Incentives and Stability in Large Two-Sided Matching Markets*

Fuhito Kojima[†] Parag A. Pathak[‡]

May 13, 2006

Abstract

The paper analyzes the scope for manipulation in many-to-one matching markets (college admission problems) under the student-optimal stable mechanism when the number of participants is large and the length of the preference list is bounded. Under a mild independence assumption on the distribution of preferences for students, the fraction of colleges that have incentives to misrepresent their preferences approaches zero as the market becomes large. We show that truthful reporting is an approximate equilibrium under the student-optimal stable mechanism in large markets that are *sufficiently thick*, a condition that allows for certain types of heterogeneity in the distribution of student preferences.

Keywords: Large markets, stability, two-sided matching.
JEL Class: C78, D61, D78.

---

*We appreciate helpful discussions with Attila Ambrus, Drew Fudenberg, Akihiko Matsui, Alvin E. Roth, Tayfun Sönmez, Satoru Takahashi and seminar participants at Harvard University, Hitotsubashi University, Kyoto University, Osaka University and University of Tokyo. Discussions with Michihiro Kandori and Yohei Sekiguchi were especially important for the current version of the paper. For financial support, Pathak is grateful to the National Science Foundation, the Spencer Foundation, and the Division of Research at Harvard Business School.

[†]Department of Economics, Harvard University, Cambridge, MA 02138, e-mail: kojima@fas.harvard.edu.

[‡]Department of Economics, Harvard University, Cambridge, MA 02138, e-mail: ppathak@fas.harvard.edu.

1 Introduction

The theory of two-sided matching has influenced the design of several entry-level labor markets and student assignment systems.[1] A central notion in the theory is stability: a matching is stable if there is no individual agent or pair of agents who prefer to be assigned to each other than their allocation in a matching. In real world applications, empirical studies have shown that stable mechanisms often succeed whereas unstable ones often fail.[2]

Although stable mechanisms have a number of virtues, they are susceptible to various types of strategic behavior before and during the match. Dubins and Freedman (1981) and Roth (1982) show that any stable mechanism is manipulable via preference lists: reporting a preference list that does not reflect the true underlying preferences may be a best response for some participants. In many-to-one markets, Sönmez (1997b) and Sönmez (1999) show that there are also other strategic concerns. First, any stable mechanism is manipulable via capacities so that colleges may sometimes benefit by underreporting their quotas. Second, any stable mechanism is manipulable via pre-arranged matches so that a college and a student may benefit by agreeing to match before receiving their allocations from the centralized matching mechanism.

Concerns about the potential for these types of manipulation are often present in real markets. For instance, in New York City (NYC) where the Department of Education has recently adopted a stable mechanism, the Deputy Chancellor of Schools described principals concealing capacity as a major issue with their previous system (New York Times (11/19/04)):

> "Before you might have a situation where a school was going to take 100 new children for 9th grade, they might have declared only 40 seats, and then placed the other 60 outside of the process."

Roth and Rothblum (1999) discuss similar anecdotes about preference manipulation from the National Resident Matching Program (NRMP), which is an entry-level matching market for hospitals and medical school graduates in the U.S.

The aim of this paper is to understand why despite these negative results many stable mechanisms appear to work well in practice. Our results show that the size of the market makes the mechanism immune to various kinds

[1] For a survey of this theory, see Roth and Sotomayor (1990). For applications to labor markets, see Roth (1984a) and Roth and Peranson (1999). For applications to student assignment, see for example Abdulkadiroğlu and Sönmez (2003), Abdulkadiroglu, Pathak, Roth, and Sönmez (2005) and Abdulkadiroglu, Pathak, and Roth (2005).

[2] For a summary of this evidence, see Roth (2002).

of manipulations. In real-world two-sided matching markets, there are often a large number of participants, and each participant submits a rank order list whose length is a small fraction of the market size. For instance, in the NRMP, the length of the applicant preference list is about 15, while the number of hospital programs is 3,000 to 4,000 and the number of students is over 20,000 per year. In NYC, the maximal length of the preference list is 12, and there are about 500 school programs and over 90,000 students per year.[3] These features motivate our study of the limit as the number of participants grows, but the length of the preference lists does not.

We consider many-to-one matching markets with the student-optimal stable mechanism, where colleges have arbitrary preferences such that every student is acceptable, and students have random preferences of fixed length drawn iteratively from an arbitrary distribution. We show that the expected proportion of colleges that can manipulate the central clearing house converges to zero as the number of colleges approaches infinity. The key intuition comes from a lemma on the vanishing market power of colleges. Under our assumptions, the lemma shows that the likelihood that the sequence of chain reactions caused when a college rejects students it was assigned from the student-optimal stable matching leads to another student applying to that college is small.

We also conduct equilibrium analysis in the large market, for which we require an additional condition. We say that the market is *sufficiently thick* if there are enough ex ante desirable colleges where the number of potential applicants is less than the number of positions. To further understand the idea of this condition, consider a disruption of the market in which a student becomes unmatched. If the market is thick, such a student is likely to find a seat in another college that has room for her. Thus the condition would imply that a small disruption of the market is likely to be absorbed by a vacant seat. One example which ensures sufficient thickness is when the distribution of student preferences is such that there are no extremely popular colleges. We show that truthful reporting is an approximate equilibrium in a large market that is sufficiently thick.

We next extend our equilibrium analysis to weaken the distributional assumption on student preferences. Specifically we allow preferences of different students to be drawn from a number of different distributions. This is important for studying a number of real world markets, for participants may

[3]For data regarding NRMP, see http://www.nrmp.org/2006advdata.pdf. For data regarding New York City high school match, see http://www.nycenet.edu/Administration/mediarelations/PressReleases/2004-2005/3-23-2005-12-20-12-570.htm

often have systematically different preferences according to their residential location, academic achievement and other characteristics. For instance, in NYC, Abdulkadiroğlu, Pathak, and Roth (2006) find that the overwhelming majority of students in a borough rank a program within their own borough as their top choice. We show that truthful reporting is an approximate equilibrium in large markets that are sufficiently thick when this sort of heterogeneity is present. The environments include, among others, a market made of two regions in which students in one region have an opposite ranking over the schools from those in the other region, and a market composed of a finite number of regions where each student in a region draws their preferences from the same distribution but each region has a different ranking.

Related literature

Our paper is most closely related to Roth and Peranson (1999) and Immorlica and Mahdian (2005). Roth and Peranson (1999) conduct a series of simulations on data from the NRMP and on randomly generated data and suggested considering situations where the size of the market is large in comparison to the length of preference lists. Based on randomly generated data, their simulations show that very few students and hospitals could have benefited by submitting false preference lists or by manipulating capacity. These simulations led them to conjecture that the fraction of participants in a two-sided market with random preference lists of limited length who can manipulate tends to zero as the size of the market grows.

Immorlica and Mahdian (2005), which this paper builds upon, theoretically investigate one-to-one matching markets where each college has only one position and show that as the size of the market becomes large, the proportion of colleges that are matched with different students in different stable matchings becomes small. Since a college can manipulate via preference lists if and only if there is more than one student in a stable matching, this result implies that most colleges cannot manipulate preference lists.

While this paper is motivated by these previous studies, there are a number of crucial differences. First, our focus in this paper is on many-to-one markets, which include several real-world markets such as the NRMP and the school choice program in NYC. In such markets colleges can sometimes manipulate via preference lists even if there is only one stable matching.[4] Moreover, in many-to-one markets there exists the additional possibilities of capacity manipulation and manipulation via pre-arrangement which are not

[4]For instance, see the example in Theorem 5.10 of Roth and Sotomayor (1990).

present in a one-to-one market.[5] As a result, having only one set of stable partners in the limit is necessary but not sufficient to explain the lack of manipulability in many-to-one markets. We nevertheless show that the scope of manipulations becomes small in a large market with several technical innovations.

Second, previous research mostly focused on counting the average number of participants that can manipulate the market, assuming that others report their preferences truthfully. This leaves open the question of whether participants will behave truthfully *at the equilibrium*. A substantial part of this paper investigates this question and show that truthful reporting is an approximate equilibrium in a large market that is sufficiently thick.[6]

Finally, both Roth and Peranson (1999) and Immorlica and Mahdian (2005) assume that all student preferences are randomly drawn from the same distribution. We extend the analysis to cases where the distribution of preferences may differ across students in realistic ways, and show that truth-telling is still an equilibrium in a large market that is sufficiently thick.

The use of large market arguments like our approach here is common in the mechanism design literature. For instance, Rustichini, Satterthwaite, and Williams (1994) establish that in a k-double auction where n buyers and sellers draw private values independently and identically distributed, the symmetric, increasing differentiable equilibria are in the limit efficient and convergence is fast.[7] The proofs of these results rely on a symmetric distribution of values. Our paper allows for an arbitrary preference profiles for colleges provided that each student is acceptable to each college, and for significant preference heterogeneity for students. There is also a related literature on the asymptotic analysis of auctions including Pesendorfer and Swinkels (2000) and Swinkels (2001). Most recently, Cripps and Swinkels (2006) relax independence and establish the asymptotic efficiency of large double auctions with private values.

Finally, there is a literature that analyzes the consequences of manipulations via preference lists and capacities in complete information finite economies. See Roth (1984b), Roth (1985) and Sönmez (1997a) for games involving preference manipulation and Konishi and Ünver (2005) and Kojima

[5]Indeed, Roth and Peranson (1999) explicitly investigate the potential for capacity manipulation in their simulations.

[6]Immorlica and Mahdian (2005) claim that truth-telling is an approximate equilibrium in a one-to-one market even without sufficient thickness. In section 4, we present an example to show that this is not the case, but truth-telling is an approximate equilibrium under an additional assumption of sufficient thickness.

[7]There are a number of related papers including Gresik and Satterthwaite (1989) and Fudenberg, Mobius, and Szeidl (2006).

(2005) for games of capacity manipulation.

The paper proceeds as follows. Section 2 presents the model. Section 3 shows the proportion of colleges that can manipulate approaches zero as the market becomes large. Section 4 conducts equilibrium analysis. Section 5 analyzes situations with heterogeneous student preferences. Section 6 discusses extensions of the basic model and alternative matching mechanisms. Section 7 concludes. All proofs are in the Appendix.

2 Model

2.1 Preliminary definitions

A market is tuple $\Gamma = (S, C, P_S, \succ_C)$. S and C are finite and disjoint sets of students and colleges. $P_S = (P_s)_{s\in S}, \succ_C = (\succ_c)_{c\in C}$. For each student $s \in S$, P_s is a strict preference relation over C and being unmatched (being unmatched is denoted by s). For each college, $\succ_c$ is a strict preference relation over the set of subsets of students. If $s \succ_c \emptyset$, then s is said to be **acceptable** to c. Similarly, c is acceptable to s if $cP_s s$. Non-strict counterparts of P_s and $\succ_c$ are denoted by R_s and $\succeq_c$, respectively. Since rankings of only acceptable mates matter for our analysis, we often write only acceptable mates to denote preferences. For example,

$$s_1 : c_1, c_2,$$

means that student s_1 prefers college c_1 most, then c_2, and c_1 and c_2 are the only acceptable colleges.

For each college $c \in C$ and any positive integer q_c, its preference relation $\succ_c$ is **responsive with quota** q_c if (i) for any $s, s' \succ_c \emptyset$, and any $S' \succ_c \emptyset$ with $s, s' \notin S'$, $|S'| < q_c$ we have $s \cup S' \succeq_c s' \cup S' \Leftrightarrow s \succeq_c s'$, (ii) for any $s \in S$ and any $S' \succ_c \emptyset$ with $s \notin S'$ and $|S'| < q_c$, we have $s \cup S' \succeq_c S' \Leftrightarrow s \succeq_c \emptyset$, and (iii) for any $S' \subseteq S$ with $|S'| > q_c$ we have $\emptyset \succ_c S'$ (Roth 1985). That is, the ranking of a student is independent of her colleagues, and any set of students exceeding quota is unacceptable. Let P_c be the corresponding **preference list of college** c, which is the preference relation over singleton sets and the empty set. The non-strict counterpart is denoted by R_c. Sometimes only the preference list structure and quotas are relevant for the analysis. We therefore sometimes abuse notation and denote by $\Gamma = (S, C, P, q)$ an arbitrary market in which the preferences induce preference lists $P = (P_i)_{i\in S\cup C}$ and quotas $q = (q_c)_{c\in C}$. We say that $(S, C, P_S, \succ_C)$ induces (S, C, P, q) in such a case. We also use the following notation; $P_{-i} = (P_j)_{j\in S\cup C\setminus i}$, $q_{-c} = (q_{c'})_{c'\in C\setminus c}, P_C = (P_c)_{c\in C}, P_{C-c} = (P_{c'})_{c'\in C, c'\neq c}$ and so on.

A **matching** μ is a mapping from $C \cup S$ to $C \cup S$ such that (i) for every s, $|\mu(s)| = 1$, and $\mu(s) = s$ if $\mu(s) \notin C$, (ii) $\mu(c) \subseteq S$ for every $c \in C$, and (iii) $\mu(s) = c$ if and only if $s \in \mu(c)$. For any matchings μ and μ', we write $\mu \succ_c \mu'$ if and only if $\mu(c) \succ_c \mu'(c)$ for any $c \in C$, and $\mu P_s \mu'$ if and only if $\mu(s) P_s \mu'(s)$ for any $c \in C$ and $s \in S$.

Given a matching μ, we say that it is **blocked** by (s, c) if s prefers c to $\mu(s)$ and either (i) c prefers s to some $s' \in \mu(c)$ or (ii) $|\mu(c)| < q_c$ and s is acceptable to c. A matching μ is **individually rational** if for each student $s \in S \cup C$, $\mu(s) R_s \emptyset$ and for each $c \in C$ and each $S' \subseteq \mu(c),$, $\mu(c) \succeq_c S'$. A matching μ is **stable** if it is individually rational and is not blocked. A mechanism is a systematic way of assigning students to colleges. A stable mechanism is a mechanism that yields a stable matching for any market.

We consider the following **student optimal stable mechanism (SOSM)**, denoted by ϕ, which is analyzed by Gale and Shapley (1962).[8]

- Step 1: Each student applies to her first choice college. Each college rejects the lowest-ranking students in excess of its capacity and all unacceptable students among those who applied to it, keeping the rest of students temporarily (so students not rejected at this step may be rejected in later steps.)

In general,

- Step t: Each student who was rejected in Step (t-1) applies to her next highest choice. Each college considers these students *and* students who are temporarily held from the previous step together, and rejects the lowest-ranking students in excess of its capacity and all unacceptable students, keeping the rest of students temporarily (so students not rejected at this step may be rejected in later steps.)

The algorithm terminates either when every student is matched to a college or every unmatched student has been rejected by every acceptable college. The algorithm always terminates in a finite number of steps. Gale and Shapley (1962) show that the resulting matching is stable. It is also known that the outcome is the same for different markets $\Gamma = (S, C, (P_s)_{s \in S}, (\succ_c)_{c \in C})$ and $\Gamma' = (S, C, (P_s)_{s \in S}, (\succ'_c)_{c \in C})$ as long as $\succ$ and $\succ'$ induce the same pair of preference lists and quotas P and q. Thus we sometimes write the resulting matching by

$$\phi(S, C, (P_s)_{s \in S}, (P_c)_{c \in C}, (q_c)_{c \in C}), \text{ or } \phi(S, C, P, q).$$

[8]SOSM is known to produce a stable matching that is unanimously most preferred by every student among all stable matchings (Gale and Shapley (1962)).

$\phi(S, C, P, q)(i)$ is the assignment given to $i \in S \cup C$ under matching $\phi(S, C, P, q)$.

2.2 Random markets

To investigate "how likely can a college benefit by manipulation?", we consider the following random environment. A **random market** is a tuple $\tilde{\Gamma} = (C, S, \succ, k, \mathcal{D})$, where k is a positive integer and $\mathcal{D} = (p_c)_{c \in C}$ is a probability distribution on C. We assume that $p_c > 0$ for each $c \in C$.[9] Assume that students in S are ordered in an arbitrarily fixed manner. Each random market induces a market by randomly generating students' preferences. More specifically, for each student $s \in S$, we construct preferences of s over colleges as described below, following Immorlica and Mahdian (2005):

- Step 1: Select a college $c_{(1)}$ independently according to $\mathcal{D}$; add this college $c_{(1)}$ as the top ranked college for student s.

In general,

- Step $t \leq k$: Select college $c_{(t)}$ independently according to $\mathcal{D}$ until a college is drawn that has not been previously drawn in steps 1 through $t-1$. Add $c_{(t)}$ to the end of the preference list for student s.

P_s is constructed by the above procedure, namely,

$$s : c_{(1)}, c_{(2)}, \dots, c_{(k)}.$$

Note that the length of the preference list is a fixed number k. In other words, only k colleges are acceptable.

For example, if $\mathcal{D}$ is the uniform distribution, then the preference list is drawn from the uniform distribution over the set of all lists of size k of colleges. Without loss of generality, we assume the set of colleges C are ordered in decreasing popularity: if $c' < c$, then $p_{c'} \geq p_c$. With abuse of notation, we write $c = m$, $c > m$ and $c < m$ for $m \in \mathbb{N}$ to mean, respectively, that c is the mth college, c is ordered after mth and c is ordered before m. We sometimes write p_m, which is the probability associated with the mth college. For each realization P_S of student preferences, a market with complete information $(C, S, P_S, \succ)$ is obtained.

A sequence of random markets is denoted by $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$, where $\tilde{\Gamma}^n = (C^n, S^n, \succ^n, k^n, \mathcal{D}^n)$ is a random market with $|C^n| = n$ for any n. Consider the following regularity conditions.[10]

[9] We impose this assumption to compare our analysis with existing literature. All our analyses remain unchanged when one allows for probabilities to be zero.

[10] Unless otherwise specified, our convention is that superscripts are used for the number of colleges present in the market whereas subscripts are used for agents.

Definition 1. A sequence of random markets $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is **regular** if there exist positive integers k and $\bar{q}$ such that

(1) $k^n = k$ for all n,

(2) $q_c \leq \bar{q}$ for all n and $c \in C^n$,

(3) $|S^n| \leq \bar{q}n$ for all n, and

(4) for all n and $c \in C^n$, every $s \in S^n$ is acceptable to c.

(1) assumes that the length of students' preference lists does not grow when the number of market participants grow. (2) requires that the number of positions of each college is bounded across colleges and markets. (3) requires that the number of students does not grow much faster than that of colleges. (4) requires colleges to find acceptable any student, but preferences are otherwise arbitrary.[11]

This paper focuses on regular sequences of random markets and makes use of each condition in our arguments. For instance, the conclusion of our result is known to fail if (1) is not satisfied but instead $k^n = n$, that is, students regard every college as acceptable (Knuth, Motwani, and Pittel 1990).[12] The examples mentioned in the introduction motivate our assumption of bounded preference lists. One reason why students do not submit long preferences lists is that it may be costly for them do to so. For example, medical school students in the U.S. have to interview to be considered by residency programs, and financial and time constraints can limit the number of interviews. Likewise, in public school choice, to form preference lists, students need to learn about the programs they may choose from and in many instances they may have to interview or audition for seats.

3 Proportion of colleges that can manipulate

The literature on two-sided matching has focused on two types of manipulations: (1) those within the centralized clearinghouse and (2) those outside the centralized clearinghouse.

[11]Some of the assumptions can be relaxed. For instance, (1) can be relaxed without difficulty to state: for any n and any student, the length of her preference list is at most k, rather than exactly k. Similarly, (3) can be relaxed to state: there exists $\tilde{q}$ such that $N \leq \tilde{q}n$ for any n. We adopt $N \leq \bar{q}n$ just for simplicity.

[12]Roth and Peranson (1999) conduct simulations on random data illustrating this point.

3.1 General manipulation within the centralized market

Dubins and Freedman (1981) show that the SOSM is **manipulable via preference lists**: there exist a market inducing (S, C, P, q), $c \in C$ and some P'_c such that

$$\phi(S, C, (P'_c, P_{-c}), q) \succ_c \phi(S, C, P, q).$$

Roth (1982) further shows that any stable mechanism is manipulable in this way if both colleges and students behave strategically. Despite these negative results, Dubins and Freedman (1981) and Roth (1982) show that students cannot manipulate the SOSM ϕ. Thus it is colleges that can potentially manipulate ϕ.

When colleges have quotas of more than one, other types of manipulations are possible. Sönmez (1997b) shows that SOSM is **manipulable via capacities**: there exist $(S, C, P_s, \succ)$ inducing (S, C, P, q), $c \in C$ and some q'_c such that

$$\phi(S, C, P, (q'_c, q_{-c})) \succ_c \phi(S, C, P, q).$$

It is easy to show that q'_c should be smaller than q_c. Sönmez (1997b) further shows that any stable mechanism is manipulable via capacities.

In many real markets, colleges can manipulate both their preference lists and capacities. Therefore we consider a general manipulation of the form (P'_c, q'_c) of college $c \in C$.

Let

$$\begin{aligned}\alpha_k(n) = E|\{c \in C | \phi(S, C, (P'_c, P_{-c}), (q'_c, q_{-c})) \succ_c \phi(S, C, P, q) \\ \text{for some } (P'_c, q'_c) \text{ in a market induced by } \tilde{\Gamma}^n.\}|.\end{aligned}$$

In words, $\alpha_k(n)$ is the expected number of colleges that can manipulate in the market induced by random market $\tilde{\Gamma}^n$ under ϕ when others report preferences truthfully.

Theorem 1. *Suppose that* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *is regular. Then*

$$\lim_{n \to \infty} \alpha_k(n)/n = 0.$$

This theorem implies that manipulation of any sort within the matching mechanism becomes unprofitable to most colleges, as the number of participating colleges becomes large.

Consider the one-to-one market where each college has only one position. It is well-known that, in one-to-one matching, a college can manipulate the market if and only if it is matched to more than one student in different stable matchings. Therefore the following is an immmediate corollary of Theorem 1, whose setting satisfies our regularity conditions.

Corollary 1 (Theorem 3.1 of Immorlica and Mahdian (2005))**.** *Suppose that* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *satisfies that* $k^n = k$*,* $q_c = 1$ *for all* $c \in C^n$*,* $|S^n| = n$*, and every student is acceptable to every college for any* n*. Then the expected proportion of colleges that are matched to more than one student in different stable matchings converges to zero as* n *approaches infinity.*

All proofs are in the Appendix. Here we explain intuition. First, if a college can manipulate the market, then it can do so by a **dropping strategy**, which simply declares some acceptable students as unacceptable (Lemma 1: Dropping strategies are exhaustive). Thus when considering manipulations, we can restrict attention to a particular class of strategies. Specifically, this lemma implies that successful manipulation will be attained simply by rejecting students who are acceptable under the true preference. This is analogous to a result by Roth and Vande Vate (1991) that preference manipulations in one-to-one markets are exhausted by **truncation strategies**, which agree with the true preference list from the top student down to a certain student, and then declare every other student as unacceptable.

The second step relates dropping strategies to the **rejection chains algorithm** (Algorithm 1). The algorithm proceeds as follows: first, run SOSM under the true preferences. Next, fix a college c and a dropping strategy, say P'_c. Let c reject a set of students who are matched under SOSM but unacceptable under P'_c. Let each of these students, say s, apply to her next best choice, say c'. If s is preferred by c' to one of its current mates, say s', let c' be matched to s and reject s'. Let s' apply to yet another college, and so on. We say that Algorithm 1 **returns to** c if one of the displaced students during the algorithm applies to c. It turns out that if this chain reaction does not return to c under any dropping strategy, then c cannot profitably manipulate by a dropping strategy (Lemma 3: Rejection chains).

Finally, we complete the proof by bounding the probability that the above algorithm does return to the original college c. Suppose that there are a large number of colleges in the market. Then there are also a large number of colleges with vacant positions. We say that a college is popular if it is given a high probability in the distribution on which students preferences are drawn. Any student is much more likely to apply to one of those colleges with vacant positions rather than c unless c is extremely popular in a large market, since there are a large number of such colleges. Since every student is acceptable to any college by assumption, the algorithm terminates without returning to c in such a case. Thus, except for a small proportion of very popular colleges, the probability that the algorithm returns is very small

(Lemma 7: Vanishing market power). Note that the expected proportion of colleges that can manipulate is equal to the sum of probabilities that individual colleges can manipulate. Together with our earlier reasoning in Lemmas 1 and 3, we conclude that the expected proportion of colleges that can successfully manipulate converges to zero when the number of colleges grows.

Our technical innovation over Immorlica and Mahdian (2005) is worth noting. Immorlica and Mahdian (2005)'s argument utilizes two results in the literature. The first is that, in one-to-one markets, a college can profitably manipulate if and only if it is matched to more than one student in different stable matchings. The second is an algorithm to count the number of mates under different stable matchings proposed by Knuth, Motwani, and Pittel (1990). Immorlica and Mahdian (2005) use the latter algorithm to show that most colleges have only one stable mate in a large market, which implies that most colleges cannot manipulate the market by the former fact. While our Algorithm 1 is a natural extension of that of Knuth, Motwani, and Pittel (1990) and Immorlica and Mahdian (2005), the former equivalence does not extend to many-to-one markets. As we mentioned in the introduction, colleges can sometimes manipulate the market even when there is only one stable matching in many-to-one markets. Moreover, colleges have more options to manipulate than in one-to-one markets, not only via preference lists but also via capacities, or via combinations of both. Our proof critically relies on the fact that we only need to consider dropping strategies (Lemma 1). Also we have developed a relationship between the result of an algorithm and the scope for manipulation (Lemma 3). Finally we generalize the technique of Immorlica and Mahdian (2005) (Lemma 7) to bound the probability that each college can manipulate, thereby showing a more general result.

Both Roth and Peranson (1999) and Immorlica and Mahdian (2005) attribute the lack of manipulability to the "core-convergence" property. This means that matching markets have small cores in large markets. While this interpretation is valid in one-to-one markets, smallness of the core is necessary but not sufficient for immunity to manipulation in many-to-one markets. Instead our arguments show that lack of manipulability comes from the "vanishing market power" in the sense that the impact of strategically rejecting a student will be absorbed elsewhere and rarely affects the college that manipulated when the market is large. This intuition underlies the formal mechanics of one of our key lemmas, Lemma 7.

Roth and Peranson (1999) analyze NRMP data and argued that of the 3,000-4,000 participating programs, less than one percent could benefit by truncating preference lists or via capacities, assuming the data are true preferences. They also conducted simulations using randomly generated data in one-to-one matching, and observed that $\alpha_k(n)$ quickly approaches zero as n becomes large. The first theoretical account of this observation is given by Immorlica and Mahdian (2005), who show Corollary 1. Theorem 1 improves upon their results and fully explains observations of Roth and Peranson (1999) in the following senses: (1) it studies manipulations via preference lists in many-to-one markets, and (2) it studies manipulations via capacities, neither of which was previously investigated theoretically. Furthermore we strengthen assertions of Roth and Peranson (1999) and Immorlica and Mahdian (2005) by showing that large markets are immune to *arbitrary* manipulations and not just truncation of preference lists or misreporting capacities.

3.2 Manipulation via pre-arranged matches

When colleges seek more than one student, there is concern for manipulation not only within the matching mechanism, but also outside the formal process. Sönmez (1999) introduces the idea of manipulation via pre-arranged matches. Suppose that c and s arrange a match before the central matching mechanism is executed. Then s does not participate in the centralized matching mechanism and c participates in the centralized mechanism with the number of positions reduced by one. SOSM is **manipulable via pre-arranged matches**, or **manipulable via pre-arrangement**, that is, for some market (S, C, P, q), college $c \in C$ and student $s \in S$ we have

$$\begin{aligned} \phi(S \setminus s, C, P_{-s}, (q_c - 1, q_{-c})) \cup s \succ_c \phi(S, C, P, q), &\text{and} \\ c \succeq_s \phi(S, C, P, q). & \end{aligned}$$

In words, both parties that engage in pre-arrangement have incentives to do so: the student is at least as well off in pre-arrangement as when she is matched through the centralized mechanism, and the college strictly prefers s and the assignment of the centralized mechanism to those without pre-arrangement. Sönmez (1999) shows that any stable mechanism is manipulable via pre-arrangement.

In some markets, matching outside the centralized mechanism is discouraged or even legally prohibited. Even so, the student and college can effectively "pre-arrange" a match by listing each other on the top of their preference lists under stable mechanisms such as SOSM. Thus the scope of manipulation via pre-arrangement is potentially large.

However, we have the following positive result in large markets. Let $\beta_k(n)$ be the expected number of colleges that can manipulate via pre-arrangement in markets induced by $\tilde{\Gamma}^n$ under ϕ when others do not pre-arrange.

Theorem 2. *Suppose that* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *is regular. Then*

$$\lim_{n\to\infty} \beta_k(n)/n = 0.$$

The intuition is similar to that of Theorem 1. It can be shown that any student involved in pre-arrangement under the SOSM is strictly less preferred by c to any student who would be matched in the absence of the pre-arrangement (Lemma 8). Therefore, in order to profitably manipulate, c should be matched to a better set of students in the central matching. By a similar reasoning as Theorem 1, the probability of being matched to better students in the centralized mechanism is small in a large market for most colleges.

4 Equilibrium analysis

The last section investigated individual colleges' incentives to manipulate the market, assuming that others behave truthfully. This section allows all participants to behave strategically and investigates equilibrium behavior in large markets. This section focuses on the simplest case to highlight the analysis of equilibrium behavior. The reader can proceed to Section 5, where we give a more general analysis incorporating heterogeneous distributions of student preferences.

To investigate equilibrium behavior, we first define a normal-form game as follows. Assume that each college $c \in C$ has an additive utility function $u_c : 2^S \to \mathbb{R} \cup \{-\infty\}$ on sets of students. More specifically, we assume that there exists $\hat{u}_c : S \to \mathbb{R} \cup \{-\infty\}$ such that

$$u_c(S') = \begin{cases} \sum_{s\in S'} \hat{u}_c(s) \text{ if } |S'| \leq q_c, \\ -\infty \text{ otherwise.} \end{cases}$$

We have that $sP_cs' \iff \hat{u}_c(s) > \hat{u}_c(s')$. If s is acceptable to c, $\hat{u}_c(s) > 0$. If s is unacceptable, $\hat{u}_c(s) = -\infty$.[13] Further suppose that $\sup_{n\in\mathbb{N}, s\in S^n, c\in C^n} \hat{u}_c(s)$ is finite.

The normal-form game is specified by a random market $\tilde{\Gamma}$ coupled with utility functions $(u_c)_{c\in C}$ and defined as follows. The set of players is C, with von Neumann-Morgenstern expected utility functions induced by the above

[13] We set utility for individually irrational matchings at negative infinity for simplicity. All the results are unchanged as long as the payoff for an unacceptable student is negative and payoff of matchings exceeding the quota is lower than a matching given by a subset meeting the quota.

utility functions. All the colleges move simultaneously. College c submits a preference list and quota pair (P'_c, q'_c) with $0 \leq q'_c \leq q_c$. After the preference profile is submitted, random preferences of students are realized according to the given distribution $\mathcal{D}$. The outcome is the assignment resulting from ϕ under reported preferences of colleges and realized students preferences. We assume that college preferences and distributions of student random preferences are common knowledge, but colleges do not know realizations of student preferences when they submit their preferences.[14] Note that we assume that students are passive players and always submit their preferences truthfully. A justification for this assumption is that truthful reporting is weakly dominant for students under ϕ (Dubins and Freedman 1981, Roth 1982).

Given $\varepsilon > 0$, a profile of preferences $(P'_C, q'_C) = (P'_c, q'_c)_{c \in C}$ is an ε-**Nash equilibrium** if there is no $c \in C$ and (P''_c, q''_c) such that

$$Eu_c(\phi(S, C, (P_S, P''_c, P'_{C-c}), (q''_c, q'_{-c}))) > Eu_c(\phi(S, C, (P_S, P'_C), q')) + \varepsilon,$$

where the expectation is taken with respect to random preference lists of students.

Is truthful reporting an approximate equilibrium in a large market for an arbitrary regular sequence of random markets? The answer is negative, as shown by the following examples.[15]

Example 1. Consider the following market $\tilde{\Gamma}_n$ for any n. $|C^n| = |S^n| = n$. $q_c = 1$ for each $c \in C^n$. Preference lists of c_1 and c_2 are given as follows:[16]

$$c_1 : s_2, s_1, \ldots,$$
$$c_2 : s_1, s_2, \ldots.$$

Suppose that $p^n_{c_1} = p^n_{c_2} = 1/3$ and $p^n_c = 1/3(n-2)$ for any $n \geq 3$ and each $c \neq c_1, c_2$. With probability $[p^n_{c_1} p^n_{c_2}/(1-p^n_{c_1})] \times [p^n_{c_1} p^n_{c_2}/(1-p^n_{c_2})] = 1/36$, preferences of s_1 and s_2 are given by

$$s_1 : c_1, c_2, \ldots,$$
$$s_2 : c_2, c_1, \ldots.$$

[14] Consider a game with incomplete information, in which each college knows other colleges' preferences only probabilistically. The analysis can be easily modified for this environment.

[15] These examples show that Claims 3.1 and 3.3 in Immorlica and Mahdian (2005) are not correct.

[16] "..." in a preference list means that the rest of the preference list is arbitrary after those written explicitly.

Under the student-optimal matching μ, we have $\mu(c_1) = s_1$ and $\mu(c_2) = s_2$. Now, suppose that c_1 submits the following preference list:

$$c_1 : s_2.$$

Then, under the new matching μ', c_1 is matched to $\mu'(c_1) = s_2$, which is preferred to $\mu(c_1) = s_1$. Since the probability of preference profiles where this occurs is $1/36 > 0$, regardless of $n \geq 3$, the opportunity for preference manipulation for c_1 does not vanish even when n becomes large. Therefore truth-telling is not an ε-Nash equilibrium if $\varepsilon > 0$ is sufficiently small and $\hat{u}_{c_1}(s_2)$ is sufficiently higher than $\hat{u}_{c_1}(s_1)$, as c_1 has an incentive to deviate.

The above example shows that, while the *proportion* of colleges who can manipulate via preferences becomes small, *for an individual college* the opportunity for such manipulation may remain large. Note on the other hand that this is consistent with Theorems 1 and 2 since the scope for manipulation becomes small for any $c \neq c_1, c_2$.

The next example shows that, under the same assumptions, manipulations via capacities or pre-arrangement may also be profitable for some colleges even in a large market.

Example 2. Consider the following market $\tilde{\Gamma}_n$ for any n. $|C^n| = |S^n| = n$. $q_{c_1} = 2$ and $q_c = 1$ for each $c \neq c_1$. c_1's preference list is

$$c_1 : s_1, s_2, s_3, s_4, \ldots,$$

and $s_1 \succ_{c_1} \{s_2, s_3\}$.

c_2's preferences are

$$c_2 : s_3, s_1, s_2, \ldots.$$

Further suppose that $p^n_{c_1} = p^n_{c_2} = 1/3$ and $p^n_c = 1/3(n-2)$ for any n and each $c \neq c_1, c_2$.

With the above setup, with probability $[p^n_{c_1} p^n_{c_2}/(1 - p^n_{c_1})] \times [p^n_{c_1} p^n_{c_2}/(1 - p^n_{c_2})]^3 = 1/6^4$, students preferences are given by

$$s_1 : c_2, c_1, \ldots,$$
$$s_2 : c_1, c_2, \ldots,$$
$$s_3 : c_1, c_2, \ldots,$$
$$s_4 : c_1, c_2, \ldots.$$

If everyone is truthful, then c_1 is matched to $\{s_2, s_3\}$. Now

(1) Suppose that c_1 reports a quota of one. Then c_1 is matched to s_1, which is preferred to $\{s_2, s_3\}$.

(2) Suppose that c_1 pre-arranges a match with s_4. Then c_1 is matched to $\{s_1, s_4\}$, which is preferred to $\{s_2, s_3\}$.

Since the probability of preference profiles where this occurs is $1/6^4 > 0$ regardless of $n \geq 3$, the opportunity for manipulations via capacities or pre-arrangement for c_1 does not vanish when n becomes large.[17]

A natural question is under what conditions one can expect a positive result. Let

$$X^*(n;T) = \{c \in C^n | p_1^n/p_c^n \leq T, |\{s \in S^n | cP_s s\}| < q_c\},$$
$$Y^*(n;T) = |X^*(n;T)|,$$

$X^*(n;T)$ is a random set which denotes the set of colleges sufficiently popular ex ante ($p_1^n/p_c^n \leq T$) where there are less potential applicants than number of positions ($|\{s \in S^n | cP_s s\}| < q_c$). $Y^*(n;T)$ is a random variable giving the number of such colleges.

Definition 2. A sequence of random markets is **sufficiently thick** if there exists $T \in \mathbb{R}$ such that

$$E[Y^*(n;T)] \to \infty,$$

as $n \to \infty$.

The condition requires that the expected number of colleges that are desirable enough, yet have fewer potential applicants than seats grows fast enough as the market becomes large. Consider a disruption of the market in which a student becomes unmatched. If the market is thick, such a student is likely to find a seat in another college that has room for her. Thus the condition would imply that a small disruption of the market is likely to be absorbed by a vacant seats.[18] The following is a leading example of thickness.

Example 3 (Nonvanishing proportion of popular colleges)**.** The sequence of random markets is said to have **nonvanishing proportion of popular colleges** if there exists $T \in \mathbb{R}$ and $a \in (0,1)$ such that

$$p_1^n/p_{[an]}^n \leq T,$$

[17]Manipulation via preference list is also possible in this example. Suppose c_1 reports preferences

$$c_1 : s_1, s_4, \ldots.$$

Then c_1 is matched to $\{s_1, s_4\}$, which is preferred to $\{s_2, s_3\}$.

[18]This condition refers to the limit as the size of the market becomes large, so this notion is not relevant to a particular finite market. In particular, thickness and the size of the market are not related. It is even possible that the market does not become "thick" even when the market becomes large, in the sense that the limit in the above definition is finite as n goes to infinity.

where $[x]$ denotes the largest integer that does not exceed x. This condition is satisfied if there are not a small number of colleges which are much more popular than all of the other colleges.

Proposition 1. *The sequence of random markets in Example 3 is sufficiently thick.*

The intuition for this proposition is the following. By assumption, there are a large number of ex ante popular colleges. With high probability, a substantial part of the positions of these colleges will be vacant. This makes the market thick by having a large number of vacant positions in fairly popular colleges in expectation.

There are even thick random markets where the proportion of popular colleges converges to zero, provided that the convergence is sufficiently slow. For example, if there exists $T \in \mathbb{R}$ such that

$$p_1^n / p_{[\gamma n/\ln n]}^n \leq T$$

where $\gamma > 0$ is a sufficiently large constant, the sequence satisfies sufficient thickness.[19]

The above examples are meant to be suggestive and thick markets include other examples of interest, some of which we describe in the next section. The uniform distribution environment examined by Roth and Peranson (1999) with simulations is a special case of Example 3.

Theorem 3. *Suppose that* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *is regular and sufficiently thick. For any* $\varepsilon > 0$*, there exists* n_0 *such that truth-telling is an* ε*-Nash equilibrium for any game with* $n > n_0$*.*

The proof of Theorem 3 is similar to that of Theorem 1 except for one point. We show that when the market is thick, the probability that Algorithm 1 returns is small for every college (Lemma 10: Uniform vanishing market power), as opposed to only for unpopular ones as in Lemma 7. This is because in a large market there are many vacant positions that are popular enough for students to apply to with a high probability and hence terminate the algorithm. In other words, the key difference between Lemma 10 and Lemma 7 is that the former gives an upper bound of manipulability for *every* college, while the latter gives an upper bound only for unpopular colleges. Given this uniform bound, the rest of the proof is analogous to Theorem 1.

A conclusion similar in spirit to Theorem 3 can be derived for prearrangement. Suppose that $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is regular and sufficiently thick. Consider the SOSM. For any $\varepsilon > 0$, there exists n_0 such that for any $n > n_0$ and

[19] The proof is analogous to Example 3 and omitted.

$c \in C^n$, the probability that c can profitably manipulate via pre-arrangement is smaller than ε. The proof is similar to Theorems 2 and 3 and hence omitted.[20]

5 Heterogeneous student preference distributions

So far we have focused on a simple case in which student preferences are drawn from the same distribution. This section extends our analysis to cases in which student preferences are drawn from a number of different distributions.

The model is the same as before except for how student preferences are drawn. We let $\tilde{\Gamma}^n = (C^n, S^n, \succ, k^n, (\mathcal{D}^n(r))_{r=1}^{R^n})$, where R^n is a positive integer. Each random market is endowed with R^n different distributions. To represent student preferences, we partition students into R^n regions, where each student is a member of one and only one region.[21] Write $\mathcal{D}^n(r) = (p_c^n(r))_{c \in C^n}$ as the probability distribution on C^n for students in region r. For each student $s \in S^n$ in region r, we construct preferences of s over colleges as described below:

- Step 1: Select a college $c_{(1)}$ independently according to $\mathcal{D}^n(r)$; add this college as the top ranked college for student s.

In general,

- Step $t \leq k$: Select college $c_{(t)}$ independently according to $\mathcal{D}^n(r)$ until a college is drawn that has not been previously drawn in steps 1 through $t-1$. Add this $c_{(t)}$ to the end of the preference list for student s.

The case with $R^n = 1$ corresponds to our earlier model with one distribution.

Our regularity assumptions extend naturally: in addition to conditions in the previous definition, we also require that, for some positive integer R, $R^n = R$ for every n in a regular market. Finally, our assumption of sufficient thickness generalizes easily to the current environment. Let

$$X^*(n;T) = \{c \in C^n | p_1^n(r)/p_c^n(r) \leq T \quad \text{for all } r, |\{s \in S^n | cP_s s\}| < q_c\},$$
$$Y^*(n;T) = |X^*(n;T)|.$$

[20] The result on pre-arrangement can also be generalized to cases where student preferences are drawn from several distributions, as discussed in Section 5.

[21] We frame the heterogeneity of student preferences in terms of multiple regions where students live. Of course alternative interpretations are possible, such as heterogeneity depending on gender, race or academic performance or combinations of these characteristics.

Definition 3. A sequence of random markets is **sufficiently thick** if there exists $T \in \mathbb{R}$ such that

$$E[Y^*(n;T)] \to \infty,$$

as $n \to \infty$.

Definition 3 is a multi-region generalization of sufficient thickness for one-region setting (Definition 2). The following examples satisfy this version of sufficient thickness.

Example 4 (Two regions with opposite popularity)**.** There are two regions, $R = \{1, 2\}$. $C^n = \{1, 2, \ldots, n\}$ and the probability distributions are:

$$p_c^1(1) = \frac{n - c + 1}{\sum_{c' \in C^n} n - c' + 1} = \frac{n - c + 1}{\frac{n(n+1)}{2}},$$
$$p_c^1(2) = \frac{c}{\sum_{c' \in C^n} c'} = \frac{c}{\frac{n(n+1)}{2}}.$$

Students in the first region prefer the first college over the second college and so forth on average, while students in the second region have the opposite preferences. There is an extreme form of differences in preferences in this market.

Example 5 (Multiple regions with within-region symmetry)**.** Assume there are R regions, $R \geq 2$. Each college is based in one of the regions. Let $r(c)$ be the region in which college c is. Let $\tilde{p}_m(r)$, $r, m \in \{1, \ldots, R\}$ be strictly positive for every r, m. From this, we define the probability $p_c^n(r)$ for any $n \in \mathbb{N}$ as follows:

$$p_c^n(r) = \frac{\tilde{p}_{r(c)}(r)}{\sum_{m \in R} \tilde{p}_m(r) \nu_m^n},$$

where $\nu_m^n = \{c \in C^n | r(c) = m\}$ denotes the number of colleges in $\tilde{\Gamma}^n$ that is based in region m.

This environment has the following interpretation. Each college is based in one of the regions, and each student lives in one region. Colleges in a given region are equivalent to one another. The "base popularity" of a college in region m for a student living in region r is given by $\tilde{p}_m(r)$. Then we normalize these to obtain $p_c^n(r)$ by the above equation. For any pair of colleges c and c' and region r, we have that

$$p_c^n(r)/p_{c'}^n(r) = \tilde{p}_{r(c)}(r)/\tilde{p}_{r(c')}(r).$$

Such heterogeneous preferences may be present in labor markets or in large urban school districts, where students in the same region have similar preferences while substantial differences are present across regions.

Proposition 2. *Sequences of random markets in Examples 4 and 5 are sufficiently thick.*

These are among the simplest examples incorporating heterogeneity. The two region case shows that directly opposing preferences satisfy sufficient thickness. The multiple region case illustrates that a great deal of heterogeneity in student preferences are allowed.

The equilibrium analysis in the one-region setting (Theorem 3) extends to heterogeneous preference distributions such that the market is sufficiently thick.

Theorem 4. *Suppose that $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is regular and sufficiently thick. For any $\varepsilon > 0$, there exists n_0 such that truth-telling is an ε-Nash equilibrium for any game with $n > n_0$.*

6 Discussion

6.1 Manipulations by coalitions

The basic model shows that individual colleges have little opportunity to manipulate a large market. One natural question is whether coalitions of colleges can manipulate by coordinating their reports. Formally, a coalition $C' \subseteq C$ manipulates the market (S, C, P, q) if there exists $(P'_{C'}, q'_{C'}) = (P'_c, q'_c)_{c \in C'}$ such that

$$\phi(S, C, (P'_{C'}, P_{-C'}), (q'_{C'}, q_{-C'})) \succ_c \phi(S, C, P, q),$$

for some $c \in C'$.

Theorem 5. *Suppose that $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is regular and sufficiently thick.*[22] *Consider the SOSM. Then, for any positive integer m and any $\varepsilon > 0$, there exists n_0 such that for any $n > n_0$ and $C' \subseteq C^n$ with $|C'| \leq m$, the probability that C' can profitably manipulate is smaller than ε.*

The notion of coalitional manipulation we consider allows for a broad range of coalitions, for a coalition is said to manipulate even if only some of its members are made strictly better off and others in the coalition are made

[22] Without the assumption of sufficient thickness, we can obtain a correspondingly weaker conclusion.

strictly worse off when they misreport their preferences jointly. Our result shows that successful coalitional manipulation is rare: with high probability, not a single college in the coalition is made strictly better off. Thus it is hard for coalitions to manipulate even when monetary transfers are possible among colleges.

6.2 The Boston mechanism

Our results have established the virtues of the student optimal stable mechanism in a large market. This finding may serve as one criterion to support its use as a market design, since other mechanisms may not share the same properties in a large market. To see this point, consider the so-called **Boston mechanism** (Abdulkadiroğlu and Sönmez 2003), which is often used for real-life matching markets. The Boston mechanism is a priority matching mechanism, where school priorities are interpreted as preferences.[23] When colleges are asked to rank students, the mechanism proceeds as follows:

- Step 1: Each student applies to her first choice college. Each college rejects the lowest-ranking students in excess of its capacity and all unacceptable students.

In general,

- Step t: Each student who was rejected in the last step proposes to her next highest choice. Each college considers these students, *only as long as* there are vacant positions not filled by students who are already matched by the previous steps, and rejects the lowest- ranking students in excess of its capacity and all unacceptable students.

The algorithm terminates either when every student is matched to a college or every unmatched student has been rejected by every acceptable college. The algorithm always terminates in a finite number of steps.

Note the difference between this mechanism and SOSM. At each step of the Boston mechanism, students who are not rejected are *guaranteed* positions; the match of these students and colleges are permanent rather than temporary, unlike in the student-optimal stable mechanism.

Under the Boston mechanism, it turns out that colleges have no incentive to manipulate via preferences nor via capacity *even in a small market with an arbitrary preference profile*. More specifically,

[23] With slight abuse of terminology we will refer to this class of priority mechanisms where colleges rank students as the Boston mechanism even though the Boston mechanism was introduced as a one-sided matching mechanism.

Remark 1. *Suppose that the Boston mechanism is employed, and preferences are drawn arbitrarily. Then no college can manipulate by reporting* (P'_c, q'_c) *different from its true preference* (P_c, q_c)*, regardless of what other colleges do.*

This is a slight extension of the result shown by Ergin and Sönmez (2006), who show that colleges cannot manipulate via preference lists under the Boston mechanism.[24] While colleges have incentives to report their true preferences, we argue that this mechanism performs badly both in small and large markets. The problem is that students have incentives to misrepresent their preferences, and there is evidence that some participants react to these incentives (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2006). The following example shows that students have incentives to manipulate the mechanism even in large markets.

Example 6. Consider market $\tilde{\Gamma}^n$, where $|S^n| = |C^n| = n$ for each n. $q_c = 1$ for every n and $c \in C^n$. Preference lists are common among colleges and given by

$$c : s_1, s_2, \dots, s_n,$$

for every $c \in C^n$.

$p^n_{c_1} = (1/2)^{1/n}, p^n_{c_2} = (1-(1/2)^{1/n})(1/2)^{1/n}$, and $p^n_c = (1-p^n_{c_1}-p^n_{c_2})/(n-2)$ for each $c \neq c_1, c_2$. Then, with probability $[p^n_{c_1}p^n_{c_2}/(1-p^n_{c_1})]^n = 1/4$, students preferences are

$$s : c_1, c_2, \dots,$$

for each $s \in S^n$. If every student is truth-telling, then s_1 and s_2 are matched to c_1 and c_2, respectively, and other students are matched to their third or less preferred choices. If $s \neq s_1, s_2$ deviates from truth-telling unilaterally and reports preference list

$$s : c_2, \dots,$$

then s is matched to her second choice c_2, which is preferred to the match under truth-telling. This occurs with probability of at least 1/4, and every student except s_1 and s_2 has an incentive not to be truth-telling.

7 Conclusion

Why do many stable matching mechanisms work in practice even though the theory suggests that they can be manipulated in many ways? This paper established that the fraction of participants who can profitably manipulate

[24] The proof of this assertion is a straightforward extension of Ergin and Sönmez (2006) and hence omitted.

the student-optimal stable mechanism in a large two-sided matching market is small under some regularity conditions. We further showed that truthful reporting is an approximate equilibrium in large markets that are sufficiently thick. Since a stable matching is efficient, this paper suggests that large matching markets achieve a high level of efficiency.

Is convergence fast enough for our results to be useful in applications? The answer to this question will depend on the particular institutional features of the market and the distribution of preferences. Proofs in Appendix establish that the order of convergence of the probability of profitable manipulation is $O(1/n)$ in several interesting cases (Examples 3, 4 and 5). Further investigation of the rate of convergence is left for future research.

Most of our analysis focused on the student-optimal stable mechanism, the mechanism that forms the core of the NRMP and the NYC high school choice plan. It would be interesting to know what types of limit results apply to other stable mechanisms. Our analysis can be extended to show that the scope of manipulations by colleges decreases as the market becomes large in any stable mechanism. However, since students may also have incentives to manipulate in an arbitrary stable mechanism, this type of extension would need to examine when the opportunity for students to successfully manipulate is small. Such a result might provide some guidance in the selection of a matching mechanism within the class of stable ones.

A Appendix: Proofs

A.1 Proof of Theorem 1

We prove Theorem 1 through several steps. Specifically, we prove three key lemmas, Lemmas 1, 3 and 7 and then use them to show the theorem.

A.1.1 Lemma 1: Dropping strategies are exhaustive

Let (P_c, q_c) be a pair of the true preference list and true quota of college c. A report (P'_c, q'_c) is said to be a **dropping strategy** if (i) $q'_c = q_c$, (ii) sP_cs' and $sP'_c\emptyset$ imply sP'_cs', and (iii) $\emptyset P_c s$ implies $\emptyset P'_c s$. In words, a dropping strategy simply declares some students who are acceptable under P_c as unacceptable. In particular, it does not change quotas or change the relative ordering of acceptable students or declare unacceptable students as acceptable.

Lemma 1 (Dropping strategies are exhaustive). *Consider an arbitrary stable mechanism. Fix preferences of colleges other than c. Suppose the mechanism produces μ under some arbitrary report of c. Then there exists a dropping* *strategy* (P'_c, q_c) *such that* $\mu' \succeq_c \mu$ *according to the true preferences of* c*, where* μ' *is the matching induced by* (P'_c, q_c) *under the stable mechanism.*

Proof. Denote the strategy profile of colleges other than c by (P_{-c}, q_{-c}). Let dropping strategy (P'_c, q_c) be such that P'_c lists all the students in $\mu(c) \cap \{s \in S | s P_c \emptyset\}$ as acceptable in the same relative order as in P_c, and report every other student as unacceptable. We will show that $\mu'(c) = \mu(c) \cap \{s \in S | s P_c \emptyset\}$ where μ' is the matching under SOSM with this dropping strategy. This implies that $\mu' \succeq_c \mu$, since $\mu'(c)$ can only differ from $\mu(c)$ in having no unacceptable students under the true preference list P_c. Let $(P', q) = (P'_c, P_{-c}, q)$.

Consider a matching μ'' defined by

$$\mu''(c') = \begin{cases} \mu(c) \cap \{s \in S | s P_s \emptyset\} & c' = c, \\ \mu(c') & c' \neq c. \end{cases}$$

That is, μ'' is a matching obtained by letting c reject students in $\mu(c)$ that are unacceptable under P_c. Consider its property under (P', q). Clearly μ'' is individually rational under (P', q). There is no blocking pair involving c, since $\mu''(c)$ is exactly the set of students acceptable under P'_c. Since μ is stable under (P, q) and preferences are unchanged between (P, q) and (P', q) for any $c' \neq c$, the only blocking pairs for μ'' involves $\mu(c) \setminus \mu''(c)$, who are unmatched under μ''. These imply that μ'' is student-quasi-stable (Blum, Roth, and Rothblum 1997) or, equivalently, simple (Sotomayor 1996). Theorem 4.3 of Blum, Roth, and Rothblum (1997) implies that given any student-quasi-stable matching, there exists a stable matching that is weakly preferred by every college. Let μ''' be one such stable matching under (P', q). Since $\mu''(c)$ exhausts all the acceptable students under P'_c and hence $\mu''(c)$ is the most preferred matching for c under (P'_c, q_c), $\mu'''(c) = \mu''(c)$.

So far we have shown that c is matched to $\mu''(c)$ under a matching μ''', which is stable under (P', q). Roth (1984a) shows that, for any college, the same number of students are matched to it across different stable matchings. By this result, c is matched to the same number of students under μ' as in μ''', both of which are stable under (P', q). Since there are just $|\mu'''(c)|$ acceptable students under P'_c, this implies that $\mu'(c) = \mu'''(c) = \mu''(c)$, completing the proof. $\square$

Lemma 1 simplifies the analysis by enabling us to focus on a particular class of strategies to investigate manipulations. This lemma is analogous to a result by Roth and Vande Vate (1991) that an agent can manipulate by truncation strategies whenever she can do so by some strategy in one-to-one markets. Lemma 1 is of independent interest, for this is a first result on restricting profitable strategies in many-to-one settings, and the conclusion holds for an arbitrary stable mechanism while most of our analysis focuses on a particular stable mechanism, SOSM.

A.1.2 Lemma 3: Rejection chains

Suppose that SOSM is run and the stable matching μ is obtained. Let B_c^1 be an arbitrary subset of $\mu(c)$. The **rejection chains** associated with B_c^1 is defined as follows.

Algorithm 1. Rejection Chains

(1) Initialization:

   (a) Let the student-optimal stable matching μ be the initial match of the algorithm. Let B_c^1 be a given subset of $\mu(c)$. Let $i = 0$. Let c reject all the students in B_c^1.

(2) Increment i by one.

   (a) If $B_c^i = \emptyset$, then terminate the algorithm.

   (b) If not, let s be the least preferred student by c among B_c^i, and let $B_c^{i+1} = B_c^i \setminus s$.

   (c) Iterate the following steps (call this iteration "Round i".)

      i. Choosing the applied:

         A. If s has already applied to every acceptable college, then finish the iteration and go back to the beginning of Step 2.

         B. If not, let c' be the most preferred college of s among those which s has not yet applied while running SOSM or previously within this algorithm. If $c' = c$, terminate the algorithm.

      ii. Acceptance and/or rejection:

         A. If c' prefers each of its current mates to s and there is no vacant position, then c' rejects s; go back to the beginning of Step 2c.

         B. If c' has a vacant position or it prefers s to one of its current mates, then c' accepts s. Now if c' had no vacant position before accepting s, then c' rejects the least preferred student among those who were matched to c. Let this rejected student be s and go back to the beginning of Step 2c. If c' had a vacant position, then finish the iteration and go back to the beginning of Step 2.

Algorithm 1 terminates either at Step 2a or at Step 2(c)iB. We say that Algorithm 1 **returns to c** if it terminates at Step 2(c)iB and **does not return to c** if it terminates at Step 2a.

Suppose that realization of students preferences are such that Algorithm 1 associated with every initial B_c^1 fails to return to c. Given any dropping strategy (P_c', q_c), there exists B_c^1 such that $B_c^1 = \{s \in \mu(c) | \emptyset P_c' s\}$. Let μ' be a matching obtained at the end of Algorithm 1 associated with $B_c^1 = \{s \in \mu(c) | \emptyset P_c' s\}$.

Lemma 2. *Under* (S, C, P_c', P_{-c}, q)*,*

(1) μ' *is individually rational,*

(2) *no* $c' \neq c$ *is a part of a blocking pair of* μ'*, and*

(3) *if* (s, c) *blocks* μ'*, then* $\arg\min_{P_c} \mu(c) P_c s$ *and* $\arg\min_{P_c'} \mu'(c) P_c' s$.[25]

Proof. Part (1): For $c' \neq c$, c' only accepts students who are acceptable in each step of SOSM and Algorithm 1. c rejects every student who is unacceptable under P_c' at the outset of Algorithm 1, and accepts no other student by the assumption that Algorithm 1 does not return to c. Therefore μ' is individually rational.

Part (2): Suppose that for some $s \in S$ and $c' \in C$, $c' P_s \mu'(s)$. Then, by the definition of SOSM and Algorithm 1, s is rejected by c' either during SOSM or in Algorithm 1. For any $c' \neq c$, this implies that $|\mu'(c')| = q_{c'}$ and $\arg\min_{P_{c'}} \mu'(c) P_{c'} s$, implying that s and c' does not block μ'.

Part (3): Suppose $c P_s \mu'(s)$ for some $s \in S$. As in Part (2), this implies that s is rejected by c' either during SOSM or in Algorithm 1. Since Algorithm 1 does not return to c by assumption, s is rejected either during SOSM or at the beginning of Algorithm 1. (s, c) is not a blocking pair in the latter case since s is declared unacceptable under P_c'. In the former case, the fact that s is rejected during SOSM implies that $\arg\min_{P_c} \mu(c) P_c s$. Finally, condition (ii) of the definition of dropping strategies implies that $\arg\min_{P_c'} \mu'(c) P_c' s$ since (P_c', q_c) is a dropping strategy, $\arg\min_{P_c'} \mu'(c) P_c s$ and $\arg\min_{P_c'} \mu'(c) P_c' \emptyset$. ∎

Lemma 3 (Rejection chains)**.** *For any market and any* $c \in C$*, if Algorithm 1 does not return to c for any* $B_c^1 \subseteq \mu(c)$*, then c cannot profitably manipulate by a dropping strategy.*

[25] For any binary relation R on X and and $X' \subseteq X$, $\arg\min_R X' = \{x \in X' | yRx \text{ for any } y \in X'\}$.

Proof. Consider an arbitrary dropping strategy (P'_c, q_c). By assumption, Algorithm 1 associated with $B^1_c = \{s \in \mu(c) | \emptyset P'_c s\}$ does not return to c. Let μ' be the matching resulting from Algorithm 1 associated with $B^1_c = \{s \in \mu(c) | \emptyset P'_c s\}$.

Let c block μ' and form a new matching μ'' by admitting its most preferred students who are willing to be matched, that is,

$$\mu''(c') = \begin{cases} \mu'(c) \cup \arg\max_{(P'_c, q_c - |\mu'(c)|)} \{s \in S | c P_s \mu'(s)\} & c' = c, \\ \mu'(c') \setminus \mu''(c) & c' \neq c, \end{cases}$$

where $\arg\max_{(P'_c, q_c - |\mu'(c)|)} X$ denotes $q_c - |\mu'(c)|$ students that are most preferred under P'_c in set X. Lemma 2 implies that μ'' is firm-quasi-stable matching (Blum, Roth, and Rothblum 1997), such that c is not a part of a blocking pair for μ''. Theorem 5.6 of Blum, Roth, and Rothblum (1997) implies that there exists a stable matching μ''' such that every position of colleges is matched to a weakly less preferred student than at μ'', if the college is not a blocking pair for μ''. Since the matching produced by SOSM is the weakly least preferred by every college among stable matchings, $\phi(S, C, P'_c, P_{-c}, q)(c)$ is composed of even less preferred students than $\mu'''(c)$.

Now recall that $\mu'(c) \subseteq \mu(c)$ by definition. Moreover, Part (3) of Lemma 2 shows that, under μ'', $|\mu'(c)|$ positions of c are filled with $\mu'(c)$ and the remaining $|\mu(c) - \mu'(c)|$ positions are filled with students less preferred to $\arg\min_{P_c} \mu(c)$. As shown above, $\mu'''(c)$ and $\phi(S, C, P'_c, P_{-c}, q)(c)$ are even less preferred under P'_c, and (since every matched student is acceptable) under P_c. These imply that $\mu \succeq_c \phi(S, C, P'_c, P_{-c}, q)$, showing that (P'_c, q_c) is not a profitable strategy. □

A.1.3 Lemma 7: Vanishing market power

We are interested in how often the algorithm ends at Step 2(c)iB, as a student draws c from distribution $\mathcal{D}^n$. Let

$$\pi_c = \Pr[\text{Algorithm 1 returns to c for some } B^1_c \subseteq \mu(c)].$$

Since Algorithm 1 returns to c for some B^1_c whenever c can manipulate SOSM (Lemmas 1 and 3), π_c gives an upper bound of the probability that c can manipulate SOSM when others are truthful conditional on μ being realized as the matching under SOSM. Here we will show Lemma 7, which bounds π_c for most colleges in large markets.

Consider the following algorithm, which is a stochastic variant of the SOSM.[26]

[26] To be more precise this is a stochastic version of the algorithm proposed by McVitie

Algorithm 2. Stochastic Student-Optimal Gale-Shapley Algorithm

(1) Initialization: Let $l = 1$. For every $s \in S$, let $A_s = \emptyset$.

(2) Choosing the applicant:

  (a) If $l \leq N$, then let s be the l'th student and increment l by one.[27]

  (b) If not, then terminate the algorithm.

(3) Choosing the applied:

  (a) If $|A_s| \geq k$, then return to Step 2.

  (b) If not, select c randomly from distribution $\mathcal{D}^n$ until $c \notin A_s$, and add c to A_s.

(4) Acceptance and/or rejection:

  (a) If c prefers each of her current mates to s and there is no vacant position, then c rejects s. Go back to Step 3.

  (b) If c has a vacant position or it prefers s to one of its current mates, then c accepts s. Now if c had no vacant position before accepting s, then c rejects the least preferred student among those who were matched to c. Let this student be s and go back to Step 3. If c had a vacant position, then go back to Step 2.

A_s records colleges that s has already drawn from $\mathcal{D}^n$. When $|A_s| = k$ is reached, A_s is the set of colleges acceptable to s.

Under SOSM, a student's application to her tth most preferred college is independent of her preferences after $(t+1)$th choice on. Therefore the above algorithm terminates, producing the student-optimal stable matching of any realized preference profile which would follow from completing the draws for random preferences. Let μ be the student-optimal stable matching obtained by the above algorithm.

Suppose that Algorithm 2 is run and the stable matching μ is obtained. Now fix a college $c \in C$ and let B_c^1 be an arbitrary subset of $\mu(c)$. The **stochastic rejection chains** associated with B_c^1 is defined as follows. As the name suggests, this is a stochastic version of Algorithm 1.

Algorithm 3. Stochastic Rejection Chains

and Wilson (1970), which they show produces the same matching as the original SOSM proposed by Gale and Shapley (1962).

[27] Recall that students are ordered in an arbitrarily fixed manner.

(1) Initialization:

   (a) Keep all the preference lists generated in Algorithm 2, that is, for each $s \in S$, let A_s be the one generated at the end of Algorithm 2. Let the student-optimal match μ be the initial match of the algorithm. Let B_c^1 be a given subset of $\mu(c)$. Let $i = 0$. Let c reject all the students in B_c^1.

(2) Increment i by one.

   (a) If $B_c^i = \emptyset$, then terminate the algorithm.

   (b) If not, let s be the least preferred student by c among B_c^i, and let $B_c^{i+1} = B_c^i \setminus s$.

   (c) Iterate the following steps (call this iteration "Round i".)

      i. Choosing the applied:

         A. If $|A_s| \geq k$, then finish the iteration and go back to the beginning of Step 2.

         B. If not, select c' randomly from distribution $\mathcal{D}^n$ until $c' \notin A_s$, and add c' to A_s. If c is selected, terminate the algorithm.

      ii. Acceptance and/or rejection:

         A. If c' prefers each of its current mates to s and there is no vacant position, then c' rejects s; go back to the beginning of Step 2c.

         B. If c' has a vacant position or it prefers s to one of its current mates, then c' accepts s. Now if c' had no vacant position before accepting s, then c' rejects the least preferred student among those who were matched to c. Let this rejected student be s and go back to the beginning of Step 2c. If c' had a vacant position, then finish the iteration and go back to the beginning of Step 2.

Algorithm 3 terminates either at Step 2a or at Step 2(c)iB. Similarly to Algorithm 1, we say that Algorithm 1 **returns to c** if it terminates at Step 2(c)iB and **does not return to c** if it terminates at Step 2a.

We are interested in how often the algorithm returns to c, as a student draws c from distribution $\mathcal{D}^n$. It is clear that the probability that Algorithm 1 returns to c is equal to the probability that Algorithm 3 returns to c. That is,

$$\pi_c = \Pr[\text{Algorithm 3 returns to } c \text{ for some } B_c^1 \subseteq \mu(c)].$$

This latter expression is useful since we can investigate the procedure step by step, utilizing conditional probabilities and conditional expectations. Let

$X_c = \{c' \in C | c' \leq c, c' \notin A_s$ for every $s \in S$ at the end of Algorithm 2$\}$, and
$Y_c = |X_c|$.

X_c is a random set of colleges that are more popular than c ex ante but listed on no student's preference list at the end of Algorithm 2. Y_c is a random variable indicating the number of such colleges.[28]

Lemma 4. *For any $c > 4k$, we have*

$$E[Y_c] \geq \frac{c}{2} e^{-\frac{8\bar{q}nk}{c}}.$$

Proof. Let $Q^n = \sum_{c=1}^{k} p_c^n$. Then the probability that c' is not a student's i'th choice given her first $(i-1)$ choices $c_{(1)}, \ldots, c_{(i-1)}$ is bounded as follows;

$$1 - \frac{p_{c'}^n}{1 - \sum_{j=1}^{i-1} p_{c_{(j)}}^n} \geq 1 - \frac{p_{c'}}{1 - Q^n}.$$

Let $E_{c'}$ be the event that $c' \notin A_s$ for every $s \in S$. Since there are at most $\bar{q}nk$ draws from $\mathcal{D}^n$ in Algorithm 2, the above inequality implies that

$$\Pr(E_{c'}) \geq \left(1 - \frac{p_{c'}}{1 - Q^n}\right)^{\bar{q}nk}.$$

Now if $c' > 2k$ we have

$$p_{c'}^n \leq \frac{1 - Q^n}{c' - k}.$$

Therefore for any $c' > 2k$ we have

$$\Pr(E_{c'}) \geq \left(1 - \frac{1}{c' - k}\right)^{\bar{q}nk} \geq e^{-\frac{2\bar{q}nk}{c'-k}} \geq e^{-\frac{4\bar{q}nk}{c'}}.$$

Combining these inequalities, for any $c > 4k$, we have

$$E[Y_c] \geq \sum_{c'=1}^{c} \Pr(E_{c'}) \geq \sum_{c'=2k}^{c} e^{-\frac{4\bar{q}nk}{c'}} \geq \sum_{c'=c/2}^{c} e^{-\frac{8\bar{q}nk}{c}} = \frac{c}{2} e^{-\frac{8\bar{q}nk}{c}}.$$

□

[28] We abuse notation and denote random variable and its realization by the same letter when there is no confusion.

For $B_c^1 \subseteq \mu(c)$, let

$$\pi_c^{B_c^1} = \Pr[\text{Algorithm 3 associated with } B_c^1 \text{ returns to c}|Y_c > EY_c/2, \mu].$$

$\pi_c^{B_c^1}$ gives an upper bound of the probability that c can manipulate SOSM when others are truthful, conditional on μ being realized and there are not too small a number of colleges ($Y_c > EY_c/2$) that are more popular than c and appear nowhere on students' preference lists at the end of Algorithm 2.

Let $c^*(n) = 16\bar{q}nk/\ln(\bar{q}n)$. As it turns out in the sequel, $c^*(n)$ is a number of "very popular colleges" in a market with n colleges. Note that $c^*(n)/n$ converges to zero as $n \to \infty$, so the proportion of such colleges goes to zero. Except for these $c^*(n)$ colleges, the following lemma gives a useful upper bound for manipulability in a large market.

Lemma 5. *Suppose that n is sufficiently large and $c > c^*(n)$. Then we have*

$$\pi_c^{B_c^1} \leq \frac{4\bar{q}}{EY_c},$$

for any $B_c^1 \subseteq \mu(c)$.

Proof. Consider Round 1, beginning with the least preferred student s of $B_c^1 \subseteq \mu(c)$ (if $B_c^1 = \emptyset$, then the inequality is obvious since $\pi_c^{B_c^1} = 0$.). Since $p_{c'}^n \geq p_c^n$ for any $c' \in X_c$, Round 1 ends at 2(c)iiB as a student applies to some college with vacant positions, at least with probability $1 - \frac{1}{Y_c/2+1} > 1 - \frac{1}{EY_c/2+1}$.

Now assume that all Rounds $1, \ldots, i$ end at Step 2(c)iiB. Then there are still at least $Y_c - i$ colleges more popular than c and with a vacant position, since at most i colleges in X_c have had their positions filled at Rounds $1, \ldots, i$. Therefore Round $(i+1)$ initiated by the least preferred student in B_c^{i+1} ends at Step 2(c)iiB with probability of at least $1 - \frac{1}{EY_c/2-i+1}$. Since there are at most $\bar{q}$ rounds, Algorithm 3 fails to return to c with probability of at least

$$\prod_{i=1}^{\bar{q}} \left(1 - \frac{1}{EY_c/2 - (i-1) + 1}\right) \geq \left(1 - \frac{1}{EY_c/4}\right)^{\bar{q}},$$

for every sufficiently n, since $EY_c/2 - (i-1) + 1 \geq EY_c/4 > 0$ for any sufficiently large n and $c > c^*(n)$ by Lemma 4. Therefore we have that

$$\pi_c^{B_c^1} \leq 1 - \left(1 - \frac{1}{EY_c/4}\right)^{\bar{q}} \leq \frac{4\bar{q}}{EY_c},$$

where the last inequality holds since $1-(1-x)^y \leq yx$ for any $x \in (0,1)$ and $y \geq 1$.[29] $\square$

We state without proof the following lemma (this is a straightforward generalization of Lemma 4.4 of Immorlica and Mahdian (2005)).

Lemma 6. *For every c, we have $Var(Y_c) \leq E[Y_c]$.*

Now we are ready to present and prove the last of the three key lemmas.

Lemma 7 (Vanishing market power)**.** *If n is sufficiently large and $c > c^*(n)$, then*

$$\pi_c \leq \frac{[\bar{q}(2^{\bar{q}}-1)+1]\ln(\bar{q}n)}{2k\sqrt{\bar{q}n}}.$$

Proof. By the Chebychev inequality, Lemma 6 and the fact that any probability is less than or equal to one, we have

$$\Pr\left[Y_c \leq \frac{EY_c}{2}\right] \leq \Pr\left[|Y_c - E[Y_c]| \geq \frac{EY_c}{2}\right] \leq \frac{Var(Y_c)}{(E[Y_c]/2)^2} \leq \frac{4}{E[Y_c]}.$$

Since the probability of a union of events is at most the sum of the probabilities of individual events (Boole's inequality), Lemma 5 implies

$$\begin{aligned}
&\Pr[\text{Algorithm 3 returns to c for some } B_c^1 \subseteq \mu(c) | Y_c \geq EY_c/2, \mu] \\
&\leq \sum_{B_c^1 \subseteq \mu(c)} \pi_c^{B_c^1} \\
&\leq 4\bar{q}(2^{\bar{q}}-1)/EY_c.
\end{aligned}$$

This inequality holds for any matching μ under SOSM. Therefore, we have the same upper bound for probability conditional on $Y_c > EY_c/2$ but unconditional on μ, that is,

$$\begin{aligned}
&\Pr[\text{Algorithm 3 returns to c for some } B_c^1 \subseteq \mu(c) | Y_c \geq EY_c/2] \\
&\leq 4\bar{q}(2^{\bar{q}}-1)/EY_c.
\end{aligned}$$

By the above inequalities and the fact that probabilities get value of at most one,

$$\begin{aligned}
\pi_c &\leq \Pr[Y_c \leq EY_c/2] + \Pr[Y_c > EY_c/2] \times 4\bar{q}(2^{\bar{q}}-1)/EY_c \\
&\leq \frac{4}{EY_c} + 4\bar{q}(2^{\bar{q}}-1)/EY_c \\
&\leq \frac{4(\bar{q}(2^{\bar{q}}-1)+1)}{EY_c}.
\end{aligned}$$

[29] Note that conditions for this inequality is satisfied since $4/EY_c \in (0,1)$ for any sufficiently large n and $c > c^*(n)$.

Applying Lemma 4 and noting that EY_c is increasing in c so $EY_{c^*(n)} \leq EY_c$ for any $c > c^*(n)$, we complete the proof of Lemma 7.[30] $\square$

A.1.4 Theorem 1

Now we prove Theorem 1. By Lemma 1, it suffices to consider dropping strategies. By Lemma 3, the probability that $c \in C$ can successfully manipulate by some dropping strategy is at most π_c. Using Lemma 7, we obtain

$$\begin{aligned}
\alpha_k(n) &= \sum_{c \in C^n} \Pr[c \text{ successfully manipulates}] \\
&\leq c^*(n) + \sum_{c \geq c^*(n)}^{n} \pi_c \\
&\leq \frac{16\bar{q}nk}{\ln(\bar{q}n)} + \frac{(\bar{q}(2^{\bar{q}} - 1) + 1)\ln(\bar{q}n)\sqrt{\bar{q}n}}{2\bar{q}k} \\
&= o(n),
\end{aligned}$$

completing the proof.

A.2 Proof of Theorem 2

A.2.1 Lemma 8

The following lemma says that a student that is involved in pre-arrangement is less preferred by the college to any student who is matched to it without pre-arrangement.

Lemma 8. *If $c \in C$ can manipulate via pre-arrangement with $s \in S$, then*

$$s' P_c s \text{ for every } s' \in \phi(S, C, P, q)(c).$$

Proof. Let $\mu(c) = \phi(S, C, P, q)(c)$. Theorem 2 of Sönmez (1999) shows that, for any stable mechanism, if c can manipulate via pre-arrangement with student s, then either $s \in \mu(c)$ or $s' P_c s$ for every $s' \in \mu(c)$. To show $s \notin \mu(c)$, suppose on the contrary that $s \in \mu(c)$. Consider matching μ' given by

$$\mu'(c') = \begin{cases} \mu(c) \setminus s & \text{if } c' = c, \\ \mu(c') & \text{otherwise.} \end{cases}$$

[30] Note that Lemma 4 can be applied since for sufficiently large n and $c \geq c^*(n)$, we have $c > 4k$.

It is easy to see, from stability of μ in (S, C, P, q), that μ' is stable in $(S \setminus s, C, P_{-s}, q_c - 1, q_{-c})$.

Since the matching under SOSM is weakly less preferred to any stable matching by colleges (attributed to Conway in Knuth (1976)) and preferences are responsive,

$$\begin{aligned}\phi(S, C, P, q) &= \mu'(c') \cup s \\ &\succeq_c \phi(S \setminus s, C, P_{-s}, q_c - 1, q_{-c}) \cup s.\end{aligned}$$

Therefore c cannot manipulate. This is a contradiction, completing the proof. $\square$

Therefore, in order to profitably manipulate, a college has to pre-arrange a match with a strictly less preferred student. Then the disadvantage of being matched with a less desirable student should be compensated by matching to a better set of students in the centralized matching mechanism after pre-arrangement.

A.2.2 Theorem 2

Now we prove Theorem 2. Since every college is made weakly better off under SOSM when the set of participating students increases (Gale and Sotomayor 1985), we obtain

$$\phi(S, C, P, q_c - 1, q_{-c}) \succeq_c \phi(S \setminus s, C, P_{-s}, q_c - 1, q_{-c}).$$

By Lemma 3 we have

$$\phi(S, C, P, q_c - 1, q_{-c}) = \begin{cases} \phi(S, C, P, q), \text{ or} \\ \phi(S, C, P, q) \setminus \arg\min_{P_c} \phi(S, C, P, q), \end{cases}$$

with probability of at least $1 - \pi_c$. In the former case it is clear that c cannot manipulate. In the latter case we have

$$\begin{aligned}\phi(S, C, P, q) &= \phi(S, C, P, q_c - 1, q_{-c}) \cup \arg\min_{P_c} \phi(S, C, P, q) \\ &\succeq_c \phi(S \setminus s, C, P_{-s}, q_c - 1, q_{-c}) \cup s,\end{aligned}$$

where the last comparison holds by responsiveness of preferences, the above preference relation in the beginning of the proof and Lemma 8. Therefore the probability that c benefits via pre-arrangement is at most π_c. Finally, by Lemma 7 we complete the proof (this last argument is similar to the one for Theorem 1 and hence omitted).

A.3 Proof of Theorems 3 and 4

Since Theorem 4 is a multi-region generalization of Theorem 3, we prove only the former.

A.3.1 Lemma 10: Uniform vanishing market power

We have a variant of Lemma 7 under the sufficient thickness assumption, which plays a crucial role in the proof of the theorems.

For $B_c^1 \subseteq \mu(c)$, let

$\pi_c^{B_c^1} =$
$\Pr[\text{Algorithm 3 associated with } B_c^1 \text{ returns to c}|Y^*(n;T) > EY^*(n;T)/2, \mu]$.

First we show a variant of Lemma 5.

Lemma 9. *Suppose* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *is regular and sufficiently thick. Let* T *be such that* $EY^*(n;T) \to \infty$ *as* $n \to \infty$. *Suppose that* n *is sufficiently large. Then we have*

$$\pi_c^{B_c^1} \leq \frac{4T\bar{q}}{EY^*(n;T)},$$

for any c *and* $B_c^1 \subseteq \mu(c)$.

Proof. Consider Round 1, beginning with the least preferred student s of $B_c^1 \subseteq \mu(c)$ (if $B_c^1 = \emptyset$, then the inequality is obvious since $\pi_c^{B_c^1} = 0$.). Since $p_{c'}^n(r) \geq p_c^n(r)/T$ for any $c' \in X^*(n;T)$ and $r = 1, \dots, R$, Round 1 ends at 2(c)iiB as a student applies to some college with vacant positions, at least with probability $1 - \frac{1}{Y^*(n;T)/2T+1} > 1 - \frac{1}{EY^*(n;T)/2T+1}$.

Now assume that all Rounds $1, \dots, i$ ends at Step 2(c)iiB. Then there are still at least $Y^*(n;T) - i$ colleges more popular than c and with a vacant position, since at most i colleges in $X^*(n;T)$ have had their positions filled at Rounds $1, \dots, i$. Therefore Round $(i+1)$ initiated by the least preferred student in B_c^{i+1} ends at Step 2(c)iiB with probability of at least $1 - \frac{1}{EY^*(n;T)/2T-i+1}$. Since there are at most $\bar{q}$ rounds, Algorithm 3 fails to return to c with probability of at least

$$\prod_{i=1}^{\bar{q}} \left(1 - \frac{1}{EY^*(n;T)/2T - (i-1) + 1}\right) \geq \left(1 - \frac{1}{EY^*(n;T)/2T - \bar{q} + 2}\right)^{\bar{q}}$$
$$\geq \left(1 - \frac{1}{EY^*(n;T)/4T}\right)^{\bar{q}}.$$

The first inequality follows since $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is sufficiently thick, n is sufficiently large and $i \leq \bar{q}$ for each i. The second inequality holds since $EY^*(n;T)/2 - \bar{q} \geq EY^*(n;T)/4 > 0$, which follows since $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ is sufficiently thick and n is sufficiently large. Therefore we have that

$$\begin{aligned}\pi_c^{B_c^1} &\leq 1 - (1 - \frac{1}{EY^*(n;T)/4T})^{\bar{q}} \\ &\leq \frac{4T\bar{q}}{EY^*(n;T)},\end{aligned}$$

where the last inequality holds since $1-(1-x)^y \leq yx$ for any $x \in (0,1)$ and $y \geq 1$. $\square$

Lemma 10 (Uniform vanishing market power)**.** *Suppose that* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \dots)$ *is regular and sufficiently thick. For any sufficiently large* n *and any* $c \in C$*, we have*

$$\pi_c \leq \frac{4\left[T\bar{q}(2^{\bar{q}}-1)+1\right]}{EY^*(n;T)}.$$

Proof. By Lemma 9 and an argument similar to Lemma 7, we obtain

$$\Pr[\text{Algorithm 3 returns to c}|Y^*(n;T) > EY^*(n;T)/2] \leq \frac{4T\bar{q}(2^{\bar{q}}-1)}{EY^*(n;T)}.$$

Therefore we have

$$\begin{aligned}\pi_c &\leq \Pr[Y^*(n;T) \leq EY^*(n;T)/2] + \Pr[Y^*(n;T) > EY^*(n;T)/2] \times \frac{4T\bar{q}(2^{\bar{q}}-1)}{EY^*(n;T)} \\ &\leq \frac{4}{EY^*(n;T)} + \frac{4T\bar{q}(2^{\bar{q}}-1)}{EY^*(n;T)} \\ &\leq \frac{4\left[T\bar{q}(2^{\bar{q}}-1)+1\right]}{EY^*(n;T)},\end{aligned}$$

completing the proof. $\square$

A.3.2 Theorems 3 and 4

We only prove Theorem 4, since Theorems 3 is a special case when $R = 1$. Suppose that colleges other than c are truth-telling, that is, any $c' \neq c$ reports $(P_{c'}, q_{c'})$. Lemmas 1 and 3 apply here since they do not rely on assumptions about student preferences. These lemmas imply that the probability that c profitably manipulates is at most π_c. By Lemma 10 and sufficient thickness, for any $\varepsilon > 0$, there exists n_0 such that for any market Γ_n with $n > n_0$, we have

$$\Pr\left[u(\phi(S,C,P'_c,P_{-c},q)(c)) > u(\phi(S,C,P,q)(c)) \text{ for some } P'_c\right] < \frac{\varepsilon}{\bar{q}\sup_{n\in\mathbb{N},s\in S^n,c\in C^n}\hat{u}_c(s)}.$$

Such n_0 can be chosen independent of $c \in C^n$. For any $n > n_0$,

$$\begin{aligned}
&Eu_c(\phi(S,C,(P'_c,P_{-c}),(q'_c,q_{-c}))(c)) - Eu_c(\phi(S,C,P,q)) \\
<&\Pr\left[u_c(\phi(S,C,(P'_c,P_{-c}),(q'_c,q_{-c}))(c) > u_c(\phi(S,C,P,q)(c))\right]\bar{q}\sup_{n\in\mathbb{N},s\in S^n,c\in C^n}\hat{u}_c(s) \\
<&\varepsilon,
\end{aligned}$$

which implies that truthful reporting is an ε-Nash equilibrium.

A.4 Proofs of Propositions 1 and 2

A.4.1 Proposition 1

Let a and T satisfy the condition of nonvanishing proportion of popular colleges. Let $c = [an]$. Then it is obvious that $X_c \subseteq X^*(n;T)$ and hence $Y_c \leq Y^*(n;T)$. For sufficiently large n, Lemma 4 shows that

$$E[Y^*(n;T)] \geq E[Y_c] \geq \frac{c}{2}e^{-\frac{8\bar{q}nk}{c}}.$$

$c = [an]$ implies that $\frac{c}{2}e^{-\frac{8\bar{q}nk}{c}} \to \infty$ as $n \to \infty$ with the order of convergence $O(n)$. Therefore $E[Y^*(n;T)] \to \infty$ as $n \to \infty$, completing the proof.

A.4.2 Proposition 2

Let

$$X^{**}(n;T) = \{c \in C^n | p_1^n(r)/p_c^n(r) \leq T \quad \text{for all } r\}.$$

Then we have $X^*(n;T) = \{c \in X^{**}(n;T) || \{s \in S^n | cP_s s\}| < q_c\}$. Let $\eta_r(c) = |\{c' \in C^n | p_c^n(r) \leq p_{c'}^n(r)\}|$ be the order of c with respect to popularity in distribution $D^n(r)$. For example, if college c is the most popular among students in region 1 and the least popular among those in region 2, then $\eta_1(c) = 1$ and $\eta_2(c) = n$.

Part (1): Example 4. Let $T = 4$, for example. Then, $X^{**}(n;4) = \{n/4, n/4+1, \ldots, 3n/4\}$. Consider any college $c \in X^{**}(n;4)$. Let s belong to region $r \in \{1,2\}$. Since s picks colleges k times according to $D^n(r)$, the probability that c does not appear in the preference list of student s, denoted by $\Pr(F_{c,s})$, is bounded as follows:

$$\Pr(F_{c,s}) \geq \left(1 - \frac{p_c^n(r)}{1 - Q^n(r)}\right)^k,$$

where

$$Q^n(r) = \sum_{c:\eta_r(c)\le k} p_c^n(r).$$

For any sufficiently large n, we have that $\eta_r(c) > 2k$ for any $c \in X^{**}(n;4)$ and $r = 1,2$ since $n - 2k > 3n/4 > c > n/4 > 2k$. For such colleges,

$$p_c^n(r) \le \frac{1 - Q^n(r)}{\eta_r(c) - k} \le \frac{1 - Q^n(r)}{\eta_r(c)/2}.$$

So

$$\Pr(F_{c,s}) \ge \left(1 - \frac{2}{\eta_r(c)}\right)^k.$$

Since $\eta_r(c) \ge n/4$ for any $c \in X^{**}(n;4)$ and any $r = 1,2$, we have

$$\Pr(F_{c,s}) \ge \left(1 - \frac{8}{n}\right)^k.$$

Let E_c be the event that c is not listed by any student. Then, since students draw colleges independently, we have

$$\Pr(E_c) = \prod_{s\in S^n} \Pr(F_{c,s}) \ge (1 - 8/n)^{k\bar{q}n} \to e^{-8k\bar{q}},$$

as $n \to \infty$. Therefore,

$$E[Y^*(n;T)] = \sum_{c\in X^{**}(n;4)} \Pr(E_c) \ge \frac{n}{2}(1 - 8/n)^{k\bar{q}n} \to \infty,$$

as $n \to \infty$ (with the order of convergence being $O(n)$), completing the proof.

Part (2): Example 5. As discussed in Example 5, for any colleges c and c' and region r, we have that

$$p_c^n(r)/p_{c'}^n(r) = \tilde{p}_{r(c)}(r)/\tilde{p}_{r(c')}(r) > 0.$$

Since there are only finite regions, $X^{**}(n;T) = C^n$ for any sufficiently large T. Fix such T.

As in the proof of Part (1), for any c and s we have

$$\Pr(F_{c,s}) \ge \left[1 - \frac{p_c^n(r(s))}{1 - Q^n(r(s))}\right]^k.$$

Since we have that $p^n_c(r)/p^n_{c'}(r) < T$ for any $c, c' \in C^n$,

$$\frac{p^n_c(r)}{1 - Q^n(r)} \leq \frac{p^n_c(r)}{(n-k)p^n_c(r)/T} \leq \frac{2T}{n},$$

for any sufficiently large n. So we have

$$\Pr(E_c) = \prod_{s \in S^n} \Pr(F_{c,s}) \geq \prod_{s \in S^n} (1 - \frac{2T}{n})^k \geq (1 - 2T/n)^{k\bar{q}n} \rightarrow e^{-2k\bar{q}T},$$

as $n \rightarrow \infty$. Therefore

$$E[Y^*(n;T)] = \sum_{c \in C^n} \Pr(E_c)$$

approaches infinity with the order $O(n)$, completing the proof.

Remark 2. In Examples 3, 4 and 5, the order of convergence of $EY^*(n;T)$ is $O(n)$. This implies that, by Lemma 10, the order of convergence of the probability of profitable manipulation is $O(1/n)$. This is the same order as in the uniform distribution case, analyzed by Roth and Peranson (1999) and Immorlica and Mahdian (2005).

A.5 Proof of Theorem 5

The proof is based on a series of arguments similar to the one for Theorem 1. First, dropping strategies are still exhaustive.

Lemma 11 (Dropping strategies are exhaustive for coalitional manipulations)**.** *Consider an arbitrary stable mechanism. Fix preferences of colleges other than $C' \subseteq C$. Suppose the mechanism produces μ under some arbitrary report $(P_{C'}, q_{C'})$. Then there exists $(P_{C''}, q_C)$, where (P''_c, q_c) is a dropping strategy for each $c \in C'$ such that $\mu' \succeq_c \mu$ where μ' is the matching induced by (P''_c, q_c) under the stable mechanism.*

Then consider a variant of Algorithm 1 where $B^1_{C'} \subseteq \bigcup_{c \in C'} \mu(c)$ replaces B^1_c and terminates either when $B^i_{C'} = \emptyset$ (the rejection chains do not return to C') or a student draws a college $c \in C'$ (the rejection chains return to C'). Then Lemmas 3 and 10 extend naturally to this modified algorithm: a manipulation is profitable only if the rejection chains returns to C', and the probability that the rejection chains return can be bounded appropriately. Finally we combine these to prove Theorem 5 in much the same way as in proving Theorem 3.

References

Abdulkadiroglu, A., P. Pathak, and A. E. Roth (2005): "The New York City High School Match," *American Economic Review, Papers and Proceedings*, 95, 364–367.

Abdulkadiroglu, A., P. Pathak, A. E. Roth, and T. Sönmez (2005): "The Boston Public School Match," *American Economic Review, Papers and Proceedings*, 95, 368–371.

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2006): "Redesigning the New York City High School Match," In progress.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2006): "Changing the Boston School Choice Mechanism," NBER Working paper, 11965.

Abdulkadiroğlu, A., and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Blum, Y., A. Roth, and U. Rothblum (1997): "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, 76, 362–411.

Cripps, M., and J. Swinkels (2006): "Depth and Efficiency of Large Double Auctions," *Econometrica*, 74, 47–92.

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

Ergin, H., and T. Sönmez (2006): "Games of school choice under the Boston mechanism," *Journal of Public Economics*, 90, 215–237.

Fudenberg, D., M. Mobius, and A. Szeidl (2006): "Existence of Equilibria in Large Double Auctions," forthcoming, *Journal of Economic Theory.*

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

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

Gresik, T., and M. Satterthwaite (1989): "The Rate at Which a Simple Market Converges to Effiency as the Number of Traders Increases," *Journal of Economic Theory*, 48, 304–332.

Immorlica, N., and M. Mahdian (2005): "Marriage, Honesty, and Stability," *SODA 2005*, pp. 53–62.

Knuth, D. E. (1976): *Marriage Stables.* Les Presse de l'Université de Montreal, Montreal.

Knuth, D. E., R. Motwani, and B. Pittel (1990): "Stable Husbands," *Random Structures and Algorithms*, 1, 1–14.

Kojima, F. (2005): "Mixed Strategies in Games of Capacity Manipulation in Hospital-Intern Markets," forthcoming, *Social Choice and Welfare*.

Konishi, H., and U. Ünver (2005): "Games of Capacity Manipulation in the Hospital-Intern Market," forthcoming, *Social Choice and Welfare*.

McVitie, D. G., and L. Wilson (1970): "Stable marriage assignments for unequal sets," *BIT*, 10, 295–309.

Pesendorfer, W., and J. Swinkels (2000): "Efficiency and Information Aggregation in Auctions," *American Economic Review*, 90, 499–525.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

——— (1984a): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991–1016.

——— (1984b): "Misrepresentation and stability in the marriage problem," *Journal of Economic Theory*, 34, 383–387.

——— (1985): "The college dmission problem is not equivalent to the marriage problem," *Journal of Economic Theory*, 36, 277–288.

——— (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics, Fisher-Schultz Lecture," *Econometrica*, 70, 1341–1378.

Roth, A. E., and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89, 748–780.

Roth, A. E., and U. Rothblum (1999): "Truncation Strategies in Matching Markets: In Search of Advice for Participants," *Econometrica*, 67, 21–43.

Roth, A. E., and M. O. Sotomayor (1990): *Two-sided matching: a study in game-theoretic modeling and analysis*. Econometric Society monographs, Cambridge.

Roth, A. E., and J. Vande Vate (1991): "Incentives in two-sided matching with random stable mechanisms," *Economic Theory*, 1, 31–44.

Rustichini, A., M. Satterthwaite, and S. Williams (1994): "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica*, 62, 1041–1064.

Sönmez, T. (1997a): "Games of Manipulation in Marriage Problems," *Games and Economic Behavior*, 20, 169–176.

——— (1997b): "Manipulation via Capacities in Two-Sided Matching Markets," *Journal of Economic Theory*, 77, 197–204.

——— (1999): "Can Pre-arranged Matches be Avoided in Two-Sided Matching Markets?," *Journal of Economic Theory*, 86, 148–156.

Sotomayor, M. A. O. (1996): "A non-constructive elementary proof of the existence of stable marriages," *Games and Economic Behavior*, 13, 135–137.

Swinkels, J. (2001): "Efficiency of Large Private Value Auctions," *Econometrica*, 69, 37–68.