China

The matching process differs across provinces in two dimensions: timing of preference submission and matching mechanism. All provinces used to have students submit their preferences before taking the CEEs (ex-ante preference submission). Over time, some provinces changed to preference submission after taking the CEEs but before knowing the scores (ex-interim preference submission), and some changed to preference submission after knowing the scores (ex-post preference submission). The left panel in Figure 2 shows the variations in preference submission timing over time. Between 2005 and 2011, ten provinces changed from either ex-ante or ex-interim to ex-post submission. By 2011, most provinces had adopted ex-post preference submission.

Before 2003, all provinces used the Immediate Acceptance (IA) mechanism. Under IA, almost all Tier I universities are filled after considering students' first choices, so the first choice is crucial. If a student aims too high and fails to be admitted to their firstchoice university, this student could end up being assigned to a university in a lower tier, or not being assigned to any university at all. These students often repeat a year of high school to go through the college admissions process

[14]Some provinces have experimented with the track system in recent years, not following the conventional binary track system. During our sample period between 2005 and 2011, the traditional two-track system was adopted across most provinces. Jiangsu and Guangdong in 2005 and 2006 and Liaoning in 2005 adopted an alternative track system, therefore we exclude them from our sample.

again, which is highly costly. Because of the risks previously discussed, some students are overly conservative in their choices and significantly undershoot, which results in undermatching. In 2003, Hunan Province was the first to adopt the parallel mechanism. Following the positive feedback from early reform provinces, the Ministry of Education issued a recommendation in 2008, which arguably generated exogenous adoption of the parallel mechanism. The right panel of Figure 2 shows that the majority of the take-up happened in 2008 and 2009, following the Ministry's recommendation. By 2011, only seven provinces were still using the IA mechanism in their Tier I university assignments.

3.2 Plausibility of homogeneous preferences assumption

To measure justified envy (JE), we need to know (1) university preferences over students and (2) student preferences over universities. Universities rank students by their College Entrance Examination scores, observed in the data, with two exceptions. First, a few colleges admit students to special programs such as art, music, acting, and sports with another interview or exam process that is also considered as part of their selection criteria. The other exception is the *"zizhu zhaosheng"* policy, or the college-specific bonus points policy, where a small number of students can earn bonus points (usually 10 to 20 out of 750 points) for a specific college. These bonus points are not applied to other universities, which generates heterogeneous preferences by universities. However, this policy is normally restricted by the Ministry of Education to less than 5% of the admission quota and only started to become more widely adopted after 2011, the end of our sample period.[15]

While we observe university preferences through student test scores, we do not have data on student preferences, which is a common obstacle in the literature.[16] Fortunately, in our setting, student preferences are relatively homogeneous due to the highly stratified higher education system and mostly single-dimensional comparison among universities in China. Since the 1990s, the Chinese government has poured resources into a select group of universities, aiming to build world-class universities by concentrating resources on a small subset. University prestige and selectivity are highly stable over time, especially within each province, the level at which the matching games take place. In addition, unlike the American higher education system, where students have more dimensions to consider, such as private versus public, small liberal arts versus large research universities, sports, and in-state tuition, these factors matter little in the Chinese context. All Tier I universities we focus on are large, public research universities.[17] College sports leagues receive very little attention in China. There is also no tuition difference for in-province students, and the tuition rates are regulated by the Ministry of Education (5000 yuan, or roughly 800 USD per year) and have not changed for years.

[15] See `http://www.moe.gov.cn/srcsite/A15/moe_776/s3110/201111/t20111115_127339.html`, retrieved on 10/31/2018).

[16] Some papers use student submitted choice lists to estimate student preferences by making assumptions about stability under DA or truthful reporting of preferences (Fack, Grenet and He, 2019), and a few papers, such as De Haan et al. (2021), surveyed students on their preferences.

[17] They are the only choices for high-achieving students, except for studying in Hong Kong or abroad. Those who plan to attend colleges abroad make their decision before the CEE takes place in June at the end of 12th grade, so they do not participate in the CEE or the matching process.

Other factors contributing to heterogeneous student preferences include distance to college, localized reputation, and localized employment premium differences due to alumni networks. To this end, the admissions process is administered at the province level, which means that all students within each matching game are from the same province. Therefore, they are likely to have similar distances to universities, enjoy similar employer premiums for local universities if they would like to come home after college, and have been exposed to similar localized reputations. This strengthens our confidence in the homogeneous preference assumption.

Finally, some groups of universities have rankings and reputations so similar to each other that students may be indifferent between them, or one student prefers one while a different student prefers another within the group (for example, Harvard and Yale in the U.S., or Tsinghua and Peking University in China). Thus, instead of analyzing JE over school-specific ranks, we divide universities into groups of schools, allowing for heterogeneous preferences within each group and only assuming homogeneous preferences across groups.

3.3 Constructing university rank groups

We assign universities to rank groups (RGs) based on students' revealed preferences: average admission score percentile (AASP) within the track-province-year cell, i.e. the matching game, during the current year. This allows university preference to vary over time within a province and to differ across provinces. To be consistent with the nature of higher education rankings in China, the group rank ranges are set to be wider as groups get less selective. In particular, Rank Group (RG) one includes universities with an AASP in the top 10%, and RGs two through four include universities with an AASP of 10% to 30%, 30% to 60%, and 60% to 100%, respectively.

Since the assignment of universities into RGs requires cutoffs, there are two potential concerns. The first concern is regarding discontinuity at the cutoffs. The lowest-ranked school in RG1 may be quite similar to the highest-ranked school in RG2, and students may have ambiguous preferences towards them. For example, suppose school A ranked at the 9.9th percentile and school B ranked at the 11th percentile. School A is assigned to RG1, while school B is assigned to RG2. However, some students may be indifferent between A and B, causing misassignment of JE. Therefore, we add a condition to JE in addition to $r_i < r_j$ and $T_i > T_j$: student i's admitting university must be 10% lower in ranking than that of j, $univpct_i - univpct_j > 10\%$. We also conduct analyses without this additional condition, and under a looser condition of $univpct_i - univpct_j > 5\%$.

The other concern is the sensitivity to the choices of cutoffs. To test whether our results are robust to the rank group assignment of universities, we use several other measures to define student homogeneous preferences in schools. Instead of assigning universities to four RGs using the AASP in the current year, we use the AASP from the previous year and AASPs through the entire sample period (2005-2011). Further, we use these three different ways of calculating AASPs to divide schools into eight RGs, with the ranking percentile cutoffs at 5%, 10%, 20%, 30%, 45%, 60%, and 80%.

In addition to the local revealed-preference approach, we also group universities based on the

985 Project and 211 Project university designations in a robustness check. In China, the most prestigious group is the thirty-nine 985 Project universities, which include Peking University and Tsinghua University as the widely agreed-upon top two, followed by other 211 Project universities.[18] Graduates from 985 and 211 Project universities enjoy a wage premium and are favored by employers in hiring processes (Li et al., 2012; Zheng, Zhang and Zhu, 2021). In this fixed RG assignment, we assume student preferences to be: Peking and Tsinghua universities >985 Project universities >211 Project universities >Rest of Tier I. We prefer the local revealed preference approach over the fixed RG assignment, because the former allows for provincial and time differences in student preferences while the latter does not. There are also some university outliers in the fixed RG method that could cause misassignment of JE. For example, Jinan University in Guangdong Province is not a 985 Project university yet is highly regarded by employers and students in Guangdong and nearby provinces.[19] We test whether conclusions are similar across these different ways of assigning rank groups to universities in the Robustness Checks section.

Despite the suitable institutional environment and the steps we take to make the homogeneous preference assumption plausible, our justified envy measures may still be higher than true values. For example, a high-scoring student assigned to a low-tier university could be due to a stronger local preference than other local students or due to preferences formed with limited information. We may mislabel such a student as having justified envy, while in reality that was indeed the school they wanted. That being said, our measures are still useful in comparing across JE measures and in comparing across mechanisms. It is implausible that student preferences towards universities changed systematically as their province switched to the parallel mechanism. As long as the heterogeneous preference distribution does not change simultaneously with mechanism reform, the change in justified envy we capture would be estimating the change due to the reform.

4 Data

We use a unique administrative data set of college admissions records between 2005 and 2011. It includes anonymous individual-level information on the year and province in which a student took the CEE, the CEE track (STEM or non-STEM), the CEE score, and the university they were admitted to. The data set spans across years when most provinces reformed from the IA to parallel mechanism and experimented with the number of parallel choices, allowing us to make causal inferences and compare across parallel choice bandwidths. In addition, the universe of individual records allows us to study distributional patterns.

The key explanatory variable of interest is the matching mechanism a particular province adopts in a given year. We collected information on the timing of mechanism reforms for each province as

[18] All 985 Project universities are also 211 Project universities, and all 211 Project universities are Tier I universities.

[19] For a more detailed example here, suppose that in Guangdong Province, student A is assigned to Jinan University with a CEE score of 640, and student B is assigned to a 985 Project University with a CEE score of 630. If we use the 985-211 grouping method, we would deem that student A justifiably envies student B; however, in reality, it could be that student A genuinely prefers their assigned school, and no JE is present.

well as the timing of students submitting their university preferences.[20] Since the submission timing also could influence matching results, we control for the timing reforms in all regressions. We also collected information on the number of parallel choices in the first choice band. We focus on the first choice band because most schools in Tier I are filled after the first choice band is considered and admission assignments are final.[21]

We restrict our sample to students admitted to a Tier I university, with CEE scores above the province-track-year-specific Tier I admission cutoffs. We focus on Tier I universities for two main reasons. First, only Tier I universities admit students from most provinces across years in China, which is important for us to compare the mechanisms across provinces. Second, and more importantly, to define JE, we assume that within each province, students have homogeneous preferences towards groups of universities, and this assumption is most likely satisfied among Tier I institutions. Lower-tier institution rankings are less unanimous and less stable. We also exclude students who attend a university in Hong Kong and those who were admitted through art, music, and sports programs because the CEE score was not the sole criteria for those admissions.[22] Our final national sample includes 3.73 million students; among them, 1.3 million students took the CEE before the parallel reform.

4.1 Descriptive statistics

The individual-level data summary statistics are presented in Table 1. For each student, we construct seven JE measures. Four of them correspond to the four measures discussed earlier. Whether a student has any justified envy, a binary measure, sums up to the number of blocking students in a matching. The number of unique university RGs blocked sums up to the number of blocking pairs, the number of envied students sums up to the number of blocking triplets, and adding up the tridimensional envy for each envied student over all students yields tridimensional envy in a game.

The three additional measures are the maximum degree of envy in the university RG dimension, student rank dimension, and diagonal distance dimension, which are graphically equivalent to the maximum lengths of the JE vectors in the horizontal, vertical, and diagonal directions in Figure 1. We construct these additional measures because it is conceivable that what matters to a student is the least fair case, with the largest difference in student rank, school quality, or both. The raw student rank difference and school RG difference are on very different scales. To make the two dimensions comparable to each other, we normalize student rank to a continuous scale from zero to four.[23]

[20] Details are presented in Table A1.

[21] In this context, a parallel mechanism with six parallel choices in each of three choice bands is roughly equivalent to the DA mechanism with six choices.

[22] Universities in Hong Kong often conduct an interview; art, music, and sports programs admit university-specific interviews and evaluations, which we do not observe their scores. The raw sample includes 11.74 million observations. Students with CEE scores above the Tier I admission cutoffs are 33.5% of the raw sample. Conditional on having a higher than Tier I cutoff score, about 9.12% were admitted to universities in Hong Kong or admitted through art, music, and sports programs.

[23] The mean for the raw rank difference between a student with JE and their envied students is around six thousand.

As shown by Table 1, 80.6% of students have justified envy. A high percentage of JE can be caused by one student with a very low CEE score being admitted into a top university, in which case a large number of students admitted to a lower-ranked university would have a higher CEE score and justifiably envy this one student.[24] One can easily see that, as long as one low-scoring top-admitted student exists, changing student assignments below the top rank group does not alter the number of blocking students at all. As mentioned in subsection 3.2, the homogeneous preference assumption may overestimate the justified envy, but it is extremely unlikely that all 80.6% of students deviate from common school ranking preference by an entire rank group and prefer their lower-ranked school assignment.

A student justifiably envies 357 other students on average. For JE in the dimension of the university RG admitted, the average maximum degree of envy is 1.47, with the maximum being three since there are only four RGs. Turning to the student rank dimension of JE, the lowest-ranked individual a student justifiably envies is ranked equivalent to 0.87 RGs lower than themselves on average. This is slightly smaller than that on the university RG dimension. Lastly, the diagonal degree of envy has, on average, a maximum degree of 1.74 and a sum degree of 456.

Table 2 presents the summary statistics at the matching game level, which is also the level at which we conduct our main analysis. The top panel shows characteristics of the matching game, including the mechanism and timing of submission adopted. Out of the 398 matching games at the track-province-year level, half are STEM track. The majority of the games (63.8%) adopt ex-post preference submission, and 33.7% adopt the parallel mechanism. The average game size is around 9,000, with a minimum of 66 and a maximum of 48,840.

The middle and bottom panels present the statistics for the outcome measures. Across matching games, around 78% of students have at least some justified envy, with a minimum of 36.2% and a maximum of 94.4%. The average number of BSs is around 7500, that of BPs is around 14 thousand, and the average number of BTs and the degree of TE are much larger, at around 3.3 million and 4.3 million respectively.

Moving on to the standardized justified envy measures in the bottom panel, we can read them as how far or how close these numbers are to zero, the stable matching, and one, the least stable matching. The averages for BS and BP are 0.864 and 0.687, suggesting that they are very far away from the stable matching, with standardized BS and BP being 0. However, as we mentioned above, these can be caused by just a handful of low-score students admitted to a top school. The averages for BT and TE are 0.078 and 0.049, which suggest that the matching games are much closer to the stable matching outcome than BS and BP measures do.

[24] Chen, Jiang and Kesten (2020) also construct a measure for percent of blocking students using data from one county, and they find it to be about 10%. Instead of assuming homogeneous preferences, they assign justified envy to students who have a listed college above her admitted college (within the same tier) whose cutoff score is lower than her test score, or those admitted to a Tier II college despite a score higher than Tier I cutoff. Although they were better able to rule out mistaken justified envy given data on choice order lists, their definition ignores the possibility that students often omit top schools with low success rates to best utilize the limited choice list (Chen and Pereyra, 2019). Students who do not include Peking University in their choice list could still justifiably envy others with a lower score but got admitted. The true JE percentage is likely somewhere in between their statistics and ours.

5 Empirical Strategy

Exploiting the geographic and timing variations in the college admissions mechanism reforms, we use a generalized difference-in-differences framework to identify the causal effect of changing from the IA to the parallel mechanism on JE measures.

$$\begin{aligned} Y_{jst} &= \beta_0 + \beta_1 Parallel_{jt} + \beta_2 Ex\text{-}interim_{jt} + \beta_3 Ex\text{-}post_{jt} \\ &\quad + \beta_4 N_{jst} + \alpha_j + \gamma_s + \delta_t + u_{jst} \end{aligned} \tag{1}$$

Y_{jst} indicates the set of JE measures for province j track s year t. $Parallel_{jt}$ is the independent variable of interest, indicating whether province j adopted the parallel mechanism in year t. If the parallel mechanism reduces the incidence and degree of JE, the coefficients of β_1 should be negative and statistically significant. $Ex\text{-}interim_{jt}$ and $Ex\text{-}post_{jt}$ are important control indicators for the preference submission timing. α_j, γ_s, and δ_t are province, track, and year fixed effects, controlling for provincial time-invariant characteristics, time-invariant differences between tracks, and nationwide yearly shocks. we also control for the matching game size N_{jst} that varies across time within province and track. The standard errors are clustered at the provincial level.

Since this is a two-way fixed effects model with switchers at different timings, the estimated average treatment effect is a weighted sum of that in each group and period. We follow recent econometric literature and use methods developed by de Chaisemartin and d'Haultfoeuille (2020) to address the negative weights and follow Goodman-Bacon (2021) to see how the comparisons between the switchers, the always treated, and the never treated contribute to the average treatment effect. Results are discussed in the robustness check subsection.

While it is important to find out the average effects on JE, it is also worth investigating whether the effects persist over time. We adopt an event study model and add a set of dummy variables to indicate the number of years before or after the college admissions reform to the parallel mechanism. The regression model is shown in Equation 2.

$$\begin{aligned} Y_{jst} &= \beta_0 + \sum_{k=-9}^{+9} \beta_1^k D_{jt}^k + \beta_2 Ex\text{-}interim_{jt} + \beta_3 Ex\text{-}post_{jt} \\ &\quad + \beta_4 N_{jst} + \alpha_j + \gamma_s + \delta_t + u_{jst} \end{aligned} \tag{2}$$

D_{jt}^k is a dummy variable set to 1 if province j experienced the reform in year $t - m$ and 0 otherwise. If province j shifted from the IA to the parallel mechanism k years before year t, k is positive. Otherwise, k is negative. The omitted year is $m = 0$, that is, the last year province j used the IA mechanism. This exercise also allows us to check if the parallel trend assumption in the difference-in-differences estimation is satisfied.

Finally, we also look at whether the effect sizes are larger for provinces that adopted a larger number of parallel choices in their first choice band. As shown in Chen and Kesten (2017), stability should improve as the number of choices in each parallel choice band increases. They show theoretically that a wider parallel choice band should result in more stable matching outcomes: a

parallel mechanism ϕ^6 with six parallel choices is more stable than a parallel mechanism ϕ^3 with three parallel choices. "More stable" is defined such that given any matching game, if ϕ^6 produces an unstable matching then so does ϕ^3; but not the other way around. This finding on the binary stability measure does not always extend to the comparisons of intensive margin stability across mechanisms (Bonkoungou and Nesterov, 2020). Given that in large matching games in real life, it is unlikely to achieve a stable matching outcome, intensive margin comparison is important. Further research is needed to understand intensive margin comparisons across mechanisms.

To explore the relationship between the effect sizes on JE and parallel choice bandwidth, we replace the parallel dummy in Equation 1 with several dummy variables, each indicating a specific number of parallel choices in the first parallel choice band. Since provinces adopted three to six parallel choices in the first parallel choice band in our sample period, we compare the matching outcomes in these province-years with those that used IA for their college admissions. Similar controls to Equation 1 are included.

$$\begin{aligned} Y_{jst} &= \beta_0 + \beta_1\phi^3_{jt} + \beta_2\phi^4_{jt} + \beta_3\phi^5_{jt} + \beta_4\phi^6_{jt} + \beta_5 Ex\text{-}interim_{jt} + \beta_6 Ex\text{-}post_{jt} \\ &\quad + \beta_7 N_{jst} + \alpha_j + \gamma_s + \delta_t + u_{jst} \end{aligned} \tag{3}$$

6 Effects of Parallel Reform on Justified Envy Measures

6.1 Average treatment effects

Table 3 presents the main results from the generalized difference-in-differences regressions specified by Equation 1 on various measures of JE. Panel A presents the results on non-standardized JE measures. The first column shows that changing from IA to the parallel mechanism reduced the number of blocking students (BSs) by 489, or 6.5% of the sample mean (hereafter relative effect). The reform also reduced the number of blocking pairs (BPs) by 4027, or a relative effect of 29.4%. The improvement in terms of relative effect is even larger for the number of blocking triplets (BTs), with a reduction of 2.26 million BTs, 67.5% of the sample mean. Column (4) shows the results for tridimensional envy, which captures how much better of a university the envied students got admitted to, how much lower the envied students ranked in their CEE scores, and how many BTs there are. The reduction in this most comprehensive JE measure is around 3.2 million, 74.9% of the sample mean.

Panel B presents the results for standardized JE measures, dividing the raw measures by the measures in the worst matching under the given matching problem. Even though the coefficients are quite different between the two panels, the relative effect sizes are very similar: around 8% for BS, 26% for BP, 56% for BT, and 61% for TE. Overall, BT and TE suggest that the Chinese college admissions outcomes are much further away from the worst matching outcome than BS and BP would suggest, and the effects of the parallel reform are much larger for BT and TE.

As mentioned earlier in subsection 3.3, we test the sensitivity of results to alternative ways of categorizing universities into rank groups, including varying the cutoff score from current year

to the previous year or the sample average, varying from four rank groups to eight rank groups, and using a fixed rank group based on 985 and 211 project universities. Results are quite robust as presented in Table 4. The relative effect sizes are very similar when we use alternative local revealed preference-based approaches, as shown by Panels A through E. When categorizing schools into eight rank groups, the relative effect sizes on standardized BS are slightly smaller, while the other effect sizes remain similar. Panel F shows that the justified envy measures are on average slightly smaller on standardized BP when we use the fixed tiers based on 985 and 211 project universities. The relative effect sizes on BT and TE are mostly comparable with the other rank group constructions. Despite these minor magnitude differences from the main results, it remains true that BT and TE suggest that the matchings are much further away from the worst matching outcome than BS and BP would, and the effects are much larger for BT and TE as well.

We also test the robustness of the RG cutoff restrictions. To account for indifference around the RG cutoffs, say between the top university in RG_2 and the lowest-ranked university in RG_1, we have required that the percentile rank of the blocked university and the admitted university has to be greater than 10%. Here, we test whether results would differ if we do not impose this condition or loosen this condition. Table 5 presents the results using data constructed without this condition, or with a similar condition but only requiring a 5% university rank difference. Overall, results are similar to the main result in Table 3 and consistently support our hypothesis: compared with IA, the parallel mechanism significantly improved matching stability on the intensive margin by reducing JE.

Finally, we also conduct the analysis using individual-level data. The regression specifications are almost identical to the game level, except that we add a control for individual student score percentile. In Table 6, columns (1) to (4) present results for the four outcome measures corresponding to our main results at the game level. The parallel reform reduced the likelihood of having any justified envy by about 4.61% (column 1) and reduced the number of envied students by 208 (column 3). The relative effect sizes compared to the sample mean are very similar to the main results. The relative effect is smallest for the binary measure of any justified envy, corresponding to the number of blocking students. The relative effect magnitude is much larger, at around 60% of the sample mean, for the number of envied students and tridimensional envy, corresponding to the number of blocking triplets and tridimensional envy at the game level. The three additional outcomes in columns 5 through 7 measure the maximum degree of envy in either or both dimensions: school RG and student rank. The relative effect magnitudes are around 21 to 28% of the sample mean, comparable to that of the unique school tiers blocked, which corresponds to the number of blocking pairs at the game level.

6.2 Effects over time: event study results

The eight graphs in Figure 3 present the event study results from Equation 2 on the four raw and four standardized JE measures. Each figure plots the point estimates and the 95% confidence intervals of the relative year dummies. The graphs show that there are no pre-trends prior to the

parallel mechanism reform. Overall, the coefficients of the first three to four years after mechanism change are similar to the average effects found in the difference-in-differences specification, reflecting the fact that many provinces adopted the parallel mechanism in 2008 or 2009, giving us the most statistical power for three or four years post-reform. Starting from the fifth year post-reform, the effect sizes generally increase and, in blocking triplets' case, continue to increase in the eighth and ninth years. Note that these longer-term effect estimates come from a few provinces that adopted reforms earlier than the rest, limiting our ability to make definitive claims on whether the effect strengthens over time. Overall, these results support that we do not observe significant pre-trends and that the effects of the mechanism change persist.

6.3 Effects and number of parallel choices

Table 7 shows the results from estimating Equation 3, comparing the effect of adopting different numbers of parallel choices in the first choice band. As discussed in section 5, Chen and Kesten (2017) shows that stability, defined in a binary sense, should improve as the number of choices in each parallel choice band increases. While this may not extend to the intensive margin comparisons, we start with the analysis with an expectation that more parallel choices could perhaps result in larger reductions in JE.[25]

Despite the raw and standardized JE measures showing similar results so far, results here differ when we consider raw or standardized outcomes. For example, the coefficients of BT and TE are larger for ϕ^3 than ϕ^4 for the raw measures, but the opposite is true for the standardized measures. This is because we are particularly interested in comparing outcomes in different provinces that adopted different numbers of parallel choices here, while they also differ in the number of students. We prefer the set of results using the standardized measures, since we account for the nonlinear relationship between the number of participating students and JE measures.

In columns (5) to (8) showing results for standardized measures, all but one coefficient are statistically significant and negative. The only statistically insignificant coefficient is on ϕ^3 when the outcome variable is the standardized BS. Interestingly, the effect size does not always linearly increase with choice bandwidth or even increase monotonically. For a linear relationship, the coefficient for ϕ^6 should be double the size of that for ϕ^3, yet this is not supported by our results. In fact, changing from IA (ϕ^1) to ϕ^3 decreases the standardized BP by 0.13, but the effect of changing to ϕ^6 is only 0.11; the two are not statistically significantly different. In general, the coefficients increases with the number of parallel choices from three to five. But matching games adopting six parallel choices show surprisingly weaker effects for several JE outcomes, especially when compared with ϕ^5 on the standardized BS and BP. Adopting ϕ^6 produces a similar or larger

[25] We test the pre-reform characteristics of the reform provinces to see if those with more justified envy are likely to adopt a wider choice band. Results are presented in Table A7.First, it seems that provinces with a larger number of students are more likely to adopt three parallel choices as opposed to four or six parallel choices. Second, while the raw JE measures differ for provinces adopting various numbers of choices, the standardized justified envy measures are not statistically significantly different. Note that the standardized justified envy measures consider the game size and account for the nonlinear relationship between JE and game size, and thus are our preferred measures.

reduction in the number of students one justifiably envies and the total degree of envy in both dimensions.

The provinces that changed to ϕ^5 seem to have experienced the largest effects in BS and BP. To explore potential explanations, we compare how provinces with different parallel choice bandwidths differ in their game size, changes in game size over the sample period, the timing of preference submissions, variations in university cutoff scores across years, and the number of Tier I universities admitting any students from the province. We find that provinces with five parallel choices are more likely to have also adopted the ex-interim or ex-post preference submission, which suggests that there could be a compounding effect between more complete information on priority order and the mechanism change. This naturally leads to an empirical test for regressing JE measures on the controls and a full set of interaction terms between mechanism and submission timing. However, we do not find statistically significant results suggesting that a combination of parallel with ex-interim or ex-post submission reduces JE much more than a combination of parallel with ex-ante submission.[26]

6.4 Robustness Checks

Average treatment effect: negative weights and decomposition

As discussed in section 5, we follow recent econometric literature to test whether our results are sensitive to addressing potential negative weights in the two-way fixed effects model using methods developed by de Chaisemartin and d'Haultfoeuille (2020). We use the Stata command *did_multiplegt* to implement the model developed in de Chaisemartin and d'Haultfoeuille (2020), and results are presented in Panel A in Table 8. None of the placebo tests turn out to be significant, which we omit from the table but can be found in the appendix Table A4. The estimates here are similar in magnitude to and not statistically different from those in the main results.

We also implement the decomposition of our average treatment effect estimate (Goodman-Bacon, 2021) to understand how the comparisons between the switchers, the always treated, and the never treated contribute to the average treatment effect.[27] The comparison between the always treated and switchers carries a weight of merely 0.08, and that between the always treated and the never treated carries a weight smaller than 0.005. The weights are mainly distributed between the other two comparisons: within switchers, and between the never treated and switchers. The effects estimated from comparison within switchers are often smaller than those from comparing the never treated and switchers. Such difference is larger for the blocking students and much smaller for the other three JE measures.

Controlling for dosage effect of the ex-post submission reform

As we have discussed in the policy background section, during the parallel reform process, some provinces also switched the timing of preference submission from ex-ante and ex-interim to ex-post. Provinces differ in their processes to adopt this reform. Even though we have already controlled

[26] See Table A6 for the detailed results.

[27] *bacondecomp* is used to implement the decomposition. Detailed results are presented in Table A5.

for the type of preference submission timing in our model, the parallel coefficient may capture the dosage effect of the ex-post submission reform. To take into account the cumulative effect of ex-post submission reform, we add a variable in our regression model to control for how many years it has been since each province adopted the ex-post preference submission reform. We report the results in Panel B in Table 8. Adding a dosage control does not affect our main results, that the parallel reform reduced JE, and the relative comparisons across columns stay the same.

Controlling for partial parallel reform

At the beginning of the reform period, some provinces adopted a partial parallel matching mechanism. For example, they continued to use the IA mechanism for students' first choice for Tier I admission while allowing for a parallel mechanism for the second and third choices of Tier I, or for all the choices of Tier II admission. In our main analysis, we only consider the matching mechanism for the first choice band in Tier I admission, since the first choice band is the most important for admissions to Tier I universities. To test if the partial parallel reforms in lower tier or higher choice bands matter, we add a binary control variable in our regression to indicate whether the province adopted a partial parallel mechanism. Panel C in Table 8 shows that results remain robust.

Tier 0 and Early Admissions

In China, before the admissions process of Tier I universities occurs, military programs, police training programs, sports-related programs, and government-funded teacher training programs conduct their admission processes. Although only a very small percentage of high-achieving students choose to participate in this Tier 0 admissions process, the Tier 0 matching mechanism may influence the matching outcomes in Tier I admission. We collect information on how each province conducted its Tier 0 admissions during the period from 2005 to 2011. Anhui Province was the only province that experienced changes in its Tier 0 matching mechanism, switching from the IA mechanism to the parallel mechanism in either 2008 or 2009.[28] We rerun our main regressions without Anhui province and show the results in Panel D of Table 8. They remain similar to the main results.

Major-university trade-off

Since Chinese students need to choose both university and major at the time of choice submission, one may worry that the major-university trade-off could influence measures of JE. For example, University B may be ranked lower than University A on average, but University B may have a very popular major that has a high cutoff score. If a student is admitted to this most-popular major in University B, they may not envy another student in University A. To test how much the outlier majors could be driving our results, we conduct two robustness checks. First, we drop all individuals admitted to the majors with the highest and lowest average CEE scores from each university and then reconstruct the JE measures. Results are presented in Panel E of

[28] Jiangsu Province adopted the parallel mechanism for Tier 0 admissions during the entire sample period.

Table 8. The magnitudes of coefficients are slightly smaller than those in the main results, but the relative effect sizes are almost all quite similar. The only slightly different coefficient is the one for the standardized BS, with a relative effect size of 14% of the mean, slightly larger than that in the main result, 8% of the mean.

Second, we reconstruct the justified envy measure to allow students to have flexible preferences over university-major combinations, instead of preferences over universities. We then use the lowest score for each university-major to measure its popularity and divide them into four rank groups, similar to our main analysis. The results are presented in Table A8. Again, the relative effect sizes compared with sample means for all the justified envy measures are similar to our main results.

7 Distribution of Justified Envy across Student Performance

The universe of admissions data at the individual level allows us to study the distribution of justified envy across student performance, as well as the heterogeneous effects of the parallel mechanism change on students in different parts of the distribution.

7.1 Hump-shaped distributions

In Figure 4, we collapse the individual-level data to the percentile-mechanism level to plot the distributions of JE measures over student performance and compare across two mechanisms. In all eight figures, we plot two hump-shaped curves, one for IA and one for parallel. The hump shape suggests that those with the highest and lowest CEE scores among those admitted by Tier I schools are less likely to have JE. It is easy to understand a lower level of JE for low-performing students because, by definition, they are less likely to encounter another student with a lower test score than themselves to justifiably envy them. The more interesting part is that the high-scoring students have a lot of "potential" to envy others if the assignment is unfair, however, they fare better than the medium-ranked students.

7.2 Discussion on potential explanations

One possible explanation for the above phenomenon is that the highest-ranked students face less congestion and less uncertainty: they know more precisely their chances of being admitted to the best colleges. The medium-ranked students have a harder time gauging their chances of getting the last few spots left in the top colleges versus the risk of not being admitted to any of their parallel choices or their only choice in the IA mechanism. This is related to how many universities may be horizontally similar to one another, which creates congestion. In particular, it could be that students' preferences for top universities are clearer, while students may be indifferent towards a group of mid-ranked universities. The limited number of choices in the parallel mechanism and the lack of coordination on which mid-ranked universities to include in choice lists lead to larger variations in the cutoff scores among mid-ranked universities.

An anecdotal phenomenon called "big-small-years" portrays the struggle some students face when deciding which mid-ranked universities to include in their choice lists. It describes the fact that some universities have surprisingly high or low cutoff scores. Quite a few universities are of similar rankings; when many students choose one particular university, it results in a high admission cutoff, "a big year" for that university. Next year, students may want to avoid this university and all choose other similarly ranked universities, resulting in low demand and a low cutoff –"a small year" for that university.

If congestion explains the distribution of JE, we should expect to see greater variation in admission cutoffs for the mid-ranked universities and less variation for the highest and lowest ranked universities in our data. Indeed, we find such a relationship in Figure 5, which plots the relationship between university selectivity on the X-axis using average cutoff scores and the variability of the cutoff scores on the Y-axis using university-level data. Each dot on the scatter plots marks one university's average and standard deviation of its cutoff scores between 2005 and 2011. The left panel shows the overall data. Some may worry that the low standard deviation in the right tail may be caused by low-ranked universities coming in and out of the unbalanced sample. Reassuringly, when we restrict to a balanced panel of universities as shown on the right, the aforementioned pattern still holds.

Another potential explanation for the hump-shaped pattern is the distribution of sophistication types over student performance. If top students have a more compressed and systematically higher distribution of strategic ability in the choice listing process, whereas the students ranked in the middle have a wider range and a lower average sophistication level, this could also generate a higher level of JE for middle-ranked students than top students.

7.3 Heterogeneous effects of mechanism change on JE across distributions

Turning to the comparison between two lines within graphs, we can see that the IA mechanism produces a higher level of JE for most measures at all student performance levels than the parallel mechanism. We estimate the statistical significance of the difference and plot the coefficients in Figure A2. Most of the time, the gaps between IA and the parallel mechanism in Figure 4 are statistically significant. The only insignificant differences occur over students in the lower and middle percentiles for the number of blocking students. For BS, changing from IA to the parallel mechanism benefits mostly the top 30% of the Tier I students. For BP, the parallel mechanism improves this metric for most students. For BT and TE, both mechanisms produce a normal distribution and the improvement also follows a roughly normal distribution. This suggests that the improvement on the intensive margin is larger for students in the middle, who had higher initial JE.

8 Conclusions

Using millions of individual-level data on Chinese college admissions, this paper provides novel empirical evidence on stability comparisons on the intensive margin. We find that changing from the IA to the parallel mechanism reduced (1) the number of blocking students (BS), which corresponds to the proportion of students with any JE, (2) the number of blocking pairs (BP), a classic theoretical concept in matching stability, (3) the number of blocking triplets (BT), which sums the number of students each student justifiably envies over all students, and (4) tridimensional envy (TE), a novel measure accounting for the degree of envy in student rank difference, school rank difference, and how many students one envies. These measures do not have a monotonic relationship: one measure increasing does not imply another measure would also increase. BS and BP capture less granular information from a matching and paint a more pessimistic picture of matching outcomes in Chinese college admissions, suggesting that the outcomes are closer to the least stable matching. BT and TE, on the other hand, suggest that the matching outcomes are far from the least stable matching. Utilizing the staggered reform from the IA to parallel mechanism at the province level, we estimate that the reduction in BS is relatively small at 7.1 percentage points (8.3%), the reduction in BP is moderate, at 26% of the sample mean, while the reduction in BT and TE is more substantial, at around 56% to 60% of the sample mean. Another intriguing finding is that the reduction in JE measures does not always increase with the number of parallel choices in the first choice band. Finally, distributional analyses show higher levels of JE and larger improvements from the mechanism reform for students ranked in the middle.

This paper leaves some unanswered questions. Fairness in this paper is only considered in terms of college admissions matching priority, which is based on test scores. One might argue that the parallel reform may have worsened income inequality despite improved test-score-based justified envy measures if test scores and socioeconomic status are highly correlated. However, the upside of the parallel reform is that it eased the burden of strategic choice in the college admissions process. The more advantaged students were better able to make strategic choice lists under the IA mechanism and thus placed better given the same test score. Future research should explore whether the parallel mechanism, while improving meritocratic fairness, also improved social equity.

Despite the suitable institutional environment and steps we take to make the homogeneous preference assumption plausible, the justified envy measures here could be overestimated. Although we are confident that the estimated effects on JE are most likely unbiased, since the heterogeneous preference distribution is unlikely to change simultaneously with the mechanism reform, the JE measures themselves are likely larger than the true values. In addition, we cannot disentangle the direct improvement from matching mechanisms and the indirect improvement from students changing their preference submissions. Due to a lack of data on students' ranked order lists, true preferences, and abilities, we cannot speak to strategy-proofness, efficiency, or ex-ante fairness. Future studies with richer data on students' lists of choices before and after the mechanism change, along with measures of students' true abilities, would be able to answer these important questions. To our knowledge, such data does not exist at the national level, thus these questions may be better

explored in future studies combining survey data and administrative data at a smaller scale or in a laboratory setting.

Finally, the paper provides a myriad of empirical findings that do not yet have their theoretical counterparts or clear explanations. Are there conditions under which we can compare mechanisms' stability on the intensive margin? Why does the improvement in justified envy measures not increase with parallel choice bands? Why do students in the middle of the distribution experience higher levels of justified envy than the students at the top? These empirical findings from the arguably largest matching market in the world will hopefully inspire further explorations.

Acknowledgment

We would like to thank Alper Arslan, Caterina Calsamiglia, Yan Chen, Isla Globus-Harris, Aram Grigoryan, David Ong, Emil Temnyalov, Binzhen Wu, Jie Zheng, and conference and seminar participants at 6th IZA Workshop on Economics of Education, ASSA 2021, Western Economic Association 2020 Annual Meeting, Southern Economic Association 2020 Annual Meeting, 12th International Symposium on Human Capital and Labor Markets, Colgate University, Jinan University, and SUNY Buffalo for helpful comments. We thank Shuai Zhao, Chenggang Hu, and Yuhao Deng for their research assistance. Wei Ha gratefully acknowledges funding from the National Natural Science Foundation of China (Grant 72074010).

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *American Economic Review*, 93(3): 729–747.

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak. 2017. "The welfare effects of coordinated assignment: Evidence from the New York City high school match." *American Economic Review*, 107(12): 3635–89.

Abdulkadiroğlu, Atila, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. 2020. "Efficiency, justified envy, and incentives in priority-based matching." *American Economic Review: Insights*, 2(4): 425–42.

Agarwal, Nikhil, and Paulo Somaini. 2018. "Demand Analysis using Strategic Reports: An application to a school choice mechanism." *Econometrica*, 86(2): 391–444.

Akbarpour, Mohammad, Adam Kapor, Christopher Neilson, Winnie van Dijk, and Seth Zimmerman. 2022. "Centralized School choice with unequal outside options." *Journal of Public Economics*, 210: 104644.

Bleemer, Zachary. 2022. "Affirmative action, mismatch, and economic mobility after Californiaâs Proposition 209." *The Quarterly Journal of Economics*, 137(1): 115–160.

Bonkoungou, Somouaoga, and Alexander Nesterov. 2020. "Reforms meet fairness concerns in school and college admissions." *Higher School of Economics Research Paper No. WP BRP*, 234.

Calsamglia, Caterina, Chao Fu, and Maia Güell. 2020. "Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives." *Journal of Political Economy*, 238(2): 642–680.

Chen, Li, and Juan Sebastián Pereyra. 2019. "Self-selection in school choice." *Games and Economic Behavior*, 117: 59–81.

Chen, Yan, and Onur Kesten. 2017. "Chinese college admissions and school choice reforms: A theoretical analysis." *Journal of Political Economy*, 125(1): 99–139.

Chen, Yan, and Onur Kesten. 2019. "Chinese college admissions and school choice reforms: An experimental study." *Games and Economic Behavior*, 115: 83–100.

Chen, Yan, Ming Jiang, and Onur Kesten. 2020. "An empirical evaluation of Chinese college admissions reforms through a natural experiment." *Proceedings of the National Academy of Sciences*.

de Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–2996.

De Haan, Monique, Pieter A Gautier, Hessel Oosterbeek, and Bas Van der Klaauw. 2021. "The performance of school assignment mechanisms in practice."

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2020. "The consequences of academic match between students and colleges." *Journal of Human Resources*, 55(3): 767–808.

Dogan, Battal, and Lars Ehlers. 2021. "Minimally Unstable Pareto Improvements over Deferred Acceptance." *Theoretical Economics*, 16(4): 1249–1279.

Dogan, Battal, and Lars Ehlers. 2022. "Robust Minimal Instability of the Top Trading Cycles Mechanism." *American Economic Journal: Microeconomics*.

Fack, Gabrielle, Julien Grenet, and Yinghua He. 2019. "Beyond truth-telling: preference estimation with centralized school choice." *American Economic Review*, 109(4): 1486–1529.

Gale, David, and Lloyd S Shapley. 1962. "College admissions and the stability of marriage." *American Mathematical Monthly*, 69(1): 9–15.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*.

Ha, Wei, Le Kang, and Yang Song. 2020. "College matching mechanisms and matching stability: Evidence from a natural experiment in China." *Journal of Economic Behavior & Organization*, 175: 206–226.

Kang, Le, Wei Ha, Yang Song, and Sen Zhou. 2020. "Matching Mechanisms, Justified Envy, and College Admissions Outcomes."

Kapor, Adam, Christopher A Neilson, and Seth D Zimmerman. 2018. "Heterogeneous beliefs and school choice mechanisms." National Bureau of Economic Research.

Kwon, Hyukjun, and Ran I Shorrer. 2020. "Justified-Envy-Minimal Efficient Mechanisms for Priority-Based Matching." *Available at SSRN 3495266*.

Levitt, Steven D, and John A List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic perspectives*, 21(2): 153–174.

Li, Bingbing, Stephanie Wang, and Xiaohan Zhong. 2016. "Timing is everything? An experimental investigation of incomplete information in centralized matching mechanisms." *Manuscript, Univ. Pittsburgh. http://www. pitt. edu/˜ swwang/papers/Matching. pdf.*

Lien, Jaimie W, Jie Zheng, and Xiaohan Zhong. 2016. "Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory." *Experimental Economics*, 19(1): 116–150.

Lien, Jaimie W, Jie Zheng, and Xiaohan Zhong. 2017. "Ex-ante fairness in the Boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission." *Games and Economic Behavior*, 101: 98–120.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu. 2012. "Does attending elite colleges pay in China?" *Journal of Comparative Economics*, 40(1): 78–88.

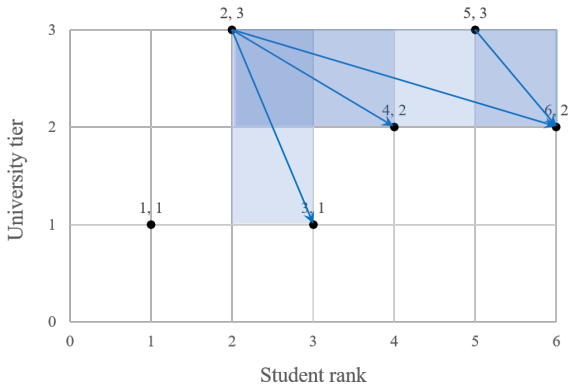
Romm, Assaf, Alvin E Roth, and Ran I Shorrer. 2020. "Stability vs. no justified envy."

Terrier, Camille, Parag A Pathak, and Kevin Ren. 2021. "From immediate acceptance to deferred acceptance: effects on school admissions and achievement in England." National Bureau of Economic Research.

Wu, Binzhen, and Xiaohan Zhong. 2014. "Matching mechanisms and matching quality: Evidence from a top university in China." *Games and Economic Behavior*, 84: 196–215.

Zheng, Yanqiao, Xiaoqi Zhang, and Yu Zhu. 2021. "Overeducation, major mismatch, and return to higher education tiers: Evidence from novel data source of a major online recruitment platform in China." *China Economic Review*, 66: 101584.

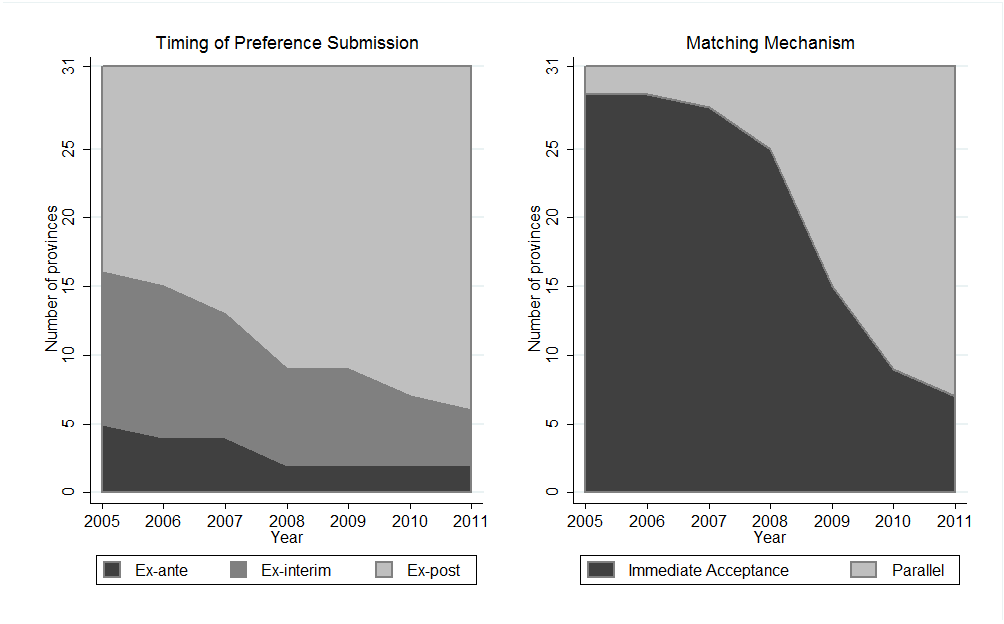
Figure 1: Illustrative Example for Justified Envy Measures



	BS = 2	BP = 3	BT = 4	TE = 10
STD:	BS = 0.5	BP = 0.5	BT =0.33	TE = 0.278

Notes: The figure illustrates one matching outcome in a game with three universities and six students. Smaller numbers indicate higher ranks. The shaded areas indicate the areas with justified envy. The arrows point from the students with justified envy to the student(s) they justifiably envy. BS: # blocking students, BP: # of blocking pairs, BT: # of blocking triplets, TE: tridimensional envy measure. The second row presents standardized measures, equaling to the measure divided by that measure in the most unfair matching outcome given the matching game.

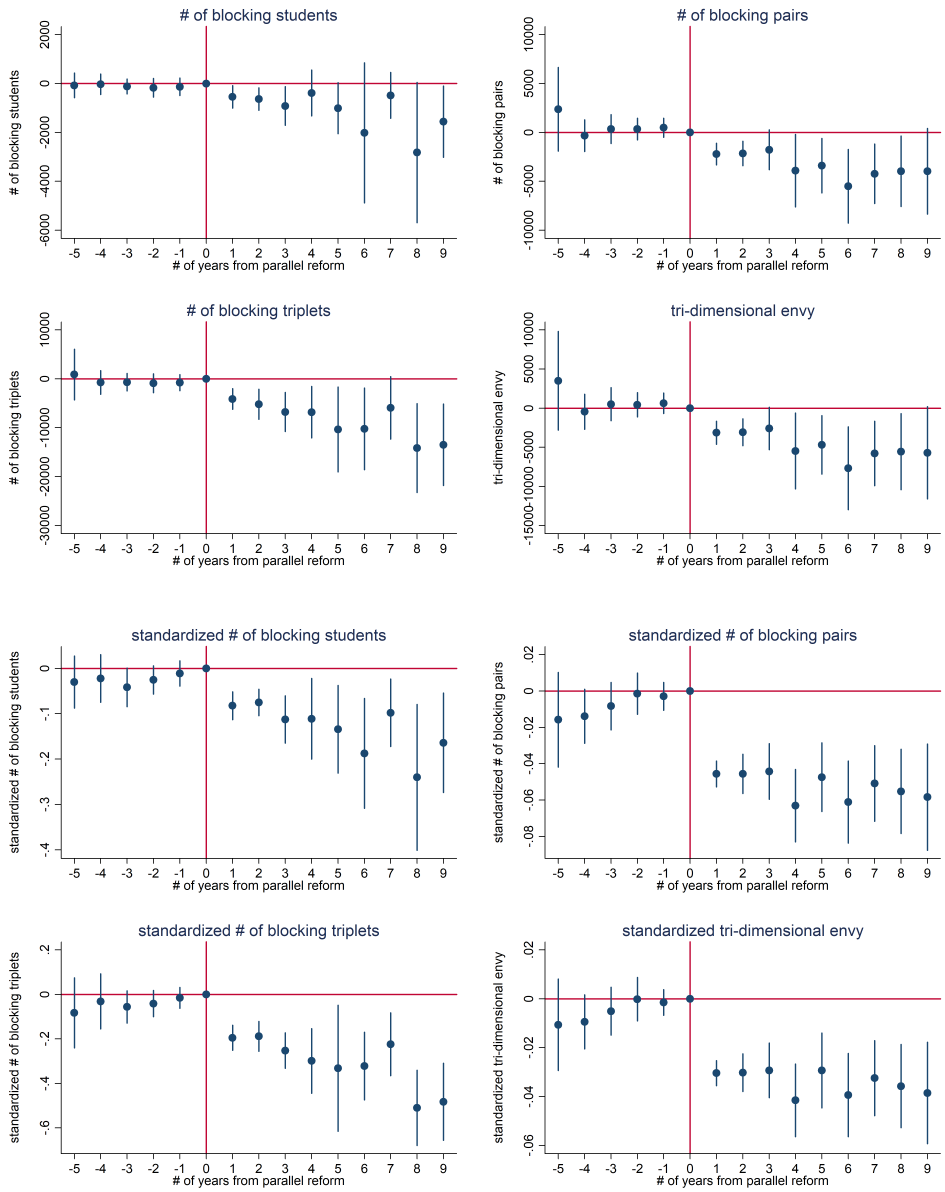
Figure 2: Provincial Variations in the Timing of College Admissions Reform



Notes: The left panel shows the number of provinces that adopted choice submission after knowing test scores (Ex-post), choice submission after the exam without knowing test scores (Ex-interim), and choice submission before the exam (Ex-ante) across years. The right panel shows the number of provinces that used the IA versus parallel mechanism in our sample.

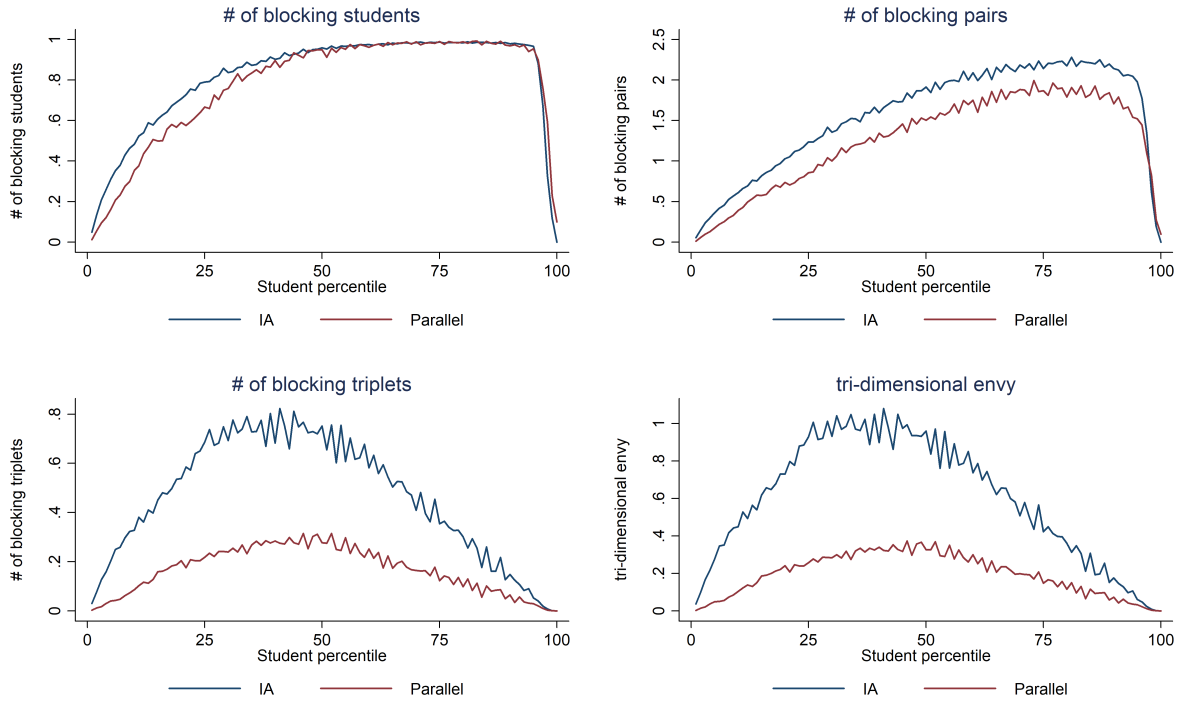
Source: Ha, Kang and Song (2020).

Figure 3: Event Study Results



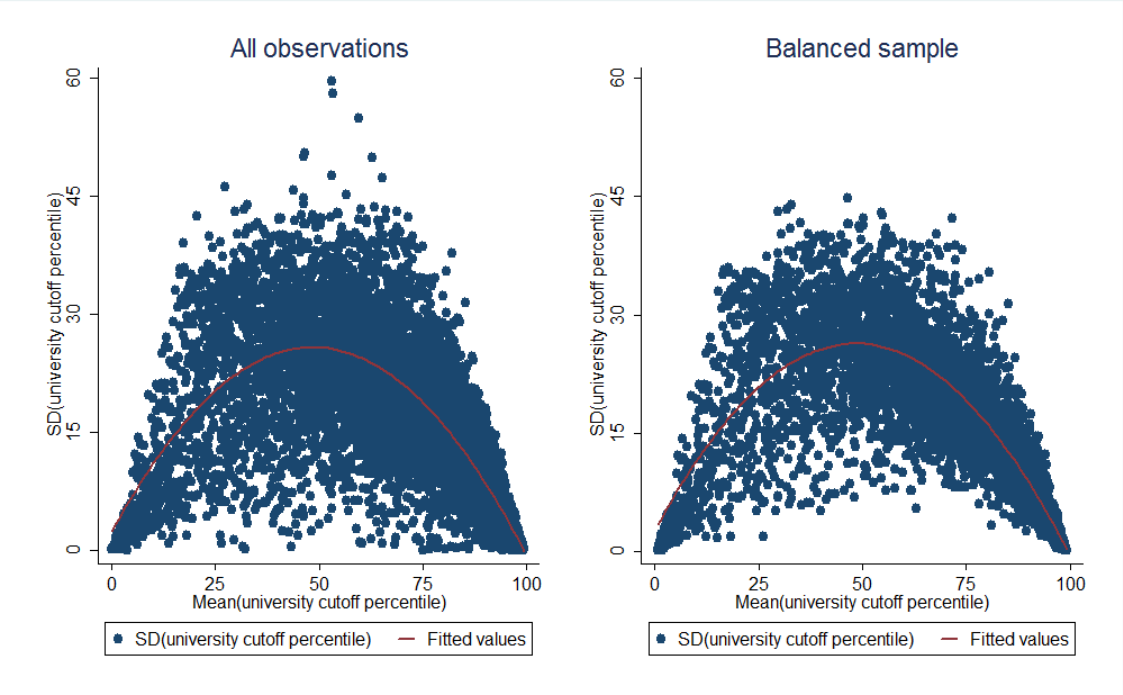
Notes: These figures present the event study coefficients of Equation 2 for the eight outcomes as labeled in the figures, with the top four for raw justified envy measures and the bottom four for standardized justified envy measures. Each figure plots the point estimates and the 95% confidence intervals of the relative year dummies. The graphs show that there are no pre-trends prior to the college admissions reform.

Figure 4: Distribution of Justified Envy



Notes: These figures present the average of various JE measures along the CEE score distribution for matching games with IA versus parallel mechanisms. A lower percentile means a higher CEE score. The sample is limited to those admitted to a Tier I university with a CEE score above the Tier I cutoff score.

Figure 5: Distribution of Congestion at the University Level



Notes: These figures present the relationship between the standard deviation and mean of university cutoff score percentile across the sample years. The left figure shows data for all universities that admit students in Tier I admissions; the right figure shows data for universities that consistently admit students in Tier I admissions between 2005 and 2011. They show that universities that are ranked the highest and lowest among the Tier I universities have low variations of cutoff score changes across years, indicating less congestion for students whose scores fall into these ranges.

Table 1: Summary Statistics at the Individual Level

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
year	3.730e+06	2,008	1.933	2,005	2,011
parallel	3.730e+06	0.373	0.484	0	1
ex-ante preference submission	3.730e+06	0.0827	0.275	0	1
ex-interim preference submission	3.730e+06	0.249	0.433	0	1
ex-post preference submission	3.730e+06	0.668	0.471	0	1
STEM track	3.730e+06	0.816	0.388	0	1
any justified envy (0/1)	3.730e+06	0.806	0.396	0	1
# of unique tiers blocked	3.730e+06	1.463	0.957	0	3
# of envied students (thousand)	3.730e+06	0.357	0.952	0	38.41
max envy on school tier dimension	3.730e+06	1.467	0.959	0	3
max envy on student rank dimension	3.730e+06	0.870	0.740	0	2.966
max tridimensional envy per triplet	3.730e+06	1.742	1.032	0	4.196
sum tridimensional envy (thousand)	3.730e+06	0.456	1.327	0	82.89

Notes: This table reports summary statistics at the individual level for all students admitted to Tier I universities with a score higher than Tier I cutoff between 2005 and 2011. More details on the definition of justified envy measures can be found in section 2, and details on constructing the measures in the Chinese context can be found in section 3.

Table 2: Summary Statistics at the Game level

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
number of participants	398	9,371	8,817	66	48,840
year	398	2,008	1.983	2,005	2,011
parallel	398	0.337	0.473	0	1
ex-ante preference submission	398	0.0955	0.294	0	1
ex-interim preference submission	398	0.266	0.443	0	1
ex-post preference submission	398	0.638	0.481	0	1
STEM track	398	0.500	0.501	0	1
Justified envy measures:					
% of blocking students	398	0.784	0.0945	0.362	0.944
# of blocking students	398	7,549	7,327	25	41,229
# of blocking pairs	398	13,706	14,691	30	100,023
# of blocking triplets (thousand)	398	3,347	6,346	0.085	53,102
tridimensional envy (thousand)	398	4,273	8,272	0.101	65,770
Standardized JE measures:					
Standardized # of blocking students	398	0.864	0.0973	0.403	0.990
Standardized # of blocking pairs	398	0.687	0.183	0.222	0.984
Standardized # of blocking triplets	398	0.0784	0.0373	0.0106	0.211
Standardized tridimensional envy	398	0.0489	0.0253	0.00597	0.157

Notes: This table reports summary statistics of game-level statistics for Tier I college admissions. Each track-province-year consist of a separate matching game. More details on the definition of justified envy measures can be found in section 2, and details on constructing the measures in the Chinese context can be found in section 3.

Table 3: Does Changing from IA to Parallel Mechanism Reduce Justified Envy?

VARIABLES	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tridimensional envy (1000)
		Not standardized		
parallel	-488.6**	-4,027***	-2,261***	-3,202***
	(217.9)	(1,135)	(543.3)	(745.0)
Observations	398	398	398	398
R-squared	0.984	0.926	0.796	0.787
Mean	7549	13706	3347	4273
SD	7327	14691	6346	8272
effect/mean	-0.0647	-0.294	-0.676	-0.749
effect/sd	-0.0667	-0.274	-0.356	-0.387
		Standardized outcome measures		
parallel	-0.0713***	-0.176***	-0.0440***	-0.0295***
	(0.0144)	(0.0257)	(0.00374)	(0.00276)
Observations	398	398	398	398
R-squared	0.531	0.610	0.796	0.769
Mean	0.864	0.687	0.0784	0.0489
SD	0.0973	0.183	0.0373	0.0253
effect/mean	-0.0825	-0.256	-0.562	-0.605
effect/sd	-0.733	-0.962	-1.180	-1.165

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression using the sample of 398 matching games at the track-year-province level in Tier I college admissions. $N = 398$ in all regressions. Controls include track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Results using Alternative Ways to Assign Universities into Rank Groups

Dependant variable	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tridimensional envy (1000)	(5) standardized # of blocking students	(6) standardized # of blocking pairs	(7) standardized # of blocking triplets	(8) standardized tridimensional envy
Panel A. 4 Tiers Based on Previous Year's School Rank ($N = 340$)								
parallel	-459.0**	-4,366***	-2,186***	-3,405***	-0.0463***	-0.173***	-0.0389***	-0.0306***
	(212.2)	(1,130)	(632.7)	(940.2)	(0.0122)	(0.0304)	(0.00586)	(0.00477)
Mean	7805	15692	3697	5062	0.919	0.812	0.106	0.0716
effect/mean	-0.0588	-0.278	-0.591	-0.673	-0.0503	-0.213	-0.368	-0.427
Panel B. 4 Tier Based on Average Rank 2005-2011 ($N = 398$)								
parallel	-589.0***	-5,024***	-2,287***	-3,470***	-0.0616***	-0.188***	-0.0444***	-0.0321***
	(216.0)	(1,136)	(630.9)	(864.7)	(0.0114)	(0.0254)	(0.00391)	(0.00300)
Mean	7744	14775	3891	5124	0.902	0.779	0.0958	0.0630
effect/mean	-0.0761	-0.340	-0.588	-0.677	-0.0683	-0.241	-0.464	-0.509
Panel C. 8 Tiers Based on Current Year's School Rank ($N = 398$)								
parallel	-224.8	-7,922***	-3,120***	-8,126***	-0.0449***	-0.151***	-0.0490***	-0.0314***
	(139.9)	(2,137)	(746.2)	(1,987)	(0.00967)	(0.0216)	(0.00429)	(0.00304)
Mean	7783	27521	5016	10846	0.897	0.626	0.0982	0.0534
effect/mean	-0.0289	-0.288	-0.622	-0.749	-0.0501	-0.241	-0.499	-0.587
Panel D. 8 Tiers Based on Previous Year's School Rank ($N = 340$)								
parallel	-202.0	-11,039***	-3,285***	-9,492***	-0.0225**	-0.197***	-0.0456***	-0.0343***
	(130.3)	(2,390)	(839.6)	(2,476)	(0.00857)	(0.0244)	(0.00534)	(0.00463)
Mean	7964	31865	5570	13219	0.936	0.734	0.126	0.0781
effect/mean	-0.0254	-0.346	-0.590	-0.718	-0.0241	-0.268	-0.361	-0.440
Panel E. 8 Tiers Based on Average Rank 2005-2011 ($N = 398$)								
parallel	-300.8*	-10,708***	-3,073***	-8,747***	-0.0346***	-0.186***	-0.0510***	-0.0355***
	(151.0)	(2,397)	(742.0)	(2,090)	(0.00826)	(0.0204)	(0.00462)	(0.00340)
Mean	7892	30050	5694	12923	0.920	0.706	0.116	0.0682
effect/mean	-0.0381	-0.356	-0.540	-0.677	-0.0376	-0.264	-0.440	-0.521
Panel F. Fixed tier based on Peking&Tsinghua > 985 > 211 > Non-211 ($N = 398$)								
parallel	-465.4**	-1,232**	-1,722***	-2,442***	-0.0788***	-0.0722***	-0.0268***	-0.0193***
	(201.6)	(530.8)	(448.1)	(631.9)	(0.0178)	(0.0196)	(0.00266)	(0.00199)
Mean	6579	10317	2557	3526	0.751	0.525	0.0597	0.0425
effect/mean	-0.0707	-0.119	-0.674	-0.692	-0.105	-0.138	-0.450	-0.455

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression using the sample of matching games at the track-year-province level in Tier I college admissions. The justified envy measures are redefined based on the group preferences described in the panel titles, instead of the four-tier preference based on current year admission cutoffs in the main results. Controls include track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Results using Alternative Restrictions around University Rank Group Cutoffs

Dependant variable	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tridimensional envy (1000)	(5) standardized # of blocking students	(6) standardized # of blocking pairs	(7) standardized # of blocking triplets	(8) standardized tridimensional envy
Panel A. Percentile Rank (University Best) - Percentile Rank (University Admitted) > 5%								
parallel	-424.2**	-3,961***	-2,340***	-3,314***	-0.0672***	-0.174***	-0.0449***	-0.0306***
	(208.8)	(1,107)	(574.6)	(779.3)	(0.0138)	(0.0253)	(0.00383)	(0.00288)
R-squared	0.988	0.932	0.813	0.802	0.536	0.617	0.802	0.774
Mean	7954	14113	3830	4828	0.903	0.707	0.0851	0.0534
effect/mean	-0.0533	-0.281	-0.611	-0.686	-0.0745	-0.246	-0.527	-0.572
Panel B. No restriction on Percentile Rank (University Best) - Percentile Rank (University Admitted)								
parallel	-426.9**	-3,963***	-2,313***	-3,287***	-0.0638***	-0.173***	-0.0448***	-0.0307***
	(202.4)	(1,101)	(594.7)	(797.4)	(0.0126)	(0.0248)	(0.00390)	(0.00294)
R-squared	0.989	0.934	0.816	0.805	0.517	0.614	0.796	0.770
Mean	8033	14192	4092	5129	0.912	0.711	0.0887	0.0558
effect/mean	-0.0531	-0.279	-0.565	-0.641	-0.0700	-0.243	-0.505	-0.551

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression using the sample of matching games at the track-year-province level in Tier I college admissions. Instead of restricting at least ten percentile difference to substantiate justified envy as in the main results, here we relax the restriction around rank group cutoffs to five percentile or no restriction at all. Controls include track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Matching Mechanism and Justified Envy: Results at the Individual Level

Dependant variable	(1) any justified envy (0/1)	(2) # of unique tiers blocked	(3) # of envied students (thousand)	(4) sum tridimensional envy (thousand)	(5) max envy on school tier dimension	(6) max envy on student rank dimension	(7) max tridimensional envy per triplet
parallel	-0.0461**	-0.422***	-0.208***	-0.296***	-0.415***	-0.182***	-0.489***
	(0.0199)	(0.0842)	(0.0242)	(0.0315)	(0.0848)	(0.0447)	(0.101)
exinterim	-0.0803***	-0.354**	0.188	0.250	-0.378**	-0.192*	-0.395*
	(0.0289)	(0.140)	(0.116)	(0.157)	(0.145)	(0.0971)	(0.206)
expost	-0.0495	-0.00549	-0.109*	-0.138*	-0.00827	-0.0327	0.0503
	(0.0297)	(0.133)	(0.0572)	(0.0811)	(0.133)	(0.0981)	(0.200)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.199	0.334	0.077	0.070	0.337	0.097	0.224
mean	0.806	1.463	0.357	0.456	1.467	0.870	1.742
sd	0.396	0.957	0.952	1.327	0.959	0.740	1.032
effect/mean	-0.0572	-0.289	-0.583	-0.649	-0.283	-0.210	-0.281
effect/sd	-0.116	-0.441	-0.219	-0.223	-0.432	-0.247	-0.474

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Does a Larger Number of Parallel Choices Increase the Reduction in Justified Envy?

Dependant variable	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tri dimensional envy (1000)	(5) standardized # of blocking students	(6) standardized # of blocking pairs	(7) standardized # of blocking triplets	(8) standardized tridimensional envy
ϕ^3	-142.9	-4,453***	-3,396***	-4,753***	-0.0138	-0.131***	-0.0404***	-0.0270***
	(269.1)	(1,440)	(1,101)	(1,511)	(0.0198)	(0.0267)	(0.00545)	(0.00372)
ϕ^4	-338.9	-2,990***	-2,150**	-3,057***	-0.0753***	-0.176***	-0.0515***	-0.0347***
	(244.8)	(1,048)	(829.7)	(1,140)	(0.0197)	(0.0305)	(0.00604)	(0.00450)
ϕ^5	-1,287***	-7,710***	-1,892**	-2,724**	-0.120***	-0.267***	-0.0356***	-0.0238***
	(395.0)	(2,165)	(851.4)	(1,146)	(0.0190)	(0.0363)	(0.00366)	(0.00245)
ϕ^6	-298.1	-1,230	-1,460	-2,046	-0.0755*	-0.112**	-0.0402***	-0.0268***
	(193.7)	(1,193)	(1,543)	(2,032)	(0.0426)	(0.0430)	(0.00799)	(0.00579)
Obs	398	398	398	398	398	398	398	398
R-squared	0.984	0.930	0.797	0.789	0.557	0.631	0.802	0.775
Mean	7549	13706	3347	4273	0.864	0.687	0.0784	0.0489
SD	7327	14691	6346	8272	0.0973	0.183	0.0373	0.0253

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

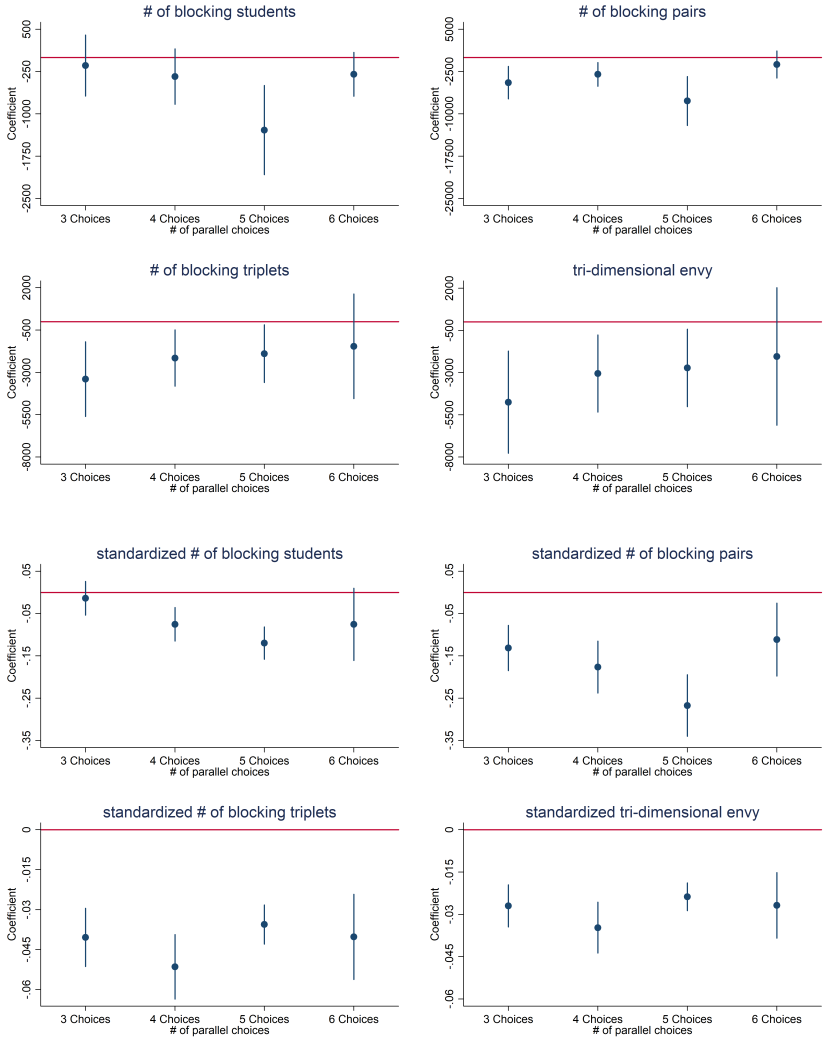
Table 8: Matching Mechanisms and Justified Envy: Robustness Checks

Dependant variable	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tridimensional envy (1000)	(5) standardized # of blocking students	(6) standardized # of blocking pairs	(7) standardized # of blocking triplets	(8) standardized tridimensional envy
Mean	7549	13706	3347	4273	0.864	0.687	0.0784	0.0489
SD	7327	14691	6346	8272	0.0973	0.183	0.0373	0.0253
Panel A. Difference in Differences Estimator with Multiple Groups and Periods								
Parallel	-593.7***	-4,135***	-2,436***	-3,396***	-0.0966***	-0.2109***	-0.0489***	-0.0323***
	(224.0)	(1,042.8)	(711.6)	(813.8)	(0.0154)	(0.0276)	(0.00441)	(0.00272)
Panel B. Control for Timing Reform Dosage								
Parallel	-446.2**	-3,967***	-2,246***	-3,221***	-0.0681***	-0.174***	-0.0468***	-0.0316***
	(222.7)	(1,154)	(562.4)	(778.1)	(0.0146)	(0.0273)	(0.00403)	(0.00304)
Panel C. Control for Partial Parallel								
Parallel	-491.6*	-4,011***	-2,074***	-2,978***	-0.0812***	-0.175***	-0.0438***	-0.0295***
	(246.9)	(1,174)	(681.5)	(938.9)	(0.0190)	(0.0307)	(0.00445)	(0.00328)
Panel D. Exclude Provinces and Years with Tier 0 Pre-Admission Mechanism Reform								
Parallel	-536.9**	-4,186***	-2,177***	-3,094***	-0.0751***	-0.181***	-0.0433***	-0.0291***
	(229.1)	(1,197)	(564.9)	(768.1)	(0.0147)	(0.0266)	(0.00388)	(0.00287)
Panel E. Drop the Most and Least Selective Majors from Each University								
Parallel	-489.5**	-3,121***	-1,496***	-2,098***	-0.113***	-0.188***	-0.0342***	-0.0216***
	(214.5)	(960.5)	(439.9)	(591.1)	(0.0222)	(0.0287)	(0.00577)	(0.00340)
effect/mean	-0.077	-0.291	-0.653	-0.710	-0.141	-0.331	-0.547	-0.607
Mean	6326	10735	2291	2954	0.801	0.568	0.0625	0.0356
SD	6670	12948	4886	6508	0.156	0.200	0.0357	0.0223

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. The mean and standard deviations for each outcome variable for the sample in Panel A, B, and C are presented at the top of this table. For Panel D, we present the new sample statistics below the coefficients. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

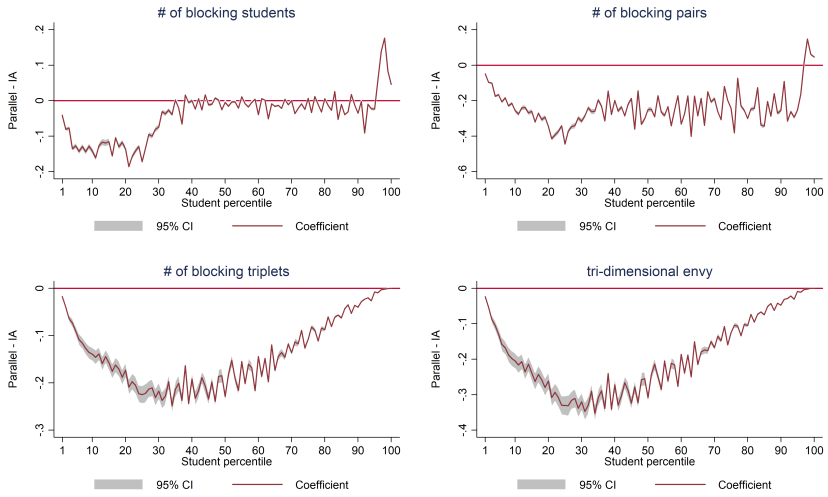
Appendix: Supplementary Figures and Tables

Figure A1: Effects on justified envy across bandwidths



Notes: These figures present the coefficients of Equation 3, comparing matching outcomes from games adopting different numbers of parallel choices with the IA mechanism. Each figure presents results from a separate OLS regression, regressing the relevant outcome on the choice bandwidth indicators, track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. Standard errors are clustered at the province level. The extension lines show the 95% confidence interval.

Figure A2: Differences in Justified Envy Measures across Academic Performance Distribution: Parallel - IA



Notes: These figures present the differences in the average values of various justified envy measures along the CEE score distribution for matching games with IA versus parallel mechanisms. The solid line connects 100 coefficients, each from comparing the justified envy measure for students with CEE scores within a one percentile sample corresponding to the X-axis values. The grey areas show the 95% confidence interval. Note that in all our analyses, we limit our sample to those with a CEE score above the Tier I cutoff score. A lower percentile means a higher CEE score.

Table A1: Timing of policy reform in each province

Province	parallel	ex-interim	ex-post
Beijing	2014	–	2015
Tianjin	2010	before 2005	2011
Hebei	2009	–	1999
Shanxi	2012	before 2005	2012
Inner Mongolia	–	–	2002
Liaoning	2008	before 2005	2014
Jilin	2009	–	2008
Heilongjiang	2013	before 2005	2013
Shanghai	2008	–	2017
Jiangsu	2005	–	2003
Zhejiang	2007	–	1999
Anhui	2008	before 2005	2007
Fujian	2009	–	2005
Jiangxi	2009	before 2005	2007
Shandong	2013	–	1998
Henan	2010	before 2005	2010
Hubei	2011	–	2014
Hunan	2003	–	2001
Guangdong	2010	–	2008
Guangxi	2009	–	2004
Hainan	2009	–	2002
Chongqing	2010	–	2006
Sichuan	2009	–	2005
Guizhou	2009	before 2005	2008
Yunnan	2009	–	2004
Tibet	2010	–	1996
Shaanxi	2010	before 2005	2010
Gansu	2015	before 2005	2008
Qinghai	2018	–	1996
Ningxia	2009	–	1999
Xinjiang	2011	before 2005	2015

Source: Ha, Kang and Song (2020).

Table A2: Results from Excluding Top-admitted and Bottom-scoring Students

Dependant variable	(1) # of blocking students	(2) # of blocking pairs	(3) # of blocking triplets (1000)	(4) tridimen-sional envy (1000)	(5) standardized # of blocking students	(6) standardized # of blocking pairs	(7) standardized # of blocking triplets	(8) standardized tridimen-sional envy
Parallel	-661.8***	-4,364***	-2,406***	-3,389***	-0.0711***	-0.175***	-0.0439***	-0.0294***
	(222.4)	(1,130)	(567.7)	(774.4)	(0.0144)	(0.0257)	(0.00374)	(0.00277)
Ex-interim	-612.5	-2,382	2,811	3,723	-0.0593*	-0.137*	0.0141*	0.00958*
	(480.8)	(1,683)	(2,087)	(2,813)	(0.0329)	(0.0815)	(0.00755)	(0.00515)
Ex-post	-484.8	-304.7	-1,051	-1,336	-0.0407	-0.0359	-0.0232***	-0.0144***
	(471.8)	(1,669)	(928.2)	(1,287)	(0.0300)	(0.0789)	(0.00410)	(0.00324)
Observations	398	398	398	398	398	398	398	398
R-squared	0.986	0.931	0.771	0.764	0.531	0.611	0.796	0.769
Mean	7549	13706	3347	4273	0.864	0.687	0.0784	0.0489
SD	7327	14691	6346	8272	0.0973	0.183	0.0373	0.0253
Effect/mean	-0.0877	-0.318	-0.719	-0.793	-0.0823	-0.255	-0.560	-0.603
Effect/SD	-0.0903	-0.297	-0.379	-0.410	-0.731	-0.960	-1.177	-1.162

Notes: This table reports a robustness check when we exclude individuals who are admitted to the top rank group and those who scored the lowest in each matching game, and then construct the justified envy measures. Those excluded individuals, by definition, would not have justified envy, since they were already admitted to the best rank group university or do not have anyone scoring lower than them to justifiably envy. Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Nonlinearity in Parallel Choice Bandwidth: Results using Individual-level Data

VARIABLES	(1) any justified envy (0/1)	(2) # of unique tiers blocked	(3) # of envied students (1000)	(4) sum tridimen-sional envy (1000)	(5) max envy on school tier dimension	(6) max envy on student rank dimension	(7) max tridimen-sional envy per triplet
3 Parallel Choices	-0.00604	-0.330***	-0.235***	-0.329***	-0.328***	-0.131***	-0.412***
	(0.0120)	(0.0484)	(0.0307)	(0.0411)	(0.0494)	(0.0348)	(0.0840)
4 Parallel Choices	-0.0301	-0.330***	-0.198***	-0.286***	-0.307***	-0.117***	-0.356***
	(0.0259)	(0.0883)	(0.0418)	(0.0568)	(0.0874)	(0.0416)	(0.0986)
5 Parallel Choices	-0.119***	-0.712***	-0.181***	-0.260***	-0.714***	-0.352***	-0.819***
	(0.0205)	(0.105)	(0.0328)	(0.0415)	(0.105)	(0.0626)	(0.131)
6 Parallel Choices	-0.0347	-0.224**	-0.240*	-0.333**	-0.241**	-0.113*	-0.285***
	(0.0217)	(0.0992)	(0.125)	(0.165)	(0.0936)	(0.0620)	(0.0986)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.202	0.341	0.077	0.070	0.345	0.101	0.232
Mean	0.806	1.463	0.357	0.456	1.467	0.870	1.742
SD	0.396	0.957	0.952	1.327	0.959	0.740	1.032

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Results for DID Estimator with Multiple Groups and Periods

Dependant Variable		Estimate	SE	95% CI Lower Bound	95% CI Upper Bound	N	Switchers
# of blocking students	Effect_0	-593.7	224.0	-1032.8	-154.6	200	44
	Placebo_1	15.2	189.8	-356.8	387.2	196	44
# of blocking pairs	Effect_0	-4135.2	1042.8	-6179.1	-2091.3	200	44
	Placebo_1	242.8	953.9	-1626.8	2112.5	196	44
# of blocking triplets	Effect_0	-2435.8	711.6	-3830.7	-1041.0	200	44
	Placebo_1	-281.1	437.9	-1139.3	577.1	196	44
tridimensional envy	Effect_0	-3396.3	813.8	-4991.3	-1801.3	200	44
	Placebo_1	-433.8	475.0	-1364.7	497.2	196	44
Std # of blocking students	Effect_0	-0.097	0.015	-0.127	-0.067	200	44
	Placebo_1	0.006	0.016	-0.025	0.037	196	44
Std # of blocking pairs	Effect_0	-0.211	0.028	-0.265	-0.157	200	44
	Placebo_1	0.004	0.028	-0.049	0.058	196	44
Std # of blocking triplets	Effect_0	-0.049	0.004	-0.058	-0.040	200	44
	Placebo_1	0.002	0.004	-0.007	0.010	196	44
Std tridimensional envy	Effect_0	-0.032	0.003	-0.038	-0.027	200	44
	Placebo_1	0.001	0.003	-0.005	0.006	196	44

Notes: This table presents detailed results of a Difference in Differences estimator with multiple groups and periods following De Chaisemartin and d'Haultfoeuille (2020) using Stata package did_multiplegt. We control for timing of preference submission and game size, in addition to two-way fixed effects, same as Table 2. Standard errors are clustered at the province level, bootstraped 50 times. Placebo estimates are comparing control and treated one year before treatment (i.e. changing from IA to parallel).

Table A5: Bacon Decomposition

	Weight	# of blocking students	# of blocking pairs	# of blocking triplets (1000)	Tridi-mensional envy (1000)	Std # of blockings student	Std # of blocking pairs	Std # of blocking triplets	Std tridi-mensional envy
Timing groups	0.35	-74.62	-2419.08	-1979.47	-2694.31	-0.0833	-0.1777	-0.0540	-0.0361
Always v timing	0.08	1507.73	2031.66	-1694.82	-2471.20	-0.0369	-0.0402	-0.0539	-0.0358
Never v timing	0.53	-617.37	-4498.82	-2662.35	-3838.12	-0.0871	-0.1939	-0.0509	-0.0341
Always v never	0.00	-47142.68	-128021.95	-40116.63	-54021.25	-1.4453	-2.8628	0.0723	0.0448
Within	0.04	-7550.03	-17771.50	6990.44	9463.58	-0.0343	-0.2589	0.1320	0.0831
coef (balanced)		-555.17	-3,821.06	-1,938.07	-2,767.02	-0.0798	-0.1808	-0.0444	-0.0299
coef (unbalanced)		-488.58	-4027.21	-2261.26	-3021.53	-0.0713	-0.1758	-0.0440	-0.0295

Notes: This table presents the bacon weights decomposition (Goodman-Bacon, 2021) using Stata command bacondecomp. All specifications are the same as Table 2 the main results.

Table A6: Interaction between Submission Timing and Matching Mechanism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw counts as outcome				Standardized measures as outcome			
VARIABLES	# of blocking students	# of blocking pairs	# of blocking triplets (1000)	tridi-mensional envy (1000)	# of blocking students	# of blocking pairs	# of blocking triplets	tridi-mensional envy
Omitted: IA X ex-ante								
IA X ex-interim	-695.2	-2,605	2,862	3,778	-0.0608*	-0.140*	0.0141*	0.00946*
	(458.6)	(1,656)	(2,102)	(2,829)	(0.0334)	(0.0824)	(0.00749)	(0.00510)
IA X ex-post	-364.0	135.3	-975.3	-1,249	-0.0336	-0.0262	-0.0267***	-0.0168***
	(474.9)	(1,719)	(923.4)	(1,278)	(0.0318)	(0.0810)	(0.00328)	(0.00256)
parallel X ex-ante	596.9	-600.6	-1,851*	-2,928*	-0.0626**	-0.212***	-0.0973***	-0.0703***
	(437.7)	(613.8)	(1,012)	(1,605)	(0.0310)	(0.0767)	(0.00846)	(0.00756)
parallel X ex-interim	-748.0	-4,346**	813.5	932.6	-0.0575	-0.181**	-0.0317***	-0.0190***
	(500.0)	(1,939)	(2,142)	(2,905)	(0.0359)	(0.0872)	(0.00779)	(0.00528)
parallel X ex-post	-999.5**	-4,478**	-3,299***	-4,521***	-0.116***	-0.217***	-0.0661***	-0.0432***
	(433.3)	(1,697)	(1,179)	(1,626)	(0.0280)	(0.0782)	(0.00433)	(0.00323)
Observations	398	398	398	398	398	398	398	398
R-squared	0.984	0.927	0.796	0.787	0.541	0.620	0.815	0.793

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression using the sample of 398 matching games at the track-year-province level in Tier I college admissions. $N = 398$ in all regressions. Controls include track fixed effects, province fixed effects, year fixed effects, and the total number of students participating in CEE in the track-province that year. Robust standard errors clustered at the province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Parallel choice bandwidth: Pre-reform balance test

VARIABLES	(1) game size	(2) # of blocking students	(3) # of block-ing pairs	(4) # of blocking triplets	(5) tri-dimensional envy	(6) standardized # of blocking students	(7) standardized # of block-ing pairs	(8) standardized # of block-ing triplets	(9) standardized tri-dimensional envy
Omitted group: 3 parallel choices									
4 Parallel Choices	-5,041***	-3,081**	-4,468	-3,042	-3,656	0.0723	0.126	0.0127	0.00888
	(1,594)	(1,381)	(3,890)	(2,221)	(2,782)	(0.0437)	(0.0888)	(0.0105)	(0.00768)
5 Parallel Choices	-2,627	-1,360	-2,084	-2,375	-2,845	0.0427	0.147	-0.00311	-0.000889
	(1,807)	(1,481)	(3,981)	(2,162)	(2,758)	(0.0437)	(0.0898)	(0.0106)	(0.00769)
6 Parallel Choices	-10,480***	-7,525***	-13,912***	-6,593***	-8,266***	-0.0314	-0.0922	-0.0120	-0.00924
	(2,572)	(2,115)	(4,946)	(2,307)	(2,896)	(0.0538)	(0.117)	(0.0136)	(0.00950)
STEM track	12,162***	9,738***	19,357***	7,675***	9,825***	0.0184	0.0829*	-0.0212***	-0.0152***
	(1,393)	(1,189)	(2,578)	(1,474)	(1,910)	(0.0186)	(0.0447)	(0.00728)	(0.00521)
Observations	46	46	46	46	46	46	46	46	46
R-squared	0.696	0.661	0.617	0.443	0.435	0.284	0.297	0.265	0.260

Note: This table presents the OLS results comparing the pre-reform year characteristics among the reform provinces to see if provinces adopted more parallel choices are systematically different from others. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8: Robustness Check: Major-by-university level justified envy analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw counts as outcome				Standardized measures as outcome			
Dependant variable	# of blocking students	# of blocking pairs	# of blocking triplets	tri-dimensional envy	standardized # of blocking students	standardized # of blocking pairs	standardized # of blocking triplets	standardized tri-dimensional envy
parallel	-452.9**	-4,148***	-1,111***	-1,384***	-0.0536***	-0.149***	-0.0204***	-0.0121***
	(184.2)	(1,193)	(300.7)	(374.5)	(0.0187)	(0.0250)	(0.00286)	(0.00186)
ex-interim	-217.2	551.0	1,419	1,754	-0.0358	-0.0179	0.00922*	0.00615**
	(267.9)	(1,762)	(1,051)	(1,285)	(0.0269)	(0.0627)	(0.00483)	(0.00307)
ex-post	-162.4	-474.7	-523.8	-632.1	-0.0235	-8.18e-05	-0.0110***	-0.00573**
	(226.0)	(1,329)	(465.6)	(576.8)	(0.0234)	(0.0567)	(0.00329)	(0.00228)
game size	0.863***	1.627***	0.403***	0.501***	7.31e-07	1.07e-06	3.21e-08	2.65e-08
	(0.0242)	(0.140)	(0.0913)	(0.123)	(1.24e-06)	(2.12e-06)	(1.62e-07)	(1.09e-07)
Observations	398	398	398	398	398	398	398	398
R-squared	0.990	0.921	0.767	0.759	0.548	0.626	0.758	0.730
mean	7577	12714	1562	1881	0.827	0.554	0.0349	0.0201
effect/mean	-0.0598	-0.326	-0.711	-0.736	-0.0649	-0.269	-0.585	-0.603
sd	7500	14405	3341	4178	0.124	0.190	0.0202	0.0125
effect/sd	-0.0604	-0.288	-0.333	-0.331	-0.432	-0.786	-1.009	-0.970

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression using the sample of matching games at the track-year-province level in Tier I college admissions. Controls include track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. University-majors are divided into four rank groups using current year cutoffs. Justified envy is restricted to exclude those student pairs with a rank difference within 10 percentile. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1