WHEN DO REFORMS MEET FAIRNESS CONCERNS IN SCHOOL ADMISSIONS?

SOMOUAOGA BONKOUNGOU* AND ALEXANDER NESTEROV†

Abstract. We study a series of reforms of school admissions mechanisms motivated, among other reasons, by fairness concerns and vulnerability to manipulation. Before the reforms and after, the mechanisms were vulnerable to preference manipulation and induced *blocking students*: students who miss desired schools despite having higher priority or seats left empty. We demonstrate that some of these reforms improved fairness by adopting mechanisms with fewer blocking students compared to the preexisting ones, while several others did not. We identify preexisting mechanisms where fairness consideration was more of an issue than vulnerability to manipulation and those where it is the reverse.

Keywords: market design, school choice, college admissions, fairness, stability, manipulability

JEL Classification: C78, D47, D78, D82

1. Introduction

In the last two decades, there has been a wave of reforms of school admissions mechanisms around the world (Pathak and Sönmez, 2013). The surprising fact is that after such reforms most matching mechanisms present the same poor properties that could have arguably justified the policy changes. For example, despite some evidence that vulnerability to manipulation and fairness concerns mostly drove the changes, most newly adopted matching mechanisms still suffer from these two deficiencies.

Fairness is at the forefront of the concerns that led to the policy changes. Perhaps the most vivid example is the 2007 major reform in England, which covers 146 local school

*Faculty of Business and Economics, University of Lausanne, Internef, CH-1015 Lausanne, Switzerland, bkgsom@gmail.com

†Department of Economics and Game Theory Lab, Higher School of Economics, 16, Soyuza Pechatnikov st., 190121, St. Petersburg, Russia, nesterovu@gmail.com

We are grateful to Bettina Klaus, Battal Doğan, Lars Ehlers, and Rustamdjan Hakimov for their suggestions. We thank Camille Terrier, Inacio Bo, Madhav Raghavan, and other participants of the online seminar of the Lausanne Market Design group for their comments, as well as the participants of the online Conference on Mechanism and Institution Design, and the 2022 EEA-ESEM meeting in Milano.

Acknowledgments: The paper was prepared within the framework of the HSE University Basic Research Program. Bonkoungou acknowledges financial support from the Swiss National Science Foundation, project number 100018_207722 and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 890648. Results in Section 4 have been obtained under support of the RSF grant No. 24-28-00608.

admissions systems. According to the then-Secretary of State, Alan Johnson, the reform aimed to *"ensure that admission authorities – whether local authorities or schools – operate in a fair way"* (Department for Education and Skills, 2007). Among other things, the reform prohibited the practice of giving *"priority to children according to the order of other schools named as preference by their parents,"* known as the *first-preference-first principle.* This principle states that a student who ranks a school higher in her list receives a higher admission priority at this school compared to students who rank it lower. Before the reform, as many as one-third of schools in England used this principle.

In 2009, Chicago education authorities reformed their Selective High School admission system. They replaced the so-called Boston mechanism that used the first-preference-first principle for each school, arguing that, due to this principle *"high-scoring kids were being rejected simply because of the order in which they listed their [schools] preferences"* (Pathak and Sönmez, 2013). The same Boston mechanism has also been used for college admissions in several provinces in China, and it raised similar complaints. For example, one parent said: *"My child has been among the best students in his school and school district. He achieved a score of 632 on the college entrance exam last year. Unfortunately, he was not accepted by his first choice. After his first choice rejected him, his second and third choices were already full. My child had no choice but to repeat his senior year"* (Chen and Kesten, 2017; Nie, 2007). In 2003, more than 3 million students, representing half of the annual intake, were matched to significantly worse colleges than what their grades allowed (Wu and Zhong, 2020).

These examples illustrate fairness concerns with the old mechanisms: they can induce a matching with a so-called *blocking student*, that is, a student who missed a school while at least one seat at that school has been assigned to a student with a lower grade or priority or even left empty. The blocking student is entitled to this seat, yet she has not been assigned to it. It is important to note that we define the concept of fairness concerning true preferences and not reported preferences. A matching with no blocking student is *stable* and is viewed as a fair outcome as it eliminates "justified envy", a situation in which a student prefers a school that is assigned to another student with lower admission priority (Abdulkadiroğlu and Sönmez, 2003).[1] Gale and Shapley (1962) show that for any instance there is a student-optimal stable matching, a matching that every student finds at least as good as any other stable matching. This stable matching can be reached by the student-proposing deferred acceptance algorithm by Gale and Shapley (1962). We refer to it as the Gale-Shapley mechanism.

Apart from the first-preference-first principle, many mechanisms induce blocking students because they have ranking constraints. In such a mechanism each student is allowed to rank-list only a limited number of schools, typically between 3 and 5 (Pathak and Sönmez,

[1]In general, the relation between stability and fairness is more nuanced, see Romm et al. (2020).

2013). Even in New York City, where the ranking constraint is 12 and there are more than 1700 schools, around 25% of students report a complete list of 12 schools, while only 5% report 9, 10, or 11 schools, suggesting that around 20% of students in New York City could not list all acceptable schools (Abdulkadiroğlu et al., 2009). Students who missed all their listed schools but could have been admitted to unlisted schools will be dissatisfied with the admissions system and deem it unfair. We consider all blocking students, whether it concerns listed schools for which admissions authorities can verify priority violations or unlisted schools that lead to dissatisfaction (see Calsamiglia et al., 2010).

Our first goal is to investigate whether the reforms led to *more fair* matching mechanisms. Our second goal is to investigate the relative importance of vulnerability to manipulation and fairness concerns in the preexisting mechanisms. Chen and Kesten (2017) propose to compare mechanisms by set inclusion of problems where they produce stable outcomes. However, this notion cannot further distinguish mechanisms in each instance where they are not stable. A finer and complementary notion is useful, in particular, when the compared mechanisms are not stable or a large fraction of instances or real-life instances lie in this domain. We indeed illustrate that for many of our compared mechanisms, real-life instances are likely to be in a domain where the compared mechanisms are not stable (Example 1).

To address this problem, we count and compare the number of blocking students across mechanisms. In an instance where two mechanisms are not stable, they can still be contrasted using the number of blocking students.[2] Our investigation led to a result that supports an important kind of reform. Broadly, these reforms involve extending ranking constraints in the Gale-Shapley mechanism. The Gale-Shapley mechanism with a relaxed constraint has weakly fewer blocking students than the restricted counterpart and there are instances where it has fewer blocking students. This took place in Chicago (2010), in Ghana (2007, 2008), in Newcastle (2010), and in Surrey (2010) (Pathak and Sönmez, 2013). For the remaining reforms, it is not possible to conclude by comparing the number of blocking students. We show that after those reforms the number of blocking students may increase.

We then answer the following question. Was fairness more of a concern compared to vulnerability to manipulation? We focus on blocking students and manipulating students, students who could gain by misreporting their preferences while others are truthful (as in Bonkoungou and Nesterov, 2023). We show that for any instance the constrained Gale-Shapley mechanism has weakly more blocking students than manipulating students. More precisely, any manipulating student is a blocking student of the mechanism. In contrast, for any instance, the constrained Boston mechanism has weakly more manipulating students

[2]To our knowledge, this criterion has been first used by Roth and Xing (1997). Niederle and Roth (2009), Eriksson and Häggström (2008) and Doğan and Ehlers (2021) used the criterion of counting the number of blocking pairs, which does not allow comparisons in our setting (see Remark in section 2.2). Doğan and Ehlers (2021) and Doğan and Ehlers (2022) introduced criteria for counting the number of blocking pairs and blocking students.

than blocking students. More precisely, any blocking student is a manipulating student of the mechanism. For the constrained serial dictatorship mechanism these sets coincide: each blocking student is a manipulating student and vice versa. A more subtle relationship between stability and manipulability can be seen in the reform in England. This reform did not adopt less manipulable mechanisms in all school districts (Bonkoungou and Nesterov, 2021) and it did not adopt more fair matching mechanisms by stability either (Example 3). However, the reform was successful according to at least one dimension by the following criterion: if the reform disrupted fairness — by producing an unstable matching while it was stable before the reform — the new mechanism is not vulnerable to manipulation.

Related literature. Apart from papers studying the reforms mentioned earlier (Pathak and Sönmez, 2013; Chen and Kesten, 2017; Bonkoungou and Nesterov, 2021, 2023; Imamura and Tomoeda, 2022) and papers that used the method of counting blocking agents and blocking pairs (Roth and Xing, 1997; Niederle and Roth, 2009; Eriksson and Häggström, 2008) there is recent literature interested in various ways of comparing matching mechanisms by fairness.

Among strategy-proof and Pareto efficient mechanisms, the Gale's Top Trading Cycles mechanism (Shapley and Scarf, 1974) is among the most fair by stability when each school has one seat (Abdulkadiroğlu et al., 2020). This result also holds for other fairness comparisons, such as the set of blocking students (Doğan and Ehlers, 2022) and the set of blocking triplets (i, j, s) – student i blocking the matching of school s with student j (Kwon and Shorrer, 2020). The result holds for any stability comparison that satisfies a few basic properties (Doğan and Ehlers, 2022).

Among Pareto efficient mechanisms, the Efficiency Adjusted Deferred Acceptance mechanism (EADA) due to Kesten (2010) is among the most fair in terms of blocking pairs and blocking triplets (Doğan and Ehlers, 2021; Tang and Zhang, 2021; Kwon and Shorrer, 2020). Independent from the present work, Doğan and Ehlers (2021) also introduced the fairness comparison by counting to show that among efficient mechanisms, EADA is not the most fair by counting unless the priority profile satisfies a few acyclicity conditions.

The first papers that studied constrained mechanisms are Romero-Medina (1998) and Haeringer and Klijn (2009). They study the stability of Nash equilibrium outcomes of the game induced by these mechanisms. The most important insight is that the Nash equilibrium outcomes of the constrained Boston mechanism are all stable, while the Nash equilibrium outcomes of the constrained Gale-Shapley may not all be stable.[3] Besides, the Nash equilibrium outcomes of the constrained Gale-Shapley are a subset of the Nash equilibrium outcomes of any constrained Gale-Shapley with a relaxed constraint. Therefore, when the Nash equilibrium outcomes of the constrained Gale-Shapley with a relaxed constraint are

[3]Ergin and Sönmez (2006) showed that the Nash equilibrium outcomes of the unconstrained Boston mechanism are stable.

all stable, the Nash equilibrium outcomes of the constrained Gale-Shapley with a restricted constraint are also stable.

Finally, the mechanisms we studied bear some resemblance to Preference Rank Partitioned mechanisms introduced by Ayoade and Pápai (2023) except that in constrained mechanisms schools' choice functions are not acceptant[4] as in Ayoade and Pápai (2023). In particular, the manipulation strategy in the constrained Gale-Shapley mechanism, consisting of including an unlisted school in the constrained list, resembles the strategy in the Preference Rank Partitioned mechanism, which consists of moving a school from a lower class to the first class. In addition, the manipulation strategy of the constrained Boston mechanism, which consists of moving a school from a lower rank to the highest rank has the same resemblance.

The rest of the paper is organized as follows. In Section 2, we introduce the model, the mechanisms, and the comparison criteria. In Section 3, we present fairness comparisons. In Section 4, we study the relationship between stability and manipulability. In Section 5, we conclude. We present most of the proofs in the Appendix.

2. Model

We consider the school choice problem (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003). It consists of the following elements:

- a finite set I of students,
- a finite set S of schools,
- a profile $P = (P_i)_{\in I}$ of preference relations for each student,
- a profile $\succ = (\succ_s)_{s\in S}$ of priority orders for each school, and
- a vector $q = (q_s)_{s\in S}$ of capacities for each school

where P and $\succ$ are defined as follows. For each student i, P_i is a strict preference relation P_i over $S \cup \{\emptyset\}$, where $\emptyset$ represents the outside option of being unmatched. For each school s, $\succ_s$ is a strict priority order over I. For each student i, let R_i denote the "at least as good as" relation associated with P_i.[5] School s is **acceptable** to student i if $s \; P_i \; \emptyset$; and it is **unacceptable** to student i if $\emptyset \; P_i \; s$. We extend the priority order $\succ_s$ of each school s over I to the set 2^I of subsets of students and assume that it is responsive to the priority order over I (Roth, 1985). By definition, the priority order $\succ_s$ over 2^I is responsive if for any students $i, j \in I$ and any subset $N \subset I \setminus \{i, j\}$ such that $|N| < q_s$, (i) $N \cup \{i\} \succ_s N$, and (ii) $N \cup \{i\} \succ_s N \cup \{j\}$ if and only if $i \succ_s j$. Let $k \in \{1, \ldots, |S|\}$, and P_i a preference relation where student i has x acceptable schools. The truncation of P_i after the k'th acceptable school (if any) of the preference relation P_i is a preference relation with $\min(x, k)$ acceptable

[4]A choice function is acceptant if it accepts all applicants when the capacity is not full (Kojima and Manea, 2010).

[5]That is, for each $s, s' \in S \cup \{\emptyset\}$, $s \; R_i \; s'$ if and only $s \; P_i \; s'$ or $s = s'$.

schools such that all schools are ordered as in P_i. Let P_i^k denote the truncation of P_i after the k'th acceptable school. Let $P^k = (P_i^k)_{i\in I}$. Given a proper subset $I' \subsetneq I$ of students, we will often write a preference profile as $P = (P_{I'}, P_{-I'})$ to emphasize the components for students in I'. The tuple $(I, S, P, \succ, q)$ is a school choice problem or simply a **problem**. We assume that there are more students than schools, that is, $|I| > |S|$. The set of students and the set of schools are fixed throughout the paper, and we denote the (school choice) problem by the triple $(P, \succ, q)$.

A **matching** μ is a function $\mu : I \to S \cup \{\emptyset\}$ such that for each school s, $|\mu^{-1}(s)| \leq q_s$. We say that student i is matched under μ if $\mu(i) \neq \emptyset$ and unmatched under μ if $\mu(i) = \emptyset$. Let $(P, \succ, q)$ be a problem. A matching μ is **individually rational** under P if for each student i, $\mu(i)\ R_i\ \emptyset$. A pair (i, s) of a student and a school **blocks** the matching μ under $(P, \succ, q)$ if $s\ P_i\ \mu(i)$ and either there is a student j such that $\mu(j) = s$ and $i \succ_s j$ or $|\mu^{-1}(s)| < q_s$. Student i is a **blocking student** for the matching μ under $(P, \succ, q)$ if there is a school s such that the pair (i, s) blocks μ under $(P, \succ, q)$. A matching μ is **stable** at $(P, \succ, q)$ if it is individually rational under P and has no blocking student. We often drop the problem and refer to a stable matching. A **mechanism** φ is a function that maps each problem to a matching. For each problem $(P, \succ, q)$, let $\varphi_i(P, \succ, q)$ denote the component for student i. A mechanism φ is individually rational if for each problem $(P, \succ, q)$ the matching $\varphi(P, \succ, q)$ is individually rational under P. A mechanism φ is stable if for each problem $(P, \succ, q)$ the matching $\varphi(P, \succ, q)$ is stable at $(P, \succ, q)$. We often drop the problem and say that a mechanism is stable (at the implicitly assumed problem).

2.1. **Mechanisms.** We are interested in mechanisms that were used either before or after the reforms. We first describe unconstrained versions.

Gale-Shapley. Gale and Shapley (1962) showed that for each problem, there exists a stable matching. In addition, there is a student-optimal stable matching, which is a matching that each student finds at least as good as any other stable matching. For each problem $(P, \succ, q)$, this matching can be found via the Gale and Shapley (1962) student-proposing deferred acceptance algorithm.

- *Step 1:* Each student applies to her most preferred acceptable school (if any). If a student did not rank any school acceptable, then she remains unmatched. Each school s considers its applicants at the first step denoted as I_s^1 and tentatively accepts $\min(q_s, |I_s^1|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. Let A_s^1 denote the set of students whom school s has tentatively accepted at this step.
- *Step t>1:* Each student, who is rejected at step $t-1$, applies to her most preferred acceptable school among those which have not yet rejected her (if any). If a student does not have any remaining acceptable school, then she remains unmatched. Each

school s considers the set $A_s^{t-1} \cup I_s^t$, where I_s^t are its new applicants at this step, and tentatively accepts $\min(q_s, |A_s^{t-1} \cup I_s^t|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. Let A_s^t denote the set of students whose school s has tentatively accepted at this step.

The algorithm stops when no student is rejected. The tentative acceptances become final at this step. Let GS denote this mechanism. Given $k \in \{1, ..., |S|\}$, the constrained version GS^k of the Gale-Shapley mechanism GS is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $GS(P^k, \succ, q)$. That is, $GS^k(P, \succ, q) = GS(P^k, \succ, q)$.

Serial Dictatorship. When schools have the same priority order, we call the Gale-Shapley mechanism the serial dictatorship mechanism.[6] Let SD denote this mechanism. The outcome of this mechanism can be computed via the following simplified process. Students move in sequence following the common priority order. The first student picks her most preferred acceptable school. The next student picks her most preferred acceptable school among the remaining ones, and so on. Given $k \in \{1, ..., |S|\}$, the constrained version SD^k of the Serial Dictatorship mechanism SD is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $SD(P^k, \succ, q)$. That is, $SD^k(P, \succ, q) = SD(P^k, \succ, q)$.

First-Preference-First. The schools are exogenously divided into two disjoint subsets S^{fpf} and S^{ep} such that $S^{fpf} \cup S^{ep} = S$. The set S^{eq} is a set of **equal-preference schools** and S^{fpf} is a set of **first-preference-first** schools. The First-Preference-First mechanism (FPF) assigns to each problem $(P, \succ, q)$, the matching $GS(P, \hat{\succ}, q)$ where $\hat{\succ}$ is obtained as follows. The priority order of each equal-preference school is maintained intact while the priority order of each first-preference-first school is adjusted according to the rank that students have assigned to it. Formally, the priority profile $\hat{\succ}$ is obtained as follows:

1. for each equal-preference school $s \in S^{ep}$, $\hat{\succ}_s = \succ_s$ and
2. for each first-preference-first school $s \in S^{fpf}$, $\hat{\succ}_s$ is defined as follows. Let $I^1(s)$ be the set of students who have ranked school s first under P, $I^2(s)$ the set of students who have ranked school s second under P, and so on. Note that we count the ranking of $\emptyset$ as well.

- For each $\ell, k \in \{1, \ldots, |S|+1\}$ such that $\ell > k$ and each students i, j such that $i \in I^k(s)$ and $j \in I^\ell(s)$, $i \mathrel{\hat{\succ}_s} j$.
- For each $k \in \{1, \ldots, |S|+1\}$ and each $i, j \in I^k(s)$, $i \mathrel{\hat{\succ}_s} j$ if and only if $i \succ_s j$.

Let FPF denote this mechanism. Given $k \in \{1, ..., |S|\}$, the constrained version FPF^k of the First-Preference-First mechanism FPF is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $FPF(P^k, \succ, q)$. That is, $FPF^k(P, \succ, q) = FPF(P^k, \succ, q)$.

[6]This is a slight abuse of our definition since the domain of a mechanism is the set of all problems — including problems where schools have different priorities.

Boston. Until 2005, the Boston public school system was using an immediate acceptance mechanism called the Boston mechanism (Abdulkadiroğlu and Sönmez, 2003). This mechanism assigns to each problem $(P, \succ, q)$, the matching as described in the following algorithm.

- *Step 1:* Each student applies to her most preferred acceptable school (if any). Each school s, considers its applicants at the first step denoted as I_s^1 and immediately accepts $\min(q_s, |I_s^1|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. For each school s, let $q_s^1 = q_s - \min(q_s, |I_s^1|)$ denote its remaining capacity after this step.
- *Step t>1:* Each student who is rejected at step $t-1$, applies to her most-preferred acceptable school among those who have not yet rejected her (if any). Each school s considers its new applicants I_s^t at this step and immediately accepts $\min(q_s^{t-1}, |I_s^t|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. For each school s, let $q_s^t = q^{t-1} - \min(q_s^{t-1}, |I_s^t|)$ denote its remaining capacity after this step.

The algorithm stops when every student is either accepted at some step or has applied to all of her acceptable schools. Let β denote this mechanism. Given $k \in \{1, ..., |S|\}$, the constrained version β^k of the Boston mechanism β is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $\beta(P^k, \succ, q)$. That is, $\beta^k(P, \succ, q) = \beta(P^k, \succ, q)$.

Remark. *In the (algorithm of the) Boston mechanism, students applying to the same school at each step have assigned the same rank to it. Therefore, students applying to a school at a given step of the algorithm rank this school higher than those applying to it at any step after. In particular, no student could be rejected by a school while another student, who has assigned a lower rank to it, is accepted by this school. Thus, the Boston mechanism is a First-Preference-First mechanism where every school is a first-preference-first school. This result follows from the Proposition 2 of Pathak and Sönmez (2008).*

Chinese parallel. Chen and Kesten (2017) describe a parametric mechanism that many Chinese provinces have been using. The parameter $e \geq 1$ is a natural number. For each problem $(P, \succ, q)$, the outcome is a sequential application of constrained GS. In the first round, the matching is final for students who are matched under $GS^e(P, \succ, q)$, while unmatched students proceed to the next round. In the next round, each school reduces its capacity by the number of students assigned to it in the last round, each matched student replaces her preferences with a preference relation where she finds no school acceptable and the unmatched students (in the previous round) are matched according to GS^{2e} for the reduced capacities and the new preference profile. The process continues until either no school has a remaining seat or no unmatched student finds a school with a remaining seat acceptable. Let $Ch^{(e)}$ denote this mechanism.[7]

[7]This definition of the Chinese parallel mechanisms is given only for the symmetric version where each round has the same length e. See Chen and Kesten (2017) for details.

2.2. **Comparison criteria.** We start with the criterion introduced by Chen and Kesten (2017). Broadly, it is a problem-by-problem comparison such that mechanisms are compared by the set inclusion of problems where they are stable.

Definition 1. *(Chen and Kesten, 2017). Mechanism φ' is **more fair by stability** than φ if (i) at each problem where φ is stable, φ' is also stable and (ii) there exists a problem where φ' is stable but φ is not.*

There are two cases in which mechanisms cannot be compared using fairness by stability. **Case 1**: there are two problems such that in one problem, one mechanism is stable but not the second mechanism, and in the second problem, the second mechanism is stable but not the first one. **Case 2**: two mechanisms have the same set of problems where they are unstable. We illustrate the second case in a restricted domain of preferences. Indeed, for certain problems that we encounter in real life, all mechanisms described above are likely to induce unstable outcomes, and the comparisons therefore are driven by some less relevant problems. Consider the high school admissions problem in Chicago[8] where schools have a common priority (constructed from student's composite scores) and where students form block preferences as illustrated in the following example.

Example 1 (Tier preferences)**.** *Consider a problem with n students and m schools such that for each $s, s' \in S$, $\succ_s = \succ_{s'}$. We assume that students have tier preferences. The set S of schools is partitioned into two sets S_1, S_2. Each student i prefers each school in S_1 to each school in S_2.[9] We assume that each student finds each school acceptable and $n > \sum_s q_s$.*

Whenever $|S_1| \geq k$, no student ranks a school in S_2 among the top k acceptable schools. Any individually rational and k-constrained ranking mechanism has a blocking student. Indeed, if every student reports her preferences truthfully, then some students are unmatched while seats at schools in S_2 are unassigned.

However, we can still distinguish different constrained Gale-Shapley mechanisms by counting the number of blocking students within this preference domain. For example, suppose that students have the same ranking over all schools in S_1 and we compare GS^{k-1} and GS^k. A student is a blocking student if and only if she is unmatched, and GS^k matches strictly more students than GS^{k-1}. Thus GS^k has fewer blocking students than GS^{k-1} does.

While this example is simplified to illustrate the fact that fairness by stability might not be able to distinguish the mechanisms that we study in real-life instances, all that is necessary for the result is that many students rank a subset of schools such that they exhaust the listing constraint, and such that these schools do not have enough seats to accommodate all these students. This example motivates the following definition.

[8]See Bonkoungou and Nesterov (2021) for details on school admissions in Chicago.

[9]Coles et al. (2013) observed that the academic job market has this structure and referred to it as block-correlated preferences.

Definition 2. *A mechanism φ' is* ***more fair by counting*** *(the number of blocking students) than a mechanism φ if (i) for each problem, there are at least as many blocking students of the outcome of φ as there are of the outcome of φ', and (ii) there is a problem where there are more blocking students of the outcome of φ than the outcome of φ'.*

Fairness by counting is not logically related to fairness by stability, but the first can be complementary to the second. If mechanism φ' is more fair by counting than φ, then for each problem where φ induces a stable matching, i.e., there is no blocking student, φ' also necessarily induces a stable matching. In addition, in some cases, in each problem where φ' and φ are not stable, φ may have less number of blocking students than φ.

3. Comparisons of Mechanisms

In this section, we compare mechanisms according to the two fairness criteria introduced. Our main result with fairness by counting is a strengthening of the comparison between different constraints of the Gale-Shapley mechanism. We first illustrate the intuition using the example below.

Example 2. *Let $I = \{i_1, \ldots, i_5\}$ and $S = \{s_1, \ldots, s_4\}$. Let $(P, \succ, q)$ be a problem where each school has one seat, and the remaining components are specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$
s_1	s_1	s_2	s_3	s_3	i_3	i_2	i_1	i_5
s_2	s_2	s_1	s_1	s_4	i_1	i_4	i_5	$\vdots$
s_3	s_3	s_3	s_2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Let us compare the mechanisms GS^2 and GS^1. We have

$$GS^2(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ \emptyset & s_2 & s_1 & \emptyset & s_3 \end{pmatrix}$$

where student i_1 is the unique blocking student for the matching under $(P, \succ, q)$. Indeed, i_1 is unmatched, finds s_3 acceptable and has a higher priority at s_3 than i_5. Let us shorten the reported list only for student i_2. Then,

$$GS^2(P^1_{i_2}, P_{-i_2}, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_1 & \emptyset & s_2 & \emptyset & s_3 \end{pmatrix}.$$

As a result of this replacement, there are three types of students, given their status in the previous matching. First, student i_2 — who was matched — became a blocking student. Second, student i_1 — who was a blocking student — is not a blocking student for the new matching. Finally, student i_4 is a new blocking student.

The intuition of this result is that by shortening the schools listed by student i_2, she is worse off while the other students are weakly better off. First, she is a blocking student for the new matching. Second, student i_1 is not a blocking student for the new matching, though she was a blocking student for the old matching. But a new blocking student appears so there are two blocking students in total.

This example turns out to be a general pattern. When students shorten the list, the set of blocking students changes, but the size of this set never decreases. By sequentially applying this argument to all students, we get the following result.[10]

Theorem 1. *Let there be at least two schools and ℓ, k integers such that $|S| > k > \ell \geq 1$.*
(i) The constrained Gale-Shapley mechanism GS^ℓ is more fair by stability than the constrained Gale-Shapley mechanism GS^k, and
(ii) the constrained Gale-Shapley mechanism GS^ℓ is more fair by counting than the constrained Gale-Shapley mechanism GS^k.

Statement (i) easily follows from statement (ii). The proof of the latter is in the appendix. In the following Proposition, we show that the two fairness notions do not explain many changes that followed the 2007 reform in the UK, as the constrained First-Preference-First mechanism is not comparable to the constrained Gale-Shapley mechanism according to this criterion.

Proposition 1. *Let there be at least seven schools and at least five students and let $k > 3$. The constrained First-Preference-First mechanism FPF^k and the constrained Gale-Shapley mechanism GS^k*
(i) are not comparable via fairness by stability, and
(ii) are not comparable via fairness by counting.

We prove both statements using the following example.[11]

Example 3. *Let $I = \{i_1, \ldots, i_7\}$ and $S = \{s_1, \ldots, s_5\}$. Let school s_3 be the only first-preference-first school. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows. (The sign $\vdots$ indicates that the remaining part is arbitrary.)*

[10] We are very grateful to a referee for suggesting a simple technique for proving this result.
[11] Example where FPF^k has blocking students while GS^k does not is immediate and omitted. Besides, β^k is a special case of FPF^k.

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	P_{i_6}	P_{i_7}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$	$\succ_{s_5}$
s_1	s_1	s_4	s_1	s_2	s_1	s_5	i_4	i_5	i_3	i_1	i_7
s_2	s_3	s_3	s_2	s_1	s_2	s_1	$\vdots$	$\vdots$	i_1	i_6	$\vdots$
s_3	$\emptyset$	$\emptyset$	s_3	s_3	s_5	s_2			i_2	i_3	
s_4			$\emptyset$	$\emptyset$	s_3	$\emptyset$			$\vdots$	$\vdots$	
$\emptyset$					s_4						
					$\emptyset$						

The outcomes of the constrained First-Preference-First FPF^4 and the constrained Gale-Shapley GS^4 at $(P, \succ, q)$ are as follows:

$$FPF^4(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\ s_4 & \emptyset & s_3 & s_1 & s_2 & \emptyset & s_5 \end{pmatrix},$$

$$GS^4(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\ s_3 & \emptyset & s_4 & s_1 & s_2 & \emptyset & s_5 \end{pmatrix}.$$

The matching $FPF^4(P, \succ, q)$ is stable.[12] *However, the matching $GS^4(P, \succ, q)$ is not stable. Indeed, the pair (i_6, s_4) blocks this matching because student i_6 is unmatched and finds school s_4 acceptable, but student i_3 is matched to s_4 while $i_6 \succ_{s_4} i_3$. The intuition is that the constraint in GS shortened the chains of the rejections needed to reach a stable matching in the Gale-Shapley algorithm. For example, student i_3 is temporarily matched to school s_4 at some step of the algorithm. At the student-optimal stable matching for $(P, \succ, q)$, school s_4 is assigned to student i_1. However, we need an application of student i_1 at that school to displace student i_3 from s_4. This does not occur under GS^4 because no student initiates the rejection chain. However, under FPF^4, the application of student i_2 at school s_3 causes the rejection of student i_1 at s_3 (student i_2 has ranked it higher than i_1 and school s_3 is a first-preference-first school). This is the rejection needed to reach the student-optimal stable matching.*

This example illustrates how the constrained GS mechanism has shortened the chains needed to reach a stable matching. It is well known that this type of chain leads to unambiguous welfare losses. Each student in the chain is worse off, and all other students are unaffected (Kesten, 2010).[13] However, under the Boston mechanism, (where all schools are first-preference-first schools) there is no such chain, and thus the constrained Boston mechanism can be compared to the constrained Gale-Shapley mechanism.

Theorem 2. *Let there be at least two schools and an integer such that $|S| \geq k > 1$.*

[12]This matching is both the student-optimal and the school-optimal stable matching.

[13]These chains are initiated by the so-called interrupters. These are students who initiate chains of rejections that return to them (Kesten, 2010).

(i) The constrained Gale-Shapley mechanism GS^k is more fair by stability than the constrained Boston mechanism β^k, and

(ii) the constrained Gale-Shapley mechanism GS^k is not comparable to the constrained Boston mechanism β^k by the criterion of fairness by counting when $k > 2$ and there are at least seven students and five schools.

The proof of (i) is in the appendix and the following counterexample proves point (ii).

Example 4 (Constrained Boston and constrained Gale-Shapley)**.** *Let $n \geq 7$, $I = \{i_1, ..., i_n\}$ and $S = \{s_1, \ldots, s_5\}$. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	$\ldots$	$P_{i_{n-1}}$	P_{i_n}	$\succ_{s,\, s\in S}$
s_1	s_2	s_3	s_1	s_1	s_1	s_1	s_4	i_1
$\vdots$	$\vdots$	$\vdots$	s_4	s_2	s_2	s_2	s_5	i_2
			s_5	s_3	s_3	s_3	$\vdots$	i_3
			$\vdots$	s_5	s_5	s_5		i_4
				$\emptyset$	$\emptyset$	$\emptyset$		i_5
								$\vdots$
								i_n

The outcomes of β^3 and GS^3 for this problem are specified as follows:

$$\beta^3(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & \ldots & i_{n-1} & i_n \\ s_1 & s_2 & s_3 & s_5 & \emptyset & \ldots & \emptyset & s_4 \end{pmatrix}$$

and

$$GS^3(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & \ldots & i_{n-1} & i_n \\ s_1 & s_2 & s_3 & s_4 & \emptyset & \ldots & \emptyset & s_5 \end{pmatrix}.$$

Let us compare the number of blocking students for the two matchings. On one hand, student i_4 is the only blocking student for $\beta^3(P, \succ, q)$. Indeed, the pair (i_4, s_4) blocks matching $\beta^3(P, \succ, q)$ under $(P, \succ, q)$. On the other hand, students $i_5, \ldots, i_{n-1}$ are all blocking students of $GS^3(P, \succ, q)$ because they are unmatched, each of them prefers school s_5 to being unmatched, and has higher priority than i_n under $\succ_{s_5}$. Since $n \geq 7$, there are at least two blocking students of $GS^3(P, \succ, q)$. Therefore, there are more blocking students of $GS^3(P, \succ, q)$ than $\beta^3(P, \succ, q)$. By Theorem 1, there is a problem where GS^3 is stable but not β^3.

Chen and Kesten (2017) have established that any (unconstrained) Chinese mechanism Ch^e is more stable than any Chinese mechanism $Ch^{e'}$ where $e' = ke$ for $k \in \mathbb{N} \cup \{\infty\}$. Their result and ours are similar but not corollary of each other. Indeed, Ch^1 is the Boston mechanism and Ch^∞ is the Gale-Shapley mechanism such that for a problem $(P, \succ, q)$, we can write $\beta^k(P, \succ, q) = Ch^{(1)}(P^k, \succ, q)$ and $GS^k(P, \succ, q) = Ch^\infty(P^k, \succ, q)$. Our results concern constrained Chinese mechanisms where both the parameter e and the constraint k could be a source of blocking while in Chen and Kesten (2017), the parameter e is the only source of blocking.

Let us now consider the Chinese mechanisms and fairness by counting. We use the fact that $Ch^{(1)} = \beta$ and also note that for the problem $(P, \succ, q)$ specified in Example 4, $Ch^{(1)}(P, \succ, q) = \beta^3(P, \succ, q)$ and $Ch^{(3)}(P, \succ, q) = GS^3(P, \succ, q)$. According to the conclusion in Example 4, there are more blocking students for $Ch^{(3)}(P, \succ, q)$ than $Ch^{(1)}(P, \succ, q)$. According to Chen and Kesten (2017), there is a problem where $Ch^{(3)}$ produces a stable outcome but $Ch^{(1)}$ does not. We can formulate the following result.

Theorem 3. *Let $e, m \in \mathbb{N}$ and $m > 2$.*

(i) The Chinese mechanism $Ch^{(me)}$ is more fair by stability than the Chinese mechanism $Ch^{(e)}$ (Chen and Kesten, 2017), and

(ii) the Chinese mechanism $Ch^{(me)}$ is not comparable to the Chinese mechanism $Ch^{(e)}$ by the criterion of fairness by counting.

Remark. *Two other notions, comparing mechanisms by the inclusion of blocking pairs and blocking students, have also been studied by Doğan and Ehlers (2021). These criteria are stronger than fairness by counting (if the set of blocking pairs or blocking students shrinks, then the number of blocking students does as well) and will lead to negative results for our comparisons. To see this, consider Example 4. In this example, (i_5, s_5) is a blocking pair for $SD^4(P, \succ, q)$ but not for $\beta^4(P, \succ, q)$. In addition, (i_4, s_4) is a blocking pair for $\beta^4(P, \succ, q)$ but not for $SD^4(P, \succ, q)$.*

For the comparison between different constrained Gale-Shapley, consider Example 2 where (i_1, s_3) is a blocking pair for GS^2 but not GS^1. In addition, (i_2, s_2) is a blocking pair for GS^1 but not GS^2.

4. Stability and manipulability

In this section, we search for the relationship between blocking students and manipulating students, i.e., students who may benefit from misrepresenting their preferences to the mechanisms as defined below.

Definition 3. *Let φ be a mechanism. (i) Student i is a **manipulating student** of φ at $(P, \succ, q)$ if there is a preference relation P'_i such that*

$$\varphi_i(P'_i, P_{-i}, \succ, q) \; P_i \; \varphi_i(P, \succ, q).$$

*(ii) Mechanism φ is **not manipulable** at $(P, \succ, q)$ if there is no manipulating student of φ at $(P, \succ, q)$.*

It turns out that there is a relationship between blocking students and manipulating students for the constrained Boston mechanism and the constrained Gale-Shapley mechanism.

Theorem 4. *Let $k \geq 1$. For any problem $(P, \succ, q)$,*

(i) student i is a manipulating student of the constrained Boston mechanism β^k at $(P, \succ, q)$ if and only if there is a school s such that $s \; P_i \; \beta_i^k(P, \succ, q)$ and there are less than q_s number of students who ranked school s first and have higher priority than student i at s, and

(ii) every blocking student of the constrained Boston mechanism β^k is a manipulating student of β^k. If the constrained Boston mechanism β^k is not manipulable, then it is stable.

Note that there are instances where a student is not a blocking student of the constrained Boston mechanism but a manipulating student. Consider Example 4 and suppose that there, student i_n prefers school s_1 first, s_4 next, and s_5 last. The preferences of the remaining students are unchanged. Then under β^3 student i_n is matched to school s_5 and is not a blocking student. However, she is matched to it by ranking school s_4 first as in the example. That is she is a manipulating student of β^3.

Part (iii) of the theorem characterizes the set of manipulating students of the Boston mechanism by those who missed schools that are not ranked first by "enough" students who have higher priority than i at the school in question. Clearly, if student i is a blocking student of $\beta^k(P, \succ, q)$, then there is a school s such that $s \; P_i \; \beta_i^k(P, \succ, q)$ and a student j who is matched to school s under $\beta^k(P, \succ, q)$ and has lower priority than i at s (or school s has an unassigned seat under $\beta^k(P; \succ, q)$). Then, there are less than q_s number of students who ranked school s first and have higher priority than i at s. Otherwise, student j would not have been matched to school s under $\beta^k(P, \succ, q)$ or no seat would have been left unassigned. This is the reason why the set of manipulating students includes blocking students.

Interestingly, the relationship between blocking students and manipulating students as stated in the theorem above is reversed in the constrained Gale-Shapley mechanism.

Theorem 5. *Let $k > 1$. For any problem,*

(i) every manipulating student of the constrained Gale-Shapley mechanism GS^k is a blocking student of the mechanism. If the constrained Gale-Shapley mechanism GS^k is stable, then it is not manipulable, and

(ii) a student is a manipulating student of the constrained serial dictatorship mechanism SD^k if and only if it is a blocking student. The constrained serial dictatorship mechanism SD^k is not manipulable if and only if it is stable.

In general, the constrained Gale-Shapley mechanism may be unstable and not manipulable. We illustrate this in the following example.

Example 5. *Let $I = \{i_1, \ldots, i_4\}$ and $S = \{s_1, \ldots, s_4\}$. Let $(P, \succ, q)$ be a problem where each school has one seat and the following components are specified.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$
s_1	s_1	s_2	s_3	i_1	i_4	i_3	$\vdots$
$\vdots$	s_2	s_3	s_2	$\vdots$	i_3	i_2	
	s_3	$\vdots$	$\vdots$		i_2	i_4	
	$\emptyset$				i_1	i_1	

Let us consider the constrained Gale-Shapley mechanism GS^2. We have

$$GS^2(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ s_1 & \emptyset & s_2 & s_3 \end{pmatrix}.$$

This matching is not stable at $(P, \succ, q)$ because student i_2 is unmatched, finds school s_3 acceptable while student i_4 is matched to it and $i_2 \succ_{s_3} i_4$. We claim that GS^2 is not manipulable at $(P, \succ, q)$. Only student i_2 could benefit from misrepresenting her preferences to the mechanism GS^2 because each of the other students is matched to her most preferred school. Let $P_{i_2}^{s_3}$ be a preference relation where student i_2 has ranked only school s_3 acceptable. Then,

$$GS^2(P_{i_2}^{s_3}, P_{-i_2}, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ s_1 & \emptyset & s_3 & s_2 \end{pmatrix},$$

that is, student i_2 remains unmatched even by ranking school s_3 first. (It is easy to verify that any other strategy also leaves i_2 unmatched.) Therefore, GS^2 is not manipulable at $(P, \succ, q)$. The intuition is that this ranking initiates a chain of rejections which returns to this student. Student i_2 becomes a so-called "interrupter" when she ranks school s_3 first (Kesten, 2010).

The contrasting results between the constrained Boston mechanism (Theorem 4 (i)) and the constrained Gale-Shapley (Theorem 5 (i)) can be traced back to the immediate versus deferred acceptance features and the constraint. By the deferred acceptance feature, students who are matched are neither blocking students nor manipulating students of the constrained Gale-Shapley mechanism. Blocking students and manipulating students are all students for whom the constraint is binding. That is, they are unmatched and have more acceptable schools than the ranking constraint. Because of the deferred acceptance feature, the priorities matter when a student contemplates a manipulation. No priority is violated for schools ranked within the constraint. If no student's priority is violated in any school that she has missed, then she cannot obtain a seat at any of these schools by ranking it within the constraint. To understand this conclusion, note that for any unmatched student who replaces one acceptable school with a school where her priority is not violated, the original matching remains stable under the new problem. Since she is unmatched in a stable matching, she

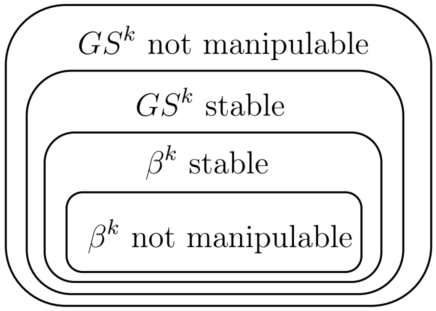

Figure 1. Set inclusion of problems for GS^k and β^k.

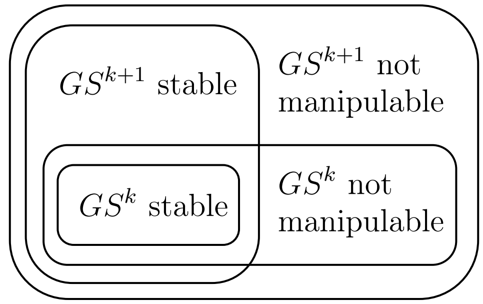

Figure 2. Set inclusion of problems for GS^k and GS^{k+1}.

will be unmatched in any other stable matching of the new problem (Roth, 1986). Then, all manipulating students are also blocking students.

These results have important implications for the relation between manipulability and stability. To see this, suppose that there is no manipulating student of the constrained Boston mechanism β^k. Then, by Theorem 4 (i), there is no blocking student for $\beta^k(P, \succ, q)$. Since β^k is individually rational, then $\beta^k(P, \succ, q)$ is stable. Suppose now that there is no blocking student for $GS^k(P, \succ, q)$. Since GS^k is individually rational, this means that $GS^k(P, \succ, q)$ is stable. Then, by Theorem 5 (i), there is no manipulating student of GS^k. We summarize these results in the following Figures 1 and 2.

The manipulation strategy under the constrained GS is to include an unlisted acceptable school in the list. But when the constrained GS is stable, all the seats of such a school are assigned to higher-priority students, and such manipulation does not help. This implies that constrained the serial dictatorship mechanism is non-manipulable and stable for the same set of problems.

Proposition 2. *Let $(P, \succ, q)$ be a problem and $k > 1$. If the constrained First-Preference-First mechanism FPF^k is stable, then the constrained Gale-Shapley mechanism GS^k is not manipulable.*

The above proposition is a surprising interplay between the two concepts for the compared mechanisms. Note that the constrained First-Preference-First mechanism and the constrained Gale-Shapley mechanism are not comparable via manipulability (Bonkoungou and Nesterov, 2021) and via fairness by stability (Example 3). However, if at some profile $(P, \succ, q)$, FPF^k is stable (while GS^k might not), then GS^k is not manipulable at $(P, \succ, q)$.

5. Conclusions

In response to various concerns, many school districts around the world have recently reformed their admissions systems. The reforms essentially contain two major changes. First, they replaced the immediate acceptance procedure (as in the Boston mechanism) with

Gale-Shapley's student-proposing deferred acceptance procedure while maintaining ranking constraints. Second, some school districts kept using the Gale-Shapley mechanism but extended the number of schools each student could report. Anecdotal evidence points to the vulnerability to manipulation and fairness as reasons for these reforms. We showed that the immediate acceptance procedure has weakly more manipulating students than blocking students and the reverse for Gale-Shapley's deferred acceptance procedure. We demonstrated that extending ranking constraints in the Gale-Shapley mechanism led to fewer blocking students.

The fact that constrained Gale-Shapley has relatively more blocking students (than it has manipulating students) suggests that in theory fairness of this mechanism might be of stronger concern than manipulability. Simultaneously, relaxing the constraint in this mechanism is guaranteed to decrease the number of blocking students. Whether this theoretical fact had any practical relevance for the reforms under consideration is an open question and requires further empirical research.

Overall, we found that all but some of the UK reforms adopted more fair matching mechanisms by stability. More specifically, the reform in the UK, where a constrained first-preference-first mechanism was replaced by a constrained Gale-Shapley mechanism with the same constraint, did not adopt a more fair matching mechanism by stability. However, we showed that the reform improved the system in at least one dimension (Proposition 2).

In addition, a few reforms can be justified using the new criterion, fairness by counting, which we explored in this paper. These reforms took place in Chicago (2010), in Ghana (2007, 2008), in Newcastle (2010), and in Surrey (2010). The other reforms in Chicago (2009), Denver (2012), thirteen provinces in China, and about sixty cities in the UK did not adopt more fair mechanisms by counting, which is not too surprising, since the new criterion of fairness by counting is a strong criterion for evaluating the reforms on the unrestricted domain of preferences. A promising future direction may be to investigate the particular preference patterns of these reforms on more restricted domains, and complement theoretical findings using empirical and computational methods.

References

Abdulkadiroğlu, A., Che, Y.-K., Pathak, P., Roth, A. E., and Tercieux, O. (2020). Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights*, 2(4):425–42.

Abdulkadiroğlu, A., Pathak, P., and Roth, A. E. (2009). Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review*, 99(5):1954–1978.

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

Ayoade, N. and Pápai, S. (2023). School choice with preference rank classes. *Games and Economic Behavior*, 137:317–341.

Balinski, M. and Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory*, 84(1):73–94.

Bonkoungou, S. and Nesterov, A. (2021). Comparing school choice and college admissions mechanisms by their strategic accessibility. *Theoretical Economics*, 16(3):881–909.

Bonkoungou, S. and Nesterov, A. (2023). Incentives in matching markets: Counting and comparing manipulating agents. *Theoretical Economics*, 18(3):965–991.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *American Economic Review*, 100(4):1860–74.

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

Coles, P., Kushnir, A., and Niederle, M. (2013). Preference signaling in matching markets. *American Economic Journal: Microeconomics*, 5(2):99–134.

Department for Education and Skills (2007). School admissions code. *Norwich, United Kingdom: Her Majesty's Stationery Office.*

Doğan, B. and Ehlers, L. (2021). Minimally unstable pareto improvements over deferred acceptance. *Theoretical Economics*, 16(4):1249–1279.

Doğan, B. and Ehlers, L. (2022). Robust minimal instability of the top trading cycles mechanism. *American Economic Journal: Microeconomics*, 14(4):556–582.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1):215–237.

Eriksson, K. and Häggström, O. (2008). Instability of matchings in decentralized markets with various preference structures. *International Journal of Game Theory*, 36(3-4):409–420.

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

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Imamura, K. and Tomoeda, K. (2022). Measuring manipulability of matching mechanisms. *Available at SSRN 4000419.*

Kesten, O. (2010). School choice with consent. *Quarterly Journal of Economics*, 125(3):1297–1348.

Kojima, F. and Manea, M. (2010). Axioms for deferred acceptance. *Econometrica*, 78(2):633–653.

Kwon, H. and Shorrer, R. I. (2020). Justified-envy minimal mechanisms in school choice. *Available at SSRN 3495266*.

McVitie, D. G. and Wilson, L. B. (1970). Stable marriage assignment for unequal sets. *BIT Numerical Mathematics*, 10(3):295–309.

Nie, H. (2007). A game-theoretical analysis of china's college admission mechanism. *China Economic Quarterly Beijing*, 6(3):899.

Niederle, M. and Roth, A. E. (2009). Market culture: How rules governing exploding offers affect market performance. *American Economic Journal: Microeconomics*, 1(2):199–219.

Pathak, P. and Sönmez, T. (2008). Levelling the playing field: Strategic and sophisticated players in the Boston mechanism. *American Economic Review*, 98(1):99–123.

Pathak, P. and Sönmez, T. (2013). School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, 103(1):80–106.

Romero-Medina, A. (1998). Implementation of stable solutions in a restricted matching market. *Review of Economic Design*, 3:137–147.

Romm, A., Roth, A. E., and Shorrer, R. I. (2020). Stability vs. no justified envy. *No Justified Envy (March 6, 2020)*.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628.

Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36(2):277–288.

Roth, A. E. (1986). On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society*, pages 425–427.

Roth, A. E. and Xing, X. (1997). Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy*, 105(2):284–329.

Shapley, L. and Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

Tang, Q. and Zhang, Y. (2021). Weak stability and pareto efficiency in school choice. *Economic Theory*, 71(2):533–552.

Wu, B. and Zhong, X. (2020). Matching inequality and strategic behavior under the Boston mechanism: Evidence from China's college admissions. *Games and Economic Behavior*, 123:1–21.

Appendix: Proofs

To simplify the exposition we divide the appendix into three subsections. In each subsection, we order the results in logical order. All mechanisms that we consider are individually rational. We only consider blocking pairs to check for (the violation of) stability. We first present a useful lemma.

Lemma 1 (Rural hospital theorem, Roth, 1986)**.** *Given a problem, let ν and μ be two stable matchings. Then,*

(i) the same set of students are matched under ν and μ, and

(ii) each school is matched to the same number of students under ν and μ, and every school which has an empty seat at one stable matching is matched to the same set of students under all stable matchings.

Appendix A: Proofs of Theorem 4 (i), Theorem 2 (i), and Proposition 2.

Proof of Theorem 4 (i). Suppose that student i is a manipulating student of β^k at $(P, \succ, q)$. Then there is P'_i such that

$$\beta^k_i(P'_i, P_{-i}, \succ, q) \; P_i \; \beta^k_i(P, \succ, q). \tag{1}$$

Since β^k is individually rational, then $\beta^k_i(P, \succ, q) \; R_i \; \emptyset$ and together with Equation 1, we have $\beta^k_i(P'_i, P_{-i}, \succ, q) = s$ for some school $s \in S$. Then student i did not rank school s first under P_i. Suppose that there are at least q_s students who ranked school s first and have higher priority than student i under $(P, \succ, q)$. Then all seats of school s would have been allocated in the first step of the immediate acceptance algorithm to some students who ranked school s first and have higher priority than student i. Then, $\beta^k_i(P'_i, P_{-i}, \succ, q) \neq s$, which is a contradiction. Therefore, there are less than q_s number of students who ranked school s first and have higher priority than student i at school s.

Let $(P, \succ, q)$ be a problem, i a student, and s a school, and suppose that $s \; P_i \; \beta^k(P, \succ, q)$ and that there are less than q_s students who ranked school s first and have higher priority than student i at school s. Thus $\beta^k(P^s_i, P_{-i}, \succ, q) = s$ and that student i is manipulating student of β^k at $(P, \succ, q)$. □

Proof of Proposition 2. We call on to two claims.

Claim 1. *Suppose that student i is matched to school s under $GS^k(P, \succ, q)$ and let P^s_i be a preference relation where she has ranked only school s as acceptable. Then student i is matched to school s under $GS^k(P^s_i, P_{-i}, \succ, q)$.*

By Roth (1982), $GS_i(P^k, \succ, q) = s$ implies that $GS_i(P^s_i, P^k_{-i}, \succ, q) = s$. We know that $(P^s_i)^k = P^s_i$. Thus, $GS^k_i(P^s_i, P_{-i}, \succ, q) = s$.

Claim 2 (Pathak and Sönmez, 2013). *Suppose that student i is a manipulating student of GS^k at $(P, \succ, q)$. Then she is unmatched under $GS^k(P, \succ, q)$.*

Suppose that $\mu = FPF^k(P, \succ, q)$ is stable at $(P, \succ, q)$. By Claim 2 every matched student under $GS^k(P, \succ, q)$ is not a manipulating student of GS^k at $(P, \succ, q)$. It is enough to show that no unmatched student under $GS^k(P, \succ, q)$ has a profitable misrepresentation. Because GS^k is individually rational, by Claim 1, we further need to restrict ourselves to manipulation by top-ranking schools. Since μ is stable at $(P, \succ, q)$, we claim that it is also stable at $(P^k, \succ, q)$. Since GS^k is individually rational, we need to check that there is no blocking pair. Suppose, to the contrary, that a pair (i, s) is a blocking pair for μ under $(P^k, \succ, q)$. Then, $s \; P_i^k \; \mu(i)$ and either (i) school s has an empty seat under μ or (ii) there is a student j such that $\mu(j) = s$ and $i \succ_s j$. Note that $s \; P_i^k \; \mu(i)$ implies that $s \; P_i \; \mu(i)$. Therefore, (i, s) is also a blocking pair for μ under $(P, \succ, q)$, thus contradicting our assumption that μ is stable at $(P, \succ, q)$. Therefore μ is stable at $(P^k, \succ, q)$. Since $GS(P^k, \succ, q)$ is the student-optimal stable matching under $(P^k, \succ, q)$,

$$\text{(2)} \qquad \text{for each student } i, \; GS_i(P^k, \succ, q) \; R_i^k \; \mu(i).$$

By Lemma 1 the same set of students are matched under μ and $GS(P^k, \succ, q)$. Let i be a student and s a school and suppose that i is unmatched under $GS(P^k, \succ, q)$ and that $s \; P_i \; GS_i(P^k, \succ, q)$. Then, student i is also unmatched under μ. Thus, $s \; P_i \; \mu(i) = \emptyset$. Because μ is stable at $(P, \succ, q)$ every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Let P_i^s denote a preference relation where i has ranked only school s acceptable. Since μ is stable at $(P^k, \succ, q)$ it is also stable at $(P_i^s, P_{-i}^k, \succ, q)$. By Lemma 1, the set of matched students is the same at all stable matchings. Thus, student i is also unmatched under $GS(P_i^s, P_{-i}^k, \succ, q)$. By Claim 1, there is no preference relation P_i' such that $GS_i^k(P_i', P_{-i}) = s$. Thus, GS^k is not manipulable at $(P, \succ, q)$.

□

Proof of Theorem 2 (i). The Boston mechanism is a special case of the First-Preference-First mechanism when every school is a first-preference-first school. Suppose that $\beta^k(P, \succ, q)$ is stable at $(P, \succ, q)$. By equation 2, each student finds the outcome $GS^k(P, \succ, q)$ at least as good as $\beta^k(P, \succ, q)$ under P^k. We also know that the Boston mechanism is Pareto efficient, that is, for each problem, there is no other matching that each student finds at least as good as its outcome (Abdulkadiroğlu and Sönmez, 2003). Therefore, the matching $\beta^k(P, \succ, q) = \beta(P^k, \succ, q)$ is Pareto efficient under P^k. Thus, $GS^k(P, \succ, q) = \beta^k(P, \succ, q)$ is stable at $(P, \succ, q)$.

We construct a problem where GS^k is stable but not β^k. Since there are at least two schools and more students than schools, let s_1, s_2 be two distinct schools and i_1, i_2 and i_3

three students. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows.

$P_{i\neq 3}$	P_3	$\succ_{s\in S}$
s_1	s_2	i_1
s_2	s_1	i_2
$\emptyset$	$\emptyset$	i_3
		$\vdots$

Since $k \geq 2$, $GS^k(P, \succ, q) = GS(P, \succ, q)$ is stable at $(P, \succ, q)$. However, the matching

$$\beta^k(P, \succ, q) = \begin{pmatrix} i_1 & i_3 & i \neq 1,3 \\ s_1 & s_2 & \emptyset \end{pmatrix}$$

is not stable because the pair (i_2, s_2) blocks it under $(P, \succ, q)$.

□

Appendix B: Proof of Theorem 1 (ii).

Lemma 2. *Let N be a subset of students and $\mu = GS(P^\ell_N, P^k_{-N}, \succ, q)$. Any blocking student for μ under $(P, \succ, q)$ is unmatched.*

Proof. We prove it by the contradiction. Suppose, to the contrary, that student i is a blocking student for μ under $(P, \succ, q)$ such that $\mu(i) = s$ for some school s. Then, there is a school s' such that $s'\ P_i\ \mu(i)$ and either (i) $|\mu^{-1}(s')| < q_{s'}$ or (ii) there is a student j such that $\mu(j) = s'$ and $i \succ_{s'} j$. Since $\mu(i) = s$, school s is one of the top x acceptable schools under P_i where $x = \ell$ if $i \in N$ and $x = k$ if $x \notin N$. Thus $s'\ P^x_i\ \mu(i) = s$ and (i, s') is a blocking pair of μ under $(P^\ell_N, P^k_{-N}, \succ, q)$, contradicting the stability of μ under $(P^\ell_N, P^k_{-N}, \succ, q)$.

□

Proof of Theorem 1 (ii). We call on to the sequential version of McVitie and Wilson (1970) of the deferred acceptance algorithm. This is a version where students apply one at a time according to a predetermined order such that in each step the highest-ordered student among the ones whose applications have not yet been tentatively accepted applies.

The idea of the proof is to consider close ranking constraints $k-1$ and k, where $k > 1$, and replace students' preference relations in P^{k-1} with the ones in P^k. Let $N \subsetneq I$ be a proper subset of I and $i \notin N$. Suppose that starting from P^{k-1} we have replaced all the preferences of students in N by their preferences in P^k and define $\widehat{P} = (P^k_N, P^{k-1}_{-N})$. Note that N may be empty. Next we consider student i: from $\widehat{P}$ we replace her preference relation P^{k-1}_i by P^k_i. Let $\nu = GS(P^{k-1}_i, \widehat{P}_{-i}, \succ, q)$ and $\mu = GS(P^k_i, \widehat{P}_{-i}, \succ, q)$.

If student i has less than k schools acceptable under P_i or is matched under ν, of course to one of her top $k-1$ most preferred schools, then $\mu = \nu$ and both matchings have the same number of blocking students. Without loss of generality suppose that student i has at least k schools acceptable under P_i and $\nu(i) = \emptyset$. Note that student i has applied to all her first $k-1$ acceptable schools in the algorithm for ν. Let s denote her k'th acceptable school under P_i^k. The algorithm for μ is a continuation of the one for ν by letting i apply to school s and completing the subsequent sequences of applications and rejections. Consider the following possible pointing sequences starting from the step at which student i applied to school s:

$$i \to s \to i_1 \to s_1 \to i_2 \to s_2 \to ...i_n \to s_n \to i_{n+1} \to \emptyset, \tag{3}$$

$$i \to s \to i_1 \to s_1 \to i_2 \to s_2 \to ...i_n \to s_n \to i_{n+1} \to s_{n+1} \tag{4}$$

where the pointing $i' \to s'$ means that student i' applies to school s' and $s' \to i'$ means that school s' rejects the application of student i'. In the sequence in equation 3, the pointing $i_{n+1} \to \emptyset$ means that student i_{n+1} applied to all her acceptable schools and thus remained unmatched. In the sequence in equation 4, the pointing $i_{n+1} \to s_{n+1}$ for which school s_{n+1} does not point to any student means that school s_{n+1} did not reject any student after i_{n+1}'s application. Note that there might be cycles and some students could appear several times in each sequence.

Let α and β denote the number of blocking students of μ and ν respectively, among the students in $I \setminus \{i, i_1, \ldots, i_{n+1}\}$ and x and y the number of blocking students of μ and ν respectively, among the students in $\{i, i_1, \ldots, i_{n+1}\}$.

Claim. *$\beta \geq \alpha$, and [$x = y$ or $x = y - 1$].*

We first prove that $x = y$ or $x = y - 1$. Note first that all students in $\{i, i_1, \ldots, i_{n+1}\} \setminus \{i, i_{n+1}\}$ are matched under μ and ν. By Lemma 2 they are not blocking students of μ and ν. Hence, the comparison of the number of blocking students of μ and ν among the students in $\{i, i_1, \ldots, i_{n+1}\}$ concerns students i and i_{n+1}. We consider two cases:

Case 1: $i = i_{n+1}$. If student i is not a blocking student of ν, then school s does not have an empty seat under ν and for each student $j \in \nu^{-1}(s)$, $j \succ_s i$. The sequence is the one in equation 3 but in simple version $i \to s \to i \to \emptyset$. Thus $\mu = \nu$ and student i is not a blocking student of μ. Thus $x = y$.

Case 2: $i \neq i_{n+1}$. In this case $i_1 \neq i$. We claim that student i is a blocking student of ν but not a blocking student of μ. The reason is that school s has accepted i's application and rejected student i_1. Thus $i \succ_s i_1$ and $\mu(i) = s$. Since $\nu(i) = \emptyset$ and $\nu(i_1) = s$, student i is a blocking student of ν. Since $\mu(i) = s$, by Lemma 2, she is not a blocking student of μ. Note now that student i_{n+1} is matched under ν. Thus by Lemma 2 she is not a blocking student of ν. Therefore, $0 = x = y - 1$ if i_{n+1} is not a blocking student of μ and $1 = x = y$ if i_{n+1} is a blocking student of μ. Overall, $x = y$ or $x = y - 1$.

We now prove that $\beta \geq \alpha$. We prove the claim that there is no student in $I \setminus \{i, i_1, \ldots, i_{n+1}\}$ who is not a blocking student of ν but a blocking student of μ. Let $j \notin I \setminus \{i, i_1, \ldots, i_{n+1}\}$ and note that $\nu(j) = \mu(j)$. If student j is matched under ν, then she is not a blocking student of ν and μ. Suppose that $\nu(j) = \emptyset$, and for a contradiction, that she is not a blocking student of ν but a blocking student of μ under $(P, \succ, q)$. There is a school s' such that $s' \; P_j \; \mu(j)$ and either (i) school s' has an empty seat under μ or (ii) there is a student j' such that $\mu(j') = s'$ and $j \succ_{s'} j'$. Consider (i). As above, school s' has an empty seat under ν. Therefore, student j is also a blocking student of ν, contradicting our assumption. Consider (ii). Suppose that school s' did not reject any student in any step from the step at which student i applies to school s to the end. Then $\mu(j') = \nu(j') = s'$ and student j is also a blocking student of ν. Suppose that school s' has rejected a student j'' such that $\nu(j'') = s'$. Since $\mu(j') = s'$, then $j' \succ_{s'} j''$. Thus $j \succ_{s'} j''$ and since $\nu(j'') = s'$, student j is a blocking student of ν, contradicting again our assumption. Therefore there is no student who is simultaneously not a blocking student of ν but a blocking student of μ under $(P, \succ, q)$. Thus $\alpha \geq \beta$.

By this claim, $\alpha + y \geq \beta + x$. Therefore, ν has a weakly larger number of blocking students than μ. Starting from $N = \emptyset$ and successively replacing students' preferences in any order we conclude that $GS(P^{k-1}, \succ, q)$ has weakly more number of blocking students than $GS(P^k, \succ, q)$.

Finally, we describe a problem where the outcome of GS^ℓ has more blocking students than the outcome of GS^k. Let $(P, \succ, q)$ be a problem where each school has one seat, each student has k acceptable schools, and such that students have a common ranking of schools. Then, $GS^k(P, \succ, q) = GS(P, \succ, q)$. Thus $GS^k(P, \succ, q)$ is stable at $(P, \succ, q)$. Let s be the school that students have ranked at the k'th position starting from the top. Since there are more students than schools and $k > \ell$, at least one student is not matched under $GS^\ell(P, \succ, q)$ and no student is matched to school s even though every student prefers it to be unmatched. Therefore, there are more blocking students for $GS^\ell(P, \succ, q)$ than $GS^k(P, \succ, q)$ under $(P, \succ, q)$. □

Appendix C: Proof of Theorem 4 (ii), 5 (i), 5 (ii).

Proof of Theorem 4 (ii). Let i be a blocking student of $\mu = \beta(P^k, \succ, q)$. There is a school s such that the pair (i, s) blocks μ under $(P, \succ, q)$. Then, $s \; P_i \; \mu(i)$ and either (a) school s has an empty seat under μ or (b) there is a student j such that $\mu(j) = s$ and $i \succ_s j$. We claim that student i did not rank school s first under P_i. Otherwise, school s has rejected student i at the first step of the Boston algorithm under $(P^k, \succ, q)$. This is because $k > 1$ and the top-ranked schools are considered under β^k. This contradicts the assumption that school s has an empty seat or has accepted student j with $i \succ_s j$. Let P_i^s be a preference relation where i has ranked school s first. Since s has an empty seat under $\beta^k(P, \succ, q)$ or has accepted student j with $i \succ_s j$, there are less than q_s students who have ranked school s first under P^k and have a higher priority than i under $\succ_s$. Therefore, $\beta_i^k(P_i^s, P_{-i}, \succ, q) = s$. Since $s \; P_i \; \mu(i)$, i is a manipulating student of β^k at $(P, \succ, q)$. Finally, suppose that β^k is not manipulable at $(P, \succ, q)$. Then, there is no manipulating student and thus, there is no blocking student. Since β^k is individually rational, then $\beta^k(P, \succ, q)$ is stable at $(P, \succ, q)$.

Proof of Theorem 5 (i): We prove this part by contradiction. Suppose that student i is a manipulating student of GS^k at $(P, \succ, q)$ but is not a blocking student of $\mu = GS^k(P, \succ, q)$ under $(P, \succ, q)$. By Claim 2, i is unmatched under $GS^k(P, \succ, q)$. Let s be a school such that $s \; P_i \; \mu(i)$. Then, $|\mu^{-1}(s)| = q_s$ and every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Let P_i^s be a preference relation where i has ranked only school s as an acceptable school. Since μ is stable at $(P^k, \succ, q)$, it is also stable at $(P_i^s, P_{-i}^k, \succ, q)$. This follows from the fact that $\mu(i) = \emptyset$ and that every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. By Lemma 1, the set of unmatched students is the same under μ and $GS(P_i^s, P_{-i}^k, \succ, q)$. Thus, i is also unmatched under $GS^k(P_i^s, P_{-i}, \succ, q)$. By Claim 1, there is no misreport by which student i is matched to s. Since s has been chosen arbitrarily, i is not a manipulating student of GS^k at $(P, \succ, q)$, a conclusion that contradicts our assumption. Therefore student i is a blocking student of $GS^k(P, \succ, q)$ under $(P, \succ, q)$. Finally, suppose that $GS^k(P, \succ, q)$ is stable. Then, there is no blocking student, and thus a manipulating student of GS^k at $(P, \succ, q)$. Hence, GS^k is not manipulable at $(P, \succ, q)$.

Proof of Theorem 5 (ii). By Theorem 5 (i), every manipulating student of SD^k is a blocking student of $SD^k(P, \succ, q)$. Let i be a blocking student of $\mu = SD^k(P, \succ, q)$. Then at i's turn, there is no seat left among her top k acceptable schools. She is then left unmatched while she has ranked a school s as acceptable and below the position k which still has a seat available. Let P_i^s be a preference relation where i ranks s first. Then $SD_i^k(P_i^s, P_{-i}, \succ, q) = s$. Therefore, student i is a manipulating student of SD^k at $(P, \succ, q)$. $\square$