NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Somouaoga Bonkoungou, Alexander Nesterov

INCENTIVES IN MATCHING MARKETS: COUNTING AND COMPARING MANIPULATING AGENTS

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 249/EC/2021

INCENTIVES IN MATCHING MARKETS: COUNTING AND COMPARING MANIPULATING AGENTS

SOMOUAOGA BONKOUNGOU* AND ALEXANDER NESTEROV†

ABSTRACT. Vulnerability to manipulation is a threat to successful matching market design. However, some manipulation is often inevitable and the mechanism designer wants to compare manipulable mechanisms and pick the best. Real-life examples include reforms in the entry-level medical labor market in the US (1998), school admissions systems in New York (2004), Chicago (2009-2010), Denver (2012), some cities in Ghana (2007-2008), and England (2005-2010). We provide a useful criterion for these design decisions: we count the number of agents with an incentive to manipulate each mechanism under consideration during these reforms, and show that this number decreased as a result of the reforms. Our conclusion is robust to further additional strategic assumptions.

Keywords: market design, two-sided matching, college admissions, school choice, manipulability

JEL Classification: C78, D47, D78, D82

1. INTRODUCTION

Many matching systems around the world recently underwent drastic changes to deal with the strategic manipulation of their matching rules. These changes include the US entry-level medical labor market in 1998, the New York City high school admissions in 2005, the Chicago selective high school admissions in 2009 and 2010,

*FACULTY OF BUSINESS AND ECONOMICS, UNIVERSITY OF LAUSANNE, INTERNEF, CH-1015 LAUSANNE, SWITZERLAND, bkgsom@gmail.com

†DEPARTMENT OF ECONOMICS AND GAME THEORY LAB, HSE UNIVERSITY, 16, SOYUZA PECHATNIKOV ST., 190121, ST. PETERSBURG, RUSSIA, nesterovu@gmail.com

Date: June 29, 2021.

We thank Rustadjam Hakimov, Lars Ehlers, Bettina Klaus, Parag Pathak, Marek Pycia, and Tayfun Sönmez for their valuable feedback. We are especially thankful to Bettina Klaus for her help in revising the manuscript. We thank the seminar participants at the University of Lausanne, Corvinus University of Budapest, North Carolina State University, and the St.Petersburg Economics seminar. We thank Oksana Nyrka for her excellent research assistance in preparing the simulations. Support from the Basic Research Program of the HSE University is gratefully acknowledged. The research is partially financed by RFBR grant 20-01-00687. Somouaoga Bonkoungou gratefully acknowledges financial support from the European Commission, Marie Skłodowska-Curie individual fellowship, project 890648.

the Denver public school admissions in 2012, the public school admissions in Ghana in 2007 and 2008, as well as school admissions in English cities between 2005 and 2010. In this paper, we study these reforms and provide a simple criterion for evaluating whether they were successful.

The matching systems in question are centralized markets whose outcomes are based on participants' reported private information. One of the key design objectives is to provide participants with incentives to report this information truthfully as opposed to "gaming the system" (Roth, 2008; Abdulkadiroğlu and Sönmez, 2003). *Truthful mechanisms* are simple, thereby guaranteeing easy and equal access to all participants regardless of their strategic sophistication (Pathak and Sönmez, 2008).[1] In contrast, *manipulable* mechanisms are complicated to deal with and can result in unfair and skewed outcomes. The old Boston school admission system provides a vivid illustration: in the 2001-2002 school year, at least 19% of students strategized poorly in a highly manipulable mechanism, and among them 27% ended up unassigned; while they could have been matched if they had strategized well (Abdulkadiroğlu et al., 2005b).[2] It is now commonly understood that manipulation is a major threat to the success of matching markets.

In practice, however, some manipulation is inevitable. Real markets are complex and involve constraints and policy goals that create opportunities for manipulation. For example, in medical labor markets, it is crucial to use a *stable* matching mechanism, which prevents doctors and hospitals from circumventing the system and arranging a mutually preferred match. Empirical evidence has documented that stable mechanisms tend to succeed, while unstable ones tend to fail (Roth, 1991, 2002).[3] Unfortunately, all stable mechanisms are manipulable (Roth, 1982).

To further illustrate the point, many school or college admission systems restrict the number of schools that students are allowed to apply to (Haeringer and Klijn, 2009; Pathak and Sönmez, 2013). For example, in New York City, out of 700 study programs, each student can rank only 12. The reasons behind these constraints are not yet fully understood, but they appear to be crucial for practitioners. They

[1]They also facilitate the interpretation and the evaluation of the outcome since they generate more credible policy-relevant data (Sönmez, 2013).

[2]In China the numbers are even more striking: in 2003, due to the use of a manipulable mechanism and poor strategizing, 3 million students – representing more than half of the cohort – were matched to significantly lower-ranked colleges than what they would have been matched to in a truthful mechanism. What is worse, this mismatch was stronger for rural students and female students, contributing to the education gap (Wu and Zhong, 2020).

[3]This is why several medical labor markets such as entry-level labor markets in Canada, the UK and the US use stable mechanisms.

remain widespread and bind the choice for a large share of students (Pathak, 2016). Unfortunately, any sensible mechanism that is augmented with such constraints is manipulable (Proposition 2).

While real matching markets involve manipulations, many systems have recently undergone drastic changes to reduce it. In this paper, we study whether these changes were successful. To quantify manipulations, we focus on agents that can beneficially misrepresent their private information when others are truthful.[4] We call them *manipulating agents*. Our analysis covers a wide range of settings, including recent real-life applications (see Table 1). These settings differ in a few major aspects: the set of strategic agents, the strategies these agents can use, whether stability is required and whether there is a ranking constraint.

Our results are threefold. First, we consider the college admission problem where both students and schools are strategic agents (Gale and Shapley, 1962) and schools can misreport their preferences as well as their capacities. Dubins and Freedman (1981) and Roth (1982) show that the student-proposing Gale-Shapley (GS) mechanism is not manipulable by students. It is one of the main arguments in favor of its choice for the National Resident Matching Program (NRMP). However, it also has the largest number of manipulating schools among all stable mechanisms (Pathak and Sönmez, 2013). In this paper, we show that, when all manipulations (by students as well as by schools) are considered, this mechanism has the smallest number of manipulating agents among all stable matching mechanisms (Theorem 1). This result supports its choice for medical labor markets such as the NRMP, which took place in 1998 (Roth and Peranson, 1999). What is more, even when schools can only misreport their capacities, but not their preferences, this mechanism is still the best choice (Proposition 1).

Second, we consider the same college admission problem where students and schools are strategic agents, but students face ranking constraints. A canonical example is the New York City high school admission (Abdulkadiroğlu et al., 2005a). In this setting, stable mechanisms (with respect to the reported preferences) are manipulable even by students. The result still holds: the constrained student-proposing GS mechanism has the smallest number of manipulating agents among all stable mechanisms (Theorem

[4]The number of these agents can be interpreted as a measure of potential for manipulations. It has been largely used to measure incentives in matching markets: Roth and Peranson (1999), for example, conducted a simulation on data on the NRMP and showed that the potential for manipulation is low. That is, a small number of agents could have beneficially misreported their information when other agents are truthful.

Who can manipulate?	What are the restrictions?	Design instances	Recommended design
Students, schools (rankings and capacity)	Stability	National Resident Matching Program 1998	Student-proposing GS (Theorem 1)
	Stability, ranking constraints		Student-proposing GS^k (Theorem 2 (i))
Students, schools (only capacity)	Stability		Student-proposing GS (Proposition 1)
	Stability, ranking constraints	New York 2004	Student-proposing GS^k (Theorem 2 (ii))
Only students	Ranking constraints	Brighton 2007, Chicago 2009, East Sussex 2007, Sefton 2007, Newcastle 2005	Replace $Boston^k$ by GS^k (Theorem 3)
		Chicago 2010, Ghana 2007, 2008, Newcastle 2010, Surrey 2010	Replace GS^ℓ by GS^k, $k > \ell$ (Theorem 4)
		Chicago 2009-2010, Denver 2012, Kent 2007, Newcastle 2005-2010	Replace $Boston^\ell$ by GS^k, $k > \ell$ (Corollary 2)

Table 1. Summary of the results.

Notes: The table presents strategic settings (column 1), restrictions on the set of the mechanisms (column 2), historical instances where these settings and restrictions occurred (column 3), corresponding recommendations and results (column 4).

2). This result is again robust to whether schools can misreport their rankings or only their capacities. These findings support using its use for the New York City high school admission to reduce manipulations. This choice has actually been made by a group of market design professionals in 2004 (Abdulkadiroğlu et al., 2005a).

Third, we consider the school choice problem (Abdulkadiroğlu and Sönmez, 2003) where students are the only strategic agents and also face ranking constraints. Historically, many school choice systems have used the constrained immediate acceptance (Boston) mechanism, but over time are shifting towards the constrained student-proposing GS mechanisms and relaxing the constraint. We demonstrate that these changes in admission mechanisms decrease the number of manipulating students (Theorem 3 and Theorem 4). These results rationalize recent reforms such as in Chicago, Denver, Ghana, and many English cities.

Related literature. This paper is part of the growing literature on two-sided matching started by Gale and Shapley (1962). Roth and Sotomayor (1990) provide a good account of the literature up until 1990. Our paper makes a contribution to a recent trend in the research, started by Pathak and Sönmez (2013), comparing manipulable mechanisms. The paper that first formalizes counting manipulating agents in mechanism design is Andersson et al. (2014a); they study the problem of allocating indivisible objects and money to agents and compare fair and budget-balanced mechanisms by counting manipulating agents.

The criterion has long been used in market design, although it has not yet received a systematic theoretical treatment. For example, in studying incentives in the medical labor market, Roth and Peranson (1999) count the number of medical students who could have benefited from truncating their rankings (and separately the number of hospitals who could have benefited from reducing their capacities) and used it as a measure of the potential for manipulations. In experiments, this criterion is also used (see the surveys by Chen, 2008, and by Hakimov and Kübler, 2020). Kojima and Pathak (2009) and Kojima et al. (2013) study incentives in large markets by counting the proportion of manipulating agents and their results support the student-proposing GS mechanism.

For the school choice problem, Pathak and Sönmez (2013) introduced a novel method that is weaker than counting: they compare mechanisms by the set-inclusion of problems with no manipulating agent.[5] This method became the state-of-art for comparing manipulable matching mechanisms (for example, it is used by Chen and Kesten, 2017; Dur et al., 2021; Dur, 2019; Dur et al., 2019). However, this method cannot be used to compare mechanisms at problems where they have manipulating agents, which is likely in practice. Pathak and Sönmez (2013) provide two results. First, at each problem where the constrained Boston mechanism has no manipulating student, the constrained student-proposing GS mechanism also has no manipulating student. Second, at each problem where the constrained student-proposing GS has no manipulating student, the student-proposing GS with an extended constraint also has no manipulating student. Our results on counting imply these results and generalize the comparisons to all problems.

[5]The social choice literature has suggested many other methods to compare manipulable mechanisms. For example, voting rules can be compared by counting the manipulable instances in the entire domain (Kelly, 1993; Aleskerov and Kurbanov, 1999), by finding the domains where some rules become strategy-proof while others do not (Moulin, 1980), and by set-inclusion of preference relations that admit dominant strategies (Arribillaga and Massó, 2016). Recently, Chen et al. (2016) studied matching with contracts (Hatfield and Milgrom, 2005) and introduced a manipulability notion that measures the set of contracts that each agent can obtain by misreporting her preference. They show that manipulability comparisons of stable mechanisms are equivalent to preference comparisons.

Similarly, two subsequent papers compared the constrained Boston and constrained GS mechanisms. They use different criteria that do not rely on instance-by-instance comparison. Bonkoungou and Nesterov (2021) used a criterion called strategic accessibility, and Decerf and Van der Linden (2020) used the notion of dominant preference-inclusion introduced by Arribillaga and Massó (2016). These criteria and counting are logically independent.

For the college admission problem, it is well-known that the student-proposing GS is not manipulable by students (Roth, 1982; Dubins and Freedman, 1981) while any stable mechanism is manipulable by schools. Furthermore, Pathak and Sönmez (2013) show that any manipulating student in a stable mechanism is also a manipulating student in the school-proposing GS mechanism, while any manipulating school in the school-proposing GS mechanism via preferences is also a manipulating school in any stable mechanism via preferences. Chen et al. (2016) define a notion of manipulability that compares the set of outcomes that each agent can obtain via manipulations and show that manipulability comparisons of stable mechanisms are equivalent to preference comparisons. Since the preferences of agents on the two sides over stable matchings are opposed, stable mechanisms are not comparable for all agents. Andersson et al. (2014b) also define a manipulability notion that compares each agent's maximal gain from manipulation and find least manipulable budget-balanced and envy-free mechanisms.

The rest of the paper is structured as follows. In Section 2, we study the college admission problem, in which both sides are strategic. In section 3, we study the school choice problem, in which only one side is strategic. In Section 4, we conclude. We present the proofs in the Appendix.

2. Two-sided matching

We consider the college admission problem (Gale and Shapley, 1962). There are students and schools; students have preferences over schools and schools have preferences over students. Both students and schools are strategic. Apart from college admission, this model covers numerous real-life matching markets such as the entry-level medical market in Canada, the UK, and the US.

Formally, the set $I \cup S$ of agents consists of a non-empty and finite set I of students with a generic element i and a non-empty and finite set S of schools with a generic element s. Being unmatched is denoted by $\emptyset$. Each student i has a strict preference relation P_i over the set $S \cup \{\emptyset\}$ of schools and remaining unmatched. Then $s \; P_i \; \emptyset$ means that school s is acceptable to student i. Let R_i denote the "at least as good as" relation associated with P_i.[6]

Each school s has $q_s \in \mathbb{N}$ seats called capacity and a strict preference relation P_s over $2^I \cup \{\emptyset\}$ where 2^I is the set of all non-empty subsets of students and $\emptyset$ the option of being unmatched. Let R_s denote the "at least as good as" relation associated with P_s. In particular, P_s induces a strict linear ordering over individual students which we denote by $\succ_s$, i.e, $i \succ_s j$ if and only if $\{i\} \; P_s \; \{j\}$. We assume that the preference relation P_s over groups of students is responsive to $\succ_s$, meaning that (a) admitting any acceptable student when there is an empty seat is better than leaving the seat unfilled and (b) replacing any student with a more preferred student leads to a better student body. Formally, the preference relation P_s of school s over groups of students is responsive (Roth, 1985) if (a) for each each $N \in 2^I$ such that $|N| < q_s$ and each $i \notin N$, we have $N \cup \{i\} \; P_s \; N \Leftrightarrow i \succ_s \emptyset$ and (b) for each $N \in 2^I$ and each $i, j \notin N$, we have $N \cup \{i\} \; P_s \; N \cup \{j\} \Leftrightarrow i \succ_s j$.

We denote by $P = (P_a)_{a \in I \cup S}$ the preference profile and by $q = (q_s)_{s \in S}$ the vector of capacities. Given an agent $a \in I \cup S$, let P_{-a} denote the preference profile of agents other than a. Given a school s, let q_{-s} denote the capacity vector of schools other than s. The tuple (I, S, P, q) is a college admissions problem, or simply a **problem**. We keep the sets I and S fixed and simply denote a problem by (P, q).

A **matching** is a function $\mu : I \to S \cup \{\emptyset\}$ mapping the set of students to the set of schools as well as the unmatched option such that no school is assigned to more students than it has seats for, that is, for each school s, $|\mu^{-1}(s)| \leq q_s$. Student i finds matching μ at least as good as matching μ' if and only if $\mu(i) \; R_i \; \mu'(i)$. School s finds matching μ at least as good as matching μ' if and only if $\mu^{-1}(s) \; R_s \; \mu'^{-1}(s)$. A **mechanism** φ is a function that maps each problem to a matching. If $\varphi(P, q) = \mu$ for a problem (P, q), then we denote by $\varphi_i(P, q) = \mu(i)$ the assignment of student i and by $\varphi_s(P, q) = \mu^{-1}(s)$ the set of students assigned to school s.

2.1. **National Resident Matching Program.** The entry-level labor market for doctors in the US has a centralized matching system called the National Resident Matching Program (NRMP). In 1951, after decades of chaotic experience with unraveling and congestion in a decentralized process, this market has adopted a centralized clearinghouse (Roth, 1984).

[6]For each $s, s' \in S \cup \{\emptyset\}$, $s \; R_i \; s'$ iff $s \; P_i \; s'$ or $s = s'$.

While several centralized markets have failed, NRMP has survived for half of a century with the same matching mechanism. Roth (1984) discovered an important feature of its design which may explain its success: no doctor and hospital have an opportunity to circumvent the system and mutually arrange a preferred matching. Such mechanism is said to be stable. Roth (1990, 1991) documented that stable matching mechanisms tend to operate longer while unstable ones tend to fail.

Formally, matching μ is **stable** at the problem (P, q) if (a) it is individually rational — every student is assigned to an acceptable school and every school is assigned acceptable students, and (b) it is not blocked — no student prefers a school which has an empty seat or has admitted a less preferred student. That is,

- μ is **individually rational** at (P, q): for each $s \in S$ and each $i \in \mu^{-1}(s)$, we have $s \; P_i \; \emptyset$ and $i \succ_s \emptyset$ and
- μ is **not blocked** at (P, q): there exists no school s and student $i \notin \mu^{-1}(s)$ such that $s \; P_i \; \mu(i)$ and either $[|\mu^{-1}(s)| < q_s$ and $i \succ_s \emptyset]$ or $[i \succ_s j$ for some $j \in \mu^{-1}(s)]$.

A mechanism φ is stable if for each problem (P, q) its outcome $\varphi(P, q)$ is stable at (P, q). Gale and Shapley (1962) show that for any problem there exists a stable matching. The set of stable matchings has a lattice structure such that there is an element, called student-optimal stable matching, where for each student it is at least as good as any other stable matching. Gale and Shapley (1962) develop the following algorithm for producing student-optimal stable matchings.

- *Step 1:* Each student applies to her most preferred acceptable school (if any). Let I_s^1 denote the *acceptable* applicants of school s at this step. Each school s tentatively accepts $\min(q_s, |I_s^1|)$ of the most preferred acceptable applicants among I_s^1 and rejects the remaining ones. Let A_s^1 denote the tentative acceptances of school s at this step.
- *Step t*, $t > 1$: Each student who is rejected at step $t-1$ applies to her most preferred acceptable school from those she has not yet applied to (if any). Let I_s^t denote the new acceptable applicants of school s at this step. Each school s considers the new and the previously held applicants and tentatively accepts $\min(q_s, |A_s^{t-1} \cup I_s^t|)$ of the most preferred acceptable applicants among $A_s^{t-1} \cup I_s^t$ and rejects the rest. Let A_s^t denote the tentative acceptances of school s at this step.

The algorithm stops when every student is either tentatively accepted or has been rejected by all her acceptable schools. Then the tentative acceptances at this step become the final matching and students who have been rejected by all acceptable schools remain unmatched. The student-proposing GS mechanism assigns to each problem (P, q) the matching $GS(P, q)$ obtained by this algorithm. Similarly, for each problem, there is a school-optimal stable matching which can be obtained by applying the school-proposing deferred acceptance algorithm.

Unfortunately, stable matching mechanisms are subject to various kinds of manipulations by students and schools. Students may gain by misrepresenting their preferences while schools may gain by misrepresenting their preferences or under-reporting their capacities. We formalize this as follows.

Definition 1 (Manipulation via preferences and capacities)**.**

(a) We say that ***student*** i ***is a manipulating agent*** *of mechanism* φ *at* (P, q) *if there is* $\hat{P}_i$ *such that*

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

(b) We say that ***school*** s ***is a manipulating agent*** *of mechanism* φ *at* (P, q) *if there is* $(\hat{P}_s, \hat{q}_s)$ *such that* $\hat{q}_s \leq q_s$ *and*

$$\varphi_s(\hat{P}_s, P_{-s}, (\hat{q}_s, q_{-s}))\ P_s\ \varphi_s(P, q).$$

Every stable matching mechanism is subject to manipulation by students or schools. Specifically, the student-proposing GS mechanism is not manipulable by students (Dubins and Freedman, 1981; Roth, 1982), while any stable matching mechanism is manipulable by schools (see, e.g., Sönmez, 1997). The school-proposing GS mechanism is manipulable by both students and schools.

The NRMP had been using the school-proposing GS mechanisms since its successful redesign in 1951-1952. However, among other issues, its manipulability had been observed and criticized for years (Williams, 1995; Ma, 2010), until, in 1998, the NRMP decided to switch to the Roth and Peranson's (1999) algorithm that is based on the student-proposing GS mechanism. One of the reasons, according to the American Medical Student Association and the Public Citizen Health Research Group, was that *"it would be best to choose the student-optimal algorithm to remove incentives, at least for students. In other words, within the set of stable algorithms, you either have incentives for both the hospitals and the students to misrepresent their true preferences or only for the hospitals"* (Ma, 2010).

However, it was later discovered that the student-proposing GS mechanism has weakly more manipulating schools than any other stable matching mechanism (Pathak and Sönmez, 2013).[7] Thus, there is an incentive trade-off and it was not clear if the student-proposing GS mechanism is preferable in terms of incentives for students and schools. The following theorem answers this question.

Theorem 1. *Suppose that both students and schools are strategic agents and can manipulate via preferences and capacities. For any problem, the student-proposing GS mechanism has fewer or an equal number of manipulating agents compared to any other stable matching mechanism.*

One of the implications of the theorem is that in a marriage market (where each school has one seat) the optimal stable matching mechanisms have the same number of manipulating agents.[8]

Corollary 1. *Suppose that every school has one seat. Then, for any problem, the student-proposing and the school-proposing GS mechanisms have the same number of manipulating agents.*

We discuss the main argument of the theorem now. The intuition as to why the student-proposing GS has fewer manipulating agents compared to any stable matching mechanism is that students and schools have opposing interests over stable matching mechanisms. To see this, consider a problem (P, q) and let φ be a stable matching mechanism. Note that every school finds $\varphi(P, q)$ at least as good as $GS(P, q)$. By implementing the matching $\varphi(P, q)$ instead of $GS(P, q)$ some schools receiving their more preferred stable matching do not have any interest in misrepresenting their preferences or capacities. These schools are matched with different students between these two stable matchings. An important basic result in two-sided matching, called the rural hospital theorem, implies that each such school has filled all its seats under any stable matching. Therefore, some students were matched with this school under $GS(P, q)$ but are matched to different schools under $\varphi(P, q)$. Because $GS(P, q)$ is the student-optimal stable matching, these students are worse off under $\varphi(P, q)$ compared

[7]This result follows from a comparison using a stronger criterion, namely that every manipulating school of an arbitrary stable mechanism is a manipulating school of the student-proposing GS mechanism (Pathak and Sönmez, 2013). The same manipulability criterion is used in matching with contracts framework by Chen et al. (2014, 2016).

[8]The analogy with the marriage market is limited and cannot provide a complete intuition for the college admissions problem. For example, when a marriage market has a unique stable matching, both student-proposing GS and school-proposing GS coincide and have zero manipulating agents. In a general college admissions problem, however, even if there is a unique stable matching, schools can manipulate the student-proposing GS and the school-proposing GS; but the student-proposing GS remains the least manipulable among all stable mechanisms (see Example 1).

to $GS(P, q)$. Finally, these students are manipulating agents of φ at (P, q) as each of them can truncate her preferences and get the same school as under $GS(P, q)$.

This argument crucially depends on whether or not schools can misrepresent their preferences. Next, we address applications where schools can only manipulate their capacities. This application is also relevant for college or university admissions. Their preferences, then called priorities, are often mandated by state or local laws. For example, when schools have to report the rankings of students according to their grades, no school has a say on these rankings. But schools may still under-report their capacities. Manipulation via capacities is first formalized by Sönmez (1997):

Definition 2 (Manipulation via capacities)**.** *We say that school s can manipulate its capacity under mechanism φ at problem (P, q) if there is $\hat{q}_s < q_s$ such that*

$$\varphi_s(P, (\hat{q}_s, q_{-s}))\ P_s\ \varphi_s(P, q).$$

This distinction is important because manipulations via capacities are more restrictive. For stable matching mechanisms, and when both manipulations via preferences and via capacities are possible, any manipulation via capacities can be replicated by manipulation via preferences while the reverse is not true. In particular, a manipulation via preferences by which a school removes some acceptable students and ranks the remaining ones according to its true preferences, called a dropping strategy, is exhaustive (Kojima and Pathak, 2009). This means that it can be used to improve the outcome of any strategy. When schools can manipulate their preferences, we can focus on dropping strategies without loss of generality. Nevertheless, this distinction does not change our conclusion.

Proposition 1. *Suppose that students and schools are strategic agents but schools can only manipulate their capacities. For any problem, the student-proposing GS mechanism has fewer or an equal number of manipulating agents compared to any stable matching mechanism.*

We provide the main argument here. Following Ehlers (2010), we know that only schools that have filled all their seats can manipulate the student-proposing GS mechanism via capacities. As before, any school that can manipulate the student-proposing GS via capacities but not another stable matching mechanism φ is matched with different sets of students under these stable matchings. Therefore, some students are matched to this school under the student-proposing GS mechanism but not under φ. Such students are manipulating agents of φ.

The following distinction between preference manipulations and capacity manipulations is worth noting. We previously discussed the strategic trade-off between manipulating students and manipulating schools. In contrast, when we restrict ourselves to capacity manipulations, the trade-off may disappear and the school-proposing GS mechanism may have *more* manipulating schools than the student-proposing GS mechanism. In other words, there might be schools that cannot manipulate the student-proposing GS mechanism via capacities but can manipulate the school-proposing GS mechanism (see also Ehlers, 2010). We illustrate this next.

Example 1. *Let there be two schools s_1 and s_2 and three students i_1, i_2 and i_3. Let (P, q) be a problem such that $q_{s_1} = q_{s_2} = 2$ and P is specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	$\succ_{s_1}$	$\succ_{s_2}$
s_2	s_1	s_1	i_1	i_3
s_1	s_2	s_2	i_2	i_1
			i_3	i_2

For this problem, there is only one stable matching specified as follows:

$$\begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_1 \end{pmatrix}.$$

Thus, this matching is the outcome of the student-proposing GS mechanism as well as the school-proposing GS mechanism. Next, we show that s_2 cannot manipulate the student-proposing GS mechanism via capacities but can manipulate the school-proposing GS mechanism via capacities.

Each student is matched to her most preferred school. Thus, there is no manipulating student for the school-proposing GS mechanism. We only focus on manipulating schools. Suppose that school s_1 reports capacity $q'_{s_1} = 1$ to the student-proposing GS mechanism. Then during the first step of the algorithm, s_1 will reject the application of i_3. Student i_3 applies to school s_2 in the second step and the algorithm ends. School s_1 will finally be matched to student i_2. Since P_{s_1} is responsive, then $\{i_2, i_3\}$ P_{s_1} $\{i_2\}$. Thus, school s_1 cannot manipulate the student-proposing GS mechanism via capacities. Since school s_2 received one application in the first step of the student-proposing GS algorithm, reporting $q'_{s_2} = 1$ or $q_{s_2} = 2$ does not change the outcome. Therefore, school s_2 cannot manipulate the student-proposing GS mechanism via capacities.

Suppose now that school s_2 reports $q'_{s_2} = 1$ to the school-proposing GS mechanism. Then the algorithm stops at the first step and s_2 is matched to student i_3. Since

$i_3 \succ_{s_2} i_1$, then school s_2 can manipulate the school-proposing GS mechanism via capacities.

In all our previous applications, students can rank as many schools as they wish. In other applications, however, there are restrictions on the number of schools that students can rank. Next, we study this setting.

2.2. **New York City High School Match.** Before 2003, the New York City high school matching system was decentralized and highly congested: one-third of applicants were unassigned and had to be administratively placed in schools that they did not list (Abdulkadiroğlu et al., 2005a). An important feature of this system was that schools were strategic agents. Although schools' rankings of students were based on their place of residence or whether they already have a sibling attending the school, "*a substantial number of schools apparently managed to conceal capacity from the central administration, thus preserving places that could be filled later*" (Abdulkadiroğlu et al., 2005a). These observations convinced the designers that the market was two-sided and thus required stability. One of the design decisions was whether to implement the student-proposing GS mechanism or another stable matching mechanism.

This application has two important features. First, before the reform, each student could apply only to 5 schools out of more than 600 schools. After the reform, this constraint was not eliminated but was extended to 12 schools.[9] Second, although schools are strategic agents, they can only misrepresent their capacities because their rankings of students are exogenous. In this setting, which mechanism would we recommend?

The standard argument, supporting that student-proposing GS is the best because it is not manipulable by students while any stable matching mechanism is manipulable by schools, cannot guide the choice in this application because the constraint makes the student-proposing GS manipulable by students.[10] Our criterion is also useful to distinguish manipulable mechanisms. To develop our argument further we need the following notation and terminology.

Let $k \in \{1, \ldots, |S|\}$. For each student i, the truncation after the k'th acceptable school (if any) of P_i with x acceptable schools is the preference relation P_i^k with $\min(x, k)$ acceptable schools such that all schools are ordered as in P_i.

[9] Abdulkadiroğlu et al. (2005a) documented that after the reform over 12,000 students ranked 12 schools, suggesting that the constraint was binding. The designers were aware that these constraints worsen the properties of the mechanisms, but the matching system officials had their reasons to maintain the constraints (Abdulkadiroğlu et al., 2005a; Pathak, 2016).

[10] Every sensible constrained mechanism is manipulable by students, see Proposition 2.

Definition 3. *Let $k \in \{1, \ldots, |S|\}$. The constrained version φ^k of the mechanism φ assigns to each problem (P, q) the matching $\varphi^k(P, q) = \varphi(P_I^k, P_S, q)$, where P_I^k is the profile of truncated preferences after the k'th acceptable school.*

The constrained mechanism φ^k is **constrained stable** if for each problem (P, q), $\varphi^k(P, q)$ is stable under (P^k, q). Despite additional strategic flaws due to constraints, the constrained student-proposing GS mechanism emerges as the best choice among all constrained stable matching mechanisms, including the constrained school-proposing GS. This result guides the design of the New York City high school match.

Theorem 2. *Let $k \geq 2$ and φ be a stable matching mechanism. Suppose that students can only rank up to k schools.*

(i) Suppose that schools can manipulate via preferences. Then the constrained student-proposing GS mechanism GS^k has fewer or an equal number of manipulating agents compared to the constrained stable matching mechanism φ^k.

(ii) Suppose that schools can only manipulate via capacities. Then the constrained student-proposing GS mechanism GS^k has fewer or an equal number of manipulating agents compared to the constrained stable matching mechanism φ^k.

The intuition behind the result is as follows. First, we show that for any problem (P, q), manipulating students of the constrained student-proposing GS are unmatched at the matching $GS^k(P, q)$. The rural hospital theorem (Roth, 1986) implies that these students are also unmatched at constrained stable matching $\varphi^k(P, q)$. The main part of the argument is to show that they are also manipulating students of the constrained stable matching mechanism φ^k at (P, q). The reason is that the strategy for manipulating the student-proposing GS mechanism can be replicated to constitute a manipulating strategy of φ^k — again due to the rural hospital theorem. Therefore, every (unmatched) manipulating student of the constrained student-proposing GS mechanism is also an unmatched manipulating student of any constrained stable matching mechanism φ^k. The rest follows similar ideas as the proof of Theorem 1 and Proposition 1.

Finally, we note that the results in this section do not rely on the fact that the problem having multiple stable matchings. Roth and Peranson (1999) observed that, in the NRMP, the core tends to be relatively small. This core "convergence" can be explained by the large size of the market, competition and interview requirements that restrict the number of hospitals students can rank (Roth and Peranson, 1999; Kojima and Pathak, 2009; Ashlagi et al., 2017). In markets where there is a unique stable matching, students cannot manipulate their preferences, but schools can still manipulate their preferences as well as their capacities, as Example 1 demonstrates.

3. School Choice

So far we have considered settings where both sides are strategic. In this section, we consider school admission systems with strategic students and study recent design choices in Chicago, Denver, Ghana, and the UK.

In a school choice problem (Abdulkadiroğlu and Sönmez, 2003), only students' welfare matters and they are the only strategic agents. The seats of each school are treated as objects to be distributed among students, and its preference relation is interpreted as students' priorities in distributing these seats. These priorities are often mandated by local/state laws and are based on students' characteristics such as location, grades, socioeconomic status, or lottery outcomes. In contrast to the college admission model, we assume that each student is acceptable to each school. That is, for each student i and each school s, $i \succ_s \emptyset$. Most importantly, we assume that priorities and capacities are reported truthfully and focus on student manipulations.

In their seminal paper, Abdulkadiroğlu and Sönmez (2003) describe the matching procedure used in Boston and other US cities. The procedure is commonly used around the world and is known in the literature as the Boston mechanism.[11] For each problem (P, q), it works as follows:

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

The algorithm stops when every student is either accepted at some step or has been rejected by all her acceptable schools. Every school is assigned the students that it has immediately accepted at each step. The Boston mechanism assigns to each problem

[11] It is also called the immediate acceptance mechanism.

(P, q), the matching $\beta(P, q)$ obtained by this algorithm. The Boston mechanism is individually rational but does not always produce a stable matching (Abdulkadiroğlu and Sönmez, 2003). This mechanism is also manipulable.

We also study constrained rankings in school choice. Real-life examples include the Chicago selective high school admission system, the Boston public school system before the reform in 2005, the Ghanaian primary public school system, the New York City high school match, the primary public school system in Denver, and school admissions systems in several cities in UK (Pathak and Sönmez, 2013; Fack et al., 2019). With such constraints, students need to be strategic about which schools to include in their rankings. It turns out that for any *sensible* mechanism, its constrained counterpart will be manipulable.

To be more specific, we define a sensible mechanism as one that is individually rational and satisfies the following mild efficiency condition. We say that a matching μ is **weakly non-wasteful** under P if there is no unmatched student who prefers a school with an empty seat. That is, there is no student i and a school s such that $\mu(i) = \emptyset$, $s \; P_i \; \emptyset$ and $|\mu^{-1}(s)| < q_s$. A mechanism is weakly non-wasteful if for each problem (P, q), its outcome $\varphi(P, q)$ is weakly non-wasteful under P.

Proposition 2. *Let $k \geq 1$ and suppose that there are at least as many students as schools, $|I| \geq |S|$, and $k < |S|$. Let φ be a weakly non-wasteful and individually rational mechanism. Then, the constrained mechanism φ^k is manipulable.*

In recent years, strategic concerns have motivated many school districts to reform their admissions systems (Pathak and Sönmez, 2013). Some school districts replaced the Boston mechanism with the student-proposing GS mechanism but maintained the ranking constraints. Other reforms allowed students to apply to more schools. None of these reforms eliminated manipulation, but, as we show next, they reduced the number of manipulating agents. Such a conclusion supports these reforms and could serve as a guide for future designs. Next, we describe these reforms and present our results.

3.1. **Chicago, Denver, and England: from immediate to deferred acceptance.** Several matching systems have replaced the constrained Boston mechanism with a student-proposing GS mechanism. Examples include the Chicago selective high school system in 2009 and the public school system in many cities in England (see Table 1 and Pathak and Sönmez, 2013).

Replacing the manipulable Boston mechanism with the non-manipulable student-proposing GS, is an obvious improvement. However, for constrained mechanisms, the comparison is not straightforward because some students who could not manipulate the constrained Boston mechanism can manipulate the constrained GS. To see this, consider the following example.

Example 2. *There are five students $i_1, i_2, \ldots, i_5$ and five schools $s_1, s_2, \ldots, s_5$. Let (P, q) be a problem such that 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}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$	$\succ_{s_5}$
s_1	s_1	s_2	s_3	s_3	i_4	i_5	i_2	$\vdots$	$\vdots$
s_2	s_2	s_3	s_1	s_1	i_1	i_1	i_3		
s_3	s_3	s_4	s_2	s_2	i_2	i_2	i_5		
$\vdots$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\vdots$	i_3	i_4		

Consider replacing the constrained Boston mechanism β^2 with GS^2. The outcome of β^2 is as follows:

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

Students i_2 and i_4 are manipulating students: i_2 could benefit by top-ranking s_2 and being matched to it, while i_4 could benefit by top-ranking s_1 and being matched to it. Each of the remaining students received her most preferred school and thus cannot manipulate β^2 at (P, q). But under GS^2 student i_5 becomes a manipulating student. To see this, consider the outcome of GS^2:

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

Student i_5 is unmatched. However, she is the highest priority student at s_2. If she top-ranks s_2 (or even ranks it second), then she is matched to it under the new problem: $GS^2_{i_5}(P^{s_2}_{i_5}, P_{-i_5}, q) = s_2$. Therefore, i_5 is a manipulating student of GS^2 but not β^2.

Thus, when the constrained Boston is replaced with the constrained GS, the set of manipulating students changes ambiguously. But, as we show in the following theorem, the size of this set unambiguously decreases.

Theorem 3. *Let $k > 1$. For any problem, the constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained Boston mechanism β^k.*

The main and novel part of the proof is to construct a one-to-one function between manipulating students of GS^k and a subset of manipulating students of β^k. We illustrate how we construct this function using Example 2 above.

Recall, that β^2 has two manipulating students i_2 and i_4, and GS^2 has two manipulating students i_2 and i_5. Let us ignore i_2 and focus on i_4 and i_5. (In the proof, we show that if a manipulating student of β^k is unmatched under GS^k, which is the case for student i_2, then this student remains a manipulating student of GS^k.)

We now show that, by replacing β^2 with GS^2, the manipulating statuses of students i_4 and i_5 are changed correspondingly. Note that for $\beta^2(P, q)$, student i_5 is matched to school s_3, which was assigned to student i_3 under $GS^2(P, q)$. Student i_3 is matched to school s_2, which was assigned to student i_1 under $GS^2(P, q)$. Finally, student i_1 is matched to school s_1, which was assigned to student i_4 under $GS^2(P, q)$ and student i_4 is unmatched. We draw a sequence of these links as follows:

$$i_5 \xrightarrow{s_3} i_3 \xrightarrow{s_2} i_1 \xrightarrow{s_1} i_4,$$

where every student is pointing at the student who was assigned under $GS^2(P, q)$ to the school that she is assigned to under $\beta^2(P, q)$. The last student, i_4, is not assigned under $\beta^2(P, q)$ to any school that was assigned under $GS^2(P, q)$ to any student and thus does not point at any student. Student i_4 is a manipulating student of β^2 at (P, q). Thus, the number of manipulating students of GS^2 is not greater than the number of manipulating students of β^2.

The steps of the proof involve showing the following points:

- Each manipulating student of GS^2 at (P, q) who is unmatched under $\beta^2(P, q)$ is also a manipulating student of β^2 at (P, q).
- Starting from each manipulating student of GS^2 at (P, q) who is matched under $\beta^2(P, q)$, the pointing sequence ends at a manipulating student of β^2 at (P, q).
- Two sequences lead to different manipulating students of β^2 at (P, q).

More generally, the function in question is constructed as follows (see Figure 1). The set of manipulating students of GS^k are distinguished into those who are matched under β^k, M, and those who are unmatched under β^k, $M^{\emptyset}$. Our function returns each student in $M^{\emptyset}$ to herself via an identity relation I_d and each student in M

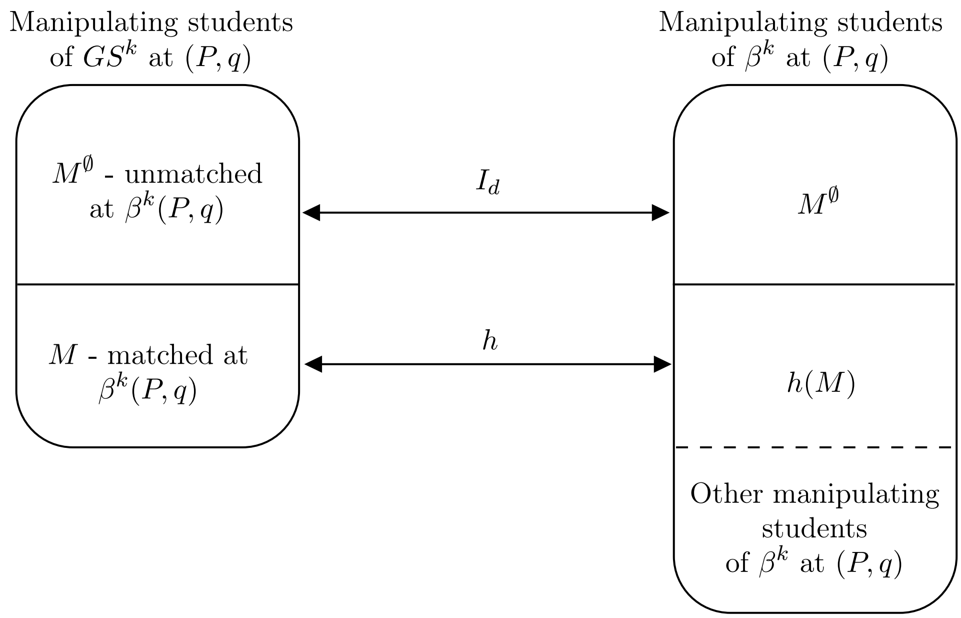

Figure 1. Relation between manipulating students of GS^k and β^k.

(initiator of a sequence) to the student closing this sequence via a relation h. The set of manipulating students of β^k includes $M^{\emptyset} \cup h(M)$ and possibly others.

3.2. **Chicago, Ghana, and England: extending constraints.** The second change involved an extension of the ranking constraint under the GS mechanism. These changes were observed in Chicago in 2010, in Ghana in 2007 and 2008, and in two cities in England (see Table 1 and Pathak and Sönmez, 2013).

The effect of these changes is also nuanced because some students who could not manipulate GS with a tighter constraint can manipulate GS with an extended constraint. To see this, consider the following example.

Example 3. *Consider the problem* (P, q) *in Example 2, and let* GS^1 *be replaced by* GS^2*. The outcome of* GS^1 *at the problem* (P, q) *is as follows:*

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

Student i_5 *received her most preferred school,* s_3*, and thus cannot manipulate* GS^1 *at* (P, q)*. But, as we saw in Example 2, student* i_5 *can manipulate* GS^2 *at* (P, q)*.*

By extending the constraint in the constrained student-proposing GS mechanism, the set of manipulating students does not change in an inclusion order. However, as the following theorem shows, the size of this set decreases.

Theorem 4. *Let $k > \ell \geq 1$. For any problem, the constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained GS mechanism GS^ℓ.*

The main and novel part of the proof involved the construction of intermediary mechanisms, in which the constraint changes for only one student. For each subset N of students, we construct a mechanism GS^N that assigns to each problem (P, q) the following matching: $GS(P_N^\ell, P_{I\setminus N}^k, P_S, q)$. That is, the constraint ℓ applies to students in N while the constraint k applies to the remaining students. Thus, $GS^\emptyset = GS^k$ and $GS^I = GS^\ell$. For each problem (P, q), we count and compare the number of manipulating students of $GS^\emptyset$, $GS^{\{i\}}, \ldots, GS^I$ at (P, q). The following example illustrates the comparison.

Example 4. *Consider the same problem as in Example 2. At problem (P, q), we compare the number of manipulating students of $GS^\emptyset = GS^2$ – where all students have an extended constraint $k = 2$, and $GS^{\{i_1\}} = GS(P_{i_1}^1, P_{-i_1}^2)$ – where student i_1 has a smaller constraint $\ell = 1$.*

Student i_1 is unmatched at the matching

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

Student i_2 is matched at $GS^{\{i_1\}}(P, q)$ and thus is not a manipulating student of $GS^{\{i_1\}}$ at (P, q). However, she was a manipulating student of GS^2 at (P, q).

Student i_1 is a manipulating student of $GS^{\{i_1\}}$ at (P, q). Indeed, if she misrepresents her preferences by ranking school s_2 first, she will be matched to it:

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

Student i_5 also remains unmatched under $GS^{\{i_1\}}(P, q)$ and is a manipulating student of $GS^{\{i_1\}}$ at (P, q). To sum up, there are two manipulating students, i_2 and i_5, of GS^2 at (P, q). One student, i_2, is no longer a manipulating student of $GS^{\{i_1\}}$ at (P, q). However, one new manipulating student, i_1, of $GS^{\{i_1\}}$ at (P, q) appears. Thus, when we replace GS^2 by $GS^{\{i_1\}}$ the number of manipulating students in the example did not decrease.

To prove the theorem, we first prove that for each proper subset N of students and each $i \notin N$, there are weakly more manipulating students of $GS^{N\cup\{i\}}$ at (P, q) compared to GS^N. The most difficult steps involve showing that:

- *there is at most one student j, who is a manipulating student of GS^N at (P, q) and who is not a manipulating student of $GS^{N\cup\{i\}}$ at (P, q) and, if such student j exists, then*
- *student i is a manipulating student of $GS^{N\cup\{i\}}$ but not a manipulating student of GS^N at (P, q), thereby "compensating" for the removal of the manipulating student j.*

We conclude that there are weakly more manipulating students of $GS^{\{i_1\}}$ at (P, q) compared to $GS^{\emptyset}$. Similarly, there are weakly more manipulating students of $GS^{\{i_1,i_2\}}$ at (P, q) compared to $GS^{\{i_1\}}$. By a repeated application of this argument, there are weakly more manipulating students of $GS^I = GS^\ell$ at (P, q) compared to $GS^{\emptyset} = GS^k$.

From these results, we get an immediate corollary that the constrained Boston mechanism has weakly more manipulating students compared to the constrained GS mechanism with a longer list. These changes occurred in one step in Kent (2007) and in Denver (2012), and in two steps in Chicago (2009 and 2010) and in Newcastle (2005 and 2010) (see Table 1 and Pathak and Sönmez, 2013).

Corollary 2. *Let $k > \ell \geq 1$. For any problem, the constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained Boston mechanism β^ℓ.*

Overall, our results — Theorem 3, Theorem 4, Corollary 2 — provide novel, stronger support for the reforms mentioned above. The state-of-the-art notion for comparing manipulable matching mechanisms, due to Pathak and Sönmez (2013), is the set inclusion of non-manipulable profiles. For each reform, they show that at each profile where the old mechanism has zero manipulating students, the new mechanism also has zero manipulating students. In other words, for the problems where the new mechanism is non-manipulable, the old mechanism has weakly more manipulating students. Thus, this notion coincides with our notion, but only in this restricted domain where the old mechanism has zero manipulating students, and our results imply the results in Pathak and Sönmez (2013). We conclude this section with a simulation that helps estimate the size of this gain.

3.3. **Numerical simulations.** We consider the school choice setting analogous to Ergin and Erdil (2008). Let there be $N = 100$ students and $K = 10$ schools each with capacity $q = 10$, uniformly randomly distributed on a 2-dimensional unit square. The priorities of schools over students are based on euclidean distance: the closer a student is located to a school – the higher is her priority at this school.

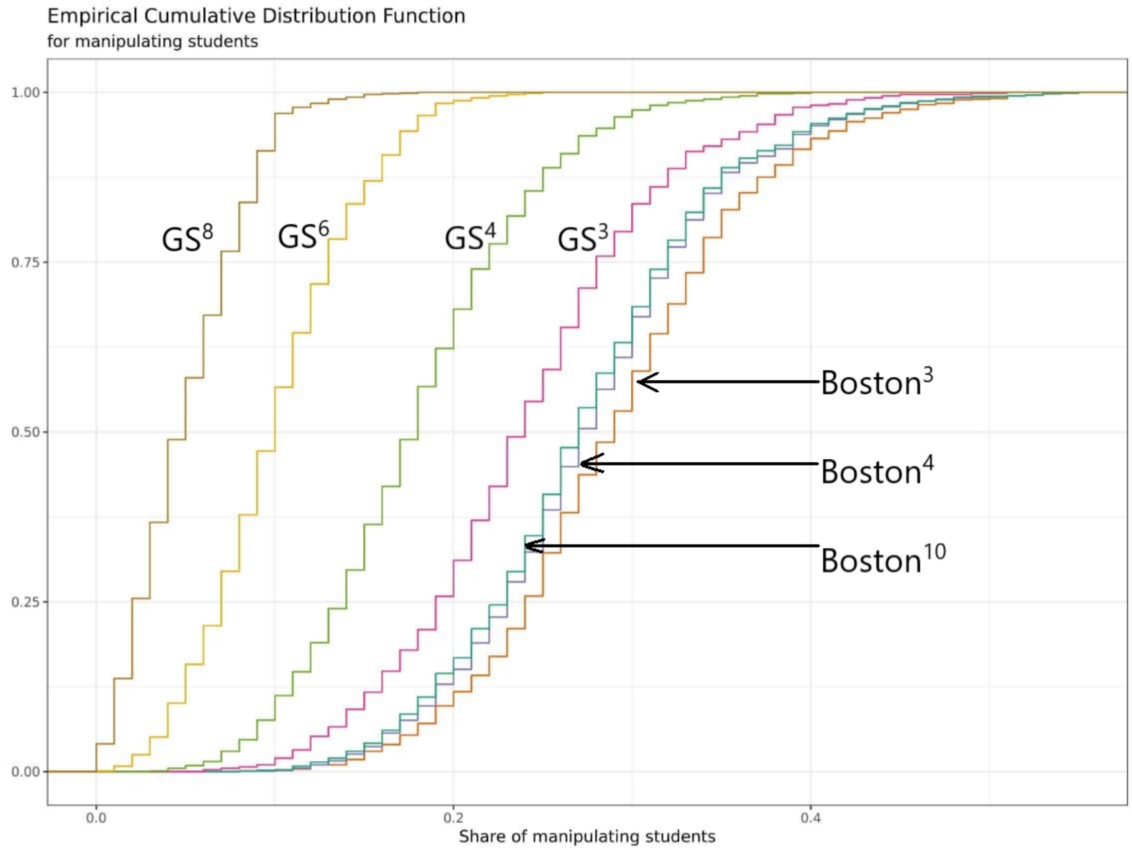

Figure 2. Comparing mechanisms by the number of manipulating students.

Notes: The diagram shows the empirical cumulative distributions of the share of manipulating students. The horizontal axis represents the share of manipulating students out of $N = 100$ students; one step of each curve is equivalent to an additional manipulating student. The vertical axis represents the share of profiles that have a particular number of manipulating students.

The utility of student i from being matched to school s is determined by three components: the euclidean distance from i to s, $d(i, s)$; the objective quality of s which is normally distributed and is common to all students, $Z_s \sim N(0, 1)$, and the intrinsic preference of i which is also normally distributed, $Z_{is} \sim N(0, 1)$:

$$U_{is} = -\beta d(i, s) + (1 - \beta)(\alpha Z_s + (1 - \alpha)Z_{is}),$$

where weights $\alpha, \beta \in [0, 1]$.

In the simulations, we consider the constrained Boston mechanisms β^3, β^4, the unconstrained Boston mechanism $\beta = \beta^{10}$, and the constrained GS mechanisms

GS^3, GS^4, GS^6, and GS^8. For each mechanism, we simulate 1000 preference profiles for $\alpha = 0.6, \beta = 0.7$, and for each profile count the number of manipulating students.

The resulting empirical distribution of the number of manipulating agents is presented in Figure 2. We see that moving from β^4 to GS^4 and further to GS^6 substantially decreases the number of manipulating agents. For example, the median number — that is the number such that half of the profiles have fewer manipulating students and half have more — decreases from 27 manipulating students for β^4 to 18 students for GS^4 and further to 10 students for GS^6. We also see that the manipulability criterion of Pathak and Sönmez (2013) applies to a small set of profiles, in which the number of manipulating students is zero: 8% of profiles for GS^8, below 1% for GS^6, and is approximately zero for the other mechanisms.

4. Conclusion

Manipulations in real matching markets are undesirable yet often inevitable due to certain restrictions. This calls for designs that keep manipulations low enough for practical use. We addressed this issue by counting the number of manipulating agents and recommending one mechanism over another in several important domains.

Our analysis demonstrates that the student-proposing GS mechanism has the smallest aggregate number of manipulating agents – students and schools – among all stable matching mechanisms. This result is surprisingly robust to changes in few key aspects of the problem: whether schools can manipulate rankings or only capacities, and whether students can rank all schools or only a limited number of schools. Overall, our results rationalize and strongly support several recent design choices.

There are many aspects of real markets that would require independent treatment. First, there might be other manipulations, such as prearrangement — when the school reduces its seats by one unit and commits to admit a student provided that she does not participate in the market. A mechanism is vulnerable to prearrangements when there is a school and a student preferring what they receive under prearrangement than what they receive without it. Sönmez (1999) shows that any stable matching mechanism is vulnerable to prearrangements.

Second, there are multiple extensions of the standard many-to-one matching model. For example, in the medical labor market, some doctors seek dual jobs as a couple. When couples are present, a stable matching may not exist (Roth, 1984). Roth and Peranson (1999) and Kojima et al. (2013) define algorithms that find a stable matching when it does exist. However, these mechanisms are manipulable by agents on both sides.

Thus, one avenue for future work is to search for a mechanism with a lower number of manipulating agents, including prearrangements and when there are couples.

Finally, it would interesting to develop a criterion that measures the amount of manipulations that involves equilibrium analysis. In this paper, we count agents with an incentive to manipulate when others report their true preferences. However, these are not necessarily equilibrium strategies. The standard game-theoretic approach would involve counting the number of agents misreporting their preferences at a Nash equilibrium. The immediate difficulty with this approach is that, in general, there might be multiple equilibria, and thus no clear conclusions.

References

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005a). The New York City high school match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005b). The Boston public school match. *American Economic Review*, 91(2):368–371.

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

Aleskerov, F. and Kurbanov, E. (1999). Degree of manipulability of social choice procedures. In *Current trends in economics*, pages 13–27. Springer.

Andersson, T., Ehlers, L., and Svensson, L.-G. (2014a). Budget balance, fairness, and minimal manipulability. *Theoretical Economics*, 9(3):753–777.

Andersson, T., Ehlers, L., and Svensson, L.-G. (2014b). Least manipulable envy-free rules in economies with indivisibilities. *Mathematical Social Sciences*, 69:43–49.

Arribillaga, R. P. and Massó, J. (2016). Comparing generalized median voter schemes according to their manipulability. *Theoretical Economics*, 11(2):547–586.

Ashlagi, I., Kanoria, Y., and Leshno, J. D. (2017). Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69–98.

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

Chen, P., Egesdal, M., Pycia, M., and Yenmez, M. B. (2014). Quantile stable mechanisms.

Chen, P., Egesdal, M., Pycia, M., and Yenmez, M. B. (2016). Manipulability of stable mechanisms. *American Economic Journal: Microeconomics*, 8(2):202–14.

Chen, Y. (2008). Incentive-compatible mechanisms for pure public goods: A survey of experimental research. *Handbook of Experimental Economics Results*, 1:625–643.

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. *Available at SSRN 2809566*.

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

Dur, U., Hammond, R. G., and Morrill, T. (2019). The secure boston mechanism: Theory and experiments. *Experimental Economics*, 22(4):918–953.

Dur, U. M. (2019). The modified boston mechanism. *Mathematical Social Sciences*, 101:31–40.

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

Ehlers, L. (2010). Manipulation via capacities revisited. *Games and Economic Behavior*, 69(2):302–311.

Ergin, H. and Erdil, H. (2008). What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review*, 98(3):669–689.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

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.

Hakimov, R. and Kübler, D. (2020). Experiments on centralized school choice and college admissions: a survey. *Experimental Economics*, pages 1–55.

Hatfield, J. W. and Milgrom, P. R. (2005). Matching with contracts. *American Economic Review*, 95(4):913–935.

Kelly, J. S. (1993). Almost all social choice rules are highly manipulable, but a few aren't. *Social choice and Welfare*, 10(2):161–175.

Kojima, F. and Pathak, P. A. (2009). Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27.

Kojima, F., Pathak, P. A., and Roth, A. E. (2013). Matching with couples: Stability and incentives in large markets. *The Quarterly Journal of Economics*, 128(4):1585–1632.

Ma, J. (2010). The singleton core in the college admissions problem and its application to the National Resident Matching Program (NRMP). *Games and Economic Behavior*, 69(1):150–164.

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455.

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. (2016). What really matters in designing school choice mechanisms. In *Preparation for Advances in Economics and Econometrics, 11th World Congress of the Econometric Society.*

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.

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

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 92(6):991–1016.

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

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

Roth, A. E. (1990). New physicians: a natural experiment in market organization. *Science*, 250(4987):1524–1528.

Roth, A. E. (1991). A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the united kingdom. *The American Economic Review*, pages 415–440.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378.

Roth, A. E. (2008). What have we learned from market design? *Innovations: Technology, Governance, Globalization*, 3(1):119–147.

Roth, A. E. and Peranson, E. (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780.

Roth, A. E. and Sotomayor, M. A. O. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.

Sönmez, T. (1997). Manipulation via capacities in two-sided matching markets. *Journal of Economic Theory*, 77(1):197–204.

Sönmez, T. (1999). Can pre-arranged matches be avoided in two-sided matching markets? *Journal of Economic Theory*, 86(1):148–156.

Sönmez, T. (2013). Bidding for army career specialties: Improving the ROTC branching mechanism. *Journal of Political Economy*, 121(1):186–219.

Williams, K. J. (1995). A reexamination of the NRMP matching algorithm. National Resident Matching Program. *Academic Medicine*, 70(6):470–6.

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

Appendix

We need the following result that is very much used in the paper. The stable set has an interesting property called the rural hospital theorem. It says that (i) each agent is matched with the same number of partners across all stable matchings and (ii) every agent which is not matched or has unfilled seats is matched to the same set of partners across all stable matchings.

Lemma 1 (Rural hospital theorem, Roth, 1986)**.** *Suppose that schools have responsive preferences. Let (P, q) be a problem and let ν and μ be two stable matchings.*

(i) Each agent is matched with the same number of partners under ν and μ.

(ii) Suppose that for some school s, $|\mu^{-1}(s)| < q_s$. Then $\mu^{-1}(s) = \nu^{-1}(s)$.

The proofs of the following Theorem 1 and Proposition 1 coincide to a large extent and are presented as one proof.

Theorem 1: *Suppose that both students and schools are strategic agents and can manipulate via preferences and capacities. For any problem, the student-proposing GS mechanism has fewer or an equal number of manipulating agents compared to any other stable matching mechanism.*

Proposition 1: *Suppose that students and schools are strategic agents but schools can only manipulate their capacities. For any problem, the student-proposing GS mechanism has fewer or an equal number of manipulating agents compared to any other stable matching mechanism.*

Proof of Theorem 1 and Proposition 1. Let φ be a stable matching mechanism and (P, q) be a problem. Let M be the set of manipulating agents of GS. Since GS is not manipulable by students (Dubins and Freedman, 1981; Roth, 1982), we have $M \subset S$. Let $M^1 \subset M$ be the subset of schools that are also manipulating agents of φ at (P, q) and let $M^2 \subset M$ be the subset of schools that are not manipulating agents of φ at (P, q). Then $M = M^1 \cup M^2$. The idea is to show that for each school $s \in M^2$, there is a subset $I(s)$ of manipulating students of φ at (P, q) such that there is no intersection between two different subsets. We divide the rest in three steps.

Step 1: each school $s \in M^2$ is matched to a different set of students: $\varphi_s(P, q) \neq GS_s(P, q)$.

The proof of this step is different depending on whether we consider Theorem 1 or Proposition 1. We consider each case separately.

Case 1: proof of Step 1 for Theorem 1. Let $s \in M^2$. We prove it by contradiction. Suppose that $GS_s(P, q) = \varphi_s(P, q)$. Because s is a manipulating school of GS at (P, q), there is (P'_s, q'_s) such that $q'_s \leq q_s$ and

$$GS_s(P'_s, P_{-s}, q'_s, q_{-s}) \; P_s \; GS_s(P, q). \tag{1}$$

Kojima and Pathak (2009, Lemma 1) show that the following manipulation strategy, called dropping strategy, is exhaustive for any stable matching mechanism; in the sense that it can be used to improve upon, according to the true preferences, the outcome of any manipulation: a dropping strategy is any strategy that declares a subset of acceptable students as not acceptable but keeps the remaining acceptable students ranked as in the original strategy. In particular, they constructed a dropping strategy P^d_s such that the acceptable students is the set of students in $GS_s(P'_s, P_{-s}, q'_s, q_{-s})$ who are acceptable under P_s. By Kojima and Pathak (2009, Lemma 1), we have

$$GS_s(P^d_s, P_{-s}, q) \; R_s \; GS_s(P'_s, P_{-s}, q'_s, q_{-s}). \tag{2}$$

Lemma 1 (i) implies that school s is matched with the same number of students under both $GS_s(P^d_s, P_{-s}, q)$ and $\varphi(P^d_s, P_{-s}, q)$. Since $|GS_s(P'_s, P_{-s}, q'_s, q_{-s})| \leq q'_s \leq q_s$, there are less than q_s or an equal number of acceptable students under P^d_s. Therefore, since φ and GS are individually rational, they match s to the same set of students,

$$\varphi_s(P^d_s, P_{-s}, q) = GS_s(P^d_s, P_{-s}, q). \tag{3}$$

By equation 2 and 3 , we have $\varphi(P^d_s, P_{-s}, q)\ R_s\ GS_s(P'_s, P_{-s}, q'_s, q_{-s})$. Now, because the preference relation P_s is transitive, this equation and equation 1 imply that $\varphi_s(P^d_s, P_{-s}, q)\ P_s\ GS_s(P, q)$. Finally, because $GS_s(P, q) = \varphi_s(P, q)$ by assumption, we have

$$\varphi_s(P^d_s, P_{-s}, q)\ P_s\ \varphi_s(P, q).$$

This equation means that school s is a manipulating agent of φ at (P, q) and thus contradicting our assumption that school s is not a manipulating agent of φ at (P, q). Therefore $GS_s(P, q) \neq \varphi_s(P, q)$.

Case 2: proof of Step 1 for Proposition 1. Let $s \in M^2$. We also prove it by contradiction. Suppose that $\varphi_s(P, q) = GS_s(P, q)$. Because school s is a manipulating agent of GS at (P, q), then there is $q'_s < q_s$ such that

$$GS_s(P, q'_s, q_{-s})\ P_s\ GS_s(P, q). \tag{4}$$

Because $GS(P, q'_s, q_{-s})$ is the school-pessimal stable matching at (P, q'_s, q_{-s}), we have

$$\varphi_s(P, q'_s, q_{-s})\ R_s\ GS_s(P, q'_s, q_{-s}). \tag{5}$$

Since R_s is transitive, equation 4 and equation 5, and the fact that $\varphi_s(P, q) = GS_s(P, q)$ imply that

$$\varphi_s(P, q'_s, q_{-s})\ P_s\ \varphi_s(P, q).$$

This equation contradicts the assumption that school s is not a manipulating agent (via capacities) of φ at (P, q). Therefore, $\varphi_s(P, q) \neq GS_s(P, q)$.

We state this result as a lemma and use it in the proof of Theorem 2 below.

Lemma 2. *Let (P, q) be a problem. Suppose that schools can manipulate via preferences or capacities or only via capacities. Let φ be a stable mechanism. Let M^2 be the subset of schools which can manipulate GS but not φ at (P, q). Then for each school $s \in M^2$, $\varphi_s(P, q) \neq GS(P, q)$.*

Step 2: For each school $s \in M^2$, there is a non-empty subset $I(s)$ of manipulating students of φ at (P, q) such that no two such subsets have a common element.

Now, since $GS_s(P,q) \neq \varphi_s(P,q)$, the contraposition of Lemma 1 (ii) implies that $|GS_s(P,q)|= q_s$. In addition, by Lemma 1, each school is matched with the same number of students under $GS(P,q)$ and $\varphi(P,q)$. Let $I(s) = \varphi_s(P,q) \setminus GS_s(P,q)$. Then, there is at least one student $i \in I(s)$. We show that i is a manipulating student of φ at (P,q).

Because $GS(P,q)$ is the student-optimal stable matching at (P,q), we have

$$GS_i(P,q)\ P_i\ \varphi_i(P,q) = s. \tag{6}$$

Since φ is individually rational, $GS_i(P,q) = s'$ for some acceptable school s' under P_i. Let $P_i^{s'}$ be a preference relation where i has ranked only s' as an acceptable school. As shown by Roth (1982), $GS_i(P_i^{s'}, P_{-i}, q) = s'$. Since student i is matched at a stable matching, Lemma 1 implies that she is also matched at any stable matching. Since φ is individually rational and s' is the only acceptable school under $P_i^{s'}$, we have $\varphi_i(P_i^{s'}, P_{-i}, q) = s'$. This, together with $s'\ P_i\ s$ and equation 6 imply that

$$s' = \varphi_i(P_i^{s'}, P_{-i}, q)\ P_i\ \varphi_i(P,q) = s.$$

This means that student i is a manipulating agent of φ at (P,q).

It is left to show that no two subsets intersect. Let $s, s' \in M^2$ be two different schools. Let $i \in I(s)$ and $j \in I(s')$. By definition of $I(s)$ and $I(s')$, $i \in \varphi_s(P,q)$ and $j \in \varphi_{s'}(P,q)$. Since $\varphi(P,q)$ is a matching and $s \neq s'$, we have $j \neq i$.

Step 3: The mechanism φ has weakly more manipulating agents than GS at (P,q).

By the previous step, for each $s, s' \in M^2$, we have $|I(s)|\geq 1$ and $I(s) \cap I(s') = \emptyset$. Thus, there are at least $|M^1|+\sum_{s''\in M^2}|I(s'')|\geq |M|$ manipulating agents of φ at (P,q). That is, φ has weakly more manipulating agents than GS at (P,q). ☐

Proposition 2: *Let $k \geq 1$ and suppose that there are at least as many students as schools, $|I| \geq |S|$, and $k < |S|$. Let φ be a weakly non-wasteful and individually rational mechanism. Then, the constrained mechanism φ^k is manipulable.*

Proof. Let φ be a weakly non-wasteful and individually rational mechanism. Let (P,q) be a problem such that each school $s \in S$ has one seat, $q_s = 1$, and let students have a common preference relation such that each school is acceptable: for each $i, j \in I$ and each $s \in S$, $P_i = P_j$ and $s\ P_i\ \emptyset$.

First, note that $\varphi^k(P,q)$ does not assign the schools not listed in P^k since this would violate individual rationality of $\varphi(P_I^k, P_S, q)$ under (P^k, P_S). Since the constraint is binding, $k < |S| \leq |I|$, and only k schools have been ranked acceptable under P^k, then at least one student i is unmatched and at least one school s has an empty seat.

Second, consider a misreport P_i^s where i only lists this school s as acceptable. Again, since s is not acceptable to any student other than i at (P_i^s, P_{-i}^k, P_S), school s cannot be assigned to a student other than i under $\varphi(P_i^s, P_{-i}^k, P_S, q)$, otherwise this would violate individual rationality of $\varphi(P_i^s, P_{-i}^k, P_S, q)$ at (P_i^s, P_{-i}^k, P_S). Similarly, at $\varphi^k(P_i^s, P_{-i}, q)$ student i cannot be matched to any school other than s as this would violate individual rationality of $\varphi(P_i^s, P_{-i}^k, P_S, q)$ under (P_i^s, P_{-i}^k, P_S). Because φ is weakly non-wasteful, $\varphi_i^k(P_i^s, P_{-i}, q) = s$, as otherwise s is unmatched and i is unmatched. Since s is acceptable to i under P, we have $\varphi_i^k(P_i^s, P_{-i}, q)\ P_i\ \varphi_i^k(P, q)$. Thus, φ^k is manipulable. □

To prove the following results, we first formulate and prove a lemma (Lemma 3).

We first define intermediary mechanisms. Note that under GS^ℓ the ranking constraint is the same for all students, as well as under GS^k. We define intermediate mechanisms where the constraint is ℓ for some students and k for the remaining students. Let $N \subseteq I$ be a subset of students. We define the mechanism GS^N that assigns to each problem (P, q) the matching

$$GS^N(P, q) = GS(P_N^\ell, P_{-N}^k, P_S, q).$$

This is the mechanism where the constraint is ℓ for students in N and the constraint is k for students in $I \setminus N$. Then $GS^k = GS^\emptyset$ and $GS^\ell = GS^I$.

We now establish that manipulating students are unmatched and any manipulating strategy can be replicated via top-ranking schools.

Lemma 3. *Let (P, q) be a problem, $i \in I$ and $s \in S$.*

(i) Suppose that student i is a manipulating student of GS^N at (P, q). Then, she is unmatched under $GS^N(P, q)$.

(ii) Suppose that $GS_i^N(P, q) = s$ and let P_i^s be a preference relation where i has ranked only school s acceptable. Then $GS_i^N(P_i^s, P_{-i}, q) = s$.

Proof. We prove (i) by contradiction. Suppose that there is a student i and a school s such that $GS_i^N(P, q) = s$, and there is a preference relation P_i' such that

$$GS_i^N(P_i', P_{-i}, q)\ P_i\ GS_i^N(P, q).$$

Because GS^N is individually rational, there is a school s' such that $GS_i^N(P_i', P_{-i}, q) = s'$. Let $\hat{P} = (P_N^\ell, P_{-N}^k, P_S)$. Then, by definition, $GS^N(P, q) = GS(\hat{P}, q)$. Suppose that $i \in N$. Then, schools s and s' are among the top ℓ acceptable schools under P_i.

Thus $s' \; P_i^\ell \; s$ and

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

This means that student i can manipulate GS at $\hat{P}$, contradicting the fact that GS is not manipulable.

Suppose that $i \notin N$. The proof is the same. Schools s and s' are among the top k schools at P_i, thus $s' P_i s$. We have

$$s' = GS_i(P_i'^k, \hat{P}_{-i}, q) \; P_i^k \; GS_i(P_i^k, \hat{P}_{-i}, q) = s,$$

and GS is manipulable at $\hat{P}$, which is a contradiction.

To prove (ii), let $\hat{P} = (P_N^\ell, P_{-N}^k, P_S)$. Then, $GS_i(\hat{P}, q) = s$. As shown by Roth (1982), $GS_i(\hat{P}, q) = s$ implies that $GS_i(P_i^s, \hat{P}_{-i}, q) = s$. Since $k > \ell \geq 1$, the truncation of P_i^s at k or ℓ is nothing but P_i^s. Thus, $GS_i^N(P_i^s, P_{-i}, q) = s$. □

Theorem 2: *Let $k \geq 2$ and φ be a stable matching mechanism. Suppose that students can only rank up to k schools.*
(i) Suppose that schools can manipulate via preferences. Then the constrained student-proposing GS mechanism GS^k has fewer or an equal number of manipulating agents compared to the constrained stable matching mechanism φ^k.
(ii) Suppose that schools can only manipulate via capacities. Then the constrained student-proposing GS mechanism GS^k has fewer or an equal number of manipulating agents compared to the constrained stable matching mechanism φ^k.

Proof. The proof is divided into three steps and is similar to the proof of Theorem 1 and Proposition 1. Let M^1 denote the set of manipulating students of GS^k and M^2 the set of manipulating schools of GS^k at (P, q).

Step 1: Every student in M^1 is a manipulating student of φ^k at (P, q). Let $i \in M^1$. By Lemma 3, student i is unmatched under $GS^k(P, q)$ and there is an acceptable school s under P_i such that $GS_i^k(P_i^s, P_{-i}, q) = s$ where school s is the only acceptable school under P_i^s. Recall that $GS(P_I^k, P_S, q)$ is stable at (P_I^k, P_S, q). By Lemma 1, student i is also unmatched under $\varphi_i^k(P, q) = \varphi(P_I^k, P_S, q)$. That is, $\varphi_i^k(P, q) = \emptyset$. Since student i is matched under $GS_i^k(P_i^s, P_{-i}, q) = s$, then by Lemma 1 she is also matched under $\varphi_i^k(P_i^s, P_{-i}, q)$. Since φ^k is individually rational and s is the only acceptable school under P_i^s, we have $\varphi_i^k(P_i^s, P_{-i}, q) = s$. Since school s is acceptable under P_i, we have

$$s = \varphi_i^k(P_i^s, P_{-i}, q) \; P_i \; \varphi_i^k(P, q) = \emptyset.$$

Therefore, i is a manipulating student of φ at (P, q).

To formulate the second step of the proof we need more notation. Divide the set of manipulating schools M^2 into $\bar{M}^2$ – the subset of schools that are also manipulating schools of φ^k at (P, q) and $\hat{M}^2$ – the subset of schools that are not manipulating schools of φ^k at (P, q). Then $M^2 = \bar{M}^2 \cup \hat{M}^2$ and $\bar{M}^2 \cap \hat{M}^2 = \emptyset$.

Step 2: For every school $s \in \hat{M}^2$, there is a subset $I(s)$ of manipulating students of φ^k at (P, q) such that no student in $I(s)$ is in M^1.

Consider the problem (P_I^k, P_S, q). By Lemma 2, for each school $s \in \hat{M}^2$, we have $\varphi_s(P_I^k, P_S, q) \neq GS_s(P_I^k, P_S, q)$. By Lemma 1, $|GS_s(P_I^k, P_S, q)| = q_s$. Let $I(s) = \varphi_s(P_I^k, P_S, q) \setminus GS_s(P_I^k, P_S, q)$. Then $I(s) \neq \emptyset$. Let $i \in I(s)$. We claim that i is a manipulating student of φ^k at (P, q). Because student i is matched under $\varphi(P_I^k, P_S, q)$, then Lemma 1 implies that she is also matched at any stable matching. Thus $GS_i(P_I^k, P_S, q) = s'$, for some school s'. Because $GS(P_I^k, P_S, q)$ is the student optimal stable matching under (P^k, P_S, q), we have

$$s' = GS_i(P_I^k, P_S, q) \; P_i^k \; \varphi_i(P_I^k, P_S, q) = s. \tag{7}$$

Therefore $s' \; P_i \; s$. Let $P_i^{s'}$ be a preference relation where school s' is the only acceptable school for student i. As shown by Roth (1982), $GS_i(P_i^{s'}, P_{I\setminus\{i\}}^k, P_S, q) = s'$. Since student i is matched at a stable matching, Lemma 1 implies that she is also matched at any stable matching, and in particular under $\varphi(P_i^{s'}, P_{I\setminus\{i\}}^k, P_S, q)$. Since φ is individually rational and s' is the only acceptable school under $P_i^{s'}$, then $\varphi_i(P_i^{s'}, P_{I\setminus\{i\}}^k, P_S, q) = s'$. Note now that because $k \geq 1$, $\varphi_i(P_i^{s'}, P_{I\setminus\{i\}}^k, P_S, q) = \varphi_i^k(P_i^{s'}, P_{-i}, q)$. This equation and equation 7 imply that

$$s' = \varphi_i^k(P_i^{s'}, P_{-i}, q) \; P_i \; \varphi_i^k(P, q) = s. \tag{8}$$

This means that student i is a manipulating student of φ^k at (P, q).

Finally, we show that no student in $I(s)$ is in M^1, that is, no student in $I(s)$ is a manipulating student of GS^k at (P, q). Let $i \in I(s)$. Because student i is matched under $\varphi(P_I^k, P_S, q)$, at a stable matching, Lemma 1 implies that she is also matched under $GS(P_I^k, P_S, q)$. By Lemma 3, student i is not a manipulating student of GS^k at (P, q) and thus $i \notin M^1$.

Step 3: φ^k has weakly more manipulating agents than GS^k at (P, q).

First, for each $s, s' \in \hat{M}^2$ such that $s \neq s'$, we show $I(s) \cap I(s') = \emptyset$. Let $i \in I(s) = \varphi_s(P_I^k, P_S, q) \setminus GS_s(P_I^k, P_S, q)$ and $j \in I(s') = \varphi_{s'}(P_I^k, P_S, q) \setminus GS_{s'}(P_I^k, P_S, q)$. Since $\varphi(P_I^k, P_S, q)$ is a matching and $s \neq s'$, then we have $i \neq j$. That is, $I(s) \cap I(s') = \emptyset$. Second, because for each school $s \in \hat{M}^2$, $|I(s)| \geq 1$, we have

$$|M^1| + |\bar{M}^2| + \sum_{s \in \hat{M}^2} |I(s)| \geq |M^1| + |\bar{M}^2| + |\hat{M}^2| \geq |M^1| + |M^2|.$$

That is, φ^k has weakly more manipulating agents than GS^k at (P, q). □

Theorem 3: *Let $k > 1$. For any problem, the constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained Boston mechanism β^k.*

Proof. We divide the proof into two parts. In the first part, we show that every manipulating student of the constrained GS mechanism who is unmatched under the constrained Boston mechanism is also a manipulating student of the constrained Boston mechanism. In the second part, we show that every manipulating student of the constrained GS mechanism who is matched under the constrained Boston mechanism induces at least one new manipulating student under the constrained Boston mechanism.

Part 1: For every problem (P, q), every manipulating student of GS^k at (P, q) who is unmatched under $\beta^k(P, q)$ is a manipulating student of β^k at (P, q).

Let $i \in I$ be a manipulating student of GS^k at (P, q) and suppose that $\beta_i^k(P, q) = \emptyset$. By Lemma 3, there is a school s such that,

$$GS_i^k(P_i^s, P_{-i}, q) = s \; P_i \; GS_i^k(P, q) = \emptyset,$$

where s is the only acceptable school under P_i^s.

First, student i did not rank school s first under P_i. Otherwise, because she is matched to school s under $GS^k(P_i^s, P_{-i}, q)$, then this matching would be stable at (P_I^k, P_S, q). By Lemma 1, the same set of students are matched at all stable matchings. Therefore, student i is also matched under $GS_i^k(P, q)$. This result contradicts the assumption that $GS_i^k(P, q) = \emptyset$.

Second, we claim that there are less than q_s students who have ranked s first under P and have higher priority than i under $\succ_s$. Otherwise, $GS_i^k(P_i^s, P_{-i}, q) = s$ would imply that at least one of these students is not matched to school s under $GS^k(P_i^s, P_{-i}, q)$. This contradicts the stability of $GS^k(P_i^s, P_{-i}, q)$ under (P_i^s, P_{-i}^k, q)

because student i is matched to school s while a student with a higher priority under $\succ_s$ prefers this school to her assignment.

Therefore, by ranking s first, i is matched to it under $\beta_i^k(P_i^s, P_{-i}, q) = s$, and

$$\beta_i^k(P_i^s, P_{-i}, q) = s \ P_i \ \beta_i^k(P, q) = \emptyset.$$

That is, student i is a manipulating student of β^k at (P, q).

Part 2: Manipulating students of GS^k at (P, q) who are matched under $\beta^k(P, q)$ can be associated in a one-to-one relation with a subset of manipulating students of β^k at (P, q) who are not manipulating students of GS^k at (P, q).

Let M denote the set of the manipulating students of GS^k at (P, q) who are matched under $\beta^k(P, q)$. For the rest of the proof, the strategy is to pair each student in M with a manipulating student of β^k at (P, q) who is not a manipulating student of GS^k at (P, q). Let

$$\mu = GS^k(P, q) \text{ and } \nu = \beta^k(P, q).$$

We label the seats of each school s into q_s different copies $s^1, ..., s^{q_s}$. Let

$$\hat{S} = \{s_1^1, ..., s_1^{q_1}, s_2^1, ..., s_2^{q_2}, ..., s_m^1, ..., s_m^{q_m}\}$$

denote the collection of these copies with a generic element x. We call them seats. We assume that each student who is matched to the same school under both μ and ν is matched to the same copy of this school. That is, for each student i and each school s such that $\mu(i) = \nu(i) = s$, then $\mu(i) = \nu(i) = s^\ell$.

To do our pairing, define a directed graph with vertices I as follows. For each students $i, j \in I$, define an edge from i to j if there is a seat $x \in \hat{S}$ such that $\nu(i) = x$ and $\mu(j) = x$. We label the edge from i to j as x. The edge $i \xrightarrow{x} j$ means that, under ν, student i has taken the seat x that was allotted to student j under μ. A **chain** in this graph is a sequence of $\kappa > 1$ different vertices $(i_1, ..., i_\kappa)$ and $\kappa - 1$ different edges $(x_1, ..., x_{\kappa-1})$ such that

(1) for each $\ell = 1, ..., \kappa - 1$, $i_\ell \xrightarrow{x_\ell} i_{\ell+1}$, and
(2) there is no outgoing edge from i_κ, that is, there is no vertex i and a seat x such that $i_\kappa \xrightarrow{x} i$.

We call the vertex i_1 the **tail** of the chain and i_κ the **head** of the chain. We establish five steps to proving the theorem.

Step 1: No loop. Suppose that there is a sequence of $\kappa > 1$ different vertices $(i_1, \ldots, i_\kappa)$ and $\kappa - 1$ different edges $(x_1, \ldots, x_{\kappa-1})$ such that $i_1 \in M$ and for each

$\ell \in \{1, \ldots, \kappa - 1\}$, $i_\ell \xrightarrow{x_\ell} i_{\ell+1}$. Then, there is no outgoing edge $i_\kappa \xrightarrow{x} j$ such that $j \in \{i_1, \ldots, i_{\kappa-1}\}$.

Suppose that there is an outgoing edge $i_\kappa \xrightarrow{x} j$ from i_κ. First, $j \neq i_1$ because $\mu(i_1) = \emptyset$ and, under ν, i_κ could not have taken a seat that was allotted to student i_1 under μ. Suppose, to the contrary, that $j = i_\ell$ for some $\ell \in \{2, \ldots, \kappa - 1\}$. Thus, $i_\kappa \xrightarrow{x} i_\ell$ and $i_{\ell-1} \xrightarrow{x_{\ell-1}} i_\ell$. By assumption, $i_{\ell-1}$ and i_κ are different vertices. Since ν is a matching, student i_κ and $i_{\ell-1}$ are allotted (if at all) different seats under ν. Then, under ν, student $i_{\ell-1}$ and student i_κ have taken seats which were allotted to student i_ℓ under μ. This conclusion contradicts the fact that μ is a matching and that i_ℓ was allotted only one seat under μ.

Step 2: Every vertex in M is the tail of a chain.

Let $i \in M$. First, there is an outgoing edge from i. To see this, recall that, by assumption, student i is matched under ν. That is, $\nu(i) = x$ for some seat $x \in \hat{S}$ while $\mu(i) = \emptyset$. Suppose that x is a seat at school s. Since the GS mechanism is individually rational, s is one of the top k acceptable schools under P_i. Thus, we have $s\ P_i^k\ \mu(i) = \emptyset$. Since $\mu = GS(P_I^k, P_S, q)$ is stable at (P_I^k, P_S, q), we have $|\mu^{-1}(s)| = q_s$. Therefore, there is a student j such that $\mu(j) = x$ and thus $i_1 \xrightarrow{x} j$. Next, there is $\kappa \geq 1$ and a sequence $(i_1, \ldots, i_{\kappa+1})$ of different vertices and different edges $(x_1, \ldots, x_\kappa)$ such that $i_1 = i$ and for each $\ell \in \{1, \ldots, \kappa\}$, $i_\ell \xrightarrow{x_\ell} i_{\ell+1}$. The sequence (i, j) and x is one of these sequences. Since there is a finite number of students, there is a finite number of these sequences. By step 1, the one with the greatest number of vertices is a chain.

Step 3: The head of each chain with a tail in M is a manipulating student of β^k at (P, q).

Let j be the head of a chain with a tail in M. There is an edge $i \xrightarrow{x} j$. Then, $\mu(j) = x$. Since there is no outgoing edge from j, either $\nu(j) = \emptyset$ or $\nu(j) = x'$ such that there is no student j' with $\mu(j') = x'$. We claim that $\mu(j)\ P_j\ \nu(j)$. Otherwise, $\nu(j)\ P_j\ \mu(j) = x$ and thus $\nu(j)\ P_j^k\ \mu(j) = x$. Because μ is individually rational under P^k, we have $\nu(j) = x'$. Suppose that x' is a copy of school s. Then $s\ P_j^k\ \mu(j)$. Since μ is stable at (P_I^k, P_S, q), we have $|\mu^{-1}(s)| = q_s$. Therefore, there is a student j' such that $\mu(j') = x'$ and $j \xrightarrow{x'} j'$. This contradicts the fact that there is no outgoing edge from j. Therefore $s = \mu(j) P_j \nu(j)$.

Next, we claim that there are less than q_s students who have ranked school s first and have higher priority than student i under P. Otherwise, the fact that $\mu(j) = s$ would imply that one of such students is not matched to school s under μ. This conclusion contradicts the fact that μ is stable at (P_I^k, P_S, q) because student i is matched to school s and a student with higher than i under $\succ_s$ prefers s to her assignment.

Finally, we claim that student j did not rank school s first under P_j. Otherwise, she would be matched to school s under $\nu = \beta^k(P, q)$ because there are less than q_s students who have ranked it first under P and have higher priority than j under $\succ_s$. Let P_j^s be a preference relation where student j has ranked only school s as acceptable. Then, $\beta_j(P_j^s, P_{-j}^k, q) = s$. Since $s = \mu(j)\ P_j\ \nu(j)$, we have

$$s = \beta_j^k(P_j^s, P_{-j}, q)\ P_j\ \beta_j^k(P, q) = \nu(j).$$

This means that j is a manipulating student of β^k at (P, q).

Step 4: The head of each chain with a tail in M is *not* a manipulating student of GS^k at (P, q).

Let i be the head of a chain with a tail in M. Then there is an edge $j \xrightarrow{x} i$. Thus $\mu(i) = x$. That is, student i is matched under $GS^k(P, q)$. By Lemma 3, student i is not a manipulating student of GS^k at (P, q).

Step 5: No two chains with different tails in M have the same head.

This follows from the fact that no two chains with tails in M have a vertex in common. Otherwise, since such chains have different tails, there are different edges $j \xrightarrow{x} i$ and $j' \xrightarrow{x'} i$ where i is one of the common vertices. Since ν is a matching, student j and j' are allotted different seats under ν. This means that both student j and j' have taken seats x and x' which were allotted to student i under μ. This conclusion contradicts the fact that μ is a matching and student i was allotted one seat under μ.

We are ready to complete the proof of the theorem (see Figure 1 for an illustration). Let (P, q) be a problem. Let $M^\emptyset$ denote the set of manipulating students of GS^k at (P, q) who are unmatched under $\beta^k(P, q)$. By part 1, every student in $M^\emptyset$ is a manipulating student of β^k at (P, q). The set $M \cup M^\emptyset$ is the set of all manipulating students of GS^k at (P, q). Let $h(M)$ denote the collection of students such that each of them is the head of a chain with a tail in M. By step 3, each student in $h(M)$ is a manipulating student of β^k at (P, q). By step 4, $M^\emptyset \cap h(M) = \emptyset$. By step 5, there are as many students in M as there are in $h(M)$. Therefore, each student in $M^\emptyset \cup h(M)$ is a manipulating student of β^k at (P, q) and $|M^\emptyset \cup M| = |M^\emptyset \cup h(M)|$. There are weakly more manipulating students of β^k than GS^k at (P, q). □

Next, we formulate and prove Lemma 4, which is the main part for proving Theorem 4 below. Recall the notation used to formulate Lemma 3 above.

Lemma 4. *Let $N \subsetneq I$ and $i \notin N$. For each problem (P, q), the mechanism $GS^{N\cup\{i\}}$ has weakly more manipulating students than GS^N at (P, q).*

Proof. Let $\hat{P} = (P^{\ell}_N, P^k_{-N}, P_S)$. Then, $GS^N(P, q) = GS(\hat{P}, q)$ and $GS^{N\cup\{i\}}(P, q) = GS(P^{\ell}_i, \hat{P}_{-i}, q)$. We compare the number of manipulating students of GS^N at (P, q) to the number of manipulating students of $GS^{N\cup\{i\}}$ at (P, q). We consider two cases depending on the matching status of student i.

Case 1: Student i is unmatched under $GS^N(P, q)$ or matched under $GS^{N\cup\{i\}}(P, q)$.

For this case, we will show that every manipulating student of GS^N at (P, q) is also a manipulating student of $GS^{N\cup\{i\}}$ at (P, q).

First, suppose that student i is unmatched under $\mu = GS^N(P, q)$. Note that because $i \notin N$, $\hat{P} = (P^k_i, \hat{P}_{-i})$ and $GS(P^k_i, \hat{P}_{-i}, q)$ is stable at $(P^k_i, \hat{P}_{-i}, q)$. Since student i is unmatched under $GS(P^k_i, \hat{P}_{-i}, q)$ and $\ell < k$, $GS(P^k_i, \hat{P}_{-i}, q)$ is also stable at $(P^{\ell}_i, \hat{P}_{-i}, q)$. By Lemma 1, the same set of students are matched in every stable matching. Therefore, the same set of students are matched under $GS(P^k_i, \hat{P}_{-i}, q)$ and at $GS(P^{\ell}_i, \hat{P}_{-i}, q)$.

Second, suppose that student i is matched under $GS(P^{\ell}_i, \hat{P}_{-i}, q)$. Since $k > \ell$, $GS(P^{\ell}_i, \hat{P}_{-i}, q)$ is also stable at $(P^k_i, \hat{P}_{-i}, q)$. By Lemma 1, the same set of students are matched under $GS(P^{\ell}_i, \hat{P}_{-i}, q)$ and $GS(P^k_i, \hat{P}_{-i}, q)$. In either case, the same set of students are matched under $GS^N(P, q) = GS(P^k_i, \hat{P}_{-i}, q)$ and $GS^{N\cup\{i\}}(P, q) = GS(P^{\ell}_i, \hat{P}_{-i}, q)$.

Let $j \in I$ be a manipulating student of GS^N at (P, q). By Lemma 3, j is unmatched under $GS^N(P, q)$ and there is a school s such that

$$s = GS^N_j(P^s_j, P_{-j}, q)\ P_j\ GS^N_j(P, q) = \emptyset,$$

where P^s_j is a preference relation where j has ranked only s as an acceptable school. Because the same set of students are matched under $GS^N(P, q)$ and $GS^{N\cup\{i\}}(P, q)$, student j is also unmatched under $GS^{N\cup\{i\}}(P, q)$. That is,

$$GS^{N\cup\{i\}}_j(P, q) = \emptyset. \tag{9}$$

First, suppose that $j = i$. Since $\ell \geq 1$, the truncation of P^s_i after the ℓ'th acceptable school is nothing but P^s_i. Therefore, $GS^{N\cup\{i\}}(P^s_i, P_{-i}, q) = GS^N(P^s_i, P_{-i}, q)$ and we have

$$s = GS^{N\cup\{i\}}_i(P^s_i, P_{-i}, q)\ P_i\ GS^{N\cup\{i\}}_i(P, q) = \emptyset.$$

This means that student i is also a manipulating student of $GS^{N\cup\{i\}}$ at (P,q).

Second, suppose that $j \neq i$. Note that since $k > \ell$, student i has extended her list of acceptable schools under P_i^k compared to P_i^ℓ. Gale and Sotomayor (1985) showed that, after such an extension, no student other than i is better off in GS. In particular,

$$GS_j(P_j^s, P_i^\ell, \hat{P}_{-\{i,j\}}, q) \; R_j^s \; GS_j(P_j^s, P_i^k, \hat{P}_{-\{i,j\}}, q) = s,$$

where the equality in the last part follows from the fact that $GS_j^N(P_j^s, P_{-j}, q) = GS_j(P_j^s, P_i^k, \hat{P}_{-\{i,j\}}, q) = s$. Since GS is individually rational, we have

$$GS_j(P_j^s, P_i^\ell, \hat{P}_{-\{i,j\}}, q) = s = GS_j^{N\cup\{i\}}(P_j^s, P_{-j}, q).$$

This equation and equation 9 yield the following relation:

$$s = GS_j^{N\cup\{i\}}(P_j^s, P_{-j}, q) \; P_j \; GS_j^{N\cup\{i\}}(P,q) = \emptyset.$$

This means that student j is a manipulating student of $GS^{N\cup\{i\}}$ at (P,q).

As a conclusion of Case 1, for each problem (P,q), each manipulating student of GS^N at (P,q) is also a manipulating student of $GS^{N\cup\{i\}}$ at (P,q). Therefore, $GS^{N\cup\{i\}}$ has weakly more manipulating students than GS^N at (P,q).

Case 2: Student i is matched under $GS^N(P,q)$ and unmatched under $GS^{N\cup\{i\}}(P,q)$.

Let $\mu = GS(\hat{P}, q)$ and $\nu = GS(P_i^\ell, \hat{P}_{-i}, q)$. Let us summarize our proof strategy in the following diagram. We divide the set of students into matched and unmatched at μ. The manipulating students of GS^N at (P,q) are unmatched under $GS^N(P,q)$. We would like to construct the set of manipulating students of $GS^{N\cup\{i\}}$ at (P,q) from the set of manipulating students of GS^N at (P,q).

First, we will show that student i joined the set of manipulating students of $GS^{N\cup\{i\}}$ at (P,q). Second, we will show that all manipulating students of GS^N at (P,q), but at most one, remain manipulating students of $GS^{N\cup\{i\}}$ at (P,q).

Step 1: Student i is a manipulating student of $GS^{N\cup\{i\}}$ at (P,q) but not a manipulating student of GS^N at (P,q).

Because student i is matched under $\mu = GS^N(P,q)$, by Lemma 3, she is not a manipulating student of GS^N at (P,q). Let $s = \mu(i)$ and let P_i^s be a preference relation where she has ranked only school s as an acceptable school. As shown by Roth (1985),

$$GS_i(\hat{P}, q) = s \;\Rightarrow\; GS_i(P_i^s, \hat{P}_{-i}, q) = s.$$

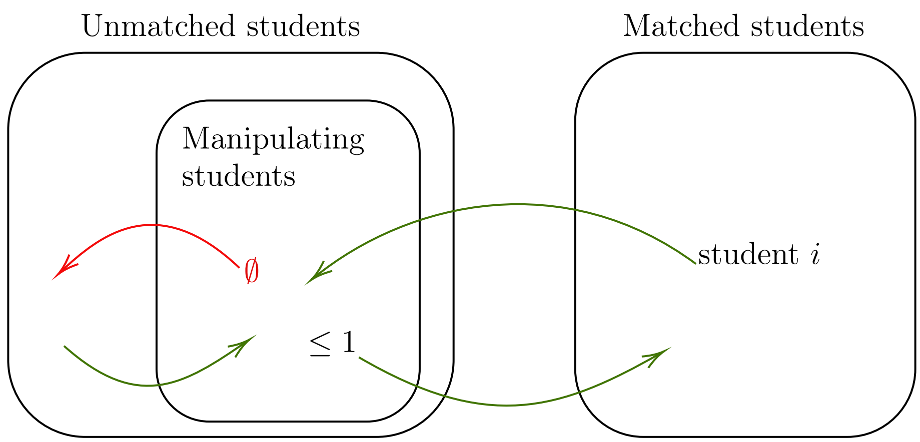

Figure 3. Matched, unmatched and manipulating students at (P, q) between GS^N and $GS^{N\cup\{i\}}$.

Notes: The diagram shows the flow of students across matched, unmatched and manipulating students when moving from mechanism GS^N to $GS^{N\cup\{i\}}$ at (P, q). The green arrows show possible flows and the red arrow shows an impossible flow. (i) At most one student can leave the set of manipulating students of GS^N at (P, q); (ii) student i, who is not a manipulating student of GS^N at (P, q) became a manipulating student of $GS^{N\cup\{i\}}$ at (P, q), and no student can leave the set of manipulating students of GS^N at (P, q) and remain unmatched under μ. While there can be new manipulating students of $GS^{N\cup\{i\}}$ that were unmatched under $GS^N(P, q)$.

Since $\ell \geq 1$, the truncation of P_i^s after the ℓ'th acceptable school is nothing but P_i^s. Therefore,

$$GS_i^N(P_i^s, P_{-i}, q) = s \;\Rightarrow\; GS_i^{N\cup\{i\}}(P_i^s, P_{-i}, q) = s.$$

Since $GS_i^{N\cup\{i\}}(P, q) = \emptyset$ and school s is an acceptable school under P_i, we have

$$GS_i^{N\cup\{i\}}(P_i^s, P_{-i}, q) = s \; P_i \; GS_i^{N\cup\{i\}}(P, q) = \emptyset.$$

This means that student i is a manipulating student of $GS^{N\cup\{i\}}$ at (P, q).

Step 2: Every manipulating student of GS^N at (P, q) who is unmatched under ν is a manipulating student of $GS^{N\cup\{i\}}$ at (P, q).

Let j be a manipulating student of GS^N at (P, q) and suppose that she is unmatched under $\nu = GS^{N\cup\{i\}}(P, q)$. Since she is a manipulating student of GS^N at (P, q), by Lemma 3, we have $GS_j^N(P, q) = \emptyset$ and there is a school s such that $s \; P_i \; GS_j^N(P, q)$ and $GS_j^N(P_j^s, P_{-j}, q) = s$. Student i has extended her list of acceptable schools under P_i^k compared to P_i^ℓ. As shown by Gale and Sotomayor (1985), no other student is better off under GS after such an extension. In particular, we have

$$GS_j(P_j^s, P_i^\ell, \hat{P}_{-\{i,j\}}, q) \; R_j^s \; GS_j(P_j^s, \hat{P}_{-j}, q) = s.$$

Since GS is individually rational, $GS_j(P_j^s, P_i^\ell, \hat{P}_{-\{i,j\}}, q) = s$. Let x be a natural number such that $x = \ell$ if $j \in N$ and $x = k$ if $j \in I \setminus N$. Since $x \geq 1$, the truncation of P_j^s after the x'th choice is nothing but P_j^s. Therefore,

$$GS_j^{N\cup\{i\}}(P_j^s, P_{-j}, q) = s.$$

Since by assumption $GS_j^{N\cup\{i\}}(P, q) = \emptyset$, we have

$$s = GS_j^{N\cup\{i\}}(P_j^s, P_{-j}, q) \; P_j \; GS_j^{N\cup\{i\}}(P, q) = \emptyset.$$

This means that student j is a manipulating student of $GS^{N\cup\{i\}}$ at (P, q).

Step 3: Every student but i who is matched under $GS^N(P, q)$ is also matched under $GS^{N\cup\{i\}}(P, q)$.

Note that student i has extended her list of acceptable schools under P_i^k compared to P_i^ℓ. As shown by Gale and Sotomayor (1985), no other student is better off in GS after such an extension. Thus,

(10) for each student $j \neq i$, $\nu(j) = GS_j(P_i^\ell, \hat{P}_{-i}, q) \; \hat{R}_j \; GS_j(P_i^k, \hat{P}_{-i}, q) = \mu(j)$.

Let $j \neq i$ be a student other than i and suppose that $\mu(j) = s$ for some school s. Since μ is individually rational under $\hat{P}$, then $\nu(j) \neq \emptyset$.

Step 4: There is at most one student who is a manipulating student of GS^N at (P, q) but not a manipulating student of $GS^{N\cup\{i\}}$ at (P, q).

By step 2, any manipulating student of GS^N at (P, q) who is not a manipulating of $GS^{N\cup\{i\}}$ at (P, q) is matched under $GS^{N\cup\{i\}}(P, q)$. We prove, more generally, that there is at most one student who is unmatched under $\mu = GS^N(P, q)$ but matched under $\nu = GS^{N\cup\{i\}}(P, q)$. To do that, we compare the number of students who are matched to each school 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. We prove that there is no student in $\nu^{-1}(s) \setminus \mu^{-1}(s)$. Suppose, to the contrary, that there is $j \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Then, because by assumption i is unmatched under ν, we have $j \neq i$. By equation 10,

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

Because school s has an empty seat under μ, by assumption, this contradicts the fact that $\mu = GS(\hat{P}, q)$ is stable at $(\hat{P}, q)$. Thus, there is no student who is 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 μ. Thus,

$$\sum_{s \in S} |\nu^{-1}(s)| \leq \sum_{s \in S} |\mu^{-1}(s)|. \tag{11}$$

Recall that by step 3, all students but student i, who are matched under μ are also matched under ν. Then inequality 11 implies that there is at most one student who is unmatched under μ but matched under ν.

To sum up, among the manipulating students of GS^N at (P, q), at most one of them is not a manipulating student of $GS^{N \cup \{i\}}$ at (P, q). By including student i, who is a manipulating student of $GS^{N \cup \{i\}}$ at (P, q), but not a manipulating student of GS^N at (P, q), there are weakly more manipulating students of $GS^{N \cup \{i\}}$ at (P, q) than GS^N at (P, q). □

Theorem 4: *Let $k > \ell \geq 1$. For any problem, the constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained GS mechanism GS^ℓ.*

Proof. Let (P, q) be a problem. For simplicity, let $I = \{1, \ldots, |I|\}$. Let $m(\varphi)$ denote the number of manipulating students of φ at (P, q). Then,

$$m(GS^{\emptyset}) \leq m(GS^{\{1\}}) \leq m(GS^{\{1,2\}}) \leq \ldots \leq m(GS^I),$$

where each inequality follows from Lemma 4. Note now that $GS^{\emptyset} = GS^k$ and $GS^I = GS^\ell$. Thus, GS^ℓ has weakly more manipulating students than GS^k at (P, q). □