On two kinds of manipulation for school choice problems

Onur Kesten

Abstract Many school districts in the US. employ centralized clearing houses to assign students to public schools. An important potential threat against any school choice mechanism is the tendency of schools to circumvent the procedure via two kinds of strategic manipulation: manipulation via capacities and manipulation via pre-arranged matches. This paper studies the extent of the vulnerability of three prominent school choice mechanisms that have been adopted (or, considered for adoption) by some school districts in the US. We find that the highly debated Boston mechanism as well as the top trading cycles mechanism are immune to manipulation via capacities, unlike the student-optimal stable mechanism (SOSM). We show that SOSM is immune to manipulation via capacities if and only if the priority structure satisfies an acyclicity condition proposed by Ergin (Econometrica 70:2489–2497, 2002). On the other hand, we show that essentially no mechanism is immune to manipulation via pre-arranged matches.

Keywords School choice · Student-optimal stable mechanism · Top trading cycles · Boston mechanism · Acyclicity

JEL Classification C71 · C78 · C79 · D61 · D71 · D78

I would like to thank William Thomson and Fuhito Kojima for useful discussions and suggestions.

O. Kesten (✉)
Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
e-mail: okesten@andrew.cmu.edu
1 Introduction

Several US states use centralized assignment programs that offer students flexibility in their choice of school. In a school choice program, each student submits a strict preference ranking over schools to a central placement authority, such as the school district, which decides which school each student will attend, after also taking into consideration the priority list for each school. Several criteria may be used for determining a priority list for a school. For example, in the current Boston public school system, a student belongs to the highest priority group for a particular school if he resides in the school’s “walk zone” and has a sibling attending that school. The remaining priority groups are similarly determined based on which one of those two criteria a student meets. Finally, a strict priority list of each school is determined using the priority groups and a random lottery draw.

A standard school choice problem (Balinski and Sönmez 1999), or simply, a problem, consists of two pieces of information: a preference profile of students and a priority profile of schools. At a matching, each student is assigned to at most one school, and the number of students assigned to a particular school does not exceed the capacity of that school. A school choice mechanism or simply, a mechanism, is a systematic way of selecting a matching for a given problem.

The Boston public school system is an example of a “one-sided” school choice problem, where schools are viewed to be passive in that priorities are mandated by local/state laws, and schools have no say on the way priority lists are determined. In one-sided school choice problems, priority lists cannot be interpreted to be representative of the preferences of schools. In other public school systems, however, such as the one in New York City, some schools are active in that their priority lists also reflect their own assessment of the students. Such a system is an example of a “two-sided” school choice problem (Gale and Shapley 1962).¹

Shortly before the school choice problems came to widespread attention, Sönmez (1997, 1999) introduced and studied two interesting kinds of strategic behavior for two-sided matching problems. The first is the possibility of “manipulation via capacities” (Sönmez 1997): A school may circumvent the centralized procedure by underreporting its capacity. Indeed, a recent study of the New York City public high school system by Abdulkadiroğlu et al. (2005a) has revealed that under the school choice mechanism that was in use between the 1990s and 2002 in New York City, a substantial number of schools concealed their capacities in order to be matched with their preferred students.² The second is the possibility of “manipulation via pre-arranged matches” (Sönmez 1999): A student and a school may commit themselves to a mutual agreement prior to the centralized procedure, according to which the student does not participate in the procedure, and in return is offered a seat at the school, which results in a strict

¹ See Roth and Sotomayor (1990) for a thorough analysis of this problem.
² Under the old NYC public high school system, the schools that underreported capacity also tried to admit new students after the centralized match, using the hidden capacities. Hence, this particular manipulation can actually be seen to be intermediate between the two notions we study in this paper.
gain for at least one of the two parties while hurting neither. It turns out that no stable\textsuperscript{3} mechanism survives any of the two manipulations in a two-sided matching context (Sönmez 1997, 1999).

In this paper, we study the two manipulation notions for school choice problems. In our formulation of the problem, we extend a standard school choice problem to contain the preferences of each school over groups of students. These preferences are now assumed to be its private information; hence, neither the central planner nor the market participants are able to observe schools’ preferences over groups of students, which we assume to be responsive (Roth 1985). On the other hand, similarly to a standard school choice problem, students’ preferences and schools’ priority lists are assumed to be publicly known. This model can accommodate both one-sided and two-sided interpretations of school choice and is in line with practical school choice applications where student assignment is based solely on students’ preferences and schools’ priority lists. To the best of our knowledge, ours is the first attempt to model and study the two manipulation notions for school choice problems.

Our analysis focuses on three prominent mechanisms that have been widely studied in the literature. The first is the so-called Boston mechanism (BM) that was in use in the Boston school district until 2006 and is still in use in several other places.\textsuperscript{4} The next are the two competing mechanisms that were advocated by Abdulkadiroğlu and Sönmez (2003) as attractive replacements for BM. The second mechanism we investigate is Gale and Shapley’s student-optimal stable mechanism (SOSM), which inherits a number of appealing properties from two-sided matching theory such as stability and strategy-proofness.\textsuperscript{5} SOSM has been adopted by both the New York City (which has the largest public school system in the country with over a million students) and the Boston (which has over 60,000 students enrolled in the public school system) school districts. Our third mechanism is the top trading cycles mechanism (TTCM), which is based on the celebrated trading procedure proposed by David Gale. TTCM has been considered for adoption by the Boston (e.g., see the initial report of the task force of the Boston Public Schools, noted in Abdulkadiroğlu et al. (2005b), recommending TTCM over SOSM) and the San Francisco (personal communication with Muriel Niederle and Alvin Roth) school districts.

We show that between the two manipulation notions, manipulation via capacities is easier to avoid, whereas it is essentially impossible to achieve immunity against manipulation via pre-arranged matches (Propositions 1 and 2). We find that while SOSM is prone to manipulation via capacities, BM and TTCM are not (Corollary 1). Motivated by its increasing popularity among school districts in the US and its theoretical appeal, we further investigate the causes of the manipulability of SOSM and identify a sufficient and necessary condition on the priority profile and the

\textsuperscript{3} A matching is stable or eliminates justified envy if (1) there is no unmatched student–school pair \((i, x)\) such that student \(i\) prefers school \(x\) to his assignment, and either (2a) school \(x\) has a vacant seat, or (2b) student \(i\) has higher priority than at least one student who is assigned a seat at school \(x\).

\textsuperscript{4} See Abdulkadiroğlu et al. (2005b), Ergin and Sönmez (2006), and Pathak and Sönmez (2008).

\textsuperscript{5} Strategy-proofness requires that no student ever gain by misrepresenting his preferences. For more on this property and its implications in indivisible good allocation, see, for example, Pápai (2000) and Kesten and Yazıcı (2010).
vector of minimum capacities (that can be reported), which ensures SOSM’s immunity to manipulation via capacities (Theorem 1). This condition turns out to be a kind of “acyclicity” restriction that was proposed by Ergin (2002).\textsuperscript{6}

Our results have some notable policy implications. Although SOSM is gaining widespread recognition because of its strong immunity to preference manipulation (i.e., strategy-proofness) along with other desirable features, it may not be completely problem free in terms of other strategic aspects. Consequently, among the three mechanisms, only TTCM turns out to survive both preference and capacity manipulation.\textsuperscript{7}

Two pieces of advice can be drawn from Theorem 1 for school districts that use SOSM. The first is regarding the debate on whether single or multiple lotteries should be used to break the ties among students who belong to the same priority group. Loosely speaking, an acyclic priority structure significantly restricts the variation across priority lists. Therefore, Theorem 1 suggests that choosing the same tie-breaking lottery across all schools may help these school districts diminish the scope of manipulation via capacities. The second is that imposing higher capacity restrictions on schools will unambiguously decrease the likelihood of success of manipulation attempts in school districts that use SOSM.

We also find that the mechanism formerly in use in Boston in fact fares better than its replacement when it comes to manipulation via capacities. Although the early literature on school choice strongly opposed the use of BM due to the complicated high-stakes game it induces (e.g., Ergin and Sönmez 2006), the recent literature has however had a more optimistic view of this mechanism (in contrast to SOSM) due to its superior (ex ante) efficiency properties (e.g., Abdulkadiroğlu et al. 2008). Thus, our study also points out an overlooked advantage of this highly debated mechanism.

Konishi and Unver (2006) analyze the capacity manipulation game under SOSM and show the non-existence of a pure strategy equilibrium. One immediate implication of Theorem 1 is that acyclicity is also a sufficient condition for the existence of a pure strategy equilibrium of this game.\textsuperscript{8} A similar line of investigation (though logically independent from the present study) is pursued by Kojima (2007). He also identifies conditions under which the two manipulations can be avoided. There are, however, two main differences between our analysis and that of Kojima (2007). First, he assumes a standard two-sided matching model in which all participants’ preferences (including schools’ preferences over groups of students) are publicly observable. Second, the conditions he proposes are on individual school preferences, as opposed to our restrictions on priority structures. Manipulation via capacities has also been studied by Kojima and Pathak (2008) in a two-sided matching market with a large number of participants. They show that under certain regularity assumptions the scope of manipulation via capacities diminishes when the market becomes large. In a recent paper, Kojima (2010) also identifies acyclicity as a sufficient and necessary condition

\textsuperscript{6} See also Kesten (2006) and Haeringer and Klijn (2009) for more on the role of the acyclicity restrictions on TTCM and SOSM.

\textsuperscript{7} Here, we have one-sided school choice in mind. In two-sided matching no stable mechanism is strategy-proof (Roth 1982). See also Pathak and Sönmez (2009) for new concepts in order to compare non-strategy-proof mechanisms.

\textsuperscript{8} See also Kojima (2006) for an analysis of the mixed strategy equilibria of this game.
for a two-sided matching mechanism to be stable, strategy-proof, and immune to a combined manipulation, i.e., survives any blockings following a possible preference misrepresentation. Under the acyclicity restriction, he shows that SOSM is the unique mechanism satisfying these requirements.

The paper is organized as follows: In the next section, we introduce the model and briefly discuss three most studied school choice mechanisms. Section 3 discusses manipulation via capacities, and Sect. 4 manipulation via pre-arranged matches. Section 5 concludes.

2 The model

Let \( N \equiv \{1, 2, \ldots, n\}, n \geq 3 \), denote a finite set of students. Let \( X \) denote a finite set of schools. Each school \( x \in X \) has a capacity \( q_x \) which is the maximum number of students it can be assigned. Let \( q \equiv (q_x)_{x \in X} \) be the capacity vector.

Each student \( i \in N \) is equipped with a complete and transitive preference relation \( R_i \) over \( X \cup \{\emptyset\} \) where \( \emptyset \) represents not being assigned to any school in \( X \). Let \( \mathcal{R}_i \) denote the class of all such preferences for student \( i \in N \). For each \( i \in N \), let \( P_i \) denote the strict relation associated with \( R_i \).

For each school, there is a complete and transitive priority list over all students, which may be determined according to state/local laws or certain criteria of school districts (such as proximity of residence and possible specific needs of a student), or may as well represent a school’s preferences. In many public school systems in the US, for each school, the priority between any two students who are identical in every relevant aspect is determined by a lottery.\(^9\) Hence, we assume that for each school the associated priority list is strict. Let \( \succ_x \) denote the priority list for school \( x \). For example, \( i \succ_x j \) means that student \( i \) has higher priority for school \( x \) than student \( j \). The collection of priority lists \( \succ \equiv (\succ_x)_{x \in X} \) is called a priority profile. The pair \( (\succ, q) \) is called a priority structure. We assume that a publicly observable priority profile is given as a primitive of the model.

A matching is a function \( \mu : N \cup X \rightarrow 2^{N \cup X} \) such that:

1. For all \( i \in N \), \( |\mu(i)| \leq 1 \) and \( \mu(i) \subseteq X \).
2. For all \( x \in X \), \( |\mu(x)| \leq q_x \) and \( \mu(x) \subseteq N \).
3. For all \( (i, x) \in N \times X \), \( \mu(i) = \{x\} \) if and only if \( i \in \mu(x) \).

Each school \( x \in X \) is equipped with a strict and responsive (Roth 1985) preference relation \( R_x \) over \( 2^N \setminus \{\emptyset\} \).\(^10\) Each school’s preferences over groups of students are its private information, and hence are not publicly observable. Preference relation \( R_x \) is responsive if for all \( M \subseteq N \) and all \( i, j \in N \setminus M \), \( \{i\} \cup M \cup \{j\} \) if and only if \( \{i\} \cup P_x \{j\} \). Let \( \mathcal{R}_x \) denote the class of all such preferences for school \( x \in X \).

\(^9\) See Erdil and Ergin (2008), Abdulkadiroğlu et al. (2009), and Kesten (2010) for the welfare consequences of this practice.

\(^10\) Note that this specification implies that each student is admissible to each school. This assumption is not needed for a “one-sided” school choice model.
We do not impose any restriction on the way priority lists and school preferences are related. If school $x$ has control over its priority list (i.e., two-sided school choice), then one can assume that $\succ_x$ coincides with the restriction of $R_x$ to singletons. If school $x$ has no control over its priority list (i.e., one-sided school choice), then one can assume that $\succ_x$ is possibly different from the restriction of $R_x$ to singletons.

A (school choice) problem is a pair $((R_s)_{s \in N \cup X}, q)$ or simply $(R, q)$, that specifies the preferences of each student, the unobservable preferences of each school, and the capacity of each school. Let $\mathcal{E} \equiv \Pi_{s \in N \cup X} R_s \times \mathbb{N}^{|X|}_+$ be the class of all problems.

Given a priority structure $(\succ, q)$, the matching outcome is determined according to the “mechanism” based only on the submitted preferences of students. A (school choice) mechanism $\varphi$ associates with each (school choice) problem $(R, q)$ a matching $\varphi(R, q)$ such that $\varphi(R, q) = \varphi(\tilde{R}, q)$ for any profile $\tilde{R}$ satisfying $R^N = \tilde{R}^N$. For ease of notation, we will sometimes use $\varphi_i(R, q)$ to refer to $\varphi(R, q)(i)$. We next describe three prominent school choice mechanisms from the recent literature.

2.1 Boston mechanism

Our first mechanism is a widespread real-life mechanism. The Boston mechanism (BM) and its slight variants are currently in use in a number of places such as Seattle, Minneapolis, Lee County, and Florida. Its outcome can be calculated via the following algorithm for a given problem:

**Step 1:** For each school $x$, consider only those students admissible to it who have listed it as their first acceptable choice. Those $q_x$ students among them with the highest priority for school $x$ (all students if fewer than $q_x$) are placed to school $x$.

**Step $k$, $k \geq 2$:** Consider the remaining students. For each school $x$ with $q^k_x$ available seats, consider only those students admissible to it who have listed it as their $k$-th acceptable choice. Those $q^k_x$ students among them with the highest priority for school $x$ (all students if fewer than $q^k_x$) are placed to school $x$.

The algorithm terminates when there are no students or schools left. Any student who has not been placed to a school acceptable to him remains unassigned.

An important critique of BM that was highlighted in the early school choice literature is that it gives strong incentives to students to misstate their preferences. Because a student who has high priority for a school may lose his advantage for that school if he does not list it as his first choice, BM forces students to play complicated coordination games. Due to these concerns and their implications regarding fairness issues, following intensive policy discussions, Boston Public Schools recently voted to replace BM with the next mechanism.

2.2 Student-optimal stable mechanism

Gale and Shapley (1962) proposed the student-optimal stable mechanism (SOSM) to find the stable allocation that is most favorable to each student among stable matchings for any given two-sided matching problem. Since then this mechanism (as well
as its dual for the school side) has been the central mechanism in two-sided matching.
Its outcome can be calculated via the following *deferred acceptance algorithm* for a
given problem:

**Step 1:** Each student applies to his favorite acceptable school. For each school $x$, those
$q_x$ applicants (all applicants if fewer than $q_x$) who have the highest priority for
school $x$ are tentatively placed to school $x$. The remaining applicants are rejected.

**Step $k$, $k \geq 2$:** Each student who was rejected from a school at step $k - 1$ applies
to his next favorite acceptable school. For each school $x$, those $q_x$ students who
have the highest priority for school $x$ among the new applicants and those who
were tentatively placed to it at an earlier step are tentatively placed to school $x$.
The remaining applicants are rejected.

The algorithm terminates when each student is either tentatively placed to a school
or has been rejected by every school acceptable to him. We denote the student-optimal
stable mechanism (SOSM) associated with the priority profile $\succ$ by $\phi^\succ$.

SOSM has proven quite successful not only in two-sided matching theory but also
in a number of entry-level labor markets, the biggest of which is the redesigned US.
National Resident Matching Program (Roth and Peranson 1999). One main reason
(aside from its stability) behind its success is the fact that it makes it a dominant
strategy for students to state their true preferences (Roth 1982; Dubins and Freedman
1981).11

SOSM has so far been the leading mechanism for school choice applications as
well. Since the first study on this problem by Abdulkadiroğlu and Sönmez (2003), it
has often been advocated for school choice problems. These efforts have recently paid
off, and SOSM has been adapted by both the New York City (in 2002) and Boston (in
2005) public school systems, which have been suffering from congestion (NYC) and
incentive (Boston) problems.

### 2.3 Top trading cycles mechanism

David Gale proposed a “top trading cycles” mechanism in the context of a *housing
market* where each agent initially owns a distinct indivisible good. The outcome of this
mechanism was later shown to coincide with the core12 of the corresponding housing
market (Shapley and Scarf 1974). Furthermore, it was shown to be the unique Pareto
efficient, individually rational, and strategy-proof mechanism (Ma 1994).

The *top trading cycles mechanism* (Abdulkadiroğlu and Sönmez 2003) is an adap-
tation of Gale’s top trading cycles algorithm to the school choice context. Its outcome
can be calculated via the following algorithm for a given problem:

---

11 See also Roth and Vande Vate (1991), Ma (1996), Bogomolnaia and Moulin (2002), and Klaus and Klijn
(2006) for more on incentive and fairness/stability issues.

12 If preferences are strict, this allocation is unique (Roth and Postlewaite 1977).
Step 1: Assign a counter to each school. Its initial value is the capacity of that school. Each student points to his favorite acceptable school, and each school points to the student among those admissible to it who has the highest priority for it. There is at least one cycle. Each student in a cycle is placed to the school he points to and removed. The counter of each school in a cycle is reduced by one, and if it becomes zero, the school is also removed. The counters of all other schools remain the same.

Step \( k, k \geq 2 \): Each remaining student points to his favorite acceptable school among the remaining schools, and each remaining school points to the student among those admissible to it who has the highest priority for it. There is at least one cycle. Each student in a cycle is placed to the school he points to and removed. The counter of each school in a cycle is reduced by one, and if it becomes zero, the school is also removed. The counters of all other schools remain the same.

The algorithm terminates when no school or student is left. The top trading cycle mechanism (TTCM) achieves Pareto efficiency (which SOSM lacks) by allowing students to trade their priorities. It is also strategy-proof.

3 Manipulation via capacities

Sönmez (1997) considers a two-sided matching application with medical interns on one side of the market and hospitals on the other. He introduces an interesting kind of manipulation that a hospital may engage in, namely, withholding its capacity.\(^\text{13}\) We consider the same possibility for the school choice problem where preferences of each school are now its private information, and the capacity reported by each school is the only parameter it has an influence on among those that are included in the calculation of the outcome of the central clearing house. In fact, this is not a merely theoretical concern. Indeed, Abdulkadiroğlu et al. (2005a) report that public schools in NYC were withholding\(^\text{14}\) their capacities under the old public school system\(^\text{15}\) that was in use from 1990s till 2003.

For each school \( x \in X \), we now interpret \( q_x \) as the capacity reported by school \( x \). In many school districts, each school is typically required to admit a certain number of students. Therefore, it is not realistic to assume that a school’s capacity can be arbitrarily small. We incorporate this possibility into our model via the following restriction: For each \( x \in X \), let \( q_x \) denote the minimum capacity imposed on school \( x \). That is, school \( x \) can never report a capacity less than \( q_x \). Let \( q = (q_x)_{x \in X} \) be the minimum capacity vector. Let \( \varphi \) be a mechanism:

\(^\text{13}\) Recently, Kojima and Pathak (2008) extend this analysis to large two-sided matching markets.

\(^\text{14}\) In principle, it is plausible to imagine the possibility of a manipulation attempt by a school through over-reporting its capacity. This, however, may not be a practical strategy since any additional students in excess of the capacity may not be accommodated or their presence may lead to various undesirable outcomes for the school.

\(^\text{15}\) It was a centralized system, which operated through three rounds of application processing, and subsequently suffered from congestion problems. See Abdulkadiroğlu et al. (2005a) for more.
Non-manipulability via capacities: For all \((R, q) \in \mathcal{E}, x \in X,\) and all \(q_x \leq q'_{x} < q_x,\) we have \(\varphi(R, q)(x) R_{x} \varphi(R, q'_{x}, q_{-x})(x).\)\(^{16}\)

Sönmez (1997) shows that the hospital-optimal stable mechanism that was used by the National Resident Matching Program is not immune to manipulation via capacities, and moreover, there is no stable mechanism that is non-manipulable via capacities in the two-sided matching context.

For school choice problems, it turns out that non-manipulability via capacities requires that if a student is assigned to a school when that school underreports its capacity, then the student should continue to be assigned to that school when the school truthfully reports its capacity.

Proposition 1 A school choice mechanism \(\varphi\) is non-manipulable via capacities if and only if for all \((R, q) \in \mathcal{E}, x \in X,\) and all \(q_x \leq q'_{x} < q_x,\) we have \(\varphi(R, q'_{x}, q_{-x})(x) \subset\varphi(R, q)(x).\)

Proof The ‘if’ part simply follows from the responsiveness of schools’ preferences together with the assumption that school preferences are defined over \(2^X \setminus \{\emptyset\}.\) To see the ‘only if’ part suppose there exist a mechanism \(\varphi,\) a problem \((R, q),\) a school \(x \in X,\) and \(q_x \leq q'_{x} < q_x\) such that \(\varphi(R, q'_{x}, q_{-x})(x) \not\subset\varphi(R, q)(x).\) But because schools’ preferences are unobservable by the mechanism designer, the outcome is independent of the preferences of school \(x.\) Then, one can easily construct preferences such that \(\varphi(R, q'_{x}, q_{-x})(x) \not\subset\varphi(R, q)(x).\)

\(\Box\)

Surprisingly, the most popular mechanism in two-sided matching theory, which is also gaining popularity among school districts, is not immune to manipulation via capacities. On the contrary, its closest competitor, TTCM, as well as the mechanism it recently replaced in Boston survive this manipulation test.

Corollary 1 The student-optimal stable mechanism\(^{17}\) is manipulable via capacities, whereas the top trading cycles mechanism and the Boston mechanism\(^{18}\) are not.

Proof The subsequent Example 1 shows that SOSM does not satisfy the property given in Proposition 1. To see that the BM and TTCM indeed do, simply observe that the steps and assignments of students for problems \((R, q'_{x}, q_{-x}),\) and \((R, q)\) are identical under each mechanism, until right after the last seat of school \(x\) is assigned.

\(\Box\)

\(^{16}\) We say that school \(x \in X\) manipulates \(\varphi\) via capacities at a problem \((R, q),\) if there exists \(q_x \leq q'_{x} < q_x\) such that \(\varphi(R, q)(x) P_{x} \varphi(R, q'_{x}, q_{-x})(x).\) Hence, a mechanism \(\varphi\) is non-manipulable via capacities if no school manipulates \(\varphi\) at any problem.

\(^{17}\) The manipulability of SOSM via capacities is also a corollary of Theorem 1 of Sönmez (1997) since the incompatibility of stability with non-manipulability via capacities in our school choice model is also implied by the same incompatibility in the two-sided matching model of Sönmez (1997). This incompatibility can be shown by a slight modification of the subsequent Example 1.

\(^{18}\) The Boston mechanism is a member of a large class of priority matching mechanisms studied by Roth (1991). These mechanisms have been in use in the entry-level labor markets in the UK. Any such priority matching mechanism that satisfies the monotonicity requirement of Proposition 1 is also immune to capacity manipulations.
SOSM is currently in use in a number of entry-level labor markets [see Roth and Rothblum (1999) for a list of these markets]. Due to its increasing popularity among the public school systems in the US, we next study this mechanism more closely and analyze the causes and the extent of its manipulability. We first consider an example that illustrates how SOSM can be manipulated:

**Example 1** Let $N = \{1, 2, 3\}$, $X = \{a, b\}$ where (true) capacities are $q_a = 2$ and $q_b = 1$. Also let

\[
\begin{array}{c|c|c}
\succ_a & \succ_b \\
1 & 3 \\
3 & 2 \\
2 & 1 \\
\end{array}
\]

\{1, 2, 3\} \ P_a \{1\} \ P_a \{2, 3\} \ P_a \{3\} \ P_a \{2\}, \ b \ P_1 \ a \ P_1 \emptyset, \ a \ P_2 \ b \ P_2 \emptyset, \ a \ P_3 \emptyset,

where a school not specified in a student’s preference relation is ranked lower than those specified. It is straightforward to calculate that $\phi(\succ, (R, q))(a) = \{2, 3\}$. Now let $q'_a = 1$. Observe that when school $a$ underreports its capacity as $q'_a = 1$, student 2 is now rejected from school $a$. Next, he applies to his second favorite school, which is $b$, and causes student 1 to be rejected from school $b$. This, in turn, causes school $a$ to admit student 1 and reject student 3. Thus, $\phi(\succ, (R, q'_a, q-a))(a) = \{1\}$, and school $a$ gains by underreporting its capacity.

Ergin (2002) identifies conditions under which SOSM can achieve a number of appealing properties it lacks otherwise. He introduces a notion of “acyclicity” for priority structures and shows that it is sufficient as well as necessary to recover properties such as Pareto efficiency and group strategy-proofness. We argue that acyclicity (adopted appropriately to our setting) also plays a key role in SOSM’s ability to avoid manipulation via capacities.

A given priority structure $(\succ, q)$ contains a cycle if the following two conditions are satisfied:

**Loop condition**: There are distinct $i, j, k \in N$ and $x, y \in X$ such that $i \succ_x j \succ_x k$ and $k \succ_y i$.

**Scarcity condition**: There exist (possibly empty) disjoint sets of students $N_x, N_y \subset N\setminus\{i, j, k\}$ such that $N_x \subset U_x(j), \ N_y \subset U_y(i), \ |N_x| = q_x - 1,$ and $|N_y| = q_y - 1$.

The priority structure $(\succ, q)$ is acyclical if it has no cycles.

Acyclicity restrictions (the loop condition and the scarcity condition) apply jointly on the priority profile and the minimum capacity vector. For SOSM, loosely speaking, acyclicity prevents those situations in which a school, by underreporting capacity, initiates a new rejection chain that eventually leads to the application of a completely new student to this school (recall Example 1). For example, in the presence of a cycle in the priority structure, some school say $x$, can simply refuse to be assigned some low-priority student, say $k$, by concealing an open slot. Such a refusal, however, may give rise to a subsequent rejection of some high-priority student, say $i$, from another school, say $y$, causing him to apply to school $x$. 

Springer
Two extreme examples of acyclic structures are the following: If the minimum capacity a school can report is \( n \), then school seats are always in abundance (hence, the scarcity condition is not satisfied), and there are no cycles. On the other hand, if the priority lists are identical for each school, then the loop condition is not satisfied, and the structure is again acyclic regardless of minimum capacities. In general, as minimum capacities get larger, the formation of cycles becomes harder. We refer the interested reader to Theorem 2 of Ergin (2002) for a characterization of acyclic structures.

Next is our main result: It turns out that SOSM is immune to manipulation via capacities so long as the priority structure \((\succ, q)\) is acyclical.

**Theorem 1** The student-optimal stable mechanism \( \phi^\succ \) associated with a priority profile \( \succ \) is immune to manipulation via capacities if and only if the priority structure \((\succ, q)\) is acyclical.

**Proof** (i) Acyclicity of \( \succ \implies \) Non-manipulability via capacities:

Suppose that there are \((R, q) \in \mathcal{E}, x \in X\), and \( q_x^\prime \leq q_x' < q_x \) such that \( \phi^\succ(R, q_x', q_x) \succ P_x \phi^\succ(R, q) \succ (x) \). Let \( \mu \equiv \phi^\succ(R, q) \) and \( \mu' \equiv \phi^\succ(R, q_x', q_x) \). If \( q_x > |\mu(x)| \), then \( \mu \neq \mu' \) implies \( q_x' < |\mu(x)| \).\(^{19}\) Otherwise, \( q_x = |\mu(x)| \) and since \( q_x' < q_x \), clearly \( q_x' < |\mu(x)| \). Thus, \( |\mu'(x)| \leq q_x' < |\mu(x)| \). This means \( |\mu(x)| \mu'(x) \geq 1 \). Moreover, since \( \mu'(x) \) and preferences are responsive, we have \( \mu'(x) \mu(x) \neq \emptyset \). Recall the way the outcome of \( \phi^\succ \) is calculated: Each student applies to his favorite choice, and if rejected, applies to his next favorite choice and so on. Note that the only difference between \((R, q)\) and \((R, q_x', q_x)\) is that \( q_x' < q_x \). This means that each student applies to the same schools at \((R, q_x', q_x)\) as he did at \((R, q)\), and to possibly more, since now \( q_x' < q_x \). This is also equivalent to saying that all students are made weakly worse off.\(^{20}\) Moreover, all students in the set \( (\mu'(x) \mu(x)) \cup (\mu(x) \mu'(x)) \) are made strictly worse off at \((R, q_x', q_x)\).

Theorem 1 of Ergin (2002) shows that when schools are assumed to be irrelevant for welfare considerations, SOSM is Pareto efficient for a problem \((R, q) \in \mathcal{E}\) if and only if priority structure \((\succ, q)\) is acyclical according to his definition.\(^{21}\) We prove this part by constructing a matching \( \mu''\) that Pareto dominates \( \mu'\) for the problem \((R, q_x', q_x)\) when we consider only students’ welfare. This means that priority structure \((\succ, q_x', q_x)\) has a cycle according to the definition in Ergin (2002).

Since \( q_x' \leq q_y' \) and \( q_y \leq q_y \), for any \( y \in X \setminus \{x\} \), the priority structure \((\succ, q)\) has a cycle according to our definition as well. A contradiction.

---

\(^{19}\) For if \( q_x' \geq |\mu(x)| \), then the steps of \( \phi^\succ \) are identical for problems \((R, q)\) and \((R, q_x', q_x)\); and thus \( \mu = \mu' \).

\(^{20}\) This is the “resource monotonicity” property of SOSM.

\(^{21}\) To be precise, a given priority structure \((\succ, q)\) contains a cycle according to Ergin (2002), if the following two conditions are satisfied:

**Loop condition:** There are distinct \( i, j, k \in N \) and \( x, y \in X \) such that \( i \succ_j j \succ_k k \) and \( k \succ_y i \).

**Scarcity condition:** There exist (possibly empty) disjoint sets of students \( N_x, N_y \subset N \setminus \{i, j, k\} \) such that \( N_x \subset U_y(j), N_y \subset U_y(i), |N_x| = q_x - 1 \) and \( |N_y| = q_y - 1 \).

The priority structure \((\succ, q)\) is acyclical if it has no cycles.
McVitie and Wilson (1970) showed that the outcome of SOSM is independent of the sequence according to which students make their applications. Therefore, one can alternatively calculate $\varphi^>(R, q'_x, q_{-x})$ as follows: Initially run the DA algorithm for the problem $(R, q)$. Then, decrease the quota for $x$ from $q_x$ to $q'_x$. This will cause only those $q_x - q'_x$ students with lowest $x$ priority to be rejected from $x$. Next, each rejected student will apply to his next choice, and the algorithm will eventually terminate when there are no more rejections.

Let $i_0 \in \mu'(x) \backslash \mu(x)$. Also, let $i_0 \in \mu(y_0)$ for some $y_0 \in X \backslash \{x\}$. Since student $i_0$ is worse off at $\mu'$, there exists $i_1 \in \mu'(y_0) \backslash \mu(y_0)$ such that $i_1 \neq i_0$, who did not apply to $y_0$ at $(R, q)$ (otherwise, he would be accepted by $y_0$). Let $i_1 \in \mu(y_1)$ for some $y_1 \in X \backslash \{x, y_0\}$. Since student $i_1$ is worse off at $\mu'$, there exists $i_2 \in \mu'(y_1) \backslash \mu(y_1)$ who did not apply to $y_1$ at $(R, q)$. Let $i_2 \in \mu(y_2)$ for some $y_2 \in X \backslash \{x, y_0, y_1\}$. Continuing in this fashion, there must exist $i_k \in \mu(y_k) \backslash \mu(y_{k-1})$ for some $y_{k-1} \in X \backslash \{x, y_0, \ldots, y_{k-2}\}$ with $k \geq 1$ and $x \equiv y_{k-1}$, who is the first student (according to the steps of the DA algorithm) in this sequence to be rejected from his assignment at $(R, q)$. By the above described calculation of $\varphi^>(R, q'_x, q_{-x})$, this student clearly belongs to $\mu(x) \backslash \mu'(x)$. Then, let $\mu''$ be such that if $i \in \{i_0, i_1, \ldots, i_k\}$, $\mu''(i) = \mu(i)$, and otherwise $\mu''(i) = \mu'(i)$. (i.e., $\mu''$ is obtained from $\mu'$ by simply assigning each $i_t$ with $t \in \{0, 1, \ldots, k\}$ to $\mu'(i_{t+1})$ where $i_{k+1} \equiv i_0$.) Since $k \geq 1$, $\mu''$ clearly, Pareto dominates $\mu'$ at $(R, q'_x, q_{-x})$ from the view of the student side.

(ii) Non-manipulability via capacities $\implies$ Acyclicity of $\succ$:

Suppose by contradiction that a priority structure $(\succ, q)$ contains a cycle. In particular, there are (1) $i, j, k \in N$ and $x, y \in X$ such that $i \succ_x j \succ_x k$ and $k \succ_y i$ and (2) $N_x, N_y \subset N \backslash \{i, j, k\}$ such that $N_x \subset U_x(j)$, $N_y \subset U_y(i)$, $|N_x| = q_x - 1$, $|N_y| = q_y - 1$, and $N_x \cap N_y = \emptyset$. Let $(R, q)$ be a problem satisfying the following: Let $q \geq q$ be any capacity vector with $q_x > q_x$ and $q_y = q_y$. For all $m \in N_x$ and all $z \in X \backslash \{x\}$, $x P_m z P_m \emptyset$. For all $m \in N_y$ and all $z \in X \backslash \{y\}$, $y P_m z P_m \emptyset$. For all $z \in X \backslash \{y\}$, $y P_i P_j z$. For all $z \in X \backslash \{x\}$, $x P_i \emptyset P_j z$. For all $m \in N \backslash \{N_x \cup N_y \cup \{i, j, k\}\}$ and all $z \in X$, let $\emptyset \subset P_m z$. Finally, suppose $N_x \cup \{i\} P_N N_x \cup \{j, k\}$. Clearly, $\varphi^>(R, q)(x) = N_x \cup \{j, k\}$. Now let $q'_x = q_x$. Then, $\varphi^>(R, q'_x, q_{-x})(x) = N_x \cup \{i\}$. Thus, school $x$ successfully manipulates $\varphi^>$ via underreporting its capacity at $(R, q)$. \hfill \Box

An acyclic priority structure completely eliminates manipulation via capacities. The presence of a cycle in a priority structure, however, introduces the possibility of manipulation. Consequently, as also suggested by the proof of Theorem 1, the more cycles a priority structure has the more room there is for manipulation.

Example 1 and the proof of Theorem 1 hint at a second observation about manipulation via capacities: Any school that gains by underreporting its capacity should indeed have preferences that have certain degree of ‘correlation’ with the school’s priority list. The connection between manipulation via capacities and the acyclicity of

\footnote{Note that $y_0 \neq \emptyset$ since $\varphi^>$ is individually rational. Otherwise, student $i_0$ would never have applied to school $x$ at $(R, q'_x, q_{-x})$.}
a priority structure shown by Theorem 1 also entails an important piece of practical advice, suggested by the next result.

**Corollary 2** Fix a priority profile $\succ$. Consider a school district that uses the student-optimal stable mechanism $\varphi^\succ$ and needs to choose between two possible minimum capacity vectors $q$ and $q'$ such that $q \succeq q'$. If a school manipulates $\varphi^\succ$ via capacities at a problem by underreporting its capacity when $q$ is imposed, then the same school manipulates $\varphi^\succ$ via capacities at the same problem when $q'$ is imposed. However, the converse is not necessarily true.

**Proof** Suppose $(\succ, q)$ is imposed and there exist a problem $(R, q) \in E$ and a school $x \in X$ that manipulates $\varphi^\succ$ via capacities at $(R, q)$. Then, by Theorem 1 $(\succ, q)$ has a cycle. Since $q \succeq q'$, the priority structure $(\succ, q')$ shares the same cycle. Then, since $q_x' \leq q_x$, school $x$ can manipulate $\varphi^\succ$ at $(R, q)$ in the same way if $(\succ, q')$ is imposed instead. For the converse case, consider Example 1. Note that school $a$ manipulates $\varphi^\succ$ via capacities at this problem if $q_a = q_b = 1$; however, this would not be possible if $q_a$ were raised to two. $\square$

Corollary 2 implies that the Boston and NYC school districts can in fact reduce the chances of manipulation via capacities by setting higher minimum capacity restrictions on schools.

Konishi and Unver (2006) consider capacity manipulation games in the hospital-intern market application of the model. They analyze the pure strategy equilibria of the game-form under the school-optimal (hospital-optimal in their context) and the student-optimal (intern-optimal in their context) stable mechanisms. They show that neither game may have a pure strategy equilibrium.

An admissible capacity of each school $x$ is a nonnegative integer no less than its minimum capacity $q_x$. Thus, $x$ can report $q_x' \in Q_x \equiv \{q_x, q_x + 1, \ldots, q_x\}$. Define the set of admissible capacities as $Q = \Pi_{x \in X} Q_x$. A capacity reporting game under a mechanism $\varphi$ is described by a strategic form game $(X, (Q_x, R_x)_{x \in X})$. A pure strategy Nash equilibrium of a game $(X, (Q_x, R_x)_{x \in X})$ is a strategy profile at which no school has a profitable deviation.

Theorem 1 has direct implications for capacity reporting games. We know that under the acyclicity restriction no school benefits by underreporting its capacity. Then, the following is immediate.

**Corollary 3** The capacity reporting game under the student-optimal stable mechanism $\varphi^\succ$ associated with a priority profile $\succ$ has a pure strategy Nash equilibrium if the priority structure $(\succ, q)$ is acyclical.

### 4 Manipulation via pre-arranged matches

Sönmez (1999) studies the vulnerability of two-sided matching markets to an alternative form of manipulation: manipulation via pre-arranged matches: A student–school pair agrees on an arrangement before the formal procedure according to which the student is enrolled at the school, and the student does not participate in the procedure.
This arrangement is successful if (at least) one of the two parties gains and neither one suffers as a result. Such behavior was observed in the US hospital-intern market and the Canadian lawyer market where stable mechanisms were in use. Sönmez (1999) shows for two-sided matching problems that there is no mechanism that is both stable and non-manipulable via pre-arranged matches.

We now study this second kind of manipulation for school choice problems. In the present context, since the population of students might change, a (school choice) problem is defined by a triplet \((N, R = (R_i)_{i \in N}, q)\). Given a preference profile \(R = (R_i)_{i \in N}\), let \(R_{-i}\) be the preference profile obtained from \(R\) by removing the preferences of student \(i\). Given a preference profile \(R = (R_i)_{i \in N}\), let \(R^M\) denote the restriction of profile \(R\) to the set of students in \(M \subset N\). The rest of the definitions and notations apply unchanged. Let \(\varphi\) be a mechanism:

**Non-manipulability via pre-arranged matches:** Given a problem \((N, R, q)\), there is no school–student pair \((x, i)\) such that \(x R_i \varphi(N, R, q)\).

It turns out that manipulations via pre-arranged matches are essentially impossible to avoid. The next result makes this point precise.

**Proposition 2** Suppose that there is a school \(x \in X\) such that \(n > q_x\). Then, no mechanism is immune to manipulations via pre-arranged matches.

**Proof** Consider any mechanism \(\varphi\). Let a priority structure \((>, q)\) be given. Take a school \(x \in X\) with \(n > q_x\). Let \(R\) be any preference profile satisfying the following: Choose any \(N'_x \subset N\) with \(|N'_x| = q_x + 1\) and for all \(i \in N'_x\) and all \(z \in X\{"x\}\), let \(x P_i z\). For all \(i \in N \setminus N'_x\) and all \(z \in X\), let \(\emptyset P_i z\). Let \(\mu \equiv \varphi(N, R, q)\). By feasibility, there is \(m \in N'_x\) such that \(\mu(m) \neq x\). Suppose the pair \((x, m)\) makes a pre-arrangement. Let \(\mu' \equiv \varphi(N \setminus \{m\}, R_{-m}^{N \setminus \{m\}}, q_{-x}, q_x - 1)\). Simply letting \(\{m\} \cup \mu'(x) P_x \mu(x)\) is sufficient to show that the pair \((x, m)\) can successfully manipulate at \((N, R, q)\). \(\square\)

Proposition 2 shows that mechanism design is almost helpless to deal with potential pre-arrangements of agents from the two sides of the problem. This result suggests that essentially the only way to prevent this kind of behavior would be via establishing explicit laws and regulations within school districts that discourage schools from enrolling students on their own, and avoiding systems that would allow decentralized behavior prior to the centralized procedure.

**Corollary 4** The student-optimal stable mechanism, the top trading cycles mechanism and the Boston mechanism are all manipulable via a pre-arranged match.

Despite the discouraging news of Proposition 2, unavoidability via pre-arranged matches does not appear to be a serious concern in practice. In fact, we are not aware...
of any manipulation attempts in this manner in any of the school districts studied to date.23

5 Conclusion

Our results show that immunity to manipulation via capacities is easier to achieve than to manipulation via pre-arranged matches. We have shown that while TTCM and the monotone priority matching mechanisms such as the Boston mechanism are readily immune to capacity maneuvers, certain restrictions are needed to make this possible for SOSM. At first glance, these results may seem oddly in conflict with those of Kesten (2006), where it was shown that SOSM has an edge over TTCM in terms of other desirable properties such as population/resource monotonicity24 and consistency. One way to reconcile the two results is that the resource monotonicity property of SOSM, that considers changes in student welfare due to a change in the number of school seats, translates into a chance of profitable manipulation for the school side given the stability of SOSM.25 It is also worthwhile to note that ‘cyclicity of a priority structure’ (Ergin 2002) has once again proven to be the main reason behind a vulnerable aspect of SOSM.

This study also has important policy implications. One surprising finding is that the controversial BM is superior to SOSM in terms of immunity against manipulation via capacities. This result hence supports the more optimistic take of the recent school choice literature on this mechanism. Given that acyclicity is a strong restriction [e.g., see Theorem 2 of Ergin (2002)], we have shown that manipulation via capacities may be hard to avoid under SOSM. Nonetheless, in places where SOSM is in use, school districts may be able to limit the scope of manipulation via capacities (1) by using a single tie-breaking rule for all schools as opposed to multiple school-specific tie-breaking rules and/or (2) by increasing the lower bounds on school capacities.

Furthermore, we conclude that among the three prominent mechanisms only TTCM is immune to both preference and capacity manipulation.

Acknowledgments I thank the editor, an associate editor, and two anonymous referees for helpful suggestions.

References

Abdulkadiroğlu, A., Che, Y-K., Yasuda, Y.: Resolving conflicting preferences in school choice: the Boston mechanism reconsidered. Am Econ Rev forthcoming (2008)

23 However, we note that the particular kind of manipulation observed under the old public high school system in NYC, mentioned in footnote 2, shares some similar features.

24 The resource (population) monotonicity property requires that all students be affected in the same direction in welfare terms whenever school capacities (the set of students) shrinks or expands. See also Ehlers et al. (2002).

25 More specifically, under SOSM whenever a school underreports its capacity, all students (weakly) lose by resource monotonicity; but to avoid a blocking pair at the new SOSM outcome, this manipulation may cause some schools to (strictly) gain.
Abdulkadiroğlu, A., Pathak, P., Roth, A.: The New York City high school match. Am Econ Rev P&P 95(2), 364–367 (2005a)
Abdulkadiroğlu, A., Pathak, P., Roth, A.: Strategyproofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. Am Econ Rev 99, 1954–1978 (2009)
Abdulkadiroğlu, A., Pathak, P., Roth, A., Sönmez, T.: The Boston public school match. Am Econ Rev P&P 95(2), 368–371 (2005b)
Abdulkadiroğlu, A., Sönmez, T.: School choice: a mechanism design approach. Am Econ Rev 93, 729–747 (2003)
Balinski, M., Sönmez, T.: A tale of two mechanisms: student placement. J Econ Theory 19, 623–636 (2002)
Dubins, L.E., Freedman, D.A.: Machiavelli and the Gale-Shapley algorithm. Am Math Monthly 88, 485–494 (1981)
Ehlers, L., Klaus, B., Pápai, S.: Strategy-proofness and population-monotonicity for house allocation problems. J Math Econ 83, 329–339 (2002)
Kesten, O.: On two competing mechanisms for priority based allocation problems. J Econ Theory 127, 155–171 (2006)
Kesten, O.: School choice with consent. Q J Econ 125(3), 1297–1348 (2010)
Klaus, B., Klijn, F.: Procedurally fair and stable matching. Econ Theory 27, 431–447 (2006)
Kojima, F.: Mixed strategies in games of capacity manipulation in hospital-intern markets. Soc Choice Welf 27, 25–28 (2006)
Kojima, F.: When can manipulations be avoided in two-sided matching markets? Maximal domain results. B.E. J Theor Econ (contribution), Article 32 (2007)
Kojima, F.: Robust stability in matching markets. Theor Econ, forthcoming (2010)
Kojima, F., Pathak, P.: Incentives and stability in large two-sided matching markets. Am Econ Rev 99, 608–627 (2008)
Konishi, H., Unver, U.: Games of capacity manipulation in hospital-intern markets. Soc Choice Welf 27, 3–24 (2006)
Ma, J.: Strategy-Proofness and the strict core in a market with indivisibilities. Int J Game Theory 23, 75–83 (1994)
Ma, J.: On randomized matching mechanisms. Econ Theory 8, 377–381 (1996)
McVitie, D., Wilson, B.: Stable marriage assignment for unequal sets. BIT 10, 295–309 (1970)
Pápai, S.: Strategy-proof assignment by hierarchical exchange. Econometrica 68, 1403–1433 (2000)
Pathak, P., Sönmez, T.: Leveling the playing field: sincere and sophisticated players in the Boston mechanism. Am Econ Rev 98, 1636–1652 (2008)
Pathak, P., Sönmez, T.: Comparing mechanisms by their vulnerability to manipulation. Working paper, MIT and Boston College (2009)
Roth, A.: The economics of matching: stability and incentives. Math Oper Res 7, 617–628 (1982)
Roth, A.: The college admissions problem is not equivalent to the marriage problem. J Econ Theory 36, 277–288 (1985)
Roth, A.: A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom. Am Econ Rev 81, 415–440 (1991)
Roth, A., Peranson, E.: The effects of a change in the NRMP matching algorithm. Am Econ Rev 89, 748–780 (1999)
Roth, A., Postlewaite, A.: Weak versus strong domination in a market with indivisible goods. J Math Econ 4, 131–137 (1977)
Roth, A., Rothblum, U.G.: Truncation strategies in matching markets – In search of advice for participants. Econometrica 67, 21–43 (1999)
Roth, A., Sotomayor, M.: Two-Sided Matching. New York: Cambridge University Press (1990)
Roth, A., Vande Vate, J.H.: Incentives in two-sided matching with random stable mechanisms. Econ Theory 1, 31–44 (1991)
Shapley, L.S., Scarf, H.: On cores and indivisibility. J Math Econ 1, 23–28 (1974)
Sönmez, T.: Manipulation via capacities in two-sided matching markets. J Econ Theory 77, 197–204 (1997)
Sönmez, T.: Can pre-arranged matches be avoided in two-sided matching markets. J Econ Theory 86, 148–156 (1999)