REFORMS MEET FAIRNESS CONCERNS IN SCHOOL AND COLLEGE ADMISSIONS

SOMOUAOGA BONKOUNGOU* AND ALEXANDER NESTEROV†

ABSTRACT. We evaluate a series of recent reforms of matching systems around the world. These reforms responded to concerns that the matching mechanisms in use were unfair. Surprisingly, the mechanisms remained unfair even after the reforms: in particular, the new mechanisms may induce outcomes with *blocking students* who did not receive the schools they desire and deserve.

We show, however, that the reforms introduced matching mechanisms that are *more fair* than the mechanisms they replaced. First, the new mechanisms are *more fair by stability*: whenever the old mechanism does not have a blocking student, the new mechanism does not have a blocking student either. Second, some reforms introduced mechanisms that are *more fair by counting* the blocking students: the old mechanism always has at least as many blocking students as the new mechanism. Most of the results remain true in two models where some students are sincere while others are sophisticated, and where all students are semi-sophisticated. We also show that stability and manipulability are strongly logically related.

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

JEL Classification: C78, D47, D78, D82

---

*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

Date: August 2, 2021.

The personal acknowledgements will be added in the final version of the paper.

Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged. The research is partially financed by the RFBR grants 20-01-00687, 19-01-00762. Somouaoga Bonkoungou gratefully acknowledges financial support from the European Commission, Marie Skłodowska-Curie individual fellowship, project 890648.

1. Introduction

In the past two decades, there has been a wave of reforms of matching systems around the world, ranging from college admissions systems in Chinese provinces, secondary public school admissions systems in multiple districts in Ghana, to public school admissions systems in multiple cities in the US and the UK. In this paper, we discuss the motives behind these reforms and evaluate their results.

The old matching systems were criticized because they were vulnerable to gaming and were unfair. The most vivid example is, perhaps, the 2007 major reform in England, which covers 146 local school admissions systems. According to the Secretary of State, Alan Johnson, the aim of the reform was to *"ensure that admission authorities – whether local authorities or schools – operate in a fair way"* (School Admissions Code, 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.* According to this principle, a student who ranks a school higher in her list receives a higher admission priority at this school compared to the students who rank this school lower. Before the reform, as many as one-third of the schools in England used this principle.

In 2009, the Chicago authorities implemented a similar reform in 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 college prep preferences"* (Pathak and Sönmez, 2013). The same Boston mechanism has also been used for college admissions in several provinces in China. 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 in 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 an unfairness issue with the old mechanisms: they can induce a matching with a so-called *blocking student*, that is, a student who prefers a school over her matching while at least one seat at this school has been assigned to a student with a lower priority (or even left empty). The blocking student desires and deserves this seat, yet she has not been assigned to it. A matching with no blocking student is called *stable* and is viewed as a fair outcome as they eliminate “justified envy" (Abdulkadiroğlu and Sönmez, 2003).[1]

Another reason why the old mechanisms induced matchings with blocking students are ranking constraints: each student was allowed to rank-list only a limited number of schools, typically between 3 and 5 (Pathak and Sönmez, 2013). Even in New York, where the ranking constraint is 12, around 25% of students report a complete list of 12 schools, while only 5% report 9, 10, or 11 schools, suggesting that 20% of the students in New York 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. In this paper, we consider all blocking students, whether it concerns listed schools for which admissions authorities can verify priority violation or unlisted schools which lead to dissatisfaction (see Calsamiglia et al., 2010).

Perhaps surprisingly, the new mechanisms are also unfair as each of them might induce a matching with a blocking student. We show, however, that the new mechanisms are more fair compared to the old alternatives by using the two fairness criteria.

The first fairness criterion is based on the set-inclusion of instances with stable outcomes. One mechanism is **more fair by stability** than a second mechanism if it induces a stable matching whenever the second mechanism induces a stable matching, and the reverse is not true for some instances.

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

Reforms	From	To	more fair by stability?		more fair by counting?	
			Arbitrary priority	Common priority	Arbitrary priority	Common priority
UK(54), 2007/11	FPF^k	GS^k	not comparable*,**	more	not comparable*,**	not comparable*,**
Chicago, 2009 UK(4), 2007	β^k	GS^k	more	more*	not comparable*,**	not comparable*,**
Chicago, 2010 Ghana, 2007/08 UK(2), 2010	GS^k	GS^{k+1}	more**	more*,**	more**	more*,**
China(13), 2001/12	β	$Ch^{(e)}$	more	more	not comparable*,**	not comparable*,**

Table 1. Comparison of the matching mechanisms by fairness criteria.

Notes: Each row represents a comparison of the mechanism in the third column to the mechanism in the second column according to one of the two fairness notions. Asterisks * and ** show which results are robust in strategic settings (see Section 4 for details.) The complete list of the the UK local matching systems and Chinese provinces that underwent the reforms can be found in Pathak and Sönmez (2013) and in Chen and Kesten (2017), respectively.

Our main results using this criterion support that most of the reforms have adopted matching mechanisms that are more fair by stability (see Table 1, column 4). For example, in China, this is true for half of its provinces (Chen and Kesten, 2017). In Chicago, the mechanism adopted after the 2009 reform is more fair by stability than the one previously used (Theorem 1); the one adopted after the 2010 reform is also more fair by stability than the mechanism adopted in 2009 (Theorem 2).

The only exception is the 2007 reform in England — in the districts where some but not all schools used the first-preference-first principle. For each of these districts, there are instances where the matching was stable under the old mechanism but is not stable under the new mechanism (Example 1). However, we restored the result when schools in such a district have a common priority order, e.g., based on students' grades or a single lottery (see Table 1, column 5; Proposition 1).

The second fairness criterion is stronger than fairness by stability and is based on counting the number of blocking students. A mechanism is **more fair by counting** (the number of blocking students) than a second mechanism if for each instance the second mechanism has at least as many blocking students as the first mechanism, and sometimes strictly more.[2]

Our main result for this criterion supports few reforms (see Table 1, columns 6,7). Broadly, these reforms involve extending ranking constraints in the Gale-Shapley mechanism (Theorem 4). This took place in Chicago (2010), in Ghana (2007, 2008), in Newcastle (2010), and in Surrey (2010) (Pathak and Sönmez, 2013). For all other reforms, this criterion is too strong. We show that after these reforms the number of blocking students may increase (Examples 3 and 4).

Note that so far we have considered fairness separately from strategic issues, which is standard in the literature as certain mechanisms are not stable under truthful reporting, yet are stable in equilibrium (e.g., Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006; Chen and Kesten, 2017). Yet, each of the mechanisms that we study also might give students an incentive to misreport their preferences. This poses a serious methodological difficulty on how to measure the unfairness induced by the first-preference-first principle and more generally by manipulable mechanisms. On the one hand, when students truthfully report their preferences, the set of blocking students is well understood but does not necessarily represent the blocking students of the expected outcomes. On the other hand, when students are well sophisticated, the outcome typically does not exhibit any blocking pair. To address this difficulty, we develop two settings.

In the first setting, some students are sophisticated and best respond while others are truthful (as in Pathak and Sönmez, 2008).[3] In the second setting, all students are semi-sophisticated. In these settings, unfairness arises from the inability of some students to best respond. We show that many of our

[2]To our knowledge, this criterion has been first used by Roth and Xing (1997). Niederle and Roth (2009), Eriksson and Häggström (2008) and Dogan and Ehlers (2020a) count the number of blocking pairs, which does not allow comparisons in our setting (see Remark in section 3).

[3]König et al. (2019) study this setting experimentally by assuming that some students report truthfully and allowing others to misreport.

results are robust to these new settings (see the asterisk icons in Table 1). We present this argument in section 4.

We also find a strong logical relationship between stability and manipulability. Under the constrained Gale-Shapley mechanism, the stability of the outcome at any instance implies non-manipulability at this instance; while for the constrained Boston mechanism it is reversed: manipulability implies instability (Corollary 1 and Figure 1). For the constrained serial dictatorship mechanism the two concepts are equivalent: its outcome at any instance is stable if and only if the mechanism is not manipulable at this instance (Proposition 3).

A more subtle relationship between stability and manipulability can be seen in the reform in England. After this reform, the mechanisms in most school districts did not become less manipulable (Bonkoungou and Nesterov, 2021) and they did not become more fair by stability either (Example 1 below). However, the reform was successful according to the following criterion: if the reform disrupted fairness — by producing an unstable matching while it would have been stable before the reform — the new matching is not vulnerable to gaming (Proposition 4).

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

Among the 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 (Dogan and Ehlers, 2020b) and the set of blocking triplets (i, j, s) – student i blocking the matching of school s with student j (Kwon and Shorrer, 2019). The result holds for any stability comparison that satisfies few basic properties (Dogan and Ehlers, 2020b).

Among the 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 (Dogan and Ehlers, 2020a; Tang and Zhang, 2020; Kwon and Shorrer, 2019). Independent from the present work, Dogan and Ehlers (2020a) also use the fairness by counting criterion to show that among efficient mechanisms EADA is not the most fair by counting, unless the priority profile satisfies few acyclicity conditions.

The first paper that studied the constrained mechanisms is Haeringer and Klijn (2009). They study the stability of the 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.[4] 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 longer list. Therefore, when the Nash equilibrium outcomes of the constrained Gale-Shapley with a longer list are all stable, the Nash equilibrium outcomes of the constrained Gale-Shapley with a shorter list are also stable.

The rest of the paper is organized as follows. Section 2 introduces the model and the mechanisms. Section 3 presents the fairness comparisons and section 4 extends them to strategic settings. Section 5 studies the relationship between stability and manipulability. We present most of the proofs in the appendix.

2. Model

In a school choice model (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), there is a finite and non-empty set I of students with a generic element i and a finite and non-empty set S of schools with a generic element s.

Each student i has a **strict preference relation** P_i over $S \cup \{\emptyset\}$, where $\emptyset$ represents the **outside option** for this student. For each student i, let

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

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 $\emptyset\ P_i\ s$. The list $P = (P_i)_{i \in I}$ is a **preference profile**. 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 the students in I'.

Each school s has a **strict priority order** $\succ_s$ over the set I of students, and a **capacity** q_s (a natural number indicating the number of its available seats). The list $\succ = (\succ_s)_{s \in S}$ is a **priority profile** and $q = (q_s)_{s \in S}$ is a **capacity vector**. We extend each priority order $\succ_s$ of school s to the set 2^I of subsets of students and assume that this extension is responsive to the priority order $\succ_s$ over I as follows. The priority order $\succ_s$ of school s is **responsive** (Roth, 1986) if

- for each $i, j \in I$ and each $I' \subset I \setminus \{i, j\}$ such that $|I'| < q_s - 1$, we have, (i) $I' \cup \{i\} \succ_s I'$, and (ii) $I' \cup \{i\} \succ_s I' \cup \{j\}$ if and only if $i \succ_s j$ and
- for each $I' \subset I$ such that $|I'| > q_s$, we have $\emptyset \succ_s I'$.

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.

A **mechanism** φ is a function which 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

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

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 $(P, \succ, q)$.

2.1. **Mechanisms.** We are interested in the mechanisms that were used either before or after the reforms. We first describe the 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 whom school s has tentatively accepted at this step.

The algorithm stops when no student is rejected and thus each student is either accepted at some step or has been rejected by all of her acceptable schools and is unmatched. The tentative acceptances become final at this step. Let $GS(P, \succ, q)$ denote the obtained matching.

Serial Dictatorship. When schools have the same priority order, we call the Gale-Shapley mechanism the serial dictatorship mechanism.[6] Let $SD(P, \succ, q)$

[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.

denote the matching assigned by the serial dictatorship mechanism to the problem $(P, \succ, q)$.

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 this school. 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 $FPF(P, \succ, q)$ denote the matching assigned to the problem $(P, \succ, q)$ by the First-Preference-First mechanism.

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 which 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 $\beta(P, \succ, q)$ denote the matching assigned by the Boston mechanism to the problem $(P, \succ, q)$.

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).*

Constrained mechanisms. Haeringer and Klijn (2009) first observed that in practice matching mechanisms often have **ranking constraints** and students are allowed to report only a limited number of schools. This means that schools that are listed below a certain position are not considered. Formally, let $k \in \{1, \ldots, |S|\}$ and let student i have x acceptable schools. The truncation after the k'th acceptable school (if any) of the preference relation P_i is the preference relation P_i^k with $\min(x, k)$ acceptable schools such that all schools are ordered as in P_i. Let $P^k = (P_i^k)_{i \in I}$. The constrained version φ^k of the mechanism φ is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $\varphi(P^k, \succ, q)$. That is, $\varphi^k(P, \succ, q) = \varphi(P^k, \succ, q)$.

Note that the constrained Gale-Shapley mechanism is stable at the reported preference, but not under the true preferences.

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 $Ch^{(e)}(P, \succ, q)$ denote the matching assigned by the mechanism to $(P, \succ, q)$.[7]

3. Comparison of Mechanisms

In this section we compare mechanisms according to two criteria: fairness by stability and fairness by counting.

3.1. Fairness by stability.

Our starting point is a comparison according to the set inclusion of the problems where mechanisms 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.

This criterion is less demanding in the sense that it does not take into account the problems where mechanisms produce unstable outcomes. However, it does 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. We demonstrate this in the following example.

Example 1. *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.)*

[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.

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.[8] *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.

In this example, we illustrate how the constrained GS mechanism has shortened the chains needed to reach a stable matching. It is well known that

[8]Note that this matching is both the student-optimal and the school-optimal stable matching.

this type of chains lead to unambiguous welfare losses (in terms of Pareto efficiency): each student in the chain is worse off, all other students are unaffected (Kesten, 2010).[9] However, under the Boston mechanism, (where all schools are first-preference-first schools) there is no such chain. The following result is an implication of this fact.

Theorem 1. *Suppose that there are at least two schools and let $k > 1$. The constrained Gale-Shapley mechanism GS^k is more fair by stability than the constrained Boston mechanism β^k.*

Similarly, when schools have a common priority order, there is no such chain in the Gale-Shapley mechanism. We restore the result for this case.

Proposition 1. *Suppose that there are at least two schools and at least one first-preference-first school. Let $k > 1$ and suppose that schools have a common priority. The constrained serial dictatorship mechanism SD^k is more fair by stability than the constrained First-Preference-First mechanism FPF^k.*

The constrained GS with shorter and longer lists mechanisms can also be compared according to this criterion. However, the intuition for this result is different. When the constrained Gale-Shapley with shorter lists is stable, the restriction has no effect on the outcome.

Lemma 1. *Let $(P, \succ, q)$ be a problem and $k > 1$. Then $GS^k(P, \succ, q)$ is stable if and only if $GS^k(P, \succ, q) = GS(P, \succ, q)$.*

Then, when the constraint in GS^k does not affect the outcome, longer constraints like GS^{k+1} will not affect the outcome either.

Theorem 2. *Suppose that there are at least three schools and let $k > \ell$. Then, the constrained Gale-Shapley mechanism GS^k is more fair by stability than GS^ℓ.*

Finally, we consider the Chinese mechanisms. These mechanisms are known to be comparable in terms of fairness by stability, but only in case one tier is a multiple of another (Chen and Kesten, 2017). We present this result for completeness.

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

Theorem 3 (Chen and Kesten, 2017)**.** *For each $M, M' \in \mathbb{N}$ such that $M' = mM$ for some $m \in \mathbb{N}$, the Chinese mechanism $Ch^{(M')}$ is more fair by stability than $Ch^{(M)}$.*

3.2. **Fairness by counting.** In this section we present the results for a stronger comparison criterion: the number of blocking students. With this criterion, the mechanisms can be compared at each problem (even where both induce unstable outcomes).

Definition 2. *An individually rational mechanism φ' is* ***more fair by counting*** *(the blocking students) than an individually rational 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 stronger than fairness by stability considered earlier. If a 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. Our main result with this concept is a strengthening of the comparison between different constraints of the Gale-Shapley mechanism (Theorem 2).

We 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 that there are two blocking students in total.

This turns out to be true in general. When a student shortens 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.

Theorem 4. *Suppose that there are at least two schools and let $|S| > k > \ell \geq 1$. The constrained Gale-Shapley mechanism GS^k is more fair by counting than GS^ℓ.*

Next, we show that the other comparisons do not extend to this stronger criterion. The first example shows that the constrained Boston mechanism is not comparable to the constrained GS.

Example 3 (Constrained Boston and GS)**.** *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 $\beta^3(P, \succ, q)$ 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 $\beta^3(P, \succ, q)$. Under Theorem 1, there is a problem where GS^3 is stable but not β^3.

Next, the symmetric Chinese parallel mechanisms are also not comparable in terms of fairness by counting.

Example 4 (Chinese parallel)**.** *We consider Example 3. Consider the Chinese mechanisms $Ch^{(1)} = \beta$ and $Ch^{(3)}$ and note that for the problem $(P, \succ, q)$ specified in that example, $Ch^{(1)}(P, \succ, q) = \beta^3(P, \succ, q)$ and $Ch^{(3)}(P, \succ, q) =$*

$GS^3(P,\succ,q)$. According to the conclusion in Example 3, 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.

The overall rankings with respect to the two criteria are presented in Table 1.

Remark. *Dogan and Ehlers (2020a) compare mechanisms by the inclusion of the blocking pairs and blocking students. However, these criteria are stronger than fairness by counting (if the set of blocking pairs or blocking students shrinks, so does the number of blocking students) and will lead to negative results for our comparisons. To see this, consider Example 3. In this example, (i_5,s_5) is a blocking pair for $SD^4(P,\succ,q)$ but not $\beta^4(P,\succ,q)$. In addition, (i_4,s_4) is a blocking pair for $\beta^4(P,\succ,q)$ but not $SD^4(P,\succ,q)$.*

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

4. Fairness in strategic settings

So far we evaluated the fairness of the outcomes of mechanisms when students report their preferences truthfully. This approach is standard starting with the seminal paper by Abdulkadiroğlu and Sönmez (2003), but it has a limited interpretation since the mechanisms in question also incentivize students to misreport their preferences.

We face the following methodological difficulty. On the one hand, when students sincerely report their true preferences, the set of blocking students is well understood but does not necessarily represent the blocking students of the expected outcomes.

On the other hand, when students are well sophisticated, potential blocking students may misrepresent their preferences such that the expected outcome does not exhibit any blocking pair. For example, in a game induced by the Boston mechanism, all equilibrium outcomes are stable (Ergin and Sönmez, 2006; Haeringer and Klijn, 2009), but students are very unlikely to reach an equilibrium outcome even in lab settings (Featherstone and Niederle, 2016). Thus, models of complete truthfulness and complete sophistication are both not conclusive.

To address this difficulty, we introduce a new method of studying the fairness of mechanisms. The general principal is that some students are not well sophisticated to best respond and thus report truthfully; and we study whether the mechanisms are fair to these students. We propose two settings. In the first setting, some students are completely sincere, while others are completely sophisticated. In the second setting, all students are semi-sophisticated: they are truthful but avoid very competitive schools, that are obviously infeasible and will be filled immediately by students of the highest priority.

We check if the results from the previous section are robust to this analysis. Note that the negative results (denoted as "not comparable" in Table 1) remain true in these settings.

4.1. **Sincere and Sophisticated: reforms in Chicago.** We use the reforms in Chicago as an illustration. The Chicago selective high school system called for a reform in the middle in their admissions process in 2009 and initiated another reform in 2010. The admission to each of these schools is very competitive. Each school uses a common priority based on students' composite scores. Prior to 2009, the school board was using the Boston mechanism with a ranking constraint of 4 schools. In 2009, they replaced this mechanism with the serial dictatorship with the same ranking constraint. A year after, they maintained the serial dictatorship rule but extended the ranking constraint to 6 schools. According to the Chicago school board, the motivation was that *"high-scoring kids were being rejected simply because of the order in which they listed their college prep preferences"* in the Boston mechanism, which suggests that some students couldn't figure out the optimal play so as to avoid losing their high priorities.

Following Pathak and Sönmez (2008), we consider a model with sincere students and sophisticated students. Let N denote the set of sincere students and M the set of sophisticated students such that $N \cup M = I$ and $N \cap M = \emptyset$. Sincere students always report their preferences truthfully while sophisticated students may strategically misreport their preferences. The idea is to measure unfairness with respect to the set of blocking students among sincere students.

Any mechanism φ and a pair $(\succ, q)$ induce a normal form game such that the students are the players, the strategies are the preference reports and the outcome function is $\varphi(., \succ, q)$. Let (φ, P) denote the game induced by φ. A strategy profile $P' = (P_N, P'_M)$ is a Nash equilibrium of the game (φ, P) if for each sophisticated student $i \in I$, there is no strategy P''_i such that $\varphi_i(P''_i, P'_{-i}, \succ, q)\ P_i\ \varphi_i(P', \succ, q)$.

We first consider the 2009 reform. We begin by describing the equilibrium outcome in a game induced by a constrained Boston mechanism.

Lemma 2. *Let $k > 1$ and suppose that some students are sincere while others are sophisticated. For any problem where schools have a common priority, the constrained Boston mechanism β^k has a unique Nash equilibrium outcome where sincere students play truthfully and sophisticated students best respond.*

The proof is constructive and rests on an important argument by Pathak and Sönmez (2008). Indeed, the authors show that the Nash equilibrium outcomes of the Boston mechanism are equivalent to the set of stable matchings with respect to some modified economies. These modifications are changes in the priorities such that every sincere student is ranked below every sophisticated student at every school that the sincere student did not rank first. The priority among sincere students are adjusted such that students who rank a school higher receive higher priority. Except for these changes the priorities stay the same as the original ones. Let $\hat{\succ}$ denote the priority profile obtained. The outcome can be produced in two steps as follows:

Step 1: students are matched according to the serial dictatorship mechanism for the problem $(P^1_N, P_M, \succ, q)$. Let $N' \subset N$ be the subset of sincere students who are unmatched and q' the profile of remaining capacities.

Step 2: the unmatched students N' are matched according to the Gale-Shapley mechanism for the augmented economy $(P^k_{N'}, \hat{\succ}, q')$.

The full proof of the lemma is a straightforward adjustment to consider the constraint. A result without sincere students is available in Haeringer and Klijn (2009).[10]

Let us consider the 2010 reform. Recall that the school board has maintained the serial dictatorship mechanism but extended the ranking constraint from 4 to 6 schools. As it turns out, the game under the constrained serial dictatorship mechanism also has a unique Nash equilibrium outcome.

Lemma 3. *Let $k > 1$ and suppose that there are sincere students and sophisticated students. For any problem where schools have a common priority, the constrained serial dictatorship mechanism SD^k has a unique Nash equilibrium outcome (where sincere students play truthfully and sophisticated students best respond).*

The proof of this result is constructive. It consists of showing that the matching $SD(P_N^k, P_{-N}, \succ, q)$, where the constraint only applies to sincere students, is the unique Nash equilibrium outcome of the game. The reason is that this is the unique stable matching under $(P_N^k, P_{-N}, \succ, q)$ and every sophisticated student who is part of a blocking pair with a school can obtain a seat at this school by ranking it first and picking it at her turn.

Proof. Let (P_N, P'_{-N}) be a Nash equilibrium of the game (SD^k, P) where sincere students report truthfully P_N and sophisticated students report P'_{-N}.

We claim that $SD^k(P_N, P'_{-N}, \succ, q)$ is stable at $(P_N^k, P_{-N}, \succ, q)$. Clearly, under the constrained profile, there is no blocking pair involving a sincere student. Suppose that there is a blocking pair (i, s) where i is a sophisticated student. Then at least one seat of s is either unassigned or has been assigned to a student with lower ranking according to the order in which students pick schools. Then at i's turn, school s still has a seat available. Let P_i^s be a preference relation where i has ranked school s first. Then at her turn, student i will pick school s under $(P_N^k, P_i^s, P'_{-(N\cup\{i\})})$. This contradicts the assumption that (P_N, P'_{-N}) is a Nash equilibrium of the game (SD^k, P).

[10]The detail of the proof is available upon request.

Next $SD^k(P_N, P'_{-N}, \succ, q)$ is individually rational under P. Clearly, it is individually rational to each sincere student. If it is not individually rational to a sophisticated student i under P_i, then she has ranked under P'_i a school s that is unacceptable under P_i and to which she is matched to under $SD^k(P_N, P'_{-N}, \succ, q)$. Student i is better reporting a preference relation where she finds no school acceptable. This contradicts the fact that (P_N, P'_{-N}) is a Nash equilibrium of the game (SD^k, P).

Since schools have the same priority order, there is a unique stable matching under $(P^k_N, P_{-N}, \succ, q)$, and thus the Nash equilibrium is unique. □

The game under the constrained serial dictatorship mechanism can be interpreted as if the constraint applies only to sincere students. When the constrained Boston mechanism has a stable Nash equilibrium outcome, then no sincere student loses her priority to a sophisticated or another sincere student.

Proposition 2. *Let $k > 1$ and suppose that there are more than k schools, and schools have a common priority. In equilibrium, the constrained serial dictatorship mechanism SD^k is more fair by stability than the constrained Boston mechanism β^k.*

Proof. Let $(P, \succ, q)$ be a problem and suppose that the unique Nash equilibrium outcome μ of the game (β^k, P) is stable at $(P, \succ, q)$. Then every sincere student i finds the school that she is matched to under μ (if any) acceptable under P^k_i. Therefore, μ is stable at $(P^k_N, P_{-N}, \succ, q)$ and thus μ is the unique Nash equilibrium outcome of the game (SD^k, P), which is also stable at $(P, \succ, q)$.

The proof of Theorem 1 presents a problem where SD^k is stable and β^k is not. □

A more restrictive constraint creates more blocking students.

Theorem 5. *Let $k > \ell > 1$ and suppose that there are sincere students, who play truthfully, and sophisticated students, who best respond. For any problem where schools have a common priority, the Nash equilibrium outcome of the constrained serial dictatorship mechanism SD^ℓ has at least as many blocking students as the Nash equilibrium outcome of the constrained serial dictatorship mechanism SD^k.*

Next we consider another setting, where all students can behave strategically but to a limited extent.

4.2. **Semi-sophisticated: reforms in Ghana and UK.** The school boards of the admissions system in Ghana and UK have called for a reform in 2007 and 2010, respectively. A year later, the school board of the admissions system in Ghana pursued its reform. These reforms consisted of extending the ranking constraint of schools under the Gale-Shapley mechanism.

Haeringer and Klijn (2009) show that in the constrained Gale-Shapley mechanism, students cannot do better than to select some acceptable schools and ranking them according to their true preferences. In an experiment, Calsamiglia et al. (2010) study how subjects elaborate their strategic choices under ranking constraints. They show that, in experiment, a higher proportion of subjects preserve their relative rankings for the first three schools that they have listed, in accordance with Haeringer and Klijn (2009). They also test and show that a significantly higher proportion of students lower the position of more competitive schools in their reports. Specifically, in their experiment, the student's behaviour is affected by the presence and the position of a so-called district school, where the student is among the top priority students.

Motivated by these observations, we consider a model in which each student is sophisticated but only to a limited extent. Informally, we assume that students respond by dropping schools that are obviously infeasible, but otherwise behave truthfully. More formally, we say that a student is *guaranteed her first choice* s if she is among the q_s-highest priority students. If there are at least q_s students who are guaranteed their first choice s then s is *competitive*. All students behave strategically by dropping competitive schools from their reports but list the remaining schools according to their true preferences.

Let $(P, \succ, q)$ be a problem. We say that a profile P' of strategies is a Nash equilibrium of the game (GS^k, P) with semi-sophisticated students if students behave strategically by dropping competitive schools, that is

- every student i who is guaranteed her first choice reports truthfully: $P'_i = P_i$,
- every other student i drops competitive schools when the constraint is binding:

– if there are at most k acceptable schools under P_i or all acceptable schools under P_i are competitive then $P'_i = P_i$. Otherwise,
– student i reports as acceptable only schools that are not competitive and reports them according to her true ranking: every competitive school that is acceptable under P_i is not acceptable under P'_i and all schools s, s' that are not competitive and that are acceptable under P_i are also acceptable under P'_i, and $s \; P'_i \; s'$ if and only if $s \; P_i \; s'$.

Note that a student does not need to drop all competitive schools from her acceptable schools: she keeps dropping the schools as long as the constraint remains binding. For example, a student may drop some competitive schools from her acceptable schools even though she did not drop any competitive school from her top k acceptable schools. In this case, the top k acceptable schools of the strategy P'_i is the same as the top k acceptable schools of P_i. Consider the following example.

Example 5. *Let $I = \{1, 2, 3\}$ and $S = \{s_1, s_2, s_3\}$ and suppose that each school has one seat. We specify a problem $(P, \succ, q)$ as follows:*

P_1	P_2	P_3	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$
s_1	s_1	s_2	1	2	3
s_2	s_2	s_1	2	3	1
s_3	s_3	s_3	3	1	2

Student 1 is guaranteed her first choice school s_1. Under the mechanism GS^2, students 2 and 3 are constrained and drop s_1 from their acceptable schools. The Nash equilibrium of the game (GS^2, P) is the strategy profile P' where student 2 and 3 rank s_2 first and s_3 second. The outcome is the student-optimal stable matching.

In the setting with semi-sophisticated students we are able to extend the comparison of the constrained GS with respect to fairness by counting (Theorem 4).

Theorem 6. *Let $k > \ell > 1$ and suppose that students are semi-sophisticated and behave strategically by dropping competitive schools but are sincere otherwise. For any problem, the Nash equilibrium outcome of GS^ℓ has at least as many blocking agents as the Nash equilibrium outcome of GS^k.*

This result implies the results for fairness by stability and for the case of common priorities as special cases (see Table 1). Overall, most of the comparisons remain true within at least one of the two settings.

5. Stability and manipulability

In this section, we will elucidate the relation between blocking students and manipulating students, i.e., those who may benefit from misrepresenting their preferences to the mechanisms. We provide the definitions 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 $\hat{P}_i$ such that*

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

(ii) The 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 strong relation between blocking students and manipulating students for the constrained Boston mechanism and the constrained Gale-Shapley mechanism. Interestingly, these relations for the two mechanisms are reversed.

Theorem 7. *Let $(P, \succ, q)$ be a problem and $k > 1$. Then,*

(i) every blocking student of the outcome $\beta^k(P, \succ, q)$ of the constrained Boston mechanism is a manipulating student of β^k at $(P, \succ, q)$ and

(ii) every manipulating student of the constrained Gale-Shapley mechanism GS^k at $(P, \succ, q)$ is a blocking student of $GS^k(P, \succ, q)$.

These results have important implications for the relation between manipulability and stability. To see this, suppose that there is no manipulating student for the constrained Boston mechanism β^k at $(P, \succ, q)$. Then, under

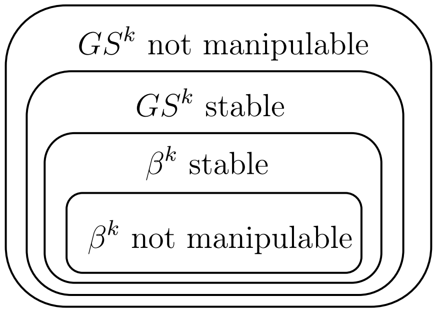

Figure 1. Set inclusion of problems for GS^k and β^k.

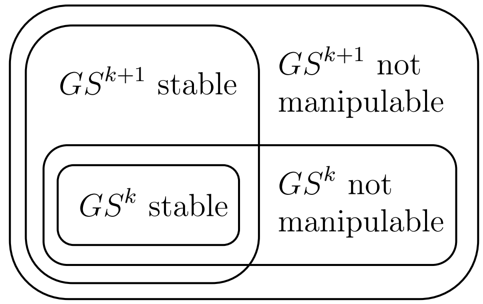

Figure 2. Set inclusion of problems for GS^k and GS^{k+1}.

part (i) of this theorem, there is no blocking student of $\beta^k(P, \succ, q)$. 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, there is no manipulating student for GS^k. We summarize these results in the following corollary and in Figures 1 and 2.

Corollary 1. *Let $(P, \succ, q)$ be a problem and $k > 1$. (i) Suppose that the constrained Boston mechanism β^k is not manipulable at $(P, \succ, q)$. Then, $\beta^k(P, \succ, q)$ is stable.*

(ii) Suppose that the constrained Gale-Shapley mechanism $GS^k(P, \succ, q)$ is stable. Then, GS^k is not manipulable at $(P, \succ, q)$.

Note that there are problems where the reverse of each of these results does not hold. See Example 6 below for the constrained Gale-Shapley mechanism. To see a counterexample of the reverse of the case (i), consider a problem $(P, \succ, q)$ where students have a common ranking of schools, have ranked k schools acceptable and where each school has one seat. Then, $\beta^k(P, \succ, q)$ is stable. However, the student who has received her third ranked school is better off top ranking the school she has ranked second as her top choice.

An implication of the latter results is a manipulability comparison introduced by Pathak and Sönmez (2013). Under part (i) of Corollary 1, when the constrained Boston mechanism is not manipulable then it is stable. By Theorem 1, the constrained Gale-Shapley mechanism is also stable. By part (ii) of Corollary 1, the constrained Gale-Shapley mechanism is not manipulable. This is the comparison established by Pathak and Sönmez (2013).

Corollary 2. *(Pathak and Sönmez, 2013). Let $(P, \succ, q)$ be a problem, $k > 1$ and suppose that the constrained Boston mechanism β^k is not manipulable at $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is not manipulable at $(P, \succ, q)$.*

Another implication is for the serial dictatorship mechanism. 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 a manipulation does not help. This implies that constrained serial dictatorship mechanism is non-manipulable and stable for the same set of problems.

Proposition 3. *Let $(P, \succ, q)$ be a problem and $k > 1$. The constrained serial dictatorship mechanism SD^k is stable if and only if it is not manipulable at $(P, \succ, q)$.*

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

Example 6. *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 remaining components are specified as follows.*

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 manipulation 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).

We also establish another direct corollary of Theorem 7 with two additional results. We show that when switching from constrained Boston to constrained GS, or when extending the list in the constrained GS, the mechanism becomes more fair by stability and less manipulable.

Corollary 3. *Let $(P, \succ, q)$ be a problem.*

(i) Let $k > 1$ and suppose that the constrained Boston mechanism β^k is stable at $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is stable and not manipulable at $(P, \succ, q)$.

(ii) Let $k > \ell > 1$ and suppose that the constrained Gale-Shapley mechanism GS^ℓ is stable at $(P, \succ, q)$. Then, the mechanism GS^k is stable and not manipulable at $(P, \succ, q)$.

Finally, we partially restore the comparisons for the First-Preference-First mechanism. Although 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 1), there is a surprising interplay between the two concepts.

Proposition 4. *Let $(P, \succ, q)$ be a problem and $k > 1$. If the constrained First-Preference-First mechanism FPF^k is stable at $(P, \succ, q)$, then the constrained Gale-Shapley mechanism GS^k is not manipulable at $(P, \succ, q)$.*

This result helps to evaluate the reforms in England, where FPF^k was replaced by GS^k. Even though for some problems the reform may increase manipulability (GS^k could be manipulable at a problem but not FPF^k) or decrease fairness (by stability) separately, it cannot be unsuccessful in both dimensions. Namely, if at some profile $(P, \succ, q)$, FPF^k is stable (while GS^k might not be stable), then GS^k is not manipulable at $(P, \succ, q)$ and there was no "increase" of manipulability due to the reform.

To sum up the results of this section, stability and manipulability are logically related, and the relationship depends on the mechanism.

6. Conclusions

In response to objections, many school districts around the world have recently reformed their admissions systems. The main reason for these objections was that the mechanisms were unfair and manipulable. Yet, the mechanisms remained unfair and manipulable even after the reforms. We showed that many reforms led to more fair matching mechanisms, first by relying on stability and second by counting and comparing the number of the blocking students. Most results remain true in strategic settings, where either some students are sophisticated while others are sincere, or all students are sophisticated but to a limited extent. Finally, we discover the inherent relationship between fairness and manipulability of the mechanisms in question.

The reforms concern essentially two major changes. First, they kept the constraint on the number of schools that each student is allowed to report but replaced the immediate acceptance procedure in the Boston mechanism (or a hybrid between Gale-Shapley and Boston mechanism) with the Gale-Shapley's student-proposing deferred acceptance procedure. Second, some school districts kept using the Gale-Shapley mechanism but extended the number of schools that each student is allowed to report.

Overall, our results provide a new justification for the reforms, complementing the existing ones. Pathak and Sönmez (2013) were the first to observe these reforms and proposed a way to explain them using a notion of *manipulability* that compares mechanisms according to the inclusion of instances where they are not vulnerable to gaming. These results were further strengthened for other mechanisms and other vulnerability criteria (Chen and Kesten, 2017; Decerf and Van der Linden, 2020; Dur et al., 2021; Bonkoungou and Nesterov, 2021). Mostly, the reforms made the systems both less vulnerable to gaming and more fair, but some reforms achieved at least one goal. Our results help better understand and evaluate these reforms and possibly help to advocate similar reforms in the future.

References

Abdulkadiroğlu, A., Che, Y.-K., Pathak, P. A., 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., Che, Y.-K., and Yasuda, Y. (2009). Resolving conflicting preferences in school choice: The "boston" mechanism reconsidered. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.1465293*.

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

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. S. (2021). Comparing school choice and college admission mechanisms by their strategic accessibility. *Theoretical Economics, forthcoming.*

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.

Decerf, B. and Van der Linden, M. (2020). Manipulability in school choice. *Journal of Economic Theory (forthcoming).*

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

Dogan, B. and Ehlers, L. (2020a). Blocking pairs versus blocking students: Stability comparisons in school choice. *Available at SSRN.*

Dogan, B. and Ehlers, L. (2020b). Robust minimal instability of the top trading cycles mechanism. *Available at SSRN*.

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88(7):485–494.

Dur, U. M., Pathak, P. A., Song, F., and Sönmez, T. (2021). Deduction dilemmas: The Taiwan assignment mechanism. *Forthcoming, American Economic Journal: Microeconomics*.

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.

Featherstone, C. R. and Niederle, M. (2016). Boston versus deferred acceptance in an interim setting: An experimental investigation. *Games and Economic Behavior*, 100:353–375.

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

Gale, D. and Sotomayor, M. (1985). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11(3):223–232.

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

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.

König, T., Kübler, D., Mechtenberg, L., and Schmacker, R. (2019). Fair procedures with naive agents: Who wants the boston mechanism? *CRC TRR 190 Discussion Paper No. 222*.

Kwon, H. and Shorrer, R. I. (2019). Justified-envy minimal mechanisms in school choice. *Available at SSRN 3495266*.

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. A. 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.

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. (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. (2020). Weak stability and pareto efficiency in school choice. *Economic Theory*, pages 1–20.

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.

Online Appendix: Proofs

We organize the appendix according to the logical order in which the results are related. To simplify the exposition, we divide the appendix in three subsections. In each subsection, we order them such that later results rely on proofs of earlier results.

We first present a lemma from the literature that will be useful throughout the appendix. This result is known as the *rural hospital theorem.*

Lemma 4 (Rural hospital theorem, Roth, 1986)**.** *Let $(P, \succ, q)$ be a problem, ν and μ 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.

Appendix A: Proof of Theorem 1 and Propositions 1, 3, 4.

Proposition 4: *Let $(P, \succ, q)$ be a problem and $k > 1$. If the constrained First-Preference-First mechanism FPF^k is stable at $(P, \succ, q)$, then the constrained Gale-Shapley mechanism GS^k is not manipulable at $(P, \succ, q)$.*

Proof. We first establish two claims.

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

Suppose that $GS_i^k(P, \succ, q) = s$ or $GS_i(P^k, \succ, q) = s$. As shown by Roth (1982), $GS_i(P^k, \succ, q) = s$ implies that $GS_i(P_i^s, P_{-i}^k, \succ, q) = s$. Since $k > 1$, the truncation of P_i^s after the k'th acceptable school is nothing but P_i^s. Thus, $GS_i^k(P_i^s, P_{-i}, \succ, q) = s$.

<u>*Claim 2:*</u> *Suppose that student i can manipulate GS^k at $(P, \succ, q)$. Then she is unmatched under $GS^k(P, \succ, q)$.*

This result follows from Pathak and Sönmez (2013).

We are now ready to prove the proposition. Let $(P, \succ, q)$ be a problem and suppose that $\mu = FPF^k(P, \succ, q)$ is stable at $(P, \succ, q)$. Under Claim 2, every matched student under $GS^k(P, \succ, q)$ cannot manipulate GS^k at $(P, \succ, q)$. It is then enough to show that no unmatched student under $GS^k(P, \succ, q)$ has a profitable misrepresentation. Because GS^k is individually rational, under

Claim 1, we further need to restrict ourselves to manipulation by top ranking a school first. 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 $(P, \succ, q)$. This conclusion contradicts 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 $(P^k, \succ, q)$,

$$\text{(1)} \qquad \text{for each student } i, \; GS_i(P^k, \succ, q) \; R_i^k \; \mu(i).$$

In line with Lemma 4, the same number 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 $(P, \succ, q)$, this implies that 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 $(P_i^s, P_{-i}^k, \succ, q)$. Under Lemma 4, the set of matched students is the same at all stable matchings. Thus, student i is also unmatched under $GS(P_i^s, P_{-i}^k, \succ, q)$. Then, under Claim 1, there is no strategy P_i' such that $GS_i^k(P_i', P_{-i}) = s$. Thus, the mechanism GS^k is not manipulable at $(P, \succ, q)$.

□

Theorem 1: *Suppose that there are at least two schools and let $k > 1$. The constrained Gale-Shapley mechanism GS^k is more fair by stability than the constrained Boston mechanism β^k*

Proof. The Boston mechanism is a special case of the First-Preference-First mechanism when every school is a first-preference-first school. Let $(P, \succ, q)$ be a problem and suppose that $\beta^k(P, \succ, q)$ is stable at $(P, \succ, q)$. As stated in equation 1, 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, $GS^k(P,\succ,q) = \beta^k(P,\succ,q)$ and consequently, $GS^k(P,\succ,q)$ 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 $(P,\succ,q)$.

$\square$

Proposition 1: *Suppose that there are at least two schools and at least one first-preference-first school. Let $k > 1$ and suppose that schools have a common priority. The constrained serial dictatorship mechanism SD^k is more fair by stability than the constrained First-Preference-First mechanism FPF^k.*

Proof. Let $(P,\succ,q)$ be a problem where schools have a common priority order and suppose that $FPF^k(P,\succ,q)$ is stable at $(P,\succ,q)$. Under equation 1, each student finds the outcome $SD^k(P,\succ,q)$ at least as good as $FPF^k(P,\succ,q)$ under P^k. With a common priority order, there is a unique stable matching under $(P,\succ,q)$ which is also Pareto efficient under P. Therefore, because $FPF^k(P,\succ,q)$ is stable at $(P,\succ,q)$, we have $FPF^k(P,\succ,q) = SD(P,\succ,q)$.

Next, every student who is matched under $SD(P, \succ, q)$ is matched to one of her top k-ranked acceptable schools. Therefore, $SD(P, \succ, q) = FPF^k(P, \succ, q)$ is also Pareto efficient under P^k. Thus, equation 1 implies that $SD^k(P, \succ, q) = FPF^k(P, \succ, q)$ and consequently, $SD^k(P, \succ, q)$ is stable at $(P, \succ, q)$.

We can adapt the example provided in the proof of Theorem 1 to show that there is a problem where SD^k is stable but not FPF^k. □

Lemma 1: *Let $(P, \succ, q)$ be a problem and $k > 1$. Then $GS^k(P, \succ, q)$ is stable if and only if $GS^k(P, \succ, q) = GS(P, \succ, q)$.*

Proof. The "if" part is straightforward because $GS(P, \succ, q) = GS^k(P, \succ, q)$ is the student-optimal stable matching under $(P, \succ, q)$.

The "only if" part. Suppose that $GS^k(P, \succ, q)$ is stable at $(P, \succ, q)$. We show that $GS^k(P, \succ, q) = GS(P, \succ, q)$. Let $N = \{i \in I | GS_i(P, \succ, q) = \emptyset\}$ denote the set of students who are unmatched under $GS(P, \succ, q)$.

Step 1: For each $i \in N$, $GS_i(P, \succ, q) = GS_i(P^k, \succ, q)$.

This follows from the assumption that $GS(P^k, \succ, q)$ is stable at $(P, \succ, q)$ and part (i) of Lemma 4.

Step 2: For each $i \in I \setminus N$, $GS_i(P, \succ, q) = GS_i(P^k, \succ, q)$.

Let $i \in I \setminus N$. Because $GS(P, \succ, q)$ is the student-optimal stable matching under $(P, \succ, q)$, we have

$$GS_i(P, \succ, q) \; R_i \; GS_i^k(P, \succ, q). \tag{2}$$

Note that for each student $j \in N$, the preference relation P_j can be interpreted as if she has extended her list of acceptable schools from P_j^k. As shown by Gale and Sotomayor (1985), when a subset of students extend their list of acceptable schools, none of the remaining students are better off. Therefore,

$$\text{for each student } j \in I \setminus N, \;\; GS_j(P_N^k, P_{-N}, \succ, q) \; R_j \; GS_j(P, \succ, q). \tag{3}$$

Because GS is individually rational under P, under equation 3, every student in $I \setminus N$ is also matched under $GS(P_N^k, P_{-N}, \succ, q)$. Next, since $GS(P^k, \succ, q)$ is stable at $(P, \succ, q)$, by assumption, Lemma 4 implies that the same set of students are matched under both $GS(P, \succ, q)$ and $GS(P^k, \succ, q)$. Therefore, i is also matched under $GS(P^k, \succ, q)$. Next, note that the students in $I \setminus N$ have extended their list of acceptable schools under (P^k_N, P_{-N}) from P^k. Then, at the end of the Gale-Shapley algorithm for the problem $(P^k, \succ, q)$, each of the students in $I \setminus N$ is accepted by a school. The school that each of them has listed below the school that has accepted her at this step of the algorithm and how she has ranked them do not affect the outcome of the algorithm. Thus,

$$GS(P^k, \succ, q) = GS(P^k_N, P_{-N}, \succ, q).$$

This equation and equation 3 imply that $GS_i(P^k, \succ, q)\ R_i\ GS_i(P, \succ, q)$. Since the preference relation P_i is strict, this relation and equation 2 imply that

$$GS_i(P^k, \succ, q) = GS_i(P, \succ, q).$$

Finally, by *Step 1* and *Step 2*, the matching is the same for each student under $GS^k(P, \succ, q)$ and $GS(P, \succ, q)$, the desired conclusion. $\square$

Proposition 3: *Let $(P, \succ, q)$ be a problem and $k > 1$. The constrained serial dictatorship mechanism SD^k is stable if and only if it is not manipulable at $(P, \succ, q)$.*

Proof. As shown by Bonkoungou and Nesterov (2021), SD^k is not manipulable at $(P, \succ, q)$ if and only if $SD^k(P, \succ, q) = SD(P, \succ, q)$. Suppose that $SD^k(P, \succ, q)$ is stable. Then, according to Lemma 1, $SD^k(P, \succ, q) = SD(P, \succ, q)$ and thus SD^k is not manipulable at $(P, \succ, q)$. Suppose that SD^k is not manipulable at $(P, \succ, q)$. Then, $SD^k(P, \succ, q) = SD(P, \succ, q)$ and thus stable. $\square$

Appendix B: Proof of Theorems 2, 4, 5, 6.

Let $\bar{\ell} = (\ell_i)_{i \in I}$ and $\bar{k} = (k_i)_{i \in I}$ be such that for each $i \in I$, $k_i \geq \ell_i > 1$. For each preference profile P, let $P^{\bar{\ell}} = (P^{\ell_i})_{i \in I}$ be such that the constraint that applies to student i is ℓ_i. We also define $P^{\bar{k}} = (P^{k_i})_{i \in I}$.

Lemma 5. *Let N be a subset of students and $\mu = GS(P^{\bar{\ell}}_N, P^{\bar{k}}_{-N}, \succ, q)$. Any blocking student for μ under $(P, \succ, q)$ is unmatched.*

Proof. We prove the lemma by the contradiction. Suppose, to the contrary, that some student i is a blocking student for $\mu = GS(P^{\bar{\ell}}_N, P^{\bar{k}}_{-N}, \succ, q)$ 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 $i \succ_{s'} j$. Let define x as follows:

$$x = \begin{cases} \ell_i & \text{if } i \in N \\ k_i & \text{otherwise} \end{cases}$$

Since $\mu(i) = s$, school s is one of the top x acceptable schools under P_i. Thus $s' \; P^x_i \; \mu(i) = s$. This relation, together with the case (i) or (ii) imply that the pair (i, s') blocks the matching μ under $(P^{\bar{\ell}}_N, P^{\bar{k}}_{-N}, \succ, q)$. This conclusion contradicts the fact that μ is stable at $(P^{\bar{\ell}}_N, P^{\bar{k}}_{-N}, \succ, q)$.

□

We next formulate a lemma, Lemma 6, that will be useful for the proof of two theorems. Let $N \subsetneq I$ be a proper subset of students and $i \notin N$. Let $\bar{\ell} = (\ell_i)_{i \in I}$ and $\bar{k} = (k_i)_{i \in I}$ and $\hat{P} = (P^{\bar{\ell}}_N, P^{\bar{k}}_{-N}, \succ, q)$. For simplicity, let $\ell_i = \ell$ and $k_i = k$. We define two matchings as follows:

$$\mu =: GS(\hat{P}, \succ, q)$$

and,

$$\nu =: GS(P^{\ell}_i, \hat{P}_{-i}, \succ, q).$$

Lemma 6. *There are at least as many blocking students for ν as for μ under $(P, \succ, q)$.*

Proof. Let n be the number of blocking students for μ under $(P, \succ, q)$. We show that there are at least n blocking students for ν under $(P, \succ, q)$.

Let us first show that every student other than i finds ν at least as good as μ under $\hat{P}$. To see this, note that student i has extended her list of acceptable schools under $\hat{P}_i = P^k_i$ from P^{ℓ}_i. As shown by Gale and Sotomayor (1985), after this extension no student, other than i, is better off. That is,

$$\text{for each student } j \neq i, \;\; \nu(j) \; \hat{R}_j \; \mu(j). \tag{4}$$

We divide the rest of the proof into two cases. In the first case, student i is unmatched under μ. For this case, we will show that any blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. In the second case, student i is matched under μ. We will show that either $\mu = \nu$ (in case student i is also matched under ν), or i is a blocking student for ν.

Case I: *Suppose that student i is unmatched under μ*, that is, $\mu(i) = \emptyset$.

We first show that $\nu(i) = \emptyset$. Suppose, to the contrary, that $\nu(i) = s$, for some school s. Then s is one of the top ℓ acceptable schools of student i under P_i. Since $k > \ell$, school s is also one of the top k acceptable schools under $\hat{P}_i = P_i^k$. Therefore,

$$GS_i(P_i^\ell, \hat{P}_{-i}, \succ, q) = s \; \hat{P}_i \; \mu(i) = GS_i(\hat{P}, \succ, q) = \emptyset.$$

This relation shows that student i is better off misrepresenting her preference to the Gale-Shapley mechanism, contradicting the fact that this mechanism is not manipulable (Dubins and Freedman, 1981; Roth, 1982). Therefore, $\nu(i) = \mu(i) = \emptyset$. This equality together with equation 4 imply that each student finds the matching ν at least as good as μ under $\hat{P}$. Because $\mu = GS(\hat{P}, \succ, q)$ is stable at $(\hat{P}, \succ, q)$, it is also stable at $(P_i^\ell, \hat{P}_{-i}, \succ, q)$. To see this, note that student i is unmatched under μ and that for each school s such that $s \; P_i^\ell \; \mu(i)$, school s does not have an empty seat under μ and every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Since $\nu = GS(P_{-i}^\ell, \hat{P}_{-i}, \succ, q)$ is also stable at $(P_{-i}^\ell, \hat{P}_{-i}, \succ, q)$, under Lemma 4, we have the following conclusion.

Conclusion: (a) the same set of students are matched (unmatched) under ν and μ and (b) every school is matched to the same number of students under both ν and μ.

Let us now prove that every blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. Let j be a blocking student for μ under $(P, \succ, q)$. There are two cases.

Case I.1: $j = i$

Then, there is a school s such that $s\ P_i\ \mu(i)$ and either (i) school s has an empty seat under μ or (ii) there is a student j' such that $\mu(j') = s$ and $i \succ_s j'$.

Consider the case (i) where school s has an empty seat under μ. Then, under part (b) of the previous conclusion, s has an empty seat under ν. Since $\nu(i) = \emptyset$, i is a blocking student for ν under $(P, \succ, q)$.

Consider the case (ii) where there is a student j' such that $\mu(j') = s$ and $i \succ_s j'$. Without loss of generality, suppose that school s does not have an empty seat under μ. Then, under part (b) of the previous conclusion, school s does not have an empty seat under ν. Suppose that $\nu(j') = s$. Since $\nu(i) = \emptyset$ and $i \succ_s j'$, then the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$. Suppose that $\nu(j') \neq s$. Since $|\nu^{-1}(s)|= q_s$, there is $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Under equation 4,

$$s = \nu(j'')\ \hat{P}_{j''}\ \mu(j'').$$

Since $\mu = GS(\hat{P}, \succ, q)$ is stable at $(\hat{P}, \succ, q)$ and $\mu(j') = s$, then $j' \succ_s j''$. Because $\succ_s$ is transitive, $i \succ_s j'$ and $j' \succ_s j''$ imply that $i \succ_s j''$. Since $s\ P_i\ \nu(i) = \emptyset$, the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$.

Case I.2: $j \neq i$

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 $j \succ_s j'$. As shown in Lemma 5, because student j is a blocking student for μ under $(P, \succ, q)$, we have $\mu(j) = \emptyset$.

Let us consider the case (i). Under Lemma 4, student j is also unmatched under ν and school s has an empty seat under ν. Thus, j is a blocking student for ν under $(P, \succ, q)$.

Second, consider (ii) where there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. If $\nu(j') = s$, then (j, s) is a blocking pair for ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$. Without loss of generality, suppose that $|\mu^{-1}(s)|= q_s$ and $\nu(j') \neq s$. According to part (b) of the above conclusion, $|\nu^{-1}(s)|= q_s$. Then, there is a student $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Because $\mu(i) = \emptyset$, we have $j'' \neq i$, and by equation 4,

$$s = \nu(j'')\ \hat{P}_{j''}\ \mu(j'').$$

Since μ is stable at $(\hat{P}, \succ, q)$ and $\mu(j') = s$, then this equation implies that $j' \succ_s j''$. Because $\succ_s$ is transitive, $j \succ_s j'$ and $j' \succ_s j''$ imply that $j \succ_s j''$. Since $s\ P_j\ \nu(j) = \emptyset$, then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

In conclusion, every blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. There are n blocking students for ν under $(P, \succ, q)$.

Case II: Student i is matched under μ.

If student i is also matched under $\nu = GS(P_i^{\ell}, \hat{P}_{-i}, \succ, q)$, then $\nu = \mu$. To see this, let $\nu(i) = s$ for some school s. School s is one of the top ℓ acceptable schools of student i under P_i. The Gale-Shapley mechanism is invariant to the modification of the preferences of the students for the part below their outcomes. We know that P_i^k is one such modification of P_i^{ℓ} below school s. Thus, $\nu = \mu$. We now consider the case where i is not matched under ν.

Suppose that student i is unmatched under ν. The strategy of the proof is to show that i is a blocking student for ν under $(P, \succ, q)$ and that there are also at least $n-1$ other blocking students for ν under $(P, \succ, q)$. We depict the flow of these students in Figure 3.

Step 1: Student i is a blocking student for ν under $(P, \succ, q)$.

Recall that we assumed that student i is matched under $\mu = GS(\hat{P}, \succ, q)$, where $\hat{P} = (P_N^{\bar{\ell}}, P_{-N}^{\bar{k}}, \succ, q)$ and $i \notin N$. Let $s = \mu(i)$. School s is one of the top k acceptable schools under P_i. Since $\nu(i) = \emptyset$, if school s has an empty seat under ν, then clearly the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$. Suppose that $|\nu^{-1}(s)| = q_s$. Since $\mu(i) = s$ and $\nu(i) = \emptyset$, there is a student $j \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. By equation 4,

$$s = \nu(j)\ \hat{P}_j\ \mu(j).$$

Since μ is stable at $(\hat{P}, \succ, q)$ and $\mu(i) = s$, we have $i \succ_s j$. Therefore, the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$.

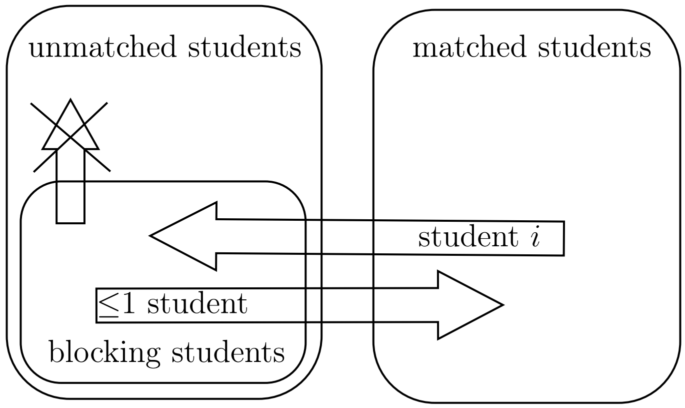

Figure 3. Case II: flow of students across matched, unmatched, and blocking status, from μ to ν: at most one student can leave the blocking status to the matched status; student i left the matched status to the blocking status, and no student can leave the blocking status and remain unmatched.

Step 2: Every blocking student for μ under $(P, \succ, q)$ who is unmatched under ν is also a blocking student for ν under $(P, \succ, q)$.

Let j be a blocking student for μ under $(P, \succ, q)$ and suppose that she is unmatched under ν. 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 $j \succ_s j'$. In addition, because j is the blocking student of μ under $(P, \succ, q)$, by Lemma 5, we have $\mu(j) = \emptyset$.

Let us consider the case (i) where school s has an empty seat under μ. We also show that s has an empty seat under ν. Assume otherwise. Then, there is $j' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. We know that student i is unmatched under ν. Thus, $j' \neq i$. Under equation 4, $s = \nu(j') \; \hat{P}_{j'} \; \mu(j')$. This contradicts the fact that μ is stable at $(\hat{P}, \succ, q)$ because s has an empty seat under μ. Therefore, s has an empty seat under ν. Then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

Let us now consider the case (ii) where there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. If school s has an empty seat under ν, then because student j is unmatched under ν, she is a blocking student for ν under $(P, \succ, q)$. Suppose that school s does not have an empty seat under ν. If $\nu(j') = s$ then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$ because $\nu(j) = \emptyset$ and $j \succ_s j'$. Suppose that $\nu(j') \neq s$. Because school s does not have an empty seat under ν, there is $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Since student i is unmatched under ν, we have $j'' \neq i$. By equation 4, we have $s = \nu(j'')\ \hat{P}_{j''}\ \mu(j'')$. Since μ is stable at $(\hat{P}, \succ, q)$, the equation and the fact that $\mu(j') = s$ imply that $j' \succ_s j''$. Because $\succ_s$ is transitive, $j \succ_s j'$ and $j' \succ_s j''$ imply that $j \succ_s j''$. Since $s\ P_j\ \nu(j) = \emptyset$, then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

Step 3: Every student but i who is matched under μ is also matched under ν.

Suppose that for some student $j \neq i$ and some school s, $\mu(j) = s$. Under equation 4, we have $\nu(j)\ \hat{R}_j\ \mu(j) = s$. Since μ is individually rational under $\hat{P}$, $\nu(j) \neq \emptyset$.

Step 4: There are at least n blocking students for ν under $(P, \succ, q)$.

Let j be a blocking student for μ under $(P, \succ, q)$ who is not a blocking student for ν under $(P, \succ, q)$. Then, j is matched under ν. Otherwise, according to step 2, she is also a blocking student for ν under $(P, \succ, q)$. We prove, more generally, that there are at most one student who is unmatched under μ but matched under ν. To do that, we compare for each school the number of students matched to it under μ and ν.

Let s be a school. Suppose that it does not have an empty seat under μ. Then, we have $|\nu^{-1}(s)| \leq |\mu^{-1}(s)| = q_s$. Suppose now that s has an empty seat under μ. Suppose that there is $j' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Then, because student i is unmatched under ν, $j' \neq i$. By equation 4,

$$s = \nu(j')\ \hat{P}_{j'}\ \mu(j').$$

This contradicts the fact that μ is stable at $(\hat{P}, \succ, q)$ because school s has an empty seat under μ. Thus, there is no student matched to school s under ν but not under μ. Therefore, $|\nu^{-1}(s)| \leq |\mu^{-1}(s)|$. We conclude that no school is matched to more students under ν than under μ. Thus,

$$\sum_{s \in S} |\nu^{-1}(s)| \leq \sum_{s \in S} |\mu^{-1}(s)|. \tag{5}$$

By step 3, all students, but student i, who are matched under μ are also matched under ν. Therefore, the set of students who are matched under ν consists of the following students:

- the students who are matched under μ, except student i and
- the students who are unmatched under μ but matched under ν.

Let x denote the number of the students who are unmatched under μ but matched under ν. Then, we have

$$\sum_{s \in S} |\nu^{-1}(s)| = \underbrace{\sum_{s \in S} |\mu^{-1}(s)| - 1}_{\substack{\text{number of students} \\ \text{matched under } \mu \text{ and } \nu}} + x,$$

where the first two expressions on the right-hand side indicate that we subtracted student i from those who are matched under μ. By rearranging this equation, we get

$$\sum_{s \in S} |\nu^{-1}(s)| - \sum_{s \in S} |\mu^{-1}(s)| = x - 1 \leq 0,$$

where the inequality follows from equation 5. Thus, there is at most one student who is unmatched under μ but matched under ν. According to Lemma 5, all blocking students for μ under $(P, \succ, q)$ are unmatched under μ. Then, there is at most one blocking student for μ under $(P, \succ, q)$ who is matched under ν. In assent with this result together with step 2, there is at most one blocking student for μ under $(P, \succ, q)$ who is not a blocking student for ν under $(P, \succ, q)$. Among the n blocking students for μ under $(P, \succ, q)$, at most one of them is not a blocking student for ν under $(P, \succ, q)$. Therefore, excluding student i, there are at least $n-1$ blocking students for ν under $(P, \succ, q)$. Since student i is also a blocking student for ν under $(P, \succ, q)$, there are at least n blocking students for ν under $(P, \succ, q)$.

$\square$

Lemma 7. *Let $\bar{k} = (k_i)_{i\in I}$ and $\bar{\ell} = (\ell_i)_{i\in I}$ be such that for each $i \in I$, $k_i \geq \ell_i > 1$. For each problem, the constrained Gale-Shapley mechanism $GS^{\bar{\ell}}$ has at least as many blocking students as the constrained Gale-Shapley mechanism $GS^{\bar{k}}$.*

Proof. The proof relies on Lemma 6, which is the main part for proving the lemma.

Let $(P, \succ, q)$ be a problem. The proof strategy is to start from $GS(P^{\bar{k}}, \succ, q)$ and replace the preference relations in $P^{\bar{k}}$ one at a time, with a preference relation in $P^{\bar{\ell}}$ for the corresponding student until we get $GS(P^{\bar{\ell}}, \succ, q)$. We prove the theorem by showing that the number of blocking students is not decreasing after each replacement.

To prove the lemma we apply Lemma 6 sequentially. Let n be the number of the blocking students for $GS(P^{\bar{k}}, \succ, q)$ under $(P, \succ, q)$. For simplicity, let $I = \{1, 2, \ldots, |I|\}$. Under Lemma 6, there are at least n blocking students for

$$\mu_1 = GS(P_1^{\bar{\ell}}, P_{-1}^{\bar{k}}, \succ, q)$$

under $(P, \succ, q)$. By the same lemma, compared to μ_1, there are at least n blocking students of the matching

$$\mu_2 = GS(P_{\{1,2\}}^{\bar{\ell}}, P_{-\{1,2\}}^{\bar{k}}, \succ, q)$$

under $(P, \succ, q)$. With a repeated replacement of the remaining components of $P^{\bar{k}}$ with their counterparts in $P^{\bar{\ell}}$, we draw the conclusion that there are at least n blocking students for $GS(P^{\bar{\ell}}, \succ, q)$ under $(P, \succ, q)$. $\square$

Theorem 4: *Suppose that there are at least two schools and let $|S| > k > \ell \geq 1$. The constrained Gale-Shapley mechanism GS^k is more fair by counting than GS^ℓ.*

Proof. The theorem is a particular case of Lemma 7 in which for each student i, $k_i = k$ and $\ell_i = \ell$. Under this lemma, for each problem, GS^ℓ has at least as many blocking students as GS^k.

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 being unmatched. Then, there is at least one blocking student for $GS^\ell(P, \succ, q)$. Therefore, there are more blocking students for $GS^\ell(P, \succ, q)$ than $GS^k(P, \succ, q)$ under $(P, \succ, q)$. □

Theorem 2: *Suppose that there are at least three schools and let $k > \ell$. Then, the constrained Gale-Shapley mechanism GS^k is more fair by stability than GS^ℓ.*

Proof. Suppose that $GS^\ell(P, \succ, q)$ is stable at $(P, \succ, q)$. Then, there is no blocking student for it under $(P, \succ, q)$. According to Theorem 4, there is no blocking student for $GS^k(P, \succ, q)$ under $(P, \succ, q)$. Since $GS^k(P, \succ, q)$ is individually rational under P, then it is stable at $(P, \succ, q)$.

We described an example in the proof of Theorem 4 where there is a blocking student (pair) for GS^ℓ but not GS^k. Since GS^k and GS^ℓ are individually rational, at this problem GS^k is stable but not GS^ℓ. □

Theorem 5: *Let $k > \ell > 1$ and suppose that there are sincere students, who play truthfully, and sophisticated students, who best respond. For any problem where schools have a common priority, the Nash equilibrium outcome of the constrained serial dictatorship mechanism SD^ℓ has at least as many blocking students as the Nash equilibrium outcome of the constrained serial dictatorship mechanism SD^k.*

Proof. Let $(P, \succ, q)$ be a problem and N the set of sincere students. Under Lemma 3, the matchings $SD(P^\ell_N, P_{-N}, \succ, q)$ and $SD(P^k_N, P_{-N}, \succ, q)$ are the unique Nash equilibrium outcomes of the games (SD^ℓ, P) and (SD^k, P), respectively. Under Lemma 7, the matching $SD(P^\ell_N, P_{-N}, \succ, q)$ has at least as many blocking students as the matching $SD(P^k_N, P_{-N}, \succ, q)$. $\square$

Lemma 8. *Let $\ell > 1$ and suppose that students are semi-sophisticated and drop competitive schools but are sincere otherwise. Let $(P, \succ, q)$ be a problem and P' a Nash equilibrium of the game (GS^ℓ, P). For each student i, let r_i denote the number of competitive schools that she has dropped but prefers to at least one of the top ℓ acceptable schools she has listed under P'_i. Then the outcome $GS^\ell(P', \succ, q)$ is the matching $GS(P^{\bar{k}}, \succ, q)$ where for each student i, $\bar{k}_i = \ell + r_i$.*

Proof. Note that under the matching $GS(P^{\bar{k}}, \succ, q)$, no student is matched to a competitive school that she has removed under P' because those schools are assigned to students who did not remove any school and who have ranked them first and have highest priority.

We prove that $GS(P'^\ell, \succ, q) = GS(P^{\bar{k}}, \succ, q)$. The idea of the proof is to start from $P^{\bar{k}}$ and sequentially replace the preference relations in $P^{\bar{k}}$ with preference relations in P'^ℓ. Note that under P'^ℓ, student i has removed r_i competitive schools from her acceptable schools compared to $P^{\bar{k}}_i$. Note also that there is no school which is not competitive and that student i has listed among the top ℓ acceptable schools under P'_i but is among the top $\bar{k}_i$ acceptable schools under P_i. Let $N \subsetneq I$ be a proper subset of students. For simplicity, let $\hat{P} = (P'^\ell_N, P^{\bar{k}}_{-N})$. Kojima and Manea (2010) show that when a student removes some schools that she did not receive under the Gale-Shapley mechanism, then no other student is worse off as a result. Let $i \notin N$. For each student $j \neq i$,

$$GS_j(P'^\ell_i, \hat{P}_{-i}, \succ, q)\ \hat{R}_j\ GS_j(\hat{P}, \succ, q). \tag{6}$$

Now note that $\mu = GS(P'^\ell_i, \hat{P}_{-i}, \succ, q)$ is stable at $(\hat{P}, \succ, q)$. This is because only student i and a school that she has listed under $P^{\bar{k}}_i$ but removed from her acceptable schools under P'_i can be a blocking pair for this matching. However, all seats of such a school have been assigned to students who have higher priority than i. Because the matching $GS(\hat{P}, \succ, q)$ is the student optimal stable matching under $(\hat{P}, \succ, q)$, for each student $j \in I$,

$$(7) \qquad GS_j(\hat{P}, \succ, q)\ \hat{R}_j\ GS_j(P_i'^{\ell}, \hat{P}_{-i}, \succ, q).$$

By equations 6 and 7, for each student $j \neq i$, $GS_j(\hat{P}, \succ, q) = GS_j(P_i'^{\ell}, \hat{P}_{-i}, \succ, q)$. In addition, we know that when a student removes some schools that she did not receive in the Gale-Shapley mechanism from her acceptable set of schools, then her outcome does not change. Therefore, $GS_i(P_i'^{\ell}, \hat{P}_{-i}, \succ, q) = GS_i(\hat{P}, \succ, q)$. Since we have considered all students, we have

$$GS(\hat{P}, \succ, q) = GS(P_i'^{\ell}, \hat{P}_{-i}, \succ, q).$$

Without loss of generality, let $I = \{1, 2, \ldots, n\}$. Consider first $N = \emptyset$. Then, we have

$$GS(P_1'^{\ell}, P_{-1}^{\bar{k}} \succ, q) = GS(P^{\bar{k}}, \succ, q).$$

Next, let $N = \{1\}$. Then we have

$$GS(P_{\{1,2\}}'^{\ell}, P_{-\{1,2\}}^{\bar{k}} \succ, q) = GS(P_1'^{\ell}, P_{-1}^{\bar{k}} \succ, q) = GS(P^{\ell}, \succ, q).$$

We repeat this argument until we replace all preference relations and get

$$GS^{\ell}(P', \succ, q) = GS(P^{\bar{k}}, \succ, q).$$

□

Theorem 6: *Let $k > \ell > 1$ and suppose that students are semi-sophisticated and behave strategically by dropping competitive schools but are sincere otherwise. For any problem, the Nash equilibrium outcome of GS^{ℓ} has at least as many blocking agents as the Nash equilibrium outcome of GS^{k}.*

Proof. The proof relies on Lemma 8 above. This lemma stipulates that the equilibrium outcome is equivalent to the outcome of Gale-Shapley for some particular individual restrictions.

Let $(P, \succ, q)$ be a problem. First every student who is guaranteed her first choice under GS^{ℓ} is also guaranteed her first choice under GS^{k} because $k > \ell > 1$. Let P' and P'' be the Nash equilibria of the games (GS^{ℓ}, P) and (GS^{k}, P), respectively. Let i be a student, r_i'' the number of competitive schools that she has dropped under P_i'' but prefers to a school she has listed among the top k acceptable schools. Similarly, let r_i' denote the number of competitive schools that she has dropped under P_i' and prefers to at least one of the schools she has listed among the top ℓ acceptable schools under P_i'. We next show that $r_i'' + k \geq r_i' + \ell$. We consider two cases:

<u>Case 1:</u> $r_i'' > 1$. Then, there are more than k acceptable schools under P_i such that they are not all competitive. Thus student i has listed only schools which are not competitive under P_i''. Since $\ell < k$, there are also more than ℓ acceptable schools under P_i such that they are not all competitive. Student i will list only schools that are not competitive among the top ℓ and k acceptable schools under P_i' and P_i'' respectively. Suppose that student i has dropped a competitive school s under P_i' that she prefers to a school s' she has listed among the top ℓ acceptable schools under P_i'. Since $k > \ell$, student i has listed school s' among the top k acceptable schools under P_i''. Therefore, she has dropped s under P_i'' and prefers it to school s'. Thus, $r_i' \leq r_i''$. Since $\ell < k$, we have $r_i' + \ell < r_i'' + k$.

<u>Case 2:</u> $r_i'' = 0$. This means that student i did not drop any competitive school that she prefers to a school listed among the top k acceptable schools under P_i''. There are three cases: (i) either all her acceptable schools under P_i are competitive or (ii) there are at most k acceptable schools under P_i or (iii) all her top k acceptable schools under P_i are not competitive. In cases (i) and (iii) student i will not drop any competitive school that she prefers to a school she has listed among the top ℓ acceptable schools under P_i'. Therefore $r_i' = 0$ and $r_i' + \ell < r_i'' + k$. In the case (ii) the number of schools that student i could drop and list under P_i' cannot exceed the number of acceptable schools under P_i. Therefore, $r_i' + \ell \leq r_i'' + k$. In either case, we have $r_i' + \ell \leq r_i'' + k$.

Under Lemma 8 the Nash equilibrium outcome of the game (GS^ℓ, P) is the matching $GS(P^{\bar{\ell}}, \succ, q)$ where for each student i, $\bar{\ell}_i = \ell + r_i'$ and the Nash equilibrium outcome of the game (GS^k, P) is $GS(P^{\bar{k}}, \succ, q)$ where for each student i, $\bar{k}_i = k + r_i''$.

Since for each student i, $\bar{k}_i \geq \bar{\ell}_i$, under Theorem 4, $GS(P^{\bar{\ell}}, \succ, q)$ has at least as many blocking students as $GS(P^{\bar{k}}, \succ, q)$ under $(P, \succ, q)$.

□

Appendix C: Proof of Theorem 7 and Corollary 1.

Theorem 7: *Let* $(P, \succ, q)$ *be a problem and* $k > 1$. *Then,*

(i) *every blocking student of the outcome* $\beta^k(P, \succ, q)$ *of the constrained Boston mechanism is a manipulating student of* β^k *at* $(P, \succ, q)$ *and*

(ii) *every manipulating student of the constrained Gale-Shapley mechanism* GS^k *at* $(P, \succ, q)$ *is a blocking student of* $GS^k(P, \succ, q)$.

Proof. Part (i). Let i be a student and suppose that she is 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, we have $s \ P_i \ \mu(i)$ 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 and $i \succ_s j$. 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 and $i \succ_s j$, 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, $\beta_i(P_i^s, P_{-i}^k, \succ, q) = s$. Since $s \ P_i \ \mu(i)$, then i is a manipulating student of β^k at $(P, \succ, q)$.

Part (ii). 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 for $\mu = GS^k(P, \succ, q)$ under $(P, \succ, q)$. By Claim 2 above, 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 a higher priority than i under $\succ_s$. According to Lemma 4, the set of unmatched students is the same under μ and $GS^k(P_i^s, P_{-i}^k, \succ, q)$. Thus, i is also unmatched under

$GS^k(P_i^s, P_{-i}^k, \succ, q)$. According to Claim 1, there is no misreport by which i is matched to s. Since s has been chosen arbitrarily, i is not a manipulating student of GS^k at $(P, \succ, q)$. This conclusion contradicts our assumption that student i is a manipulating student of GS^k at $(P, \succ, q)$. □

Corollary 1: *Let $(P, \succ, q)$ be a problem and $k > 1$.*

(i) *Suppose that the constrained Boston mechanism β^k is not manipulable at the problem $(P, \succ, q)$. Then, $\beta^k(P, \succ, q)$ is stable.*

(ii) *Suppose that the constrained Gale-Shapley mechanism GS^k is stable at the problem $(P, \succ, q)$. Then, GS^k is not manipulable at $(P, \succ, q)$.*

Proof. We prove (i) by the contraposition. Suppose that $\beta^k(P, \succ, q)$ is not stable at $(P, \succ, q)$. Since β^k is individually rational, there is a pair (i, s) of a student and a school which blocks $\beta^k(P, \succ, q)$ under $(P, \succ, q)$. Following (i) of Theorem 7, student i is a manipulating student of β^k at $(P, \succ, q)$. Thus, β^k is manipulable at $(P, \succ, q)$.

We now prove part (ii) by the contraposition. Let GS^k be stable but manipulable at $(P, \succ, q)$. Then, by Theorem 7, there is a manipulating student who is a blocking student of $GS^k(P, \succ, q)$.

□