Noname manuscript No.
(will be inserted by the editor)

Strategic Schools under the Boston Mechanism Revisited

Inácio Bó · C.-Philipp Heller

October, 2016

Abstract Ergin & Sönmez (2006) showed that for schools it is a dominant strategy to report their preferences truthfully under the Boston mechanism, and that the Nash equilibrium outcomes in undominated strategies of the induced game are stable. We show that these results rely crucially on two assumptions. First, schools need to be restricted to reporting all students as acceptable. Second, students cannot observe the preferences reported by the schools before submitting their own preferences. We show that relaxing either assumption gives schools an incentive to manipulate their reported preferences. We provide a full characterization of undominated strategies for schools and students for the simultaneous move game induced by the Boston mechanism. Nash equilibrium outcomes in undominated strategies of that game may contain unstable matchings. Furthermore, when students observe schools' preferences before submitting theirs, the subgame perfect Nash equilibria of the sequential game induced by the Boston mechanism may also contain unstable matchings. Finally, we show that schools may have an incentive to manipulate capacities only if students observe the schools' strategies before submitting their own preferences.

Keywords Mechanism Design · Two-Sided Matching · Boston Mechanism · School Choice · Market Design

JEL classification: C78, D63, D78, D82

We thank Lars Ehlers, Morimitsu Kurino, Utku Ünver, Tayfun Sönmez, Yosuke Yasuda, and two anonymous referees for helpful comments. Financial support from the Deutsche Forschungsgemeinschaft and the Einstein Stiftung (Heller) is gratefully acknowledged.

Corresponding author. Address: WZB Berlin Social Science Center, Reichpietschufer 50, D-10785 Berlin; website: http://www.inaciobo.com;e-mail: inacio.bo@wzb.eu; phone: +49 (0)30-25491-292. · Humboldt-Universität zu Berlin and WZB Berlin Social Science Center; e-mail: hellerkr@hu-berlin.de.

1 Introduction

Having their children attend a good school is an important concern for many parents. Until recently, in many countries parents did not have much choice when it came to choosing the public school their children would attend, as most children were administratively assigned to a school nearby. In their seminal contribution Abdulkadiroğlu and Sönmez (2003) analyzed school assignment procedures in Boston and other cities in the U.S., and suggested the student-proposing deferred acceptance mechanism (DA, Gale and Shapley, 1962) and the top trading cycles mechanism (TTC, Roth and Postlewaite, 1977) as methods to allocate students to schools. One argument in favor of the DA and TTC over the mechanism used in Boston at that time (referred to commonly as the *Boston mechanism*) is that reporting preferences truthfully under these mechanisms is a dominant strategy. Under the Boston mechanism, on the other hand, students may obtain better assignments by strategically manipulating the preferences over the schools they report.

The Boston mechanism lets each student apply to schools in the order of their reported preferences. Schools in each round consider all applications received in that round, accept the applications from the highest-ranked students according to the schools' preferences or some exogenous ranking over students, until their capacities are filled, and reject the remaining applicants. In subsequent rounds, rejected students apply to their next highest-ranked school. The schools which still have seats available again accept the applications from the highest-ranked students until their capacities are filled.

This paper revisits the question of the possibility of manipulations by schools under the Boston mechanism. Ergin and Sönmez (2006) showed that under the Boston mechanism it is a dominant strategy for schools to truthfully rank the students. This paper shows that for this result to hold, schools need to be restricted to finding all students acceptable, and that students cannot observe the rankings set by the schools before submitting their preferences. If schools are allowed to deem students unacceptable, we show that doing so may yield a better matching from the schools' point of view. In addition, Ergin and Sönmez (2006) showed that, taking schools' rankings as given, the set of Nash equilibrium outcomes under the Boston mechanism equals the set of stable matchings. Since the (student-proposing) DA mechanism always yields the student-optimal stable matching, this implies that the Boston mechanism is Pareto dominated by DA if all students behave strategically. We show that if schools are strategic and can deem students unacceptable then there may be Nash equilibria in undominated strategies whose outcomes are not stable with respect to the true preferences (and capacities) of students and schools. Moreover, we show that if students are allowed to observe the rankings set by the schools prior to submitting their preferences, it is also the case that schools may have an incentive to manipulate the rankings over students or their capacities, even if they cannot declare any of them to be unacceptable, and that the set of subgame perfect Nash equilibria of the induced game may contain unstable matchings.

We also provide a full characterization of the set of undominated strategies for both students and schools under the game induced by the Boston mechanism. Students' undominated strategies never involve declaring an acceptable school as unacceptable, but may change the relative ranking among acceptable schools. In contrast, schools may have incentives to declare some acceptable students as unacceptable, but never find it optimal to deviate from truthfully revealing their relative preferences over the students they declare acceptable.

The Boston mechanism and its properties have been analyzed by many papers in the literature, starting from Roth (1991), which studies a variety of matching procedures for regional medical labor markets in the UK. While some regions used stable mechanisms (Edinburgh and Cardiff) others use priority mechanisms (Newcastle, Birmingham, and Edinburgh), of which the Boston mechanism is a special case.

While the use of the Boston mechanism has been criticized beginning with Abdulkadiroğlu and Sönmez (2003) due to the possibility of gains from manipulation by students, a number of authors show that it has some desirable properties. Abdulkadiroğlu, Che and Yasuda (2011) and Miralles (2009) both argue in favor of the Boston mechanism on the grounds of *ex ante* cardinal efficiency. In particular, when schools' rankings are random and uniform and students rank schools similarly but prefer them with different intensities, the equilibrium outcome under the Boston mechanism yields higher expected welfare than the deferred acceptance mechanism. The reason is that students who put a particularly high cardinal value on a school are more likely to rank this school highly and thereby obtain a seat there. Hence, the equilibrium under the Boston mechanism makes use of information on preference intensities, while the deferred acceptance mechanism does not.

Pathak and Sönmez (2008) analyze the Boston mechanism when there are two types of students with different levels of strategic sophistication, while schools' rankings are taken as fixed. They show that sophisticated students are better off under the Boston mechanism than naïve students. For these authors this justifies the move toward non-manipulable mechanisms. Kojima and Ünver (2014) give two characterizations of the Boston mechanism while allowing for schools to deem some students unacceptable.

Mennle and Seuken (2014) distinguish between the 'naïve' Boston mechanism and the 'adaptive' Boston mechanism. They differ in that under the adaptive Boston mechanism students in later rounds never apply to schools whose slots were filled in earlier rounds. The mechanism designer thereby effectively optimizes the students' strategies on their behalf. The adaptive Boston mechanism additionally involves randomly drawing schools' rankings over the students. As such, there is no scope for strategic behavior by the schools in their paper. Dur (2015) considers the modified Boston mechanism which also involves students never applying to schools that were filled in a previous round. Unlike Mennle and Seuken (2014) he allows for schools' rankings to be exogenously given, rather than randomly generated. The author also does not consider a strategic role for the schools. Dur (2015) shows that the modified Boston mechanism is less manipulable than the Boston mechanism in the sense of Pathak and Sönmez (2013) and that the set of Nash equilibria induced by the (complete information) preference revelation game equals the set of stable matchings under the true preferences.

Pais and Pintér (2008) experimentally compare the performance of the Boston, DA, and TTC mechanisms in terms of efficiency and manipulability. As expected from the theory, the frequency of manipulation under the Boston mechanism is greater than under the other two strategy-proof mechanisms, especially when participants (taking the role of the students) are given more information.

Schools' ability to independently determine rankings over students is present in many school choice procedures currently being used. Procedures for matching students to elementary schools in Ireland (Chen, 2016) and for secondary education in Amsterdam (De Haan, Gautier, Oosterbeek and Van der Klaauw, 2015) and Berlin (Basteck, Huesmann and Nax, 2016), for example, explicitly allow for schools to determine, sometimes subject to approval by the district school board, the criteria to be used to select students when demand exceeds the number of seats. Other than Ergin and Sönmez (2006), however, to our knowledge only two other papers consider the question of the manipulation of the Boston mechanism by schools.

Kojima (2008) generalized the earlier results of Ergin and Sönmez (2006) to a model in which restrictions on the possible preferences of the schools are relaxed. Treating schools' preferences as given, he shows that if schools' preferences satisfy a substitutability condition then the Nash equilibrium outcomes under the Boston mechanism are stable. He also shows that stable matchings can be supported by Nash equilibria under more general preference structures. He further provides an example in which a school may profitably manipulate the Boston mechanism when its preferences satisfy substitutability but not responsiveness. Since we show that schools with responsive preferences may profitably manipulate the Boston mechanism by declaring some student unacceptable, this result is implied by our paper. More specifically, our result shows that it is not necessary to extend schools' preferences beyond responsiveness to obtain incentives for manipulation.

Ehlers (2008) considers manipulations of priority mechanisms (manipulations by schools under the Boston mechanism being a special case of them) under incomplete information. The author shows that when agents have symmetric (incomplete) information, any non-truncation strategy is stochastically dominated by a truncation[1] of the true preferences of that agent. That does not imply our results, however, since it relies on his specific assumptions over beliefs. This strategic behavior under symmetric information is explored experimentally by Featherstone and Mayefsky (2015). Finally, the game induced by schools' ability to manipulate their capacities is analyzed in a series of papers for the case of stable mechanisms (Ehlers, 2010; Konishi and Ünver, 2006;

[1] A truncation strategy leaves the true preference over students unchanged, but might drop some acceptable students.

Romero-Medina and Triossi, 2013). None of these results, however, imply our results on capacity manipulation.

We introduce our model and the variants of the Boston mechanism that we consider in section 2. Results concerning manipulability by schools and the stability of the Nash equilibria under the simultaneous Boston mechanism are obtained in section 3. The sequential Boston mechanism is analyzed in section 4. Proofs absent from the main text are in the appendix.

2 Model

A **two-sided matching market** consists of:

1. A finite set of **students** $I = \{i_1, \ldots, i_n\}$,
2. A finite set of **schools** $S = \{s_1, \ldots, s_m\}$,
3. A **capacity vector** $q = (q_{s_1}, \ldots, q_{s_m})$,
4. A list of strict **student preferences** $P_I = (P_{i_1}, \ldots, P_{i_n})$ and
5. A list of strict **school preferences** $P_S = (P_{s_1}, \ldots, P_{s_m})$.

We assume that $n \geqslant 2$ and $m \geqslant 2$. Preference relations P_i for students are over the set of schools and the option of remaining unmatched, that is, $S \cup \{\varnothing\}$. For $s \neq s' \in S$ if sP_is' we say that student i strictly prefers school s to school s'. Preference relations of the schools, P_s, are over sets of students. For $J \neq J' \subseteq I$ if JP_sJ' we say that school s prefers having students J over having students J'. Preferences are complete and transitive. Capacities are positive integers. For each school $s \in S$ and any positive integer q_s the preference relation P_s is **responsive with capacity** q_s if for any set of students $J \subset I$ with $|J| < q_s$ and students $i, i' \notin J$, $\{i\} P_s \{i'\}$ if and only if $J \cup \{i\} P_s J \cup \{i'\}$, $J \cup \{i\} P_s J \iff \{i\} P_s \varnothing$ and $\varnothing P_s J$ for every $J \subset I$ with $|J| > q_s$. For most of our results, the knowledge of the schools' preferences over individual students, as opposed to sets of students, is sufficient. Therefore, in the absence of ambiguity, we consider schools' preferences to be represented by the former type. Let $\mathcal{P}_I$ be the set of all possible strict preferences over schools, and the option of remaining unassigned, and $\mathcal{P}_S$ be the set of all possible responsive preferences, with capacities q, over sets of students. Moreover, for a student i, let $\mathcal{P}_{-i}$ be the set of all possible strict preference profiles over schools for students $I \backslash \{i\}$, and $\mathcal{P}_{-s}$ and q_{-s} defined accordingly for school s. We use $P_S = (P_s, P_{-s})$. From here on we abuse notation slightly by denoting singleton sets by i or s. We say that student i is **unacceptable** according to P_s for school s if $\varnothing P_s i$. Unacceptable schools are defined analogously for students. We assume that every school s has at least q_s acceptable students and that each student has at least one acceptable school. This is a very mild assumption, used for technical purposes. For any agent $x \in I \cup S$ and her preference P_x, we define the corresponding weak preference relation R_x, where $cR_xc' \iff cP_xc'$ or $c = c'$.

A **matching** μ is a function from $I \cup S$ to subsets of $I \cup S$ such that:

– $\mu(i) \in S \cup \{\varnothing\}$ and $|\mu(i)| = 1$ for every student i,[2]

[2] We abuse notation and consider $\mu(i)$ as an element of S, instead of a set with an element of S.

- $\mu(s) \subseteq I$ and $|\mu(s)| \leqslant q_s$ for every school s,
- $\mu(i) = s$ if and only if $i \in \mu(s)$.

We allow for preferences to also be defined over matchings: $\mu P_i \mu' \iff \mu(i) P_i \mu'(i)$. The set of matchings is denoted by $\mathcal{M}$. A matching is **individually rational** if for every student i, $\mu(i) P_i \varnothing$ and for every school s and every student $i' \in \mu(s)$, $i' P_s \varnothing$. A matching μ is **blocked** by a student i and school s if $s P_i \mu(i)$ and there is a set $I' \subseteq \mu(s) \cup \{i\}$ such that $i \in I'$ and $I' P_s \mu(s)$. A matching μ is **stable** with respect to (P_I, P_S, q) if it is individually rational and is not blocked under those preferences and capacities. Often we will simply say that a matching is stable. In that case we mean that it is stable with respect to the true student preferences and school preferences and capacities. A **(school choice) mechanism** Ψ is a mapping from the set of students' preferences, preferences (or rankings) over students, and schools' capacities to the set of matchings, i.e., $\Psi : \mathcal{P}_{\mathcal{I}} \times \mathcal{P}_S \times \mathbb{N}_+^m \to \mathcal{M}$. A mechanism is stable if it yields a stable matching for every profile of students' preferences and schools' preferences and capacities.

A mechanism Ψ is **manipulable** by schools if there is a school s and school preferences $P_s, P'_s \in \mathcal{P}_s$, capacities q_s, q'_s such that $q'_s \leqslant q_s$, $P_I \in \mathcal{P}_I$ and $P_{-s} \in \mathcal{P}_{-s}$, q_{-s}such that:

$$\Psi\left(P_I, \left(P'_s, P_{-s}\right), \left(q'_s, q_{-s}\right)\right) P_s \Psi\left(P_I, P_S, q\right)$$

We distinguish between three types of manipulations by schools. First, a mechanism Ψ is **manipulable by declaring students unacceptable** if it is manipulable by a pair (P'_s, q'_s) such that $q'_s = q_s$ and there exists some $i \in I$ such that $i P_s \varnothing$ and $\varnothing P'_s i$. Second, a mechanism Ψ is **manipulable by a ranking change** if it is manipulable by a pair (P'_s, q'_s) such that $q'_s = q_s$ and for all $i \in I$, $i P_s \varnothing$ implies $i P'_s \varnothing$. Third, a mechanism Ψ is **manipulable via capacities** if it is manipulable by a pair (P'_s, q'_s) such that $P'_s = P_s$ and $q'_s < q_s$. If a mechanism is not manipulable by schools we say that for that mechanism truth-telling is a dominant strategy for the schools. Manipulability by students is defined analogously, except students do not report capacities.

For a given mechanism Ψ and a school s with true preference and capacity (P_s, q_s) we say that a strategy $(\tilde{P}_s, \tilde{q}_s)$ with $\tilde{q}_s \leqslant q_s$ is **dominated** by strategy (P'_s, q'_s), with $q'_s \leqslant q$, if for all P_I, P_{-s}, q_{-s} it holds that

$$\Psi\left(P_I, \left(P'_s, P_{-s}\right), \left(q'_s, q_{-s}\right)\right) R_s \Psi\left(P_I, \left(\tilde{P}_s, P_{-s}\right), \left(\tilde{q}_s, q_{-s}\right)\right)$$

Notice that, in principle, if two strategies give exactly the same outcome for all strategies of the others, both would be dominated. In the context analyzed in this paper, however, that never happens: no two different strategies will yield the same outcome for all strategies of the others. If there is no strategy that dominates $(\tilde{P}_s, \tilde{q}_s)$, we say that $(\tilde{P}_s, \tilde{q}_s)$ is **undominated**. Undominated strategies for students are analogously defined. If a mechanism is not manipulable then it follows that reporting preferences truthfully is undominated.

For manipulable mechanisms one cannot rely on the truthful revelation of preferences by all agents. Instead, we will examine the induced complete information preference revelation game and consider its Nash equilibria.

Definition 1. A strategy profile $(\tilde{P}_I, \tilde{P}_S, \tilde{q})$, such that for all schools $s \in S$, $\tilde{q}_s \leqslant q_s$, is a **Nash equilibrium** of the game induced by the mechanism Ψ under (P_I, P_S, q) if the following conditions hold:
(i) for all $s \in S$, $\hat{q}_s \leqslant q_s$ and $\hat{P}_s \in \mathcal{P}_s$, it holds that:

$$\Psi\left(\tilde{P}_I, \tilde{P}_S, \tilde{q}\right) R_s \Psi\left(\tilde{P}_I, \left(\hat{P}_s, \tilde{P}_{-s}\right), (\hat{q}_s, \tilde{q}_{-s})\right)$$

(ii) for all $i \in I$ and $\hat{P}_i \in \mathcal{P}_i$, it holds that:

$$\Psi\left(\tilde{P}_I, \tilde{P}_S, \tilde{q}\right) R_i \Psi\left(\left(\hat{P}_i, \tilde{P}_{-i}\right), \tilde{P}_S, \tilde{q}\right)$$

We say that $\Psi\left(\tilde{P}_I, \tilde{P}_S, \tilde{q}\right)$ is a **Nash equilibrium outcome** for preferences (P_I, P_S) and capacity q if $\left(\tilde{P}_I, \tilde{P}_S, \tilde{q}\right)$ is a Nash equilibrium of the game induced by the mechanism Ψ under (P_I, P_S, q).

For some results, we consider allowing students to observe reports by the schools before submitting their own preferences. In that case the students' strategy would no longer be the choice of a preference ordering but a preference ordering for each possible combination of preferences and capacities reported by the schools. We denote by $f_i : \mathcal{P}_S \times \mathbb{N}_+^m \to \mathcal{P}_i$ a strategy for a student in a preference revelation game in which schools move first. Let $\mathcal{F}_i$ be the set of strategies for student i and $\mathcal{F}_\mathcal{I}$ be the set of all possible strategy profiles for the students. We say that a mechanism Ψ is a **sequential (school choice) mechanism** if students can observe the schools' reported preferences before themselves simultaneously submitting preferences.[3] To analyze schools' incentives in a sequential mechanism, we restrict our attention to students' strategies that are optimal against the other students' strategies given the submitted preferences and capacities of the schools. In other words, given the schools' reports the students are assumed to play Nash equilibrium strategies. Since students' strategies allow their reported preferences to arbitrarily depend on preferences and capacities reported by the schools, there are many unreasonable strategies for the students that would create artificial incentives for the schools to deviate from truth-telling.[4] Our restriction makes our results

[3] There is an alternative concept of sequential mechanisms in the school choice literature due to Dur and Kesten (2014). In their analysis schools are not strategic agents. There are two sets of schools whose seats are filled sequentially. In the first round students are matched to one of the schools in the first set, based solely on their preferences over those schools. In the second round, students who were left unmatched in the first round are matched to the second set of schools. In their case "sequential" thus refers to sequentially making an allocation decision. This allows us to use different matching rules, such as the Boston mechanism, top trading cycles or deferred acceptance for different rounds. In contrast, "sequential" in our paper refers to schools submitting their preferences before the students with a fixed mechanism.

[4] For example, all students could declare all schools unacceptable if only at least one school is truthful but otherwise report preferences truthfully. In that case there is at least one school which

on manipulations by schools stronger and more relevant, and manipulations by schools would only be profitable when students' strategies are "reasonable."

Definition 2. A strategy profile of students $f_I \in \mathcal{F}_I$ is **sequentially rational** (with respect to a sequential mechanism Ψ) if for all $\left(\tilde{P}_S, \tilde{q}\right) \in \mathcal{P}_S \times \mathbb{N}_+^m$ and for all $i \in I$ and $\hat{f}_i \in \mathcal{F}_i$ it holds that:

$$\Psi\left(f_I\left(\tilde{P}_S, \tilde{q}\right), \tilde{P}_S, \tilde{q}\right) R_i \Psi\left(\left(\hat{f}_i\left(\tilde{P}_S, \tilde{q}\right), f_{-i}\left(\tilde{P}_S, \tilde{q}\right)\right), \tilde{P}_S, \tilde{q}\right)$$

Definition 3. A sequential mechanism Ψ is **sequentially manipulable by schools** if there is a school s, a school preference $P'_s \in \mathcal{P}_s$, a capacity $q'_s \in \mathbb{N}_+^m$ such that $q'_s \leqslant q_s$, and a sequentially rational $f_I \in \mathcal{F}_I$ such that:

$$\Psi\left(f_I\left(\left(P'_s, P_{-s}\right), \left(q'_s, q_{-s}\right)\right), \left(P'_s, P_{-s}\right), \left(q'_s, q_{-s}\right)\right) P_s \Psi\left(f_I\left(P_S, q\right), P_S, q\right)$$

The above definition of manipulability of a sequential mechanism restricts the set of strategies that students can play to those that are optimal for all possible reports made by the schools and strategies chosen by the other students. More importantly, if a mechanism is sequentially manipulable by schools, it is not a dominant strategy for schools to submit their true preferences and/or capacities. While this definition is not standard, it captures the notion that schools can manipulate the sequential Boston mechanism even if students react in a rational way to their manipulations.[5] The definitions of **sequentially manipulable by declaring students unacceptable, sequentially manipulable by a ranking change** and **sequentially manipulable via capacities** are straightforwardly derived from the ones for the simultaneous game.

Definition 4. A profile $\left(f_I, \tilde{P}_S, \tilde{q}\right)$ is a **subgame perfect Nash equilibrium** (SPNE) of a mechanism Ψ under (P_I, P_S, q) if the following conditions hold:

(i) for all $s \in S$, $\tilde{q}_s \leqslant q_s$, $\hat{P}_s \in \mathcal{P}_s$ and $\hat{q}_s \leqslant q_s$ we have:

$$\Psi\left(f_I\left(\tilde{P}_S, \tilde{q}\right), \tilde{P}_S, \tilde{q}\right) R_s \Psi\left(f_I\left(\left(\hat{P}_s, \tilde{P}_{-s}\right), \left(\hat{q}_s, \tilde{q}_{-s}\right)\right), \left(\hat{P}_s, \tilde{P}_{-s}\right), \left(\hat{q}_s, \tilde{q}_{-s}\right)\right)$$

(ii) for all $i \in I$, for all $\hat{P}_S \in \mathcal{P}_S$, for all $s \in S$, $\hat{q}_s \leqslant q_s$ and for all $\hat{f}_i \in \mathcal{F}_i$ we have:

$$\Psi\left(f_I\left(\hat{P}_S, \hat{q}\right), \hat{P}_S, \hat{q}\right) R_i \Psi\left(\left(\hat{f}_i\left(\hat{P}_S, \hat{q}\right), f_{-i}\left(\hat{P}_S, \hat{q}\right)\right), \hat{P}_S, \hat{q}\right)$$

would gain by deviating from truth-telling. In our view, such strategies for the students can be considered "unreasonable."

[5] The idea behind this definition is not that it is a property that is necessarily of independent interest. Rather it is a definition that allows us to precisely discuss how the Boston mechanism may give schools an incentive to misrepresent their preferences if students observe the schools' reports before submitting theirs.

Notice that one can use the assumption of sequential rationality without considering the SPNE of the game. As we will show in section 4, that assumption allows for a more meaningful analysis of the incentives that the mechanism induces in the schools.

We focus here on one particular mechanism, which was used in the Boston Public School Match before it was changed to the student-proposing deferred acceptance mechanism of Gale and Shapley (1962) in 2005 (Abdulkadiroğlu, Pathak, Roth and Sönmez, 2005a). Many other cities have used similar mechanisms to allocate school seats to students.

The Boston mechanism: Each student and school reports a preference order over each other and the option of being unmatched. Schools additionally report their capacities. Students and schools are matched in rounds.

- **Round 1:** Each student applies to the highest-ranked school according to the reported preference order. Each school accepts the students who applied to it in order of their ranking according to the preference order reported by the school until there are either no more students who applied to the school or the school reaches its capacity. Students whose applications are unsuccessful are rejected.
- **Round** $k \geqslant 2$**:** Each student who was rejected in round $k-1$ applies to her k highest-ranked school if it is acceptable to the student. Otherwise the student is assigned to the outside option. Each school with remaining spots accepts students who applied to it in round k in order of their ranking according to the preference order reported by the school until there are either no more students who applied or the school has reached its capacity. Students whose applications are unsuccessful are rejected.

The procedure ends when all students are either assigned a spot at a school or the outside option.

The Boston mechanism as described above will be referred to as the **two-sided simultaneous Boston mechanism**. The reason is that both sides of the market, students and schools, are taken as strategic agents who report preferences simultaneously. A commonly analyzed variant is the **one-sided Boston mechanism** in which only the students are seen as strategic agents and the schools' preferences and capacities are seen as administrative rankings which are simply observed or directly chosen by the market designer. We further consider the **two-sided sequential Boston mechanism,** which only differs from the two-sided simultaneous Boston mechanism in that students observe the preferences and capacities reported by the schools before reporting their own preferences. It is often more realistic to suppose that schools move first in school choice mechanisms since schools' preferences are often set in advance and communicated to prospective applicants. For example, a school could indicate that it accepts students according to their residence location or a specified weighted average of the students' grades in an exam. Since we are concerned with the strategic behavior of schools under the Boston mechanism, we focus on the two-sided variants of the Boston mechanism, although we will make use of some existing results of the one-sided Boston mechanism.

In a setting in which schools are restricted to reporting preferences such that every student is acceptable,[6] Ergin and Sönmez (2006) prove the following results:

Theorem. **(Theorem 2 in Ergin and Sönmez, 2006)** In the two-sided (*simultaneous*) version of the Boston mechanism, it is a dominant strategy for any school s to rank students based on its true preferences P_s. Moreover, any other dominant strategy of school s is outcome equivalent to truthfully ranking students based on P_s.

Theorem. **(Theorem 1 in Ergin and Sönmez, 2006)** Let P_I be the list of true student preferences, and consider the preference revelation game induced by the (*one-sided*) Boston mechanism. The set of Nash equilibrium outcomes of this game is equal to the set of stable matchings under the true preferences P_I.

Corollary. **(Corollary 1 in Ergin and Sönmez, 2006)** In the two-sided version *(simultaneous)* of the Boston mechanism, the set of Nash equilibrium outcomes in undominated strategies is equal to the set of stable matchings under the true preferences.

In the next section we show that if schools are allowed to declare students as being unacceptable, the result in Theorem 2 in **Ergin and Sönmez, 2006** no longer holds. Combining the results from Theorem 2 and 1 in **Ergin and Sönmez, 2006** leads to their Corollary 1. Our results show that there may be Nash equilibria in undominated strategies of the two-sided simultaneous Boston mechanism that are not stable. However, it is still possible to support each stable matching with some Nash equilibrium in undominated strategies.

3 The Two-Sided Simultaneous Boston Mechanism

First, we show that schools may improve the set of students who are matched to them by declaring some students unacceptable:

Proposition 1. *The simultaneous two-sided Boston mechanism is manipulable by declaring students unacceptable.*

Proof Consider the following two-sided matching market:

$$\begin{array}{rl} I = \{i_1, i_2, i_3\} & S = \{s_1, s_2\},\ q_1 = q_2 = 1 \\ P_{i_1} : s_2\ s_1 & P_{s_1} : i_1\ i_2\ i_3 \\ P_{i_2} : s_1\ s_2 & P_{s_2} : i_3\ i_1\ i_2 \\ P_{i_3} : s_2\ s_1 & \end{array}$$

The outcome of the Boston mechanism when students and schools submit their preferences (and capacities) truthfully is μ:

[6] This rules out manipulating by declaring students unacceptable.

$$\mu = \begin{pmatrix} s_1 & s_2 & \varnothing \\ i_2 & i_3 & i_1 \end{pmatrix}$$

If school s_1 submits the preference $P'_{s_1} : i_1 \; i_3$ and the same capacity, while the strategies of the other agents remain unchanged, the outcome of the Boston mechanism is μ':

$$\mu' = \begin{pmatrix} s_1 & s_2 & \varnothing \\ i_1 & i_3 & i_2 \end{pmatrix}$$

Since s_1 prefers student i_1 to i_2, school s_1 gains from the manipulation.

The rationale behind the manipulation of school s_1 is as follows. In the first round of the Boston mechanism only student i_2 applies to it. However, in the next round there will be an application by student i_1, who school s_1 prefers to i_2. By declaring i_2 unacceptable school s_1 can prevent student i_2 from taking its only seat and can then accept student i_1 in the second round.

Manipulating by declaring some students unacceptable appears to have been featured in some real-life markets. Consider, for example, the school assignment procedure used in New York City before the change to the student-proposing deferred acceptance mechanism of Gale and Shapley (1962) in 2003. Before the change, there was no central authority coordinating the assignment. Strictly speaking, there was no use of the Boston mechanism in New York. However, the procedure used is roughly comparable to the Boston mechanism. Students could send a letter to up to five schools in the first round. Schools, upon receiving an application could decide whether to accept the student, put her on a waiting list or reject the student (Abdulkadiroğlu, Pathak and Roth, 2005b). After the change to a new mechanism it was reported that: “Before you might have a situation where a school was going to take 100 new children for ninth grade, they might have declared only 40 seats, and then placed the other 60 outside of the process” (*New York Times*, November 19, 2004). While it may appear as though schools may have manipulated capacities, this is not the case.[7] By declaring only 40 seats and then filling another 60 seats at a later stage, the schools effectively declared some students as unacceptable. Furthermore, the type of manipulation that has occurred suggests that the incentives to manipulate that we identified in Proposition 1 are of empirical relevance and importance. This suggests that the incentive to strategically withhold seats in order to accept more preferred applicants at a later stage, which we have identified under the Boston mechanism, is practically relevant. Another method by which schools could declare students as unacceptable is by setting grade thresholds below which students are not accepted or by setting other admission requirements. Both of these manipulations may be argued to be due to an objective necessity of students satisfying such requirements.

[7] When manipulating capacities, a mechanism cannot assign more students to a school than its stated capacities. In New York schools clearly received more students than was possible in their initially declared capacity quota.

3.1 Undominated Strategies

Before we analyze the equilibria of the simultaneous move game induced by the Boston mechanism we describe the undominated strategies of this game. The reason for our focus on undominated strategies is that this game, like most two-sided matching games, has many Nash equilibria. For example, every student and school declaring the option of being unmatched as the most preferred is always a Nash equilibrium, irrespective of preferences and capacities. Such equilibria are not reasonable, so we will additionally require that both students and schools play undominated strategies. Reporting that being unmatched is the most-preferred outcome is clearly a dominated strategy for any underlying preferences that have at least one acceptable school or student, respectively. Thus, the focus on undominated strategies ensures that very unreasonable Nash equilibria will be ruled out. We begin by showing some properties of the set of undominated strategies for schools.

Lemma 1. *Under the two-sided simultaneous Boston mechanism, the strategy* $(\tilde{P}_s, \tilde{q}_s)$ *is an undominated strategy for school* s *with true preferences and capacity* (P_s, q_s) *if and only if the following conditions hold:*

(i) $i\tilde{P}_s\varnothing$ *only if* $iP_s\varnothing$.
(ii) $\tilde{q}_s = q_s$.
(iii) *If* $i, j\tilde{P}_s\varnothing$ *then* $i\tilde{P}_sj \Leftrightarrow iP_sj$.
(iv) *If* i *is the most-preferred student under* P_s*, then* $i\tilde{P}_s\varnothing$.
(v) *There are at least* q_s *acceptable students under* $\tilde{P}_s$.

Strategies that satisfy conditions *(i)*, *(ii)* and *(iii)* are denoted *dropping strategies* in Kojima and Pathak (2009). Lemma 1 says that undominated strategies by the schools may involve declaring some students unacceptable, but always being truthful about the relative ranking of acceptable students and never dropping the most-preferred student. In the proof, which can be found in the appendix, we see that a strategy of a school which reports an incorrect (with respect to the true preferences) relative ranking over two acceptable students is dominated by truthfully ranking them. This implies that ranking changes as a manipulation strategy are dominated by truth-telling. Undominated strategies also involve declaring at least q_s students as acceptable, which is a consequence of the assumption that $q_s \leqslant n$.

Undominated strategies also involve no capacity manipulation. Any strategy that involves capacity manipulation is dominated by keeping the reported preferences the same, while being truthful about capacities. Hence, we have the following corollary:

Corollary 1. *The two-sided simultaneous Boston mechanism is not manipulable via capacities. The two-sided simultaneous Boston mechanism is not manipulable via ranking changes.*

The intuition behind these results is similar to that behind Theorem 1 in Ergin and Sönmez (2006). By reporting a capacity below the true capacity, a school will still receive the same sequence of applications but it can accept only a subset of the applicants. Therefore, there is no opportunity to gain from pretending to have a smaller capacity. By changing the reported preferences over students, schools may sometimes obtain a set of students that are worse than those they would have received under truth-telling. One consequence of this is that if the schools can only manipulate by a ranking change then the set of Nash equilibrium outcomes in undominated strategies of the two-sided simultaneous Boston mechanism will equal the set of stable matchings.

Since a school's true preferences satisfy conditions (i)-(v) of Lemma 1, truth-telling is an undominated strategy for the schools.

Corollary 2. *Under the two-sided simultaneous Boston mechanism, truth-telling is an undominated strategy for schools.*

We next characterize undominated strategies for the students.

Lemma 2. *Under the two-sided simultaneous Boston mechanism $\tilde{P}_i$ is an undominated strategy for student i with true preferences P_i if and only if for all schools s we have $sP_i\varnothing \Leftrightarrow s\tilde{P}_i\varnothing$.*

The Lemma above implies that for some given preferences of some student i, any other reported preference that has the same set of acceptable schools, is an undominated strategy. This implies that the set of undominated strategies for two students with different preferences is the same if and only if both consider the same set of schools acceptable. Like for schools, the true preferences satisfy the condition in Lemma 2, so we have the following corollary.

Corollary 3. *Under the two-sided simultaneous Boston mechanism, truth-telling is an undominated strategy for students.*

The type of undominated strategy differs for students and schools. Students' undominated strategies can be described as being all strategies that are truthful about which schools are acceptable, but perhaps untruthful about the relative preference for acceptable students. In contrast, undominated strategies for the schools involve being truthful about the relative preferences for students that are declared acceptable. However, schools' undominated strategies may involve not being truthful about which students are acceptable.

3.2 Nash Equilibria in Undominated Strategies under the Two-Sided Simultaneous Boston Mechanism

Before analyzing the properties of Nash equilibria in undominated strategies, we show that under the two-sided simultaneous Boston mechanism, such equilibria always exist.

Theorem 1. *For every stable matching μ, the game induced by the two-sided simultaneous Boston mechanism has a pure strategy Nash equilibrium in undominated strategies with outcome μ.*

Theorem 1 shows that some flavors of the Theorem 1 in Ergin and Sönmez (2006) remain in our setting, since every stable matching can be supported as a Nash equilibrium in undominated strategies under the two-sided simultaneous Boston mechanism. However, the reverse does not hold: there exist Nash equilibria in undominated strategies that are not stable.

Theorem 2. *The set of Nash equilibrium outcomes in undominated strategies of the game induced by the two-sided simultaneous Boston mechanism may contain unstable matchings. Moreover, the resulting equilibrium outcome may by weakly preferred to all stable matchings by the schools.*

Proof Consider the following two-sided matching market:

$$\begin{array}{rl} I = \{i_1, i_2, i_3, i_4\} & S = \{s_1, s_2, s_3\}, \; q_1 = 2, \; q_2 = q_3 = 1 \\ P_{i_1} : s_2 \; s_1 \; s_3 & P_{s_1} : i_1 \; i_2 \; i_3 \; i_4 \\ P_{i_2} : s_1 \; s_2 & P_{s_2} : i_2 \; i_1 \; i_3 \\ P_{i_3} : s_2 \; s_3 \; s_1 & P_{s_3} : i_4 \; i_1 \; i_3 \\ P_{i_4} : s_1 \; s_3 & \end{array}$$

The strategy profile $(\tilde{P}_I, \tilde{P}_S, (2, 1, 1))$ is a Nash equilibrium, where:

$$\begin{array}{rl} \tilde{P}_{i_1} : s_2 \; s_1 \; s_3 & \tilde{P}_{s_1} : i_1 \; i_3 \\ \tilde{P}_{i_2} : s_2 \; s_1 & \tilde{P}_{s_2} : i_2 \; i_1 \; i_3 \\ \tilde{P}_{i_3} : s_2 \; s_3 \; s_1 & \tilde{P}_{s_3} : i_4 \; i_1 \; i_3 \\ \tilde{P}_{i_4} : s_3 \; s_1 & \end{array}$$

The outcome in this Nash equilibrium under the Boston mechanism is the matching μ:

$$\mu = \begin{pmatrix} s_1 & s_2 & s_3 \\ i_1, i_3 & i_2 & i_4 \end{pmatrix}$$

To see that this strategy profile is a Nash equilibrium, consider the deviations that schools and students could have:

- Student i_1 cannot be accepted by s_2 at the first step, since i_2 is also applying there at that step and has a higher ranking. Since there are no other seats at s_2, i_1 cannot profitably deviate.
- Student i_2 is not acceptable to school s_1 under $\tilde{P}_{s_1}$, and therefore cannot profitably deviate.
- Student i_3 cannot be accepted at school s_3, since i_4 is applying there at the first step and has higher ranking. She also cannot be accepted at s_2, since she has a lower ranking than i_1 and i_2, who apply there at the first step. Therefore, i_3 also cannot profitably deviate.

– Student i_4 is not acceptable at school s_1, and therefore cannot profitably deviate.
– School s_1 would only be able to be better off if it got students i_1 and i_2. Since i_2 ranked school s_2 first and is accepted in the first step, the report s_1 makes does not affect where i_2 is allocated.
– Schools s_2 and s_3 have their most-preferred students, and so have no incentive to deviate.

The matching μ is not stable since school s_1 and student i_2 form a blocking pair. To show that μ is weakly preferred to all stable matchings by the schools, it suffices to show that μ is preferred to the school-optimal stable matching μ^S:

$$\mu^S = \begin{pmatrix} s_1 & s_2 & s_3 \\ i_2, i_3 & i_1 & i_4 \end{pmatrix}$$

Given the schools' preferences, $\mu(s_1) P_{s_1} \mu^S(s_1)$, $\mu(s_2) P_{s_2} \mu^S(s_2)$ and $\mu(s_3) R_{s_3} \mu^S(s_3)$.

From Lemma 1 and Corollary 2 it follows that schools' strategies are undominated. From Lemma 2 it follows that students' strategies are undominated.

While not every Nash equilibrium outcome in undominated strategies of the two-sided simultaneous Boston mechanism is stable, we can be more precise about what type of instabilities can appear. The following result shows that in Nash equilibrium outcomes in undominated strategies under the Boston mechanism, no unmatched student is part of a blocking pair.

Proposition 2. *Let μ be the outcome of a Nash equilibrium $(\tilde{P}_I, \tilde{P}_S, q)$ in undominated strategies under the two-sided simultaneous Boston mechanism. If some student i is part of a blocking pair of μ under the true preferences and capacities (P_I, P_S, q)* then $\mu(i) \in S$.

Proof Take some matching μ that is the outcome of a Nash equilibrium $(\tilde{P}_I, \tilde{P}_S, q)$ in undominated strategies under the two-sided simultaneous Boston mechanism. For a contradiction suppose that student i and school s block μ under their true preferences and that $\mu(i) = \varnothing$. This means that both are acceptable to each other and either *(a)* $|\mu(s)| < q_s$ and $i P_s \varnothing$, or *(b)* $|\mu(s)| = q_s$ and there exists $j \in \mu(s)$ such that $i P_s j$.

Suppose that *(a)* holds. Since μ is the result of a Nash equilibrium in undominated strategies, from Lemma 2 it follows that $\tilde{P}_i$ must have declared s to be acceptable. Because i is not matched to any school under μ, there is some round of the Boston mechanism in which i applies to school s and is rejected. From Lemma 1 it follows that s truthfully stated its capacities, so that i could only have been rejected by s if $\varnothing \tilde{P}_s i$. However, since s does not fill its capacity, this implies that declaring i as acceptable is a profitable deviation. This contradicts $(\tilde{P}_I, \tilde{P}_S, q)$ being a Nash equilibrium. So condition *(a)* cannot hold.

Suppose that *(b)* holds. If $i \tilde{P}_s \varnothing$ there is a contradiction. Since i is acceptable under $\tilde{P}_s$ and since undominated strategies satisfy $i \tilde{P}_s j \Leftrightarrow i P_s j$ if $i, j \tilde{P}_s \varnothing$

(Lemma 1), i can profitably deviate by ranking s first. This is because either fewer than q_s students apply to schools s in the first round of Boston or because there is some student j that applies in the first round such that $i\tilde{P}_s j$ by condition *(b)*. If $\varnothing \tilde{P}_s i$ holds then we can construct a strategy P'_s, which differs from $\tilde{P}_s$ in that it declares i acceptable but not some $j \in \mu(s)$ such that $iP_s j$. The resulting outcome of the Boston mechanism for school s is then $\{\mu(s)\backslash j\} \cup i$, since i declares s acceptable under $\tilde{P}_i$ and is unmatched under μ. Hence, school s has a profitable deviation, which is a contradiction.

One key result of the theory of two-sided matchings is that every stable matching has the same set of students who are unassigned.[8] The proof of Theorem 2 involved an example in which the same set of students are matched in the Nash equilibrium in undominated strategies under the Boston mechanism and in the stable matchings. This might suggest some connection between the number of students who are matched in different Nash equilibria in undominated strategies. However, this is not the case, as we show below.

Proposition 3. *The outcome of a Nash equilibrium* $(\tilde{P}_I, \tilde{P}_S, q)$ *in undominated strategies under the two-sided simultaneous Boston mechanism and true preferences and capacities* (P_I, P_S, q) *may match fewer or more students to schools than any matching that is stable under the true preferences.*

4 The Sequential Boston Mechanism

In many school choice applications, students who apply to schools are aware of how schools' rankings over students are formed. For example, in Boston it was well-known that students with a sibling attending a school are given higher ranking. The analysis so far has assumed that both students and schools report their preferences simultaneously. We now consider the case in which schools first submit their preferences over students and their capacities, and second, students submit their preferences over schools after having observed the schools' reported preferences and capacities. This allows students to report different preferences depending on the observed rankings reported by the schools.

Students' strategies can depend on the schools' reported preferences in arbitrary ways. For example, a student i could report some school s as unacceptable unless that school ranks student i first. If school s ranks i first then i reports that s is the most-preferred school. Given this strategy for student i, school s would have an incentive to rank i first in some circumstances, irrespective of whether student i really is the most-preferred student of school s. This strategy of student i does not seems reasonable, and would allow schools to successfully manipulate the Boston mechanism based solely on unrealistic actions taken by the students. In order to obtain more robust and realistic results on the incentives induced by the Sequential Boston mechanism, we

[8] This is the famous Rural Hospitals Theorem of Roth (1986).

restrict our analysis to sequentially rational strategies for the students. This requires them to report their preferences optimally, given the schools' reported preferences and capacities as well as the other students' reported preferences. The following lemma shows that sequentially rational strategies for the students have strong implications for the outcomes that the Boston mechanism produces.

Lemma 3. *Let $f = (f_{i_1}, \ldots, f_{i_n})$ be the sequentially rational strategies of the students with true preferences P_I. Then for all $(\tilde{P}_S, \tilde{q})$, the outcome of the sequential Boston mechanism, $\Psi^{BOS}\left(f(\tilde{P}_S, \tilde{q}), (P_S, \tilde{q}\right)$, is stable with respect to P_I and $(\tilde{P}_S, \tilde{q})$.*

Proof The result is a Corollary of Theorem 1 of Ergin and Sönmez (2006), who showed that, given schools' preferences and capacities, the set of Nash equilibria of the game induced by the Boston mechanism for the students is equivalent to the set of stable matchings.

Lemma 3 substantially simplifies the analysis of the sequential Boston mechanism. Instead of having to directly specify strategies for the students, we can use the fact that any equilibrium strategy of the students has to produce an outcome under the Boston mechanism that is stable with respect to the schools' *reported* preferences and capacities and the students' *true* preferences.

The following proposition shows that the schools can manipulate the sequential Boston mechanism without having to declare a student unacceptable. This implies that the non-manipulability of the Boston mechanism requires both that schools cannot declare students unacceptable and that the preferences and capacities of the schools and students' preferences are submitted simultaneously.

Proposition 4. *The two-sided sequential Boston mechanism is sequentially manipulable by ranking changes.*

Proof By Lemma 3, every outcome of the sequential Boston mechanism with sequentially rational strategies for the students yields an outcome that is stable with respect to schools' *reported* preferences and capacities and students' *true* preferences. Hence, we need to find preferences of the students P_I, schools' true preferences and capacities (P_S, q), and a school s with reported preferences $\tilde{P}_s$, with $\tilde{P}_s \neq P_s$ and $iP_s\varnothing \Leftrightarrow i\tilde{P}_s\varnothing$, such that some matching that is stable with respect to $(P_I, (\tilde{P}_s, P_{-s}), q)$ is preferred by s to a matching that is stable with respect to (P_I, P_S, q).

Consider the following two-sided matching market:

$$
\begin{array}{rl}
I = \{i_1, i_2, i_3, i_4, i_5\} & S = \{s_1, s_2, s_3\},\ q_1 = 3,\ q_2 = q_3 = 1 \\
P_{i_1} : s_2\ s_1\ s_3 & P_{s_1} : i_1\ i_2\ i_3\ i_4\ i_5 \\
P_{i_2} - P_{i_5} : s_1\ s_2\ s_3 & P_{s_2} : i_3\ i_2\ i_1\ i_4\ i_5 \\
 & P_{s_3} : i_1\ i_2\ i_3\ i_4\ i_5
\end{array}
$$

Consider the matching μ:

$$\mu = \begin{pmatrix} s_1 & s_2 & s_3 \\ i_2, i_3, i_4 & i_1 & i_5 \end{pmatrix}$$

The matching μ is the schools' optimal stable matching and the students' optimal stable matching, and is therefore the unique stable matching. Consider the following deviation by school s_1: it reports the same capacity and P'_{s_1} : i_1 i_2 i_4 i_5 i_3 in the first stage of the sequential Boston mechanism while the other schools report their preferences truthfully. Then the unique stable matching with respect to the schools' *reported* preferences and the students' *true* preferences (and therefore the unique equilibrium outcome of that subgame) is μ':

$$\mu' = \begin{pmatrix} s_1 & s_2 & s_3 \\ i_1, i_2, i_4 & i_3 & i_5 \end{pmatrix}$$

School s_1 therefore receives students $\mu'(s_1) = \{i_1, i_2, i_4\}$ with the deviation, which is a set of students strictly preferred to $\mu(s_1) = \{i_2, i_3, i_4\}$ for any responsive preferences consistent with the true preferences of school s_1 over the students.

Ergin and Sönmez (2006) have shown that if schools rank all students truthfully then the set of Nash equilibrium outcomes in undominated strategies under the two-sided simultaneous Boston mechanism is equal to the set of stable matchings when schools have no unacceptable students. This leaves open the possibility of unstable equilibrium outcomes which are supported by schools not reporting their preferences truthfully. The result in Proposition 4 shows that when submitting their preferences before the students, manipulations that do not involve declaring some students unacceptable may be profitable for schools. The reason is that by changing their reports, schools can affect the students' reported preferences. In the example used to prove Proposition 4 school s_1, by letting student i_3 be the least-preferred acceptable student, effectively induces some students to apply to school s_2 first. This leads student i_1 to apply first to school s_1, as otherwise s_1 would be unable to be matched to school s_1.

Theorem 3. *The set of subgame perfect Nash equilibrium outcomes of the sequential Boston mechanism may contain matchings that are not stable with respect to students' true preferences and schools' true preferences and capacities. Moreover, the resulting equilibrium may be weakly preferred by all schools to all stable matchings.*

Proof Consider the example used to prove Proposition 4 and the matching μ' in that proposition's proof, that results when s_1 reports P'_{s_1} : i_1 i_2 i_4 i_5 i_3 and the other schools report their preferences truthfully:

$$\mu' = \begin{pmatrix} s_1 & s_2 & s_3 \\ i_1, i_2, i_4 & i_3 & i_5 \end{pmatrix}$$

We show that there exists an SPNE in which schools report $\big((P'_{s_1}, P_{s_2}, P_{s_3}), (3,1,1)\big)$. To do this, we use Lemma 3, which implies that the SPNE strategies of the students are such that the outcome of the sequential Boston mechanism is stable with respect to the schools' reported preferences and capacities and the students' true preferences, both for the equilibrium strategies of the schools as well as for schools' deviations from the SPNE strategies. Hence, it is sufficient to show that no matching that is stable with respect to students' true preferences, the other schools' reported preferences and capacities, and some school's deviation from the SPNE strategies is preferred to some matching that is stable with respect to $\big((P'_{s_1}, P_{s_2}, P_{s_3}), (3,1,1)\big)$ and students' true preferences.

School s_2 gets its most-preferred student i_3 and therefore cannot gain by any deviation. School s_3 is the least liked by all the students. Suppose that for some other reported preference profile school s_3 obtains a student $i \in \{i_1, i_2, i_3, i_4\}$. If $i = i_3$, then i and s_2 constitute a blocking pair. If $i \in \{i_1, i_2, i_4\}$ then i and school s_1 constitute a blocking pair because i is among the three most-preferred students of s_1 according to P'_{s_1}. Therefore, it cannot be a continuation equilibrium for s_3 to obtain a student other than i_5 irrespective of the report submitted by s_3. Thus, s_3 has no incentive to deviate. Lastly, consider the incentives of school s_1 to deviate. To gain from a deviation s_1 needs to be matched to students $\{i_1, i_2, i_3\}$. Suppose that there is some report $(\tilde{P}_{s_1}, \tilde{q}_{s_1})$ that yields this outcome for s_1 while the other schools report truthfully. Since the outcome must be stable with respect to students' true preferences and $((\tilde{P}_{s_1}, P_{s_2}, P_{s_3}), (\tilde{q}_{s_1}, 1, 1))$, in a continuation equilibrium yielding this outcome for s_1 we would need i_4 to be matched to s_2 and i_5 to s_3. Otherwise (i_4, s_2) would be a blocking pair. However, in this case (i_1, s_2) is a blocking pair. The outcome is thus unstable. Hence there is a contradiction if school s_1 could profitably deviate from P'_{s_1}. It follows that s_1 also cannot gain by a deviation. Therefore, μ' is a subgame perfect Nash equilibrium outcome of the sequential Boston mechanism.

It remains to show that μ' is not stable with respect to schools' true preferences and capacities and students' true preferences. The pair (i_3, s_1) block the matching μ', since $s_1 P_{i_3} \mu'(i_3) = s_2$, $i_4 \in \mu'(s_1)$ and $i_3 P_{s_1} i_4$. Finally, since schools s_1 and s_2 strictly prefer μ' over μ (which is the unique stable matching with respect to the true preferences) and school s_3 is indifferent between them, μ' is weakly preferred by all schools over all stable matchings.

In our proof we relied on the fact that the set of Nash equilibrium outcomes in undominated strategies of the one-sided Boston mechanism is the set of stable matchings with respect to students' true preferences. Even the student-proposing deferred acceptance mechanism (Gale and Shapley, 1962) has equilibria in undominated strategies that are not stable (Roth and Sotomayor, 1990) when schools report preferences truthfully. In other words, while the Boston mechanism implements the set of stable matchings in undominated strategies, that is not the case for the student-proposing deferred acceptance mechanism and so the incentives that schools have under the latter does not necessarily translate into incentives in the former.

One feature of the preference relation P'_{s_1} that we use in the proofs of Proposition 4 and Theorem 3 is that it does not declare any student unacceptable. For the simultaneous Boston mechanism Ergin and Sönmez (2006) show that such manipulations cannot yield a better outcome for the schools. What our result highlights is that the timing of preference submission is a critical assumption in that result. The manipulability of the sequential Boston mechanism by declaring a student unacceptable and the instability of its equilibrium outcome can be shown by following the same steps of the proofs of Proposition 4 and Theorem 3 by school s_1 declaring student i_3 to be unacceptable.

Remark 1. The two-sided sequential Boston mechanism is sequentially manipulable by declaring students unacceptable. Moreover, the set of subgame perfect Nash equilibrium outcomes may contain matchings that are not stable with respect to students' true preferences and schools' true preferences and capacities.

If we consider the case in which students submit their preferences before the schools, Theorem 2 in Ergin and Sönmez, 2006 implies that schools will not have any incentive to misrepresent their preferences. Therefore, students will play the preference revelation game induced by the one-sided version of the Boston mechanism and the corresponding results in Ergin and Sönmez, 2006, hold.

Remark 2. If under the Boston mechanism students submit their preferences first and schools submit theirs afterward, the set of subgame perfect Nash equilibria outcomes equals the set of stable matchings under the true preferences, when all students are required to be acceptable by all schools.

For the two-sided simultaneous Boston mechanism we have shown that schools do not have an incentive to manipulate their capacities. We now show that this result also depends on the timing of the game. Specifically, when schools report their capacities before students report their preferences (and capacities are observable by the students) then schools may have an incentive to manipulate their capacities.

Proposition 5. *The two-sided sequential Boston mechanism is sequentially manipulable via capacities.*

Proof Consider the following example, where preferences for both schools are responsive.[9]

$$
\begin{array}{ll}
I = \{i_1, i_2, i_3, i_4\} & S = \{s_1, s_2\},\ q_1 = 3,\ q_2 = 2 \\
P_{i_1} : s_2\ s_1 & P_{s_1} : \{i_1, i_2, i_3\}\ \{i_1, i_2\}\ \{i_1, i_3\}\ \{i_1\}\ \{i_2, i_3\}\ \{i_2\}\ \{i_3\}\ \{i_4\} \\
P_{i_2} : s_1\ s_2 & P_{s_2} : i_4\ i_3\ i_2\ i_1 \\
P_{i_3} : s_1\ s_2 & \\
P_{i_4} : s_2\ s_1 &
\end{array}
$$

[9] For this example it is not necessary to specify the preferences of school s_2 beyond its ranking over singleton sets of students.

For school s_1 we have only shown preferences that are relevant to our analysis (that is, we do not show how the preferences of school s_1 over sets including student i_4 are). We assume that schools' reported preferences are fixed. Given the schools' reported preferences and capacities, and by sequential rationality and Lemma 3, the outcome of the subgame played by the students after schools report their capacities is stable with respect to their *true* preferences and the schools' *reported* preferences and capacities. Hence, it is sufficient to consider the resulting set of stable matchings for each possible combination of capacities reported by the schools. When the reported capacities are $(3, 2)$ it is easy to verify that the unique stable matching is given by:

$$\mu = \begin{pmatrix} s_1 & s_2 \\ i_2, i_3 & i_1, i_4 \end{pmatrix}$$

When the reported capacities are $(1, 2)$ the unique stable matching is given by:

$$\hat{\mu} = \begin{pmatrix} s_1 & s_2 \\ i_1 & i_3, i_4 \end{pmatrix}$$

Since $\{i_1\} P_{s_1} \{i_2, i_3\}$, school s_1 gains from understating its capacity by 2.

Unlike in the simultaneous Boston mechanism, schools may have incentives to misstate their capacities in the sequential Boston mechanism. The intuition is similar to preference manipulations of the sequential Boston mechanism: by misstating its capacities, a school may change the set of stable matchings.

The next theorem considers the game induced by the sequential Boston mechanism when schools' only strategic variable is their reported capacity.

Theorem 4. *Holding schools' submitted preferences fixed, the set of SPNE outcomes of the sequential Boston mechanism may have no stable matching when schools are only able to manipulate their capacities. Moreover, all SPNE outcomes may be strictly preferred by the schools over all stable matchings.*

Proof Consider the example in the proof of Proposition 5. The only stable matching, under true capacities, is μ:

$$\mu = \begin{pmatrix} s_1 & s_2 \\ i_2, i_3 & i_1, i_4 \end{pmatrix}$$

Clearly, the profile $(3, 1)$ cannot be part of an SPNE strategy profile, since school s_2 would only obtain student s_4, which is worse than the outcome under $(3, 2)$. If schools report $(1, 2)$, instead, the SPNE outcome is unique and is given by:

$$\hat{\mu} = \begin{pmatrix} s_1 & s_2 \\ i_1 & i_3, i_4 \end{pmatrix}$$

which is preferred by both schools to μ and for school s_2 is its most-preferred set of students. Therefore, $(3, 2)$ is not part of an SPNE strategy profile. Moreover, it follows that $(1,1)$ is not part of an SPNE strategy profile, as s_2 would prefer to state its capacity truthfully. Consider next what happens when the reported capacities are $(2,2)$. In that case there are two stable matchings (and thus two possible subgame perfect continuation outcomes). These are μ and $\tilde{\mu}$:

$$\tilde{\mu} = \begin{pmatrix} s_1 & s_2 \\ i_1, i_2, & i_3, i_4 \end{pmatrix}$$

Note that under $\tilde{\mu}$ both schools get their most-preferred set of two students, while μ is the unique stable outcome with respect to true preferences. There are two equilibrium outcomes for the subgame after the reported capacities $(2,2)$. When μ is the outcome then $(2,2)$ is not part of an SPNE strategy profile, since school s_1 prefers the outcome of $(1,2)$ to μ, which it can reach by deviating. If instead the outcome is $\tilde{\mu}$, then $(2,2)$ is part of an SPNE strategy profile. Lastly, $(2,1)$ is not part of an SPNE strategy profile, since s_2 only obtains student s_4, which is worse than the outcome under $(2,2)$, irrespective of the following equilibrium outcome.

Finally, note that the two possible equilibrium outcomes, $\tilde{\mu}$ and $\hat{\mu}$, are strictly preferred by both schools over the unique stable matching μ.

The proof of Theorem 4 shows that there are situations in which no stable matching is supported by a subgame perfect Nash equilibrium when schools can misrepresent their capacities, but not their preferences. By misstating their capacities, schools can obtain an outcome that all of them strictly prefer to the stable matching. The restriction that schools can only manipulate capacities has bite in the example used above: if school s_1 reports that only students i_2 and i_3 are acceptable, while s_2 reports that only i_1 and i_4 are acceptable, the outcome of the continuation game then leads to the stable matching μ. Neither s_1 nor s_2 can gain by deviating, since μ is the only matching that is stable with respect to those reported preferences and students' true preferences. This argument can be generalized:

Theorem 5. *For every stable matching μ, the game induced by the two-sided sequential Boston mechanism has a pure strategy subgame perfect Nash equilibrium with outcome μ.*

Proof Fix a stable matching μ. Let each school report $(\tilde{P}_s, q_s)$ such that $i\tilde{P}_s\varnothing$ if and only if $i \in \mu(s)$. Then any continuation subgame will be such that each student i reports $\mu(i)$ as an acceptable school, so the outcome will be μ. There are many such strategies, all with the same outcome. Students cannot gain by deviating: for each student i, there is only one school that reported i to be acceptable. That school is $\mu(i)$. Hence, declaring $\mu(i)$ a unacceptable, results in student i being unassigned.

The schools cannot gain by deviating. Suppose some school s, reported (P'_s, q'_s) instead and obtained a better outcome. Let the outcome of the subgame that results when s deviates and all other schools continue to report $(\tilde{P}_{-s}, q_{-s})$ be μ'. If the school obtained an outcome, say $\mu'(s)$ that it prefers over $\mu(s)$, then any student $i \in \mu'(s)$ must weakly prefer s to $\mu(i)$. Otherwise, in the subgame i could obtain $\mu(i)$ by ranking it first. However, this contradicts μ being stable since school s and the students in $\mu'(s)$ constitute a blocking pair.

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School Choice: A Mechanism Design Approach," *American Economic Review*, 2003, pp. 729–747. 1

_____, **Parag A Pathak, Alvin E Roth, and Tayfun Sönmez**, "The Boston public school match," *American Economic Review*, 2005, pp. 368–371. 2

_____, _____, **and** _____, "The New York City High School Match," *American Economic Review*, 2005, pp. 364–367. 3

_____, **Yeon-Koo Che, and Yosuke Yasuda**, "Resolving Conflicting Preferences in School Choice: The "Boston Mechanism" Reconsidered," *American Economic Review*, 2011, *101* (1), 399–410. 1

Basteck, Christian, Katharina Huesmann, and Heinrich Nax, "Matching Practices for secondary schools – Germany," 03 2016. 1

Chen, Li, "Matching Practices for Elementary Schools – Ireland," 03 2016. 1

Dur, Umut, "The Modified Boston Mechanism," *Working Paper*, April 2015. 1

_____ **and Onur Kesten**, "Sequential versus Simultaneous Assignment Systems and Two Applications," *Working Paper*, 2014. 3

Ehlers, Lars, "Truncation strategies in matching markets," *Mathematics of Operations Research*, 2008, *33* (2), 327–335. 1

_____, "Manipulation via capacities revisited," *Games and Economic Behavior*, 2010, *69* (2), 302–311. 1

Ergin, Haluk and Tayfun Sönmez, "Games of school choice under the Boston mechanism," *Journal of Public Economics*, 2006, *90* (1), 215–237. 1, 2, 3.1, 3.2, 4, 4, 4, 4

Featherstone, Clayton R. and Eric Mayefsky, "Why Do Some Clearinghouses Yield Stable Outcomes? Experimental Evidence on Out-of-Equilibrium Truth-Telling," *Working Paper*, 2015. 1

Gale, David and Lloyd S Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 1962, *69* (1), 9–15. 1, 2, 3, 4

Haan, Monique De, Pieter A Gautier, Hessel Oosterbeek, and Bas Van der Klaauw, "The performance of school assignment mechanisms in practice," 2015. 1

Kojima, Fuhito, "Games of school choice under the Boston mechanism with general priority structures," *Social Choice and Welfare*, 2008, *31* (3), 357–365. 1

_____ **and M Utku Ünver**, "The "Boston" School-Choice Mechanism: An Axiomatic Approach," *Economic Theory*, 2014, *55* (3), 515–544. 1

_____ **and Parag A Pathak**, "Incentives and stability in large two-sided matching markets," *American Economic Review*, 2009, pp. 608–627. 3.1

Konishi, Hideo and M Utku Ünver, "Games of capacity manipulation in hospital-intern markets," *Social Choice and Welfare*, 2006, *27* (1), 3–24. 1

Mennle, Timo and Sven Seuken, "The Naïve versus the Adaptive Boston Mechanism," *Working Paper*, 2014. 1

Miralles, Antonio, "School choice: The Case for the Boston Mechanism," in "Auctions, Market Mechanisms and Their Applications," Springer, 2009, pp. 58–60. 1

Pais, Joana and Ágnes Pintér, "School choice and information: An experimental study on matching mechanisms," *Games and Economic Behavior*, 2008, *64* (1), 303–328. 1

Pathak, Parag A and Tayfun Sönmez, "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 2008, *98* (4), 1636–1652. 1

Pathak, Parag A. and Tayfun Sönmez, "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation," *American Economic Review*, 2013, *103* (1), 80–106. 1

Romero-Medina, Antonio and Matteo Triossi, "Games with capacity manipulation: incentives and Nash equilibria," *Social Choice and Welfare*, 2013, *41* (3), 701–720. 1

Roth, Alvin E, "On the allocation of residents to rural hospitals: a general property of two-sided matching markets," *Econometrica*, 1986, pp. 425–427. 8

———, "A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom," *American Economic Review*, 1991, pp. 415–440. 1

——— **and Andrew Postlewaite**, "Weak versus strong domination in a market with indivisible goods," *Journal of Mathematical Economics*, 1977, *4* (2), 131–137. 1

——— **and Marilda A Oliveira Sotomayor**, *Two-sided matching: A study in Game-theoretic Modeling and Analysis* number 18, Cambridge University Press, 1990. 4

Appendix

Proof of Lemma 1

Necessity of the conditions.

Suppose that a strategy $\tilde{P}_s$ is undominated. We will show that if each of the conditions above fail, there is a strategy that dominates $\tilde{P}_s$.

For *(i)*, suppose that there is some student i such that $i\tilde{P}_s\varnothing$ and $\varnothing P_s i$. Consider a strategy $(P'_s, \tilde{q}_s)$ where P'_s is the same as $\tilde{P}_s$ except that for every student j such that $\varnothing P_s j$ we make $\varnothing P'_s j$ (including i). Under the simultaneous Boston mechanism, a school can't change the set of students who apply to that school in each period. As a result, under P'_s, the set of acceptable students who are matched to s under $\tilde{P}_s$ will still be matched to s under P'_s, but if some unacceptable student is matched to s under $\tilde{P}_s$, she will be replaced by either an acceptable student or an empty seat. By responsiveness, both cases are preferred by s to being matched with some unacceptable student. Therefore, P'_s dominates $\tilde{P}_s$.

For *(ii)*, suppose $\tilde{q}_s < q_s$ and consider a strategy $(\tilde{P}_s, q_s)$. For given preferences and capacities of the other schools and for given preferences of the students, each round of the Boston mechanism will proceed identically under $(\tilde{P}_s, \tilde{q}_s)$ and $(\tilde{P}_s, q_s)$ until the capacity constraint $\tilde{q}_s$ starts to bind, say in round t. This means that in round t there are more students applying to school s than the remaining capacity. However, since $q_s > \tilde{q}_s$ under $(\tilde{P}_s, q_s)$ not only will the same set of students be accepted as under $(\tilde{P}_s, \tilde{q}_s)$ but, in addition, some more students will be accepted. Since preferences are responsive and since $\tilde{P}_s$ lists as acceptable only those students that are acceptable under P_s, it follows that the outcome of the Boston mechanism under $(\tilde{P}_s, q_s)$ is weakly preferred to the outcome under $(\tilde{P}_s, \tilde{q}_s)$. Hence, only strategies with $\tilde{q}_s = q_s$ can be undominated.

For *(iii)*, consider two students, i, j such that $i, j\tilde{P}_s\varnothing$, iP_sj and $j\tilde{P}_si$. Consider the strategy $(P'_s, \tilde{q}_s)$ where P'_s is such that $k\tilde{P}_s\varnothing \Leftrightarrow kP'_s\varnothing$ and for all $k, k'P'_s\varnothing$ we have $kP'_sk' \Leftrightarrow kP_sk'$. Since under $(\tilde{P}_s, \tilde{q}_s)$ and $(P'_s, \tilde{q}_s)$ the same set of students is acceptable, each round of the Boston mechanism is equivalent under both strategies unless in some round there are more acceptable applicants than available capacity, for given preferences and capacities of other schools and preferences of the students. Because capacities are the same, this is the same round under both strategies. In that round the preferences over the student determine which students will be accepted. Under $(P'_s, \tilde{q}_s)$ the best students according to P_s will be accepted. This is not the case under $(\tilde{P}_s, \tilde{q}_s)$. Thus school s strictly prefers the outcome under $(P'_s, \tilde{q}_s)$ to the outcome under $(\tilde{P}_s, \tilde{q}_s)$ whenever cases arise in which both i and j apply in the same round, but only j is accepted under $(\tilde{P}_s, \tilde{q}_s)$. Otherwise the outcome under both strategies is the same. Hence, $(\tilde{P}_s, \tilde{q}_s)$ is dominated by $(P'_s, \tilde{q}_s)$, which is a contradiction.

For *(iv)*, let i be the most-preferred student under P_s but $\varnothing\tilde{P}_si$. We construct a strategy for school s, $(P'_s, \tilde{q}_s)$ that dominates $(\tilde{P}_s, \tilde{q}_s)$. Let P'_s be the same as $\tilde{P}_s$ except that i is now the most-preferred acceptable student. For given preferences and capacities of the other school and preferences of the students, each round of the Boston mechanism will proceed identically under both strategies until student i applies to school s. In that round, if school s still has available capacity, i will be rejected under $(\tilde{P}_s, \tilde{q}_s)$ but not under $(P'_s, \tilde{q}_s)$ and otherwise the outcome under both strategies is the same. The rejection of i under $(\tilde{P}_s, \tilde{q}_s)$ can at most lead to one additional application from an acceptable student, with other applications being the same in each subsequent round. This additional application occurs when student i in later rounds is matched to some other school s' and thereby prevents another student i' from being matched there and i' likewise prevents another student from being matched to some other school. This sequence of rejections either ends with one student being matched to a school whose capacity constraint under the Boston mechanism and $(P'_s, \tilde{q}_s)$ is not binding (this includes being left unmatched) or it ends with some student i'' applying to school s. If i'' is not acceptable, the sequence of rejections continues until it either ends or some other student applies to s. If s accepts i'' then the assignment under both strategies will differ by only a single agent. But since i is the most-preferred student under P_s, the outcome under $(P'_s, \tilde{q}_s)$ is preferred by s. Hence, $(\tilde{P}_s, \tilde{q}_s)$ is dominated by $(P'_s, \tilde{q}_s)$, which is a contradiction.

For *(v)*, let there be only $l < q_s$ acceptable students under $(\tilde{P}_s, q_s)$. Construct (P'_s, q_s) such that it is the same as $(\tilde{P}_s, q_s)$ except that one additional student i, who is acceptable under P_s is acceptable under it. This is possible since we assumed that at least q_s students are acceptable under each school's true preferences. Since only l students under $\tilde{P}_s$ are acceptable, the capacity constraint will never bind in any round of the Boston mechanism under either (P'_s, q_s) or $(\tilde{P}_s, q_s)$ for any given preferences and capacities of the other schools and preferences of the students. This means that for both strategies any application by the l students, that are acceptable under $(\tilde{P}_s, q_s)$, will be accepted. In addition, under (P'_s, q_s) student i can also be accepted. Therefore, (P'_s, q_s) never yields a worse outcome than $(\tilde{P}_s, q_s)$ but may sometimes yield a preferred outcome, evaluated under the true preferences, which contradicts $(\tilde{P}_s, q_s)$ being undominated.

Sufficiency of the conditions.

Suppose $(\tilde{P}_s, \tilde{q}_s)$ satisfies all five conditions, so that we can replace $\tilde{q}_s$ by q_s. Without loss of generality, any strategy (P'_s, q'_s) that is a candidate for a strategy that dominates $(\tilde{P}_s, q_s)$ also needs to satisfy the five conditions. Otherwise we could replace it with a strategy that dominates it. So we have $q'_s = q_s$. Furthermore, whenever two students i, i' are acceptable under $\tilde{P}_s$ and P'_s we must have $i\tilde{P}_s i' \Leftrightarrow iP'_s i'$, since both satisfy the third condition. Thus, P'_s can only meaningfully differ from $\tilde{P}_s$ in terms of the set of students who are reported to be acceptable.

Suppose first that q_s students are acceptable under $(\tilde{P}_s, q_s)$. Define the set of students that are acceptable under this strategy to be $\tilde{A} \subseteq I$. Let i_1 be the most-preferred student under P_s. We then have that $i_1 \in \tilde{A}$. It cannot be dominated by another strategy (P'_s, q_s) that declares a different set of q_s students acceptable. To see this, suppose that students in $\tilde{A}$, and only those students, rank school s first. All other students declare s to be unacceptable. In that case, under strategy $(\tilde{P}_s, q_s)$ the set $\tilde{A}$ is accepted, but only a subset of those students are accepted under (P'_s, q_s), which contradicts (P'_s, q_s) dominating $(\tilde{P}_s, q_s)$.

Now suppose that $(\tilde{P}_s, q_s)$ with q_s acceptable students is dominated by some other strategy (P'_s, q_s) that declares k students acceptable, with $q_s < k \leqslant n$. Denote this set of students as A'. We have that $\tilde{A} \subset A'$. Otherwise, let the students in $\tilde{A} \backslash A'$ be the only ones who consider s acceptable and let them rank s first. Then $(\tilde{P}_s, q_s)$ yields a better outcome to s than (P'_s, q_s).

Suppose now that a subset of $q_s - 1$ students in $\tilde{A}$ and another student i' in $A' \backslash \tilde{A}$ rank s first. Suppose further that i_1 is the student in $\tilde{A}$ who does not rank s first, but instead ranks another school first, which does not consider i_1 acceptable. Furthermore, let s be the second choice of i_1. Then the outcome of the Boston mechanism under $(\tilde{P}_s, q_s)$ differs from that under (P'_s, q_s) only in that i_1 is matched to s in the former, while i' is matched to s in the latter. Since $i_1 P_s i'$, the outcome under $(\tilde{P}_s, q_s)$ is preferred to that under (P'_s, q_s). This argument can be adapted straightforwardly to show that no strategy $(\tilde{P}_s, q_s)$ that considers $k \geqslant q_s$ students acceptable is dominated by another strategy (P'_s, q_s) that considers $k' \geqslant k$ students acceptable.

Finally, suppose that k, with $n \geqslant k > q_s$, students are acceptable under $(\tilde{P}_s, q_s)$. We will show that $(\tilde{P}_s, q_s)$ cannot be dominated by any strategy (P'_s, q_s) declaring $k' < k$ students acceptable. The associated sets of acceptable students are $\tilde{A}$ and A' as before. Since fewer students are acceptable under (P'_s, q_s) we have $\tilde{A} \backslash A' \neq \varnothing$. Now suppose all students in $\tilde{A} \backslash A'$ rank school s first and are the only students to consider i as acceptable. Then up to the q_s most-preferred students will be accepted by s under $(\tilde{P}_s, q_s)$ but no student will be matched to s under (P'_s, q_s). Thus, $(\tilde{P}_s, q_s)$ cannot be dominated by a strategy(P'_s, q_s) that considers fewer students acceptable.

Proof of Lemma 2

Necessity of $sP_i\varnothing \Leftrightarrow s\tilde{P}_i\varnothing$.

For a contradiction suppose that $\tilde{P}_i$ is an undominated strategy. Suppose that there exists some student i and school s such that $sP_i\varnothing$ but $\varnothing\tilde{P}_i s$. Let preferences P'_i be the same as $\tilde{P}_i$ except that s is the least-preferred acceptable school. For all strategies $\tilde{P}_{-i}$, $(\tilde{P}_S, \tilde{q}_s)$ of the other agents in which i is matched to some school under $\tilde{P}_i$, the assignment under P'_i has to be the same under the Boston mechanism, since in each round i applies to the same schools. If i remains unmatched under $\tilde{P}_i$ then i also remains unmatched under P'_i, unless i is accepted by school s when applying to it. Hence, the outcome under both is the same, unless i is accepted by school s, which cannot happen under $\tilde{P}_i$. One profile of preferences and capacities that has this property is one in which all schools except s consider i to be unacceptable and school s considers i and no other student to be acceptable. For this profile of preferences, under P'_i, i is matched to s, which because $sP_i\varnothing$, is strictly preferred to the outcome when reporting $\tilde{P}_i$. Hence $\tilde{P}_i$ is dominated by P'_i, a contradiction.

For any strategy $\tilde{P}_i$ that declares a school s with $\varnothing P_i s$ to be acceptable, one can find preferences for the schools and the other students so that i is matched to s. For any such preferences, a strategy that is identical to $\tilde{P}_i$, but declares s and any other school s', with $\varnothing P_i s'$, to be unacceptable, can at worst result in i being unmatched. This is preferred to being matched to an unacceptable school, so $\tilde{P}_i$ is a dominated strategy, which is a contradiction.

Sufficiency of $sP_i\varnothing \Leftrightarrow s\tilde{P}_i\varnothing$.

To see the result, we need to show that any ordering of acceptable schools constitutes an undominated strategy. We show first that no strategy $\tilde{P}_i$ that ranks some acceptable school s first can be dominated by another strategy P'_i that ranks another acceptable school s' first. We assume that both schools are acceptable under P_i. Suppose that $\tilde{P}_{-i}$ is such that q_s other students rank s first, but that i is the most-preferred student according to $\tilde{P}_s$. The other

q_s students are acceptable to s. Suppose in addition that no other school, including s', considers i to be acceptable. In that case, under $\tilde{P}_i$ student i obtains school s. Under P'_i student i remains unmatched. Hence, a strategy $\tilde{P}_i$ ranking school s first can only be dominated by a strategy that also ranks s first.

Suppose that the top k schools, with $2 \leqslant k < m$, under $\tilde{P}_i$ are ranked as follows: $s_1 \tilde{P}_i s_2 \dots s_{k-1} \tilde{P}_i s_k$. Suppose that $P'_i \neq \tilde{P}_i$ ranks the first $k-1$ schools the same way, but ranks another acceptable school $s'_k \neq s_k$ as the kth best school. The preferences over the remaining schools under both $\tilde{P}_i$ and P'_i are arbitrary. We show that no such P'_i can dominate $\tilde{P}_i$. We construct reported preferences for the other students and strategies for the schools as follows. Suppose school s reports a capacity of $\tilde{q}_{s_k} = 1$. Let there be one other student i' who reports the same preferences over the first k schools as student i. Let both i and i' be acceptable to school s_k but unacceptable to any school $s_1, s_2, \dots, s_{k-1}$. Let the preferences of other students be such that none of them apply to school s_k until after round k. Furthermore, let $i \tilde{P}_{s_k} i'$. Last, we assume that s_k is the only school that considers i to be acceptable.

Under these (incompletely) specified preferences it follows that i is matched to s_k when reporting $\tilde{P}_i$ but remains unmatched when reporting P'_i. Since s_k is acceptable, it follows that P'_i does not dominate $\tilde{P}_i$. Hence, no strategy is dominated by another strategy that is identical for the first $k-1$ schools and deviates thereafter. It also follows that for a given k and $s_1, s_2, \dots, s_k$, the strategy $\tilde{P}_i$ can also not be dominated by any strategy that differs in the ranking of the $k-1$ most-preferred schools. It follows that no strategy $\tilde{P}_i$ that satisfies for all $s \in S$, $s \tilde{P}_i \varnothing \Leftrightarrow s P_i \varnothing$ is dominated by some other strategy that also satisfies this condition.

Proof of Theorem 1

To prove the theorem, we will propose undominated strategies for the students and schools and verify that these strategies constitute a Nash equilibrium. Take some matching μ that is stable with respect to the true preferences and capacities. For all $i \in I$ such that $\mu(i) = \varnothing$ let i report preferences truthfully. For all $i \in I$ such that $\mu(i) \in S$ let i report school $\mu(i)$ as the most preferred and all other acceptable schools arbitrarily ranked below. Let all schools report preferences truthfully. We will refer to this strategy-profile as (P_I^μ, P_S, q), to make clear that only students' strategies depend on the matching μ.

Claim 1: (P_I^μ, P_S, q) consists only of undominated strategies.

From Lemma 1 and Corollary 2 it follows that schools' strategies are undominated. From Lemma 2 it follows that students' strategies are undominated.

Claim 2: (P_I^μ, P_S, q) is a Nash equilibrium.

If all agents use these strategies, the outcome under the Boston mechanism is μ. Every student i applying to $\mu(i)$ will be accepted by the school in the first round of the Boston mechanism, since μ is stable. Students unmatched under μ will be rejected by all schools to which they apply, since if they were accepted by some school, this would contradict the stability of μ given that schools are assumed to report preferences truthfully. No school s can gain from a deviation. Any student applying to a school other than s will be accepted in the first round, and so cannot be obtained by school s. The only students that s could obtain under some alternative strategy $\tilde{P}_s$ are those who apply to it in the first round and those who are unmatched under μ and who consider s acceptable. However, since μ is stable, it follows that s prefers $\mu(s)$ to obtaining some students who are unmatched under μ and who consider s acceptable. Hence, s cannot possibly gain from any type of deviation.

No student i who remains unmatched under μ can gain from a deviation. Since μ is stable and schools report preferences truthfully, any unmatched student must be less preferred by the schools than the students accepted by the school in the first round of the Boston mechanism. If a school accepts some student i who is unmatched under μ, this contradicts the stability of μ. Similarly, any student i who is matched to a school under μ cannot gain from a deviation. If such a student ranked another, preferred school s first then that school would not accept the application. Otherwise, if it accepted, this would contradict the stability of μ, since s would have chosen i in the first round even when $\mu(s)$ was available. If some other strategy resulted in i being matched to s, then i ranking s first would also do so. Hence, no other deviation can yield student i being assigned to a school s that is preferred over $\mu(i)$. Hence, no student can gain from a deviation.

Proof of Proposition 3

Fewer students matched under $(\tilde{P}_I, \tilde{P}_S, q)$ than in stable matchings.

Consider the following true preferences and capacities:

$I = \{i_1, i_2\} \quad S = \{s_1, s_2\}, q_{s_1} = q_{s_2} = 1$

$P_{i_1} : s_1 \quad P_{s_1} : i_2 \; i_1$

$P_{i_2} : s_2 \; s_1 \quad P_{s_2} : i_1 \; i_2$

It can be easily verified that the unique stable matching under the true preferences is $\{(s_1, i_1), (s_2, i_2)\}$. Consider the following strategy profile (assuming capacities are reported truthfully):

$$\tilde{P}_{i_1} : s_1 \quad \tilde{P}_{s_1} : i_2 \; i_1$$
$$\tilde{P}_{i_2} : s_1 \; s_2 \quad \tilde{P}_{s_2} : i_1$$

Using Lemmas 1 and 2, it follows that $(\tilde{P}_I, \tilde{P}_S, q)$ are undominated strategies. The outcome of the Boston mechanism under $(\tilde{P}_I, \tilde{P}_S, q)$ is $\{(s_1, i_2), (s_2, \emptyset), (\emptyset, i_1)\}$, which leaves i_1 unmatched. In contrast, in the stable matching, i_1 is matched to s_1.

It remains to show that $(\tilde{P}_I, \tilde{P}_S, q)$ constitutes a Nash equilibrium. School s_1 gets its most-preferred outcome, so have no incentive to deviate. Student i_2

is deemed unacceptable by school s_2 under $\tilde{P}_{s_2}$, and therefore cannot deviate and be matched there. Since i_2 is matched to s_1 in the first round of the Boston mechanism, there is no report that s_2 could make to obtain either i_2 or i_1. Similarly, i_1 applies to s_1 in the first round and is rejected. Student i_1 has no incentive to deviate, since s_1 is the only school that is acceptable to i_1.

More students matched under $(\tilde{P}_I, \tilde{P}_S, q)$ than in stable matchings. Consider the following true preferences and capacities:

$$\begin{array}{ll} I = \{i_1, i_2\} & S = \{s_1, s_2, s_3\}\,,\ q_1 = 2,\ q_2 = 1 \\ P_{i_1} : s_1\ s_2 & P_{s_1} : i_1\ i_2\ i_3 \\ P_{i_2} : s_1\ s_2 & P_{s_2} : i_1\ i_2 \\ P_{i_3} : s_1\ s_2 & \end{array}$$

It can be easily verified that the unique stable matching under the true preferences is $\{(s_1, (i_1, i_2)), (s_2, \varnothing), (\varnothing, i_3)\}$. Student i_3 is thus unassigned in the stable matching. Consider the following strategy profile (assuming capacities are reported truthfully):

$$\begin{array}{ll} \tilde{P}_{i_1} : s_1\ s_2 & \tilde{P}_{s_1} : i_1\ i_3 \\ \tilde{P}_{i_2} : s_2\ s_1 & \tilde{P}_{s_2} : i_1\ i_2 \\ \tilde{P}_{i_3} : s_1\ s_2 & \end{array}$$

Using Lemmas 1 and 2, it follows that $(\tilde{P}_I, \tilde{P}_S, q)$ are undominated strategies. The outcome under the Boston mechanism of strategy profile $(\tilde{P}_I, \tilde{P}_S, q)$ is $\{(s_1, (i_1, i_3)), (s_2, i_2)\}$, which has i_3 being assigned to s_1.

It remains to verify that $(\tilde{P}_I, \tilde{P}_S, q)$ constitutes a Nash equilibrium. Students i_1 and i_3 are matched to their most-preferred school, so have no incentive to deviate. Student i_2 is declared unacceptable under $\tilde{P}_{s_1}$, so being matched to s_2 is the best i_2 can achieve. School s_1 has no incentive to deviate, since i_2 is matched to s_2 in the first round of the Boston mechanism, independent of the preferences reported by s_1. Similarly, i_1 is matched to s_1 in the first round of the Boston mechanism, so s_2 cannot obtain i_1 by reporting other preferences. Hence, s_2 also has no incentive to deviate.