INCENTIVES IN MATCHING MARKETS: COUNTING AND COMPARING MANIPULATING AGENTS

SOMOUAOGA BONKOUNGOU* AND ALEXANDER NESTEROV†

Abstract. Manipulability is a threat to the successful design of centralized matching markets. However, in many applications some manipulation is inevitable and the designer wants to compare manipulable mechanisms. We count the number of agents with an incentive to manipulate and rank mechanisms by their level of manipulability. This ranking sheds a new light on practical design decisions such as the design of the entry-level medical labor market in the US, and school admissions systems in New York, Chicago, Denver and many cities in Ghana and in the UK.

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

JEL Classification: C78, D47, D78, D82

1. Introduction

Numerous matching systems around the world recently underwent drastic changes to deal with strategic issues of their matching rules. 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;

*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: November 4, 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. This paper has been presented to various seminars and we thank the participants for their comments: at the University of Lausanne (August, 2020), Corvinus University of Budapest (March, 2021), North Carolina State University (March, 2021), the St.Petersburg Economics seminar (October, 2020), Ecole Polytecnique (March, 2022), the 2nd URV Workshop on Game Theory (April, 2022), and the 12th Conference on Economic Design (June, 2022). Support from the Basic Research Program and the "Priority 2030" of the HSE University is gratefully acknowledged. The research is partially financed by RFBR grants 19-01-00762, 20-01-00687. Somouaoga Bonkoungou gratefully acknowledges financial support from the European Commission, Marie Skłodowska-Curie individual fellowship, project 890648.

Abdulkadiroğlu and Sönmez, 2003). The benefits of *truthful mechanisms* have been enumerated and praised in the literature (Vickrey, 1961; Abdulkadiroğlu and Sönmez, 2003). In the Boston K-12 admissions system, for example, they level the playing field by removing the harm that well-informed and sophisticated students could do to others (Pathak and Sönmez, 2008).[1]

Manipulability is a practical concern that has motivated numerous reforms. The US entry-level medical labor market and the Boston K-12 admissions systems are examples of matching systems that have reformed their matching rules, in part, to explicitly address manipulability. Numerous other changes have been observed worldwide (Pathak and Sönmez, 2013), arguably, motivated by manipulability. Meanwhile, in many applications some manipulation is inevitable. Real markets are complex and involve constraints and policy goals that create opportunities for manipulation. In two-sided matching, stability is considered important, yet any *stable* matching mechanism is manipulable (Roth, 1982). Similarly, many school and college admissions systems restrict the number of schools that students are allowed to apply to (Haeringer and Klijn, 2009; Pathak and Sönmez, 2013; Fack et al., 2019). Unfortunately, these constrained mechanisms are also manipulable.

We propose a notion to quantify manipulation and investigate the conjecture that matching mechanisms adopted at many design decisions are less vulnerable to manipulation than those they have replaced. In our notion, we focus on agents that can beneficially misrepresent their private information when others are truthful. We refer to them as *manipulating agents*. The number of these agents can be interpreted as a measure of the potential for manipulation since manipulation is more likely with more manipulating agents.

This number is arguably an arbitrary criterion and hard to justify. For example, in one important application it adds up agents with possibly different abilities and incentives to manipulate, such as students and schools. Nonetheless, it is intuitive and turns out to be useful. It has been occasionally used to measure incentives in matching markets: Roth and Peranson (1999), for example, conducted a simulation on NRMP data and showed that a small number of medical students could have beneficially misreported their preferences when other agents are truthful.

Our analysis covers a wide range of settings that differ in the following aspects: the set of strategic agents, the strategies these agents can use, whether stability is

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

required and whether there are ranking constraints (see Table 1). Our results are twofold. First, we consider the college admissions problem where both students and schools are strategic agents (Gale and Shapley, 1962) and schools can misreport their preferences as well as their capacities. We show that, when all manipulations (by students as well as by schools) are considered, the student-proposing Gale-Shapley (GS) mechanism has the smallest number of manipulating agents among all stable matching mechanisms (Theorem 1). Dubins and Freedman (1981) and Roth (1982) show that this mechanism is not manipulable by students. This was 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). Our result still supports its choice when all strategic agents are considered. What is more, it is still the best choice even when schools can only misreport their capacities, but not their preferences. All these conclusions carry over to the general model where, in addition, students face ranking constraints: while the student-proposing GS mechanism is now manipulable by students it is still the least manipulable mechanism.

Second, 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 shifted towards the constrained student-proposing GS mechanisms and relaxing the constraint. We demonstrate that the number of manipulating students (Theorem 2) weakly decreased as a result of these changes.

Related literature. The seminal approach by Pathak and Sönmez (2013) compares mechanisms by the set inclusion of problems with no manipulating agent.[2,3] This approach considers one mechanism less manipulable than another if the former has a larger domain, by the inclusion of problems where there is no manipulating agent, than the latter. This approach ignores the number of manipulating agents and regards two mechanisms as equal in manipulation at a problem where, for example, one mechanism has one manipulating agent and the other numerous manipulating agents.

[2]They also introduced two other approaches: the approach comparing mechanisms by the inclusion of manipulating agents and the approach comparing mechanisms based on the magnitude of gain from a manipulation.

[3]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).

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

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). The indices k, ℓ denote the ranking constraints. The detailed descriptions of the design instances can be found in Pathak and Sönmez (2013); Bonkoungou and Nesterov (2021b).

Nonetheless, it became the state-of-art method for comparing manipulable matching mechanisms (Chen and Kesten, 2017; Dur et al., 2022; Dur, 2019; Dur et al., 2019). In contrast, the counting approach is quantitative and enables subtle distinction of mechanisms at problems where they are manipulable. We will show that counting is useful in distinguishing stable matching mechanisms in two-sided matching while the approach by Pathak and Sönmez (2013) cannot distinguish them (Proposition 1). However, for the school choice problem, both approaches have been able to similarly compare all manipulable mechanisms studied.

The paper that first formalizes the idea of 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 but without 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 measuring the proportion of manipulating agents and their results support the student-proposing GS mechanism.

Three papers subsequent to Pathak and Sönmez (2013) compared the constrained Boston and constrained GS mechanisms. Bonkoungou and Nesterov (2021a) used a criterion called strategic accessibility, and Decerf and Van der Linden (2021) used the notion of dominant preference-inclusion introduced by Arribillaga and Massó (2016). These criteria and counting are logically independent. Independently from us, Imamura and Tomoeda (2022) also used the criteria of counting for comparing these mechanisms, but their comparison between the constrained Boston and the constrained student-proposing GS is proven in the one-to-one setting.

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 matching mechanisms are equivalent to preference comparisons. Since the preferences of agents on the two sides over stable matchings are opposed, stable matching 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 introduce the general framework. In Section 3, we present our results. We present proofs in the Appendix.

2. General framework

We consider the two-sided matching problem (Gale and Shapley, 1962). It consists of the following elements:

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

where P is defined as follows. Being unmatched is denoted by $\emptyset$. For each student i, P_i is a strict preference relation 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.[4] For each school s, P_s is a strict preference relation 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 $N \cup \{i\}\ P_s\ N \cup \{j\} \Leftrightarrow i \succ_s j$.

Given an agent $v \in I \cup S$, let P_{-v} denote the preference profile of agents other than v. 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, $|\mu^{-1}(s)| \leq q_s$. The student i finds matching μ at least as good as matching μ' if and only if $\mu(i)\ R_i\ \mu'(i)$. The 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. We introduce two useful definitions.

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

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

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

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

(c) We say that ***school*** s ***can manipulate its capacity*** *under mechanism* φ *at problem* (P, q) *— and a manipulating school — if there is* $\hat{q}_s < q_s$ *such that*

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

(d) We say that ***mechanism*** φ ***is manipulable*** *at* (P, q) *if there is a manipulating agent of* φ *at* (P, q)*.*

We model ranking constraints where students cannot list all schools. There is a maximum number $k \in \{1, \ldots, |S|\}$ of schools that each student can list. 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.

Definition 2. *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.*

A mechanism φ is **constrained stable** if for each problem (P, q), $\varphi(P, q)$ is stable under (P^k, q).

3. Results

We first present our results for the college admissions problem where agents on both sides are strategic. We next present our results for the school choice problem where students are strategic but not schools.

3.1. College admissions.

3.1.1. *Model and results.* We consider the college admissions problem (Gale and Shapley, 1962) where students as well as schools are strategic. Students misrepresent their preferences but schools may misrepresent their preferences or their capacities. The following notion of stability turns out important for the design of such a market. A 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 to 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 an algorithm called student-proposing Deferred Acceptance for producing the student-optimal stable matching. We denote this mechanism as GS. Similarly, for each problem, there is a school-optimal stable matching which can be obtained by applying the school-proposing deferred acceptance algorithm.

Stable matching mechanisms are subject to various kinds of manipulations by both students and schools. Every stable matching mechanism is manipulable by students and schools. Interestingly, 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. This constitutes the standard argument supporting the student-proposing GS: since schools can manipulate any stable mechanism let us remove all manipulations by students. At the same time, the student-proposing GS mechanism has weakly more manipulating schools than any other stable matching mechanism (Pathak and Sönmez, 2013) and it is unclear how to resolve this trade-off. In the following proposition, we show that the approach by Pathak and Sönmez (2013) about comparing mechanisms by the inclusion of problems with no manipulating agent cannot distinguish stable matching mechanisms.

Proposition 1. *Consider the two-sided matching problem and suppose that schools can manipulate via preferences. Let φ and ϕ be two stable matching mechanisms. For any problem, either φ and ϕ are both manipulable or none is manipulable.*

All stable matching mechanisms are manipulable via preferences on the same domain of problems. One of the approaches that can distinguish stable matching mechanisms in this domain is counting the number of manipulating agents. The approach is still useful in distinguishing them when there are ranking constraints. Note that with ranking constraints the standard argument above cannot guide the choice because these constraints make the student-proposing GS manipulable by students.[5,6] Constrained mechanisms are very common. The reasons behind these constraints are not yet fully understood, but they appear to be crucial for practitioners.

Theorem 1. *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.

3.1.2. *Discussion.* We illustrate in Table 1 practical examples of design decisions that the above theorem explains. The NRMP and the New York City high school match are examples. In the latter example, in particular, students face ranking constraints and schools can only manipulate their capacities (Abdulkadiroğlu et al., 2005).

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 the 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 constrained student-proposing GS mechanism GS^k 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. Thus we essentially need to prove the result for unconstrained mechanisms.

The intuition why the unconstrained 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

[5]Every sensible constrained mechanism is manipulable by students, see Proposition 2 in Bonkoungou and Nesterov (2021b).

[6]Note that if the constraint k is larger than the number of schools, then we are dealing with unconstrained mechanisms and compare all stable matching mechanisms.

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. The rural hospital theorem (Roth, 1986) 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 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 $GS(P, q)$.

Note that only schools that have filled all their seats can manipulate the student-proposing GS mechanism via capacities (Ehlers, 2010). However, the proof is similar.

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.

Corollary 1. *Consider the marriage market where every school has one seat. For any problem, the student-proposing and the school-proposing GS mechanisms have the same number of manipulating agents.*

Finally, when the problem has a unique stable matching the notion is still useful. 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). When there is a unique stable matching, students cannot manipulate their preferences, but schools can still manipulate their preferences as well as capacities (Ehlers, 2010).

3.2. **School Choice.**

3.2.1. *Model and results.* We consider the school choice problem (Abdulkadiroğlu and Sönmez, 2003) where students are strategic with respect to reporting their preferences but not schools. That is, each school reports its priorities and capacities truthfully. In contrast to the college admissions problem above, we assume that each student is acceptable to each school. That is, for each student i and each school s, $i \succ_s \emptyset$.

In their seminal paper, Abdulkadiroğlu and Sönmez (2003) describe a matching procedure that were used in Boston and still used in many places around the world.

The mechanism is called the Boston mechanism. We denote it as *Boston*. This mechanism is also called immediate acceptance mechanism to highlight the difference with the "deferred" acceptance of the GS mechanism. Broadly, students apply to acceptable schools one at a time in decreasing order of preferences. Schools immediate accept highest priority applicants and reduce their capacities accordingly. This mechanism is individually rational but does not always produce a stable matching. Worse, it 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, and some reforms did both. None of these reforms eliminated manipulation, but, as we show next, they reduced the number of manipulating agents.

Theorem 2. *Let $k > \ell \geq 1$. Then for any problem the following statements are true.*

(i) The constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained Boston mechanism $Boston^k$.

(ii) The constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained GS mechanism GS^ℓ.

(iii) The constrained GS mechanism GS^k has fewer or an equal number of manipulating students compared to the constrained Boston mechanism $Boston^\ell$.

3.2.2. *Discussion.* In Table 1, we document reforms that Theorem 2 explains. Statement (iii) is a straightforward corollary of the first two statements. Statements (i) and (ii) require a strong argument each. The main and novel part of the proof of statement (i) is to construct a one-to-one function between manipulating students of GS^k and a subset of manipulating students of $Boston^k$. 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 there may be some students who cannot manipulate the constrained Boston mechanism can manipulate the constrained GS. To see this, consider the following example.

Example 1. *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 $Boston^2$ with GS^2. The outcome of $Boston^2$ is as follows:

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

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 $Boston^2$ 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 $Boston^2$. Note that student i_2 is also a manipulating student of GS^2 at (P,q).

Recall, that $Boston^2$ has two manipulating students i_2 and i_4, and GS^2 has two manipulating students i_2 and i_5. In the proof, we show that if a manipulating student of $Boston^k$ is unmatched under GS^k, which is the case for student i_2, then this student remains a manipulating student of GS^k. Let us focus on i_4 and i_5. We show that, by replacing $Boston^2$ with GS^2, students i_4 and i_5's incentives to manipulate are changed correspondingly. Note that under $Boston^2(P,q)$, 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 $Boston^2(P,q)$. The last student, i_4, is

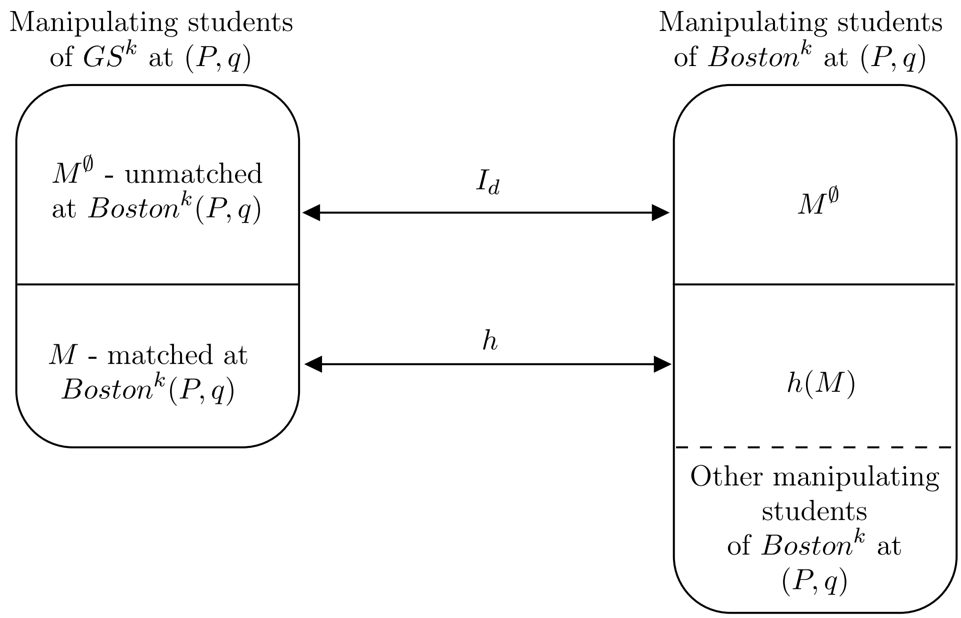

Figure 1. Relation between manipulating students of GS^k and $Boston^k$.

not assigned under $Boston^2(P, q)$ 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 $Boston^2$ at (P, q). Thus, the number of manipulating students of GS^2 is not greater than the number of manipulating students of $Boston^2$.

The steps of the proof involve showing the following points:

- Each manipulating student of GS^2 at (P, q) who is unmatched under $Boston^2(P, q)$ is also a manipulating student of $Boston^2$ at (P, q).
- Starting from each manipulating student of GS^2 at (P, q) who is matched under $Boston^2(P, q)$, the pointing sequence ends at a manipulating student of $Boston^2$ at (P, q).
- Two distinct sequences lead to distinct manipulating students of $Boston^2$ 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 $Boston^k$, M, and those who are unmatched under $Boston^k$, $M^\emptyset$. Our function returns each student in $M^\emptyset$ to herself via an identity relation I_d and each student in M (initiator of a sequence) to the student closing this sequence via a relation h. The set of manipulating students of $Boston^k$ includes $M^\emptyset \cup h(M)$ and possibly others.

For statement (ii), 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: $GS(P^\ell_N, P^k_{I \setminus N}, P_S, q)$. 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 examples illustrate the comparison.

Example 2. *Consider the problem (P, q) in Example 1, 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 1, student i_5 can manipulate GS^2 at (P, q).

The point of this example is to show that by extending the constraint in the student-proposing GS mechanism, some students may become manipulating students and thus there is no inclusion order relation. However, as the following example illustrates, the number of manipulating students does not increase.

Example 3. *Consider the same problem as in Example 1. 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 $GS^{\{i_1\}} = GS(P^1_{i_1}, P^2_{-i_1})$ – 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:

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

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 $GS^{\emptyset} = GS^k$.

References

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005). 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. (2021a). Comparing school choice and college admissions mechanisms by their strategic accessibility. *Theoretical Economics*, 16(3):881–909.

Bonkoungou, S. and Nesterov, A. (2021b). Incentives in matching markets: Counting and comparing manipulating agents. *Higher School of Economics Research Paper No. WP BRP*, 249.

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. (2021). Manipulability in school choice. *Journal of Economic Theory*, 197:105313.

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., Pathak, P. A., Song, F., and Sönmez, T. (2022). Deduction dilemmas: The taiwan assignment mechanism. *American Economic Journal: Microeconomics*, 14(1):164–85.

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

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

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.

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

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.

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

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

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 $\mu^{-1}(s) = \nu^{-1}(s)$.

Proof of Proposition 1

Proof. We show that for any problem where there are more than one stable matching, any stable matching mechanism is manipulable. For any problem where there is one stable matching, stable matching mechanisms have the same set of manipulating agents. Thus stable matching mechanisms are (non-)manipulable on the same set of problems. Let (P, q) be a problem.

Case 1: There are more than one stable matching. Let φ be a stable matching mechanism and $\mu = \varphi(P, q)$. Since there are more than one stable matching, there is a stable matching $\nu \neq \mu$. We consider two sub-cases.

Case 1.1: There is a student i such that $\nu(i)\ P_i\ \mu(i)$. Since μ is individually rational, $\nu(i) \in S$. Let $s = \nu(i)$ and P_i^s a preference relation where school s is acceptable and every other school is unacceptable. Clearly, the matching ν is stable under (P_i^s, P_{-i}, q). This is because if it is blocked under (P_i^s, P_{-i}, q), it is also blocked under (P, q). By the Lemma 1 (i), student i is matched with the same number of partners across all stable matchings. Thus she is matched under $\varphi(P_i^s, P_{-i}, q)$. Since φ is individually rational and no school other than s is acceptable under P_i^s,

$\varphi_i(P_i^s, P_{-i}, q) = s$. Since $s = \nu(i)\ P_i\ \mu(i)$, student i is a manipulating agent of φ at (P, q). Therefore, φ is manipulable at (P, q).

Case 1.2: For any student i, $\mu(i)\ R_i\ \nu(i)$. Since $\mu \neq \nu$, there is a student i such that $\mu(i)\ P_i\ \nu(i)$. Since ν is individually rational, $\mu(i) \in S$. Let $s = \mu(i)$. Since $\nu(i) \neq s$, we have $\mu^{-1}(s) \neq \nu^{-1}(s)$. First, by the Lemma 1 (ii), $|\nu^{-1}(s)| = |\mu^{-1}(s)| = q_s$. Second, we claim that $\nu^{-1}(s)\ P_s\ \mu^{-1}(s)$. Suppose, on the contrary, that $\mu^{-1}(s)\ P_s\ \nu^{-1}(s)$. By Roth and Sotomayor (1990), Theorem 5.27, for each student $k \in \mu^{-1}(s)$ and each $j \in \nu^{-1}(s) \setminus \mu^{-1}(s)$, we have $k \succ_s j$. By the Lemma 1 (i), school s is matched with the same number of students under μ and ν. Since $\mu(i) = s$ and $\nu(i) \neq s$, there is $j \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Thus $i \succ_s j$. Then the matching ν is blocked. This conclusion contracts the stability of ν at (P, q). Let P'_s be a preference relation of school s where $\nu^{-1}(s)$ is the set of acceptable students and no other student is acceptable. We can follow the same argument as above to show that $\varphi_s(P'_s, P_{-s}, q) = \nu^{-1}(s)$. Thus school s is a manipulating agent of φ at (P, q). Therefore, φ is manipulable at (P, q).

Case 2: There is one stable matching μ. Let φ and ϕ be two stable matchings. By Kojima and Pathak (2009) we focus on preference misreports. Suppose that v is a manipulating agent of φ at (P, q). There is a preference relation P'_v such that $\varphi_v(P'_v, P_{-v}, q)\ P_v\ \varphi_v(P, q)$. Similarly, we construct a preference relation P''_v and show that $\phi(P''_v, P_{-v}, q) = \varphi(P'_v, P_{-v}, q)$. Since $\phi(P, q) = \varphi(P, q)$, agent v is also a manipulating agent of ϕ at (P, q).

□

We prove the following lemmas and use them in the proof of Theorem 1 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 $s \in M^2$, $\varphi_s(P, q) \neq GS(P, q)$.*

Proof. The proof is different depending on whether we consider schools' manipulation by preferences or capacities or manipulation by capacities only.

Case 1: Manipulation by preferences or capacities. 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 in 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_s^d 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

$$GS_s(P_s^d, P_{-s}, q) \; R_s \; GS_s(P_s', P_{-s}, q_s', q_{-s}). \tag{2}$$

Lemma 1 (i) implies that school s is matched with the same number of students under both $GS_s(P_s^d, P_{-s}, q)$ and $\varphi(P_s^d, P_{-s}, q)$. 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_s^d. Therefore, since φ and GS are individually rational, they match s to the same set of students,

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

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

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

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: Manipulation by capacities. 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, $\varphi_s(P,q) \neq GS_s(P,q)$.

□

To proceed to the next lemma 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

$$GS^N(P,q) = GS(P^\ell_N, P^k_{-N}, P_S, q).$$

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 $GS^N_i(P,q) = s$ and let P^s_i be a preference relation where i has ranked only school s acceptable. Then $GS^N_i(P^s_i, P_{-i}, q) = s$.

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

$$GS^N_i(P'_i, P_{-i}, q) \; P_i \; GS^N_i(P,q).$$

Because GS^N is individually rational, there is a school s' such that $GS^N_i(P'_i, P_{-i}, q) = s'$. Let $\hat{P} = (P^\ell_N, P^k_{-N}, P_S)$. 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 $s' \; P^\ell_i \; s$ and

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

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

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

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

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

Proof of Theorem 1

Proof. The proof has three steps. 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 $GS^k_i(P^s_i, P_{-i}, q) = s$ where school s is the only acceptable school under P^s_i. Recall that $GS(P^k_I, P_S, q)$ is stable at (P^k_I, P_S, q). By Lemma 1, student i is also unmatched under $\varphi^k_i(P, q) = \varphi(P^k_I, P_S, q)$. That is, $\varphi^k_i(P, q) = \emptyset$. Since student i is matched under $GS^k_i(P^s_i, P_{-i}, q) = s$, then by Lemma 1 she is also matched under $\varphi^k_i(P^s_i, P_{-i}, q)$. Since φ^k is individually rational and s is the only acceptable school under P^s_i, we have $\varphi^k_i(P^s_i, P_{-i}, q) = s$. Since school s is acceptable under P_i, we have

$$s = \varphi^k_i(P^s_i, P_{-i}, q) \; P_i \; \varphi^k_i(P, q) = \emptyset.$$

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

$$s' = GS_i(P^k_I, P_S, q) \; P^k_i \; \varphi_i(P^k_I, P_S, q) = s. \tag{6}$$

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), $GS_i(P_i^{s'}, P^k_{I\setminus\{i\}}, P_S, q) = s'$. 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 $\varphi(P_i^{s'}, P^k_{I\setminus\{i\}}, P_S, q)$. Since φ is individually rational and s' is the only acceptable school under $P_i^{s'}$, then $\varphi_i(P_i^{s'}, P^k_{I\setminus\{i\}}, P_S, q) = s'$. Note now that because $k \geq 1$, $\varphi_i(P_i^{s'}, P^k_{I\setminus\{i\}}, P_S, q) = \varphi^k_i(P_i^{s'}, P_{-i}, q)$. This equation and equation 6 imply that

$$s' = \varphi^k_i(P_i^{s'}, P_{-i}, q) \; P_i \; \varphi^k_i(P, q) = s. \tag{7}$$

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 $\varphi(P^k_I, P_S, q)$, at a stable matching, Lemma 1 implies that she is also matched under $GS(P^k_I, P_S, q)$. 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 $i \in I(s) = \varphi_s(P^k_I, P_S, q) \setminus GS_s(P^k_I, P_S, q)$ and $j \in I(s') = \varphi_{s'}(P^k_I, P_S, q) \setminus GS_{s'}(P^k_I, P_S, q)$. Since $\varphi(P^k_I, P_S, q)$ 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). □

Proof of Theorem 2

Proof of Theorem 2 (i). 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 $Boston^k(P, q)$ is a manipulating student of $Boston^k$ at (P, q).

Let $i \in I$ be a manipulating student of GS^k at (P, q) and suppose that $Boston_i^k(P, q) = \emptyset$. 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 $GS^k(P, q)$. 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 the Boston mechanism. That is, $Boston_i^k(P_i^s, P_{-i}, q) = s$. Therefore,

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

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

Part 2: Manipulating students of GS^k at (P, q) who are matched under $Boston^k(P, q)$ can be associated in a one-to-one relation with a subset of manipulating students of $Boston^k$ 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 $Boston^k(P, q)$. For the rest of the proof, the strategy is to pair each student in M with a manipulating student of $Boston^k$ at (P, q) who is not a manipulating student of GS^k at (P, q). Let

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

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 $Boston^k$ 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 $\nu = Boston^k(P, q)$ 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, $Boston_j(P_j^s, P_{-j}^k, q) = s$. Since $s = \mu(j)\ P_j\ \nu(j)$, we have

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

This means that j is a manipulating student of $Boston^k$ 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 $Boston^k(P, q)$. By part 1, every student in $M^{\emptyset}$ is a manipulating student of $Boston^k$ 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 $Boston^k$ 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 $Boston^k$ at (P, q) and $|M^{\emptyset} \cup M| = |M^{\emptyset} \cup h(M)|$. There are weakly more manipulating students of $Boston^k$ than GS^k at (P, q). □

Next, we formulate and prove Lemma 4, which is the main part for proving Theorem 2 (ii). 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 $\hat{P} = (P_N^{\ell}, P_{-N}^k, P_S)$. Then, $GS^N(P, q) = GS(\hat{P}, q)$ and $GS^{N \cup \{i\}}(P, q) = GS(P_i^{\ell}, \hat{P}_{-i}, q)$. 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 $i \notin N$, $\hat{P} = (P_i^k, \hat{P}_{-i})$ and $GS(P_i^k, \hat{P}_{-i}, q)$ is stable at $(P_i^k, \hat{P}_{-i}, q)$. Since student i is unmatched under $GS(P_i^k, \hat{P}_{-i}, q)$ and $\ell < k$, $GS(P_i^k, \hat{P}_{-i}, q)$ is also stable at $(P_i^{\ell}, \hat{P}_{-i}, q)$. By Lemma 1, the same set of students are matched in every stable matching. Therefore, the same set of students are matched under $GS(P_i^k, \hat{P}_{-i}, q)$ and at $GS(P_i^{\ell}, \hat{P}_{-i}, q)$.

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

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

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

where P_j^s 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,

$$GS_j^{N\cup\{i\}}(P,q) = \emptyset. \tag{8}$$

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

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

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 8 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).

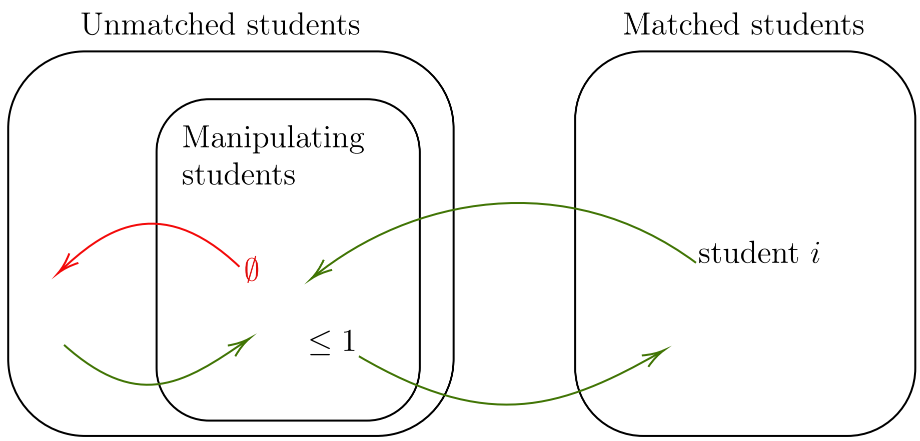

Figure 2. Matched, unmatched and manipulating students at (P, q) between GS^N and $GS^{N\cup\{i\}}$.

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 $GS^N(P, q)$.

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),

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

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,

(9) 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 9,

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

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,

$$\sum_{s\in S}|\nu^{-1}(s)|\leq \sum_{s\in S}|\mu^{-1}(s)|. \tag{10}$$

Recall that by step 3, all students but student i, who are matched under μ are also matched under ν. Then inequality 10 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). □

Proof of Theorem 2 (ii). 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). □