Incentives and Stability in Large Two-Sided Matching Markets*

Fuhito Kojima[†] Parag A. Pathak[‡]

December 4, 2005

Preliminary and Incomplete
Do not cite or distribute without permission

Abstract

The paper analyzes incentives and stability in two-sided matching markets when the number of participants is large and the length of the preference list is finite. Building on and extending Immorlica and Mahdian (2005), we first investigate the scope for manipulation in a many-to-one market. When preference lists are drawn from an arbitrary distribution for one side of the market and under a mild independence assumption on the distribution for the other side, we establish that the fraction of participants who can profitably misrepresent their preferences via truncation is small. Moreover, the scope of manipulations via capacities and pre-arranged matches approaches zero as the size of the market becomes large. With an additional bounded distribution assumption, truthful reporting is an approximate equilibrium, implying efficiency of the resulting matching.

*We appreciate helpful discussions with Attila Ambrus, Drew Fudenberg, Alvin E. Roth, Tayfun Sönmez, Satoru Takahashi and Utku Ünver. 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 is one of the most elegant and well-developed areas of game theory. A central notion is stability: there is no individual agent or pair of agents who prefer to be assigned to each other than their allocation in a matching. Matching theory also has real-world applications in entry-level labor markets, where new physicians seek positions in hospitals as residents[1] and the mechanisms inspire solutions in related contexts such as student assignment.[2] In these contexts, stable mechanisms often succeed whereas unstable ones are likely to fail.

Although stable mechanisms have a number of virtues, they are not immune 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 (1997a) and Sönmez (1999) shows 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 allocation from the match.

Concerns about the potential for these types of manipulation are often present in real world markets. For instance, in New York City 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).

The aim of this paper is to understand why despite these negative results many stable mechanisms appear to work well in practice. In the 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

[1]See Roth (1984a) and Roth and Peranson (1999).

[2]See for example Abdulkadiroğlu and Sönmez (2003), Abdulkadiroğlu, Pathak, Roth, and Sonmez (2005) and Abdulkadiroğlu, Pathak, and Roth (2005).

preference list is about 15, while the number of positions is on the order of 20,000 per year. In New York City, the maximal length of the preference list is 12, and there are over 90,000 participants per year. As a result, our focus is on understanding what happens in the limit in environments as the number of participants grows, but the length of the preference lists does not.

Our results show that the size of the market makes the mechanism immune to various kinds of manipulations. Specifically, 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 via truncating preferences or capacity or pre-arrangement 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. Immorlica and Mahdian (2005) claim that with the same set of assumptions, truth-telling is an approximate equilibrium. We present an example to show that this is not the case. We next define a condition on the distribution of preferences as allowing no superstar if there are no extremely popular colleges and no extremely unpopular colleges. Under this assumption, truth-telling is an approximate equilibria in games of preference truncation and capacity reporting. Furthermore, not engaging in pre-arrangement is approximately optimal for each college when the size of the market is large.

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 first 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 showed that very few students and hospitals could have benefitted by submitting false preference lists or by manipulating capacity. These simulations led them to conjecture that the fraction of people 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 heavily upon, was one of the first theoretical attempts to understand these results. They consider 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 our techniques are similar, our focus in this paper is on many-to-one markets like the NRMP, where this implication no longer holds. Even if there is only one stable matching, colleges can sometimes manipulate via preference lists. Moreover, in many-to-one markets there exists the additional possibility of capacity manipulation and manipulation via pre-arrangement which are not present in a one-to-one market. As a result, having only one set of stable partners in the limit is not sufficient to explain the lack of manipulability in many-to-one markets. We argue instead that the results follow from vanishing market power of colleges in the limit. This paper thus complements and extends Immorlica and Mahdian (2005) to a many-to-one market and considers an expanded set of manipulations. Such an exercise is of theoretical interest given the widespread nature of many-to-one markets. In addition, it is necessary to understand simulation evidence on manipulations presented by Roth and Peranson (1999).

The use of large society 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.[3] The proofs of these results rely on a symmetric distribution of values. In our paper, we will allow for an arbitrary distribution of values for colleges provided each student is acceptable, and independent and identically distributed values for students, allowing colleges to be ex-ante asymmetric. 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 (2005) 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 (2005) for games of capacity manipulations.

The next section presents the model, and the following section presents the main lemma on vanishing market power used to derive the main results.

[3]There are a number of related papers including Gresik and Satterthwaite (1989) and Fudenberg, Mobius, and Szeidl (2003).

Section 4 then analyzes the fraction of colleges who can manipulate, while section 5 conducts equilibrium analysis. Section 6 examines some implications of our result for another two-sided matching mechanism. The last section concludes.

2 Model

A market is tuple $\Gamma = (S, C, (P_s)_{s\in S}, (\succ_c)_{c\in C})$. S and C are finite and disjoint sets of students and colleges. Assume that students in S are ordered in an arbitrarily fixed manner.[4] 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 $cP_s\emptyset$. 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, we often write only acceptable mates to denote preferences. For example,

$$s_1 : c_1, c_2,$$

means that s_1 prefers 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 $s \cup S' \succeq_c s' \cup S' \Leftrightarrow s \succeq_i s'$, (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 denote by $\Gamma = (S, C, P, q)$ an arbitrary game in which the preferences induce preference lists $P = (P_i)_{i\in S\cup C}$ and quotas $q = (q_c)_{c\in C}$. 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 S to $C \cup S$, such that for every s, $\mu(s) \in C \cup \{s\}$. We define $\mu(c) = \{s \in S | \mu(s) = c\}$ for any $c \in C$. For any matchings μ and μ', we write $\mu \succeq_c \mu'$ if and only if $\mu(c) \succ_c \mu'(c)$ for any $c \in C$, and $\mu P_s \mu'$ if and only if $\mu(s) P_s \mu'(s)$.

[4] The order on C will be given later in this section.

Given a matching μ, we say that it is **blocked** by (c, s) 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 $s \in \mu(c)$, $s \succ_c \emptyset$. 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 gives a stable matching for any market. A generic stable mechanism is denoted by ψ.

We consider the procedure, called the **student optimal stable mechanism** (SOSM), and denoted by ϕ, which is analyzed by Gale and Shapley (1962).[5]

- 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 step. 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 they induce the same pair of preference list orders and quotas of colleges P and q. Thus we sometimes write the resulting matching by

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

Similarly, in this paper we assume that matchings produced by mechanism ψ depend only on (S, C, P, q), so we will use notation such as $\psi(S, C, P, q)$.

[5]SOSM is known to produce a stable matching that is unanimously most preferred by every agent in S.

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

For the rest of the paper, we consider a family of college admission problems $(\Gamma_1, \Gamma_2, \dots)$ with fixed positive integers $\bar{q}, k \in \mathbb{N}$. For each $n \in \mathbb{N}$, there are n colleges in Γ_n. The capacity of college $c \in C$ is a positive integer $q_c \in \{1, \dots, \bar{q}\}$ for any Γ_n (so $\bar{q}$ is a uniform upper bound of quotas across colleges and problems.) The number of students is $N \leq \bar{q}n$ in Γ_n.[6]

Constructing Preference lists

We assume that $c \in C$ has **arbitrary** responsive preferences with quota q_c with one restriction: every student is acceptable to every college. That is, colleges would rather admit any student than keep their positions vacant.

For students, we construct random preferences following Immorlica and Mahdian (2005). Let $\mathcal{D}^n$ be an arbitrary fixed distribution over the set of colleges, corresponding to Γ_n. Suppose that the probability of each college in $\mathcal{D}^n$, denoted by p_c^n, is nonzero.[7] Having a high probability in $\mathcal{D}^n$ means that the college is popular. For each Γ_n and each student, we construct preferences of students over colleges as follows:

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

In general,

- Step t: Select college $c_{(t)}$ independently according to $\mathcal{D}^n$ 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.

Let A_s be the *unordered* set of colleges in this procedure at each step, suppressing superscript t. In other words, Step t draws colleges repeatedly until a college $c \notin A_s$ is drawn. 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. One justification for this assumption is that in many real markets, it is costly to form a complete preference list

[6] This assumption can be relaxed easily to the following: there exists r such that $N \leq rn$ for any n. We adopt $N \leq \bar{q}n$ just for simplicity.

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

for participants. 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.

Let $It(k, \mathcal{D}^n)$ be the distribution over lists of size k produced by this process. If $\mathcal{D}^n$ is the uniform distribution, then $It(k, \mathcal{D}^n)$ is 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 $p^n_{c'} \geq p^n_c$ in distribution $\mathcal{D}^n$. 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^n_m, which is the probability associated with mth college in distribution $\mathcal{D}^n$.

3 Vanishing "market power" in large markets

This section presents a lemma which plays a crucial role in most of our results. We consider the following hypothetical question: Suppose that the student-optimal stable matching is reached and then a college $c \in C$ rejects all the students assigned to c. Suppose that these students apply to other colleges, which may accept them and reject some of students originally matched to them, and so on. What is the probability that such chain-reactions caused by the rejected students will at some point come back to c (that is, a student rejected during this chain reaction proposes to c)? If this does not happen, then strategically rejecting a student has no positive effect to the college. Thus the above probability is indicative of "market power" of college c, by which we mean the prospect of affecting the market to benefit itself.

To answer the above question, we use the **principle of deferred decisions**: the result of the following algorithms from a randomly drawn preference lists is the same as the result when preferences are drawn one at a time, when it is needed.

Consider the following algorithm, which is a stochastic variant of the SOSM.[8]

Algorithm 1. Stochastic S-Optimal Gale-Shapley Algorithm

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

(2) Choosing the applicant:

[8]To be more precise this is a stochastic version of the algorithm proposed by McVitie and Wilson (1970), which they show is equivalent to the original SOSM proposed by Gale and Shapley (1962).

(a) If $l \leq N$, then let s be the l'th student and increment l by one.[9]

(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 b, 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.

In the above algorithm, the principle of deferred decisions implies that for a student to make an offer to her tth most preferred college, her preferences after $(t+1)$th choice on does not matter at all. By the principle of deferred decisions, 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. Now suppose that a fixed college $c \in C$ rejects all the students in $\mu(c)$. More precisely, consider the following algorithm, beginning at the final state of the above algorithm.

Algorithm 2. STOCHASTIC REJECTION CHAINS

(1) Initialization:

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

(b) Let $B_c^1 = \mu(c)$.

(2) Increment i by one.

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

(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 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 2 terminates either at Step 2a or at Step 2(c)iB. We are interested in how often the algorithm ends at Step 2(c)iB, as a student draws c from distribution $\mathcal{D}^n$. If the algorithm terminates at Step 2a rather than Step 2(c)iB, then Algorithm 2 is guaranteed to have no effect on the students matched to c, except for that it lost students in $\mu(c)$ by rejecting them at the beginning of the algorithm. Let π_c be the probability that Algorithm 2 terminates at Step 2(c)iB. Let $c^*(n) = 16\bar{q}nk/\ln(\bar{q}n)$. The following lemma[10] gives an upper bound for π_c.[11]

Lemma 1 (Vanishing market power). *For any sufficiently large n and any $c > c^*(n)$, we have*

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

[10]Proofs of all results are contained in the appendix.

[11]Lemma 1 generalizes a technique developed by Immorlica and Mahdian (2005). One difference is that we consider an algorithm general enough to be applied to analysis of various kinds of manipulation. Another difference is that we consider cases with multiple quotas and multiple rejections.

The lemma shows that π_c is small except for very popular colleges when the size of the market is large. Observe that $\pi_c \to 0$ as $n \to \infty$ for $c > c^*(n)$. Note also $c > c^*(n)$ holds for "most" colleges for large n, since $c^*(n)/n = 16\bar{q}k/\ln(\bar{q}n) \to 0$ as $n \to \infty$. Lemma 1 suggests that, for most colleges, strategically rejecting students, even *all the students*, has no positive indirect effect on that college when the size of the market is large.

The intuition of Lemma 1 follows from the fact that there are many colleges that have some vacant positions in large markets. Therefore, it is very likely that a rejected student during the algorithm makes an offer to one of these colleges. Because every college treats a student as acceptable, such a college admits the student and it does not reject any student previously assigned to it, thus terminating the chain reactions before another student makes an offer to college that initiated the chain.

4 Proportion of colleges that can manipulate

The literature on two-sided matching has focused on three types of manipulations: (1) manipulation via preference lists, (2) manipulation via capacity, and (3) manipulation via pre-arranged matches.

We show that these manipulations are likely to be rare in large markets. More specifically, we show that the expected proportion of colleges that can profitably manipulate approaches zero as the size of the market approaches infinity, assuming that others behave truthfully. Note that this is not an equilibrium analysis: we count the number of possible deviators when others are truthful, which may or may not be in their interest. Equilibrium behavior is analyzed in Section 5.

4.1 Manipulation via preference lists

First we consider manipulation via preference lists. Mechanism ψ is **manipulable via preference lists** if there exist a market $(S, C, (P_i)_{i \in S \cup C}, (q_c)_{c \in C})$, $i \in S \cup C$ and some P'_i such that

$$\psi(S, C, P'_i, P_{-i}, q) \succ_i \psi(S, C, P, q).$$

Dubins and Freedman (1981) show that the SOSM ϕ is manipulable via preference lists. Roth (1982) further shows that any stable mechanism is manipulable in this way. 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.

We consider a particular class of manipulation via preferences. A preference list P'_c is a **truncation** of P_c if there is a student $sP_c\emptyset$ such that

$$
\begin{aligned}
s', s'' R_c s &\Rightarrow [s' P'_c s'' \iff s' P_c s''], \text{and} \\
s P_c s' &\Rightarrow \emptyset P'_c s'.
\end{aligned}
$$

In words, a truncation agrees with the original preference lists up to student s, but renders every student less preferred to s under P_c unacceptable. A college $c \in C$ can **manipulate ϕ via truncating preference lists** if there exists a truncation P'_c such that

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

When colleges have multiple positions, truncation may not be "exhaustive", in the sense that there may be other instances where preference manipulation is profitable even though truncation is not. We nevertheless focus on truncation strategies because they are a plausible starting place for types of preference manipulation. For instance, in their simulation study of randomly generated data, Roth and Peranson (1999) only consider truncation strategies. Moreover, truncation is known to be exhaustive when the colleges have only one position (Roth and Vande Vate 1991).

The following theorem shows that the expected proportion of colleges that can manipulate the market in this way converges to zero as the size of the market becomes large if others behave truthfully.

Theorem 1. *Let $\alpha_k(n)$ be the expected number of colleges that can manipulate via truncating preference lists in Γ_n under ϕ when others report preference lists truthfully. We have*

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

If truncating a preference list does not reject anyone who would not be rejected under the original preference list, then it has no effect on the matching. If truncation results in rejecting some students in $\mu(c)$, this rejection creates a chain reaction of applications, acceptances and rejections, which may give c a better offer from a student who was rejected elsewhere. The algorithm is, however, similar to Algorithm 2 that probability of such an instance is bounded by π_c. Because of the vanishing market power of colleges (Lemma 1), the probability that the additional rejections have preferable effects on c can be shown to be small for most colleges when the size of the market is large.

Consider the one-to-one market where each college has only one position. The following is a corollary of Theorem 1.

Corollary 1 (Theorem 3.1 of Immorlica and Mahdian (2005)). *Let $q_c = 1$ for any $c \in C$. Then the expected proportion of colleges who can manipulate via preference lists converges to zero as the size of the market approaches infinity.*

Roth and Peranson (1999) analyzed NRMP data and argued that of the 3,000-4,000 participating programs, less than one percent could benefit by truncating preference lists. They also conduct simulations using randomly generated data in one-to-one matching, and observe that $\alpha_k(n)$ quickly approaches zero as n becomes large. The first theoretical account is given by Immorlica and Mahdian (2005) (who show Corollary 1). The result in Theorem 1 is a simple extension to many-to-one markets such as the NRMP.

The assumption that students have a fixed length k of preference list can be relaxed to some extent: if each student has a preference list of length k or less, all the theorems in our paper holds. On the other hand, if every student has a preference list of full length so that every college is acceptable, then the conclusion of Theorem 1 does not hold. When all colleges and students have random preferences drawn from a uniform distribution Knuth, Motwani, and Pittel (1990) show that the proportion of colleges that can manipulate approaches one if every college is acceptable to every student. This pattern is confirmed by the Roth and Peranson (1999) simulations on random data.

4.2 Manipulation via capacities

When colleges have quotas of more than one, there are other types of manipulations to consider. Sönmez (1997b) formally introduced the idea of capacity manipulation. A mechanism ψ is **manipulable via capacities** if there exist (S, C, P, q), $c \in C$ and some q'_c such that

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

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

The following theorem asserts that the expected number of colleges that benefit from manipulation via capacities becomes small as the size of the market increases for fixed k.

Theorem 2. *Let $\beta_k(n)$ be the expected number of colleges that can manipulate via capacities in Γ_n under ϕ when others report truthfully. We have*

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

The proof of Theorem 2 is again based on Lemma 1. The reason that reducing capacity might benefit c is that a chain of applications, acceptances and/or rejections caused by the rejection of a student due to capacity reduction may result in more offers from desirable students to remaining positions of c. This is formally known as the "vacancy chain dynamics" (Blum, Roth, and Rothblum 1997). The process is similar to Algorithm 2 and the probability that at least one desirable student is matched to c is bounded by π_c. The argument requires defining a modification to Algorithm 2, which is presented in the appendix.

In the Roth and Peranson (1999) simulations, less than one percent of programs in the NRMP could manipulate via capacities, assuming that the preference data are truthful. Theorem 2 explains their observations by the fact that NRMP market is quite large, involving 3,000 to 4,000 programs.

4.3 Manipulation via pre-arranged matches

When colleges seek more than one student, there is yet another source of manipulations. Sönmez (1999) introduced 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. A mechanism ψ is **manipulable via pre-arranged matches**, or **manipulable via pre-arrangement**, if for some market (S, C, P, q), college $c \in C$ and student $s \in S$ we have

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

The mechanism is manipulable via pre-arrangement if both parties that engage in pre-arrangement have incentive to do so: the student should be at least as well off in pre-arrangement as when she is matched through the centralized mechanism, and the college should strictly prefer s and the assignment of the centralized mechanism to those without pre-arrangement. Note that this manipulation may be possible even if matching outside of the match is prohibited under an alternative interpretation: if pre-arranged student and college list each other on the top of their preference lists, then they will be matched for sure in most mechanisms, so "pre-arrangement" can actually be implemented within the centralized matching mechanism.[12] Sönmez (1999) shows that any stable mechanism is manipulable via pre-arrangement.

[12]The last remark is true for ϕ. There are some mechanisms that do not satisfy this property. The so-called linear programming matching mechanism, used to match interns in some British hospitals, is one such example.

However, we have the following positive result in large markets.

Theorem 3. *Let $\gamma_k(n)$ be the expected number of colleges that can manipulate via pre-arrangement in Γ_n under ϕ when others do not pre-arrange. We have*

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

Intuition is similar to those for Theorems 1 and 2. It can be shown that any student 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 9 in the Appendix.) Therefore, in order to profitably manipulate, c should be matched to a better set of students in the central matching. Now, by a similar reasoning to those for Theorems 1 and 2, the probability of being matched to better students in the centralized mechanism is bounded by π_c.

Taken together, Theorems 2 and 3 provide justification to the assertion that the student-optimal stable mechanism is immune to capacity manipulation and manipulation via pre-arrangement and may help understand why there is limited evidence of both in real matching markets using stable mechanisms.

Remark This section analyzed scope for different types of manipulations separately. In the real matching market, colleges may be able to use *combinations* of these manipulations. It is a tedious but a straightforward exercise to show that the proportion of colleges that can profitably manipulate by using combinations of these manipulations approaches zero as the market size grows.

5 Equilibrium analysis

Theorems 1, 2 and 3 show that the expected proportion of colleges who can profitably manipulate the market becomes small as the market becomes large, assuming that others behave truthfully. This section investigates equilibrium behavior of colleges in large markets.

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

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

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, $\hat{u}_c(s) = -\infty$. Further suppose that $\sup_{\Gamma_n, s \in S} \hat{u}(s)$ is finite. A game of preference truncations is a market Γ_n coupled with utility functions $(u_c)_{c \in C}$, where the set of players is C, the strategy set of player c is all truncations and the outcome is the assignment specified by ϕ under reported preference lists.[13]

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

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

where the expectation is taken with respect to random preference lists of students. A truncated preference list P'_c is ε**-dominant** if for any P'_{C-c} there is no other truncation P''_c such that

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

ε-Nash equilibrium and ε-dominant strategies are defined in analogous manners in games of capacity manipulaion and pre-arrangement.

With the above setup, Immorlica and Mahdian (2005) claim the following, attributing their argument to Corollary 1.[14]

Claim 1 (Corollary 3.1 of Immorlica and Mahdian (2005)). *Let $q_c = 1$ for any $c \in C$. Then, for any $c \in C$, the probability that c can manipulate via truncating preference lists converges to zero as $n \to \infty$.*

Claim 2 (Corollary 3.3 of Immorlica and Mahdian (2005)). *Let $q_c = 1$ for any $c \in C$. Then, for any $\varepsilon > 0$, there exists n_0 such that truth-telling P_C is an ε-Nash equilibrium for any game with $n > n_0$.*

The following example shows that Claims 1 and 2 are not correct.

[13] We assume that students are passive players and always submit their preferences truthfully since truthful reporting is weakly dominant for students under ϕ (Dubins and Freedman 1981, Roth 1982).

[14] They consider multiplicative approximate equilibrium as opposed to additive approximate equilibrium, that is, there is no other strategy yielding a better utility by a multiplicative factor. However the analysis is essentially the same and our result can be stated in terms of multiplicative approximate equilibrium. Also we use the term approximate Nash equilibrium while Immorlica and Mahdian (2005) use the term "approximate Bayes-Nash equilibrium" in a game with incomplete information, since the game can be thought of as a game of complete information where only colleges are active players and students are passive agents.

Example 1. Consider the following market Γ_n for any n. $|C| = |S| = n$. $q_c = 1$ for each $c \in C$. Preference lists of c_1 and c_2 are given as follows:[15]

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

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

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

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 of preference manipulation for c_1 does not vanish even when n becomes large. It is also clear that truth-telling is not an ε-Nash equilibrium if $\varepsilon > 0$ is sufficiently small, 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 each college* the opportunity of such manipulation may remain large.

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

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

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

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

c_2's preferences are

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

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

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

$$
\begin{aligned}
s_1 &: c_2, c_1, \ldots, \\
s_2 &: c_1, c_2, \ldots, \\
s_3 &: c_1, c_2, \ldots, \\
s_4 &: c_1, c_2, \ldots.
\end{aligned}
$$

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 of manipulations via capacities or pre-arrangement for c_1 does not vanish even when n becomes large.[16]

A natural question is under what conditions one can expect a positive result. Consider the following condition for the popularity of colleges. Let $(\mathcal{D}^n)_{n=1}^{\infty}$ be a sequence of distributions, where $\mathcal{D}^n$ is a distribution over colleges in Γ_n.

Definition. A sequence of distributions $(\mathcal{D}^n)_n$ **allows no superstar** if there exists $T \in \mathbb{R}$ such that

$$\frac{p^n_1}{p^n_n} < T,$$

for any n.

This assumption rules out extremely popular colleges.[17] Note that distributions in Examples 1 and 2 allow superstars, since $p^n_1/p^n_n = n - 2 \to \infty$ as $n \to \infty$.

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

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

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

[17]Actually we can relax this assumption to some extent. First, it is enough for the inequality to be true for each sufficiently large n. Second, the condition can relaxed to cases in which some colleges are so unpopular that the inequality is violated, if the proportion of such colleges are not too large. For example, for all theorems below to hold,

We have a variant of Lemma 1 under the above assumption, which plays a crucial role in what follows.

Lemma 2 (Uniform Vanishing Market Power). *Suppose that $(\mathcal{D}^n)_n$ allows no superstar. For any sufficiently large n and any $c \in C$, we have*

$$\pi_c \leq \frac{8(T\bar{q}+1)e^{8\bar{q}k}}{n}.$$

The key difference between Lemma 2 and Lemma 1 is that the former gives an upper bound for *every* college, while the latter gives an upper bound only for unpopular colleges. With Lemma 2, we can obtain the following results.

Theorem 4. *Suppose that $(\mathcal{D}^n)_n$ allows no superstar. Consider the SOSM ϕ.*

(1) *For any $c \in C$, the probability that c can manipulate via preferences converges to zero as $n \to \infty$.*

(2) *For any $\varepsilon > 0$, there exists n_0 such that truth-telling of preference lists is an ε-Nash equilibrium of a preference truncation game for any Γ_n with $n > n_0$.*

Theorem 4 implies that the conclusions of Claims 1 and 2 are correct with the additional assumption that $(\mathcal{D}^n)_n$ allows no superstar.

For games of capacity manipulations, we can obtain a stronger result. Truth-telling is an ε-dominant strategy. Let the strategy set of $c \in C$ be $\{1, \ldots, q_c\}$.[18]

Theorem 5. *Suppose that $(\mathcal{D}^n)_n$ allows no superstar. Consider the S-optimal stable mechanism ϕ.*

(1) *For any $c \in C$, the probability that c can manipulate via capacities converges to zero as $n \to \infty$.*

(2) *For any $\varepsilon > 0$, there exists n_0 such that for any $c \in C$, truth-telling of capacities is an ε-dominant strategy of a capacity reporting game for any Γ_n with $n > n_0$.*

---

it is enough that

$$\frac{p_1^n}{p_{[an]}^n} < T,$$

for some $a \in (0,1)$, which is independent of n, where $[x]$ denotes the largest integer less than or equal to x.

[18] We restrict our attention to strictly positive capacity less than or equal to q_c. Note that reporting $q'_c = 0$ and $q'_c > q_c$ are weakly dominated if one includes these strategies.

Finally pre-arrangement is investigated.

Theorem 6. *Suppose that $(\mathcal{D}^n)_n$ allows no superstar. Consider the SOSM phi. Then, for any $c \in C$, the probability that c can manipulate via pre-arrangement converges to zero as $n \to \infty$.*

The three theorems rely on the no superstar condition to ensure that uniformly no college wields too much market power as the size of the market grows. Since a stable matching is efficient, we can state that an efficient matching is an approximate equilibrium in a large market.

6 Discussion

The idea that a large markets solve many inefficiency and incentive problems appears throughout economics. Examples include auction markets, exchange economy, insurance markets among others. In two-sided matching markets, it is not true, however, that a large market is sufficient to promote these virtues irrespective of the mechanism. The design of the market mechanism can matter even in large markets. 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.[19] When colleges are asked to rank students, in a two-sided context, 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. Students who are not rejected are *guaranteed* positions; the match of these students and colleges are permanent rather than temporary, unlike in the S-optimal stable mechanism.

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. Students who are not rejected are *guaranteed* positions; the match of these

[19] With slight abuse of notation 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.

students and colleges are permanent rather than temporary, unlike in the student-optimal stable mechanism.

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 step.

In 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,

Remark. Suppose that the Boston mechanism is employed, and preferences are drawn arbitrarily. Then, for any n, we have the following.

(1) No college can manipulate via preferences. Therefore truthtelling is an (exact) Nash equilibrium in dominant strategies in games of preference manipulations both under perfect and imperfect information about students preferences.

(2) No college can manipulate via capacities. Therefore truthtelling is an (exact) Nash equilibrium in dominant strategies in games of capacity manipulations both under perfect and imperfect information about students preferences.

(1) is first shown by Ergin and Sonmez (2005). While colleges have incentives to behave truthfully, we argue that this mechanism performs badly both in small and large markets. The problem lies in the student side and there is evidence that some participants react to these incentives (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2005). The following example shows that students have incentives to behave dishonestly even in large markets.

Example 3. Consider markets $(\Gamma_n)_n$, where $|S| = |C| = n$ for each Γ_n. $q_c = 1$ for every $c \in C$. Preference lists are common among colleges and given by

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

for every $c \in C$.

$p^n_{c_1} = (1/2)^{1/n}, p^n_{c_2} = (1-(1/2)^{1/n})(1/2)^{1/n}$, 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 $s \in S$. s_1 and s_2 are matched to c_1 and s_2, respectively, and other students are matched to their third or less preferred choices. If $s \neq s_1, s_2$ 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.

Pre-arrangement remains profitable to many participants even when the size of the market is large as well. It is easy to construct a similar example for pre-arrangement.

7 Conclusion

The paper establishes conditions under which the fraction of participants who can profitably manipulate a large two-sided matching market is small and identified an additional distributional assumption under which truthful reporting is an approximate equilibrium, implying efficiency of the resulting matching.

There are several dimensions in which our results could be potentially generalized, which we are currently pursuing. First, there are other types of preference manipulations than truncation that are feasible in a many-to-one market. While truncation strategies are a natural class of manipulations to consider, colleges can sometimes profitably employ non-truncation strategies even when they cannot benefit by truncation.

Second, we are exploring how much and in what directions our distributional assumptions can be weakened. While college preferences are arbitrary, student preferences are generated identically and independently and students agree on the ordering of school popularity. Fully exploring the sufficiency of our distributional assumptions will be critical to understanding under which domains we can expect large markets to resolve incentive problems and remedy inefficiencies.

Appendix: Proofs

Proof of Lemma 1

Let

$X_c = \{c' \in C | c' \leq c, c' \notin A_s \text{ for every } s \in S \text{ at the end of Algorithm 1}\},$ and $Y_c = |X_c|$.

Lemma 3. *For any $c > 4k$, we have*

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

Proof. Let $Q = \sum_{c=1}^{k} p_c$. 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;

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

Let $E_{c'}$ be the event that $c' \notin A_s$ for every $s \in S$. From the above argument, we have

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

Now if $c' > 2k$ we have

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

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

$$\begin{aligned} Pr(E_{c'}) &\geq (1 - \frac{1}{c' - k})^{\bar{q}nk} \\ &\geq e^{-2qnk/(c'-k)} \\ &\geq e^{-\frac{4\bar{q}nk}{c'}}. \end{aligned}$$

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

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

$\square$

Let $\pi_c^* = Pr[\text{Algorithm 2 terminates at Step 2(c)iB}|Y_c > EY_c/2]$.

Lemma 4. *Suppose that $2kn - \ln(\bar{q}n)\sqrt{\bar{q}n} \geq 0$ and $c > c^*(n)$. Then we have*

$$\pi_c^* \leq \frac{4\bar{q}}{EY_c}.$$

Proof. Consider Round 1, beginning with the least preferred student s of $B_c^1 = \mu(c)$. Since $p_{c'} \geq p_c$ 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 $1 - \frac{1}{EY_c/2+1}$.

Now assume that all Rounds $1, \ldots, i$ ends at Step 2(c)iiB. Conditional on this assumption Round $(i+1)$, initiated by the $(i+1)^{st}$ least preferred student in $\mu(c)$, that is the least preferred one in B_c^{i+1}, ends at Step 2(c)iiB with probability of at least $1 - \frac{1}{EY_c/2-i+1}$ (at most i colleges in X_c have had their positions filled at Rounds $1, \ldots, i$.) Therefore Algorithm 2 finishes at Step 2a with probability of at least

$$\begin{aligned}
\prod_{i=1}^{q_c}(1 - \frac{1}{EY_c/2 - (i-1) + 1}) &\geq (1 - \frac{1}{EY_c/2 - \bar{q} + 2})^{q_c} \\
&\geq (1 - \frac{1}{EY_c/4})^{\bar{q}},
\end{aligned}$$

where the last inequality holds since we have $q_c \leq \bar{q}$ by assumption, and $EY_c/2 - \bar{q} \geq EY_c/4 > 0$ by Lemma 3, $2nk - \ln(\bar{q}n)\sqrt{\bar{q}n} \geq 0$ and $c \geq c^*(n)$. Therefore we have that

$$\begin{aligned}
\pi_c^* &\leq 1 - (1 - \frac{1}{EY_c/4})^{\bar{q}} \\
&\leq \frac{4\bar{q}}{EY_c},
\end{aligned}$$

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

$\square$

Now we prove Lemma 1. We state without proof the following lemma (this is a straightforward generalization of Lemma 4.4 of Immorlica and Mahdian).

Lemma 5. *For every c, we have $Var(Y_c) \leq E[Y_c]$.*

By the Chebychev inequality, Lemma 5 and the fact that any probability is less than or equal to one, we have

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

By the above inequality and Lemma 4,

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

Applying Lemma 3 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.[20]

Proof of Theorem 1

Suppose that the S-optimal stable matching is reached through Algorithm 1. Now consider the following algorithm:

Algorithm 3. STOCHASTIC PREFERENCE MANIPULATION CHAINS
Fix an arbitrary truncation P_c' of P_c. This algorithm is the same as Algorithm 2 except for Steps 1b and 2(c)iB. Replace Steps 1b and 2(c)iB of Algorithm 2 with the following:

1b' Let $B_c^1 = \{s \in S | s \in \mu(c) \wedge \emptyset P_c' s\}$.

2(c)iB' If not, select c' randomly from distribution $\mathcal{D}^n$ until $c' \notin A_s$, and add c' to A_s. If $c' = c$ and $\emptyset P_c' s$, then go back to the beginning of Step 2c. If not, proceed to the next step.

Lemma 6. *If $P_c' \leq P_c^s$, then the resulting matching of Algorithm 3 associated with P_c' is stable under (S, C, P_c', P_{-c}, q).*

[20] Note that Lemma 3 can be applied since for sufficiently large n and $c \geq c^*(n)$, we have $c > 4k$.

Proof. By construction, the set $I \equiv C \cup S \setminus B_c^1$ satisfies, under the market (S, C, P_c', P_{-c}, q), that

(1) The restriction of μ on I is stable in a market restricted to I.

(2) No agent in I have partners outside of I.

Algorithm 3 is a stochastic variant of the algorithm proposed by Roth and Vande Vate (1990). Roth and Vande Vate (1990) show that the algorithm terminates and the resulting matching is stable under conditions (1) and (2), completing the proof. $\square$

Result 1 (Gale and Shapley (1962)). *Let (S, C, P, q) be a market and μ' be a stable matching of (S, C, P, q). Then for any $c \in C$, we have*

$$\mu' \succeq_c \phi(S, C, P, q).$$

Lemma 7.

$$Pr\left[\phi(S, C, P_c', P_{-c}, q) \succ_c \mu(c) \setminus B_c^1 \text{ for some } P_c' \leq P_c\right] \leq \pi_c.$$

Proof. Compare Algorithms 2 and 3 for (S, C, P_c', P_{-c}, q). Whenever Algorithm 2 terminates at Step 2a, Algorithm 3 terminates while no new offer is given to c for any $P_c' \leq P_c$. In such a case the resulting matching of Algorithm 3 is $\mu(c) \setminus B_c^1$, which is stable by Lemma 6. By Result 1, we have $\mu(c) \succeq_c \mu(c) \setminus B_c^1 \succeq_c \phi(S, C, P_c', P_{-c}, q)$, completing the proof. $\square$

Now we prove Theorem 1. By Lemma 7, the probability that some truncation is profitable for $c \in C$ is at most π_c. Using Lemma 1, we obtain

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

completing the proof.

Proof of Corollary 1

Proof. When $q_{c'} = 1$ for every $c' \in C$, for any preference list profile P_{-c} there is a truncation of P_c that is a best response to P_{-c} (Theorem 2 of Roth and Vande Vate (1991).) This observation and Theorem 1 completes the proof. $\square$

Proof of Theorem 2

Let $B'_c(q'_c)$ be the q'_c most preferred students in $\mu(c)$. Suppose that the S-optimal stable matching is reached through Algorithm 1. Now consider the following algorithm:

Algorithm 4. Stochastic Vacancy Chains

The algorithm is the same as Algorithm 2 except for the following changes. Use $q'_c < q_c$ as the capacity of c in Step 2(c)ii. Also, replace Steps 1b and 2a of Algorithm 2 with Steps 1b" and 2a" described below.

1b" Let $B^1_c = \mu(c) \setminus B'_c(q'_c)$.

2a" If not, select c' randomly from distribution D until $c' \notin A_b$, and add c' to A_b.

Let the resulting matching be denoted by $\phi'(q'_c)$. The above algorithm is a stochastic version of the vacancy chain dynamics analyzed by Blum, Roth, and Rothblum (1997). They show that the algorithm terminates at a stable matching in the new market with the reduced positions. More specifically,

Result 2 (Theorem 4.3 of Blum, Roth, and Rothblum (1997)). $\phi'(q'_c)$ *is stable in* (S, C, P, q'_c, q_{-c}).

Lemma 8. *For any* (S, C, P, q),

$$Pr\left[\phi'(q'_c) \neq B'_c(q'_c) \text{ for some } q'_c < q_c\right] \leq \pi_c.$$

Proof. Compare Algorithms 2 and 4 for (S, C, P, q). Whenever Algorithm 2 terminates at Step 2a, Algorithm 4 terminates while no new offer is given to c for any $q'_c < q_c$. In such a case we have $\phi'(q'_c)(c) = B'_c(q'_c)$ by definition of Algorithm 4. □

Now we prove Theorem 2. By Results 1 and 2 we have that

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

for any $q'_c \leq q_c$.

By Lemma 8, with probability of at least $1 - \pi_c$ we have that

$$\phi'(q'_c) = B'_c(q'_c),$$

for any $q'_c \leq q_c$, which is less preferred to $\phi(S, C, P, q)$. In such a case we have

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

Using Lemma 1, we obtain

$$\begin{aligned}\beta_k(n) &\leq \frac{16\bar{q}nk}{\ln(\bar{q}n)} + \frac{(\bar{q}+1)\ln(\bar{q}n)\sqrt{\bar{q}n}}{2\bar{q}k} \\ &= o(n),\end{aligned}$$

completing the proof.

Proof of Theorem 3

Lemma 9. *If $c \in C$ can manipulate via pre-arrangement with $s \in S$, then*

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

Proof. First it is well known that if c can manipulate via pre-arrangement with student s, then either $s \in \phi(c)$ or $s' \succ_c s$ for every $s' \in \phi(S, C, P, q)(c)$ (Theorem 2 of Sönmez (1999).)[21] Suppose $s \in \mu(c)$. Then it is easy to see that the matching μ' given by

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

is stable in $(S \setminus s, C, P_{-s}, q_c - 1, q_{-c})$. Therefore by Result 1 and responsiveness,

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

Therefore c cannot manipulate, which completes the proof. □

Therefore, in order to profitably manipulate, a college has to pre-arrange a match with a strictly less preferred student and then compensate it by matching to a better set of students in the centralized matching procedure after pre-arrangement.

Result 3 (Gale and Sotomayor (1985)). *Let (S, C, P, q) be a market. For any student $s \in S$ and $c \in C$, we have*

$$\phi(S, C, P, q) \succeq_c \phi(S \setminus s, C, P_{-s}, q).$$

[21] This conclusion holds for any stable mechanism. See Sönmez (1999).

Now we prove Theorem 3.

By Result 3, we have

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

By Lemma 8 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, Result 3 and Lemma 9. Therefore the probability that c benefits via pre-arrangement is at most π_c. Finally, by an argument similar to those for Theorems 1 and 2 we complete the proof.

Proof of Lemma 2

Let $\pi_c^{**} = Pr[\text{Algorithm 2 terminates at Step 2a} | Y_n > EY_n/2]$. By an argument analogous to the one in the proof of Lemma 1, we obtain

$$\pi_c^{**} \leq \frac{4T\bar{q}}{EY_n},$$

for any c. Therefore we have

$$\begin{aligned} \pi_c &\leq Pr[Y_n \leq EY_n/2] + Pr[Y_n > EY_n/2]\pi_c^{**} \\ &\leq \frac{4}{EY_n} + \pi_c^{**} \\ &\leq \frac{4(T\bar{q} + 1)}{EY_c}. \end{aligned}$$

Applying Lemma 3 for $c = n$, we complete the proof.

Proof of Theorem 4

(1) Consider an arbitrary $c \in C$. As in the proof of Theorem 1, the probability that c profitably manipulates via truncating preference is at most π_c. By Lemma 2 we have $\pi_c \to 0$ as $n \to \infty$, completing the proof.

(2) Suppose that colleges other than c are truth-telling, that is, $P'_{-c} = P_{-c}$. By (1) for any $\varepsilon > 0$ there exists n_0 such that for any market Γ_n with $n > n_0$, we have

$$Pr\left[u(\phi(S, C, P'_c, P_{-c}, q)(c)) > u(\phi(S, C, P, q)(c))\right] < \frac{\varepsilon}{q_c \sup_{\Gamma_n, s \in S} \hat{u}(s)}.$$

Such n_0 can be chosen independent of $c \in C$, which implies that P_C is an ε-Nash equilibrium.

Proof of Theorem 5

(1) Consider an arbitrary $c \in C$. As in the proof of Theorem 2, the probability that c profitably manipulates via capacity is at most π_c. By Lemma 2 we have $\pi_c \to 0$ as $n \to \infty$, completing the proof.

(2) Suppose that colleges other than c are reporting q'_{-c}. By (1) for any $\varepsilon > 0$ there exists n_0 such that for any market Γ_n with $n > n_0$, we have

$$Pr\left[u(\phi(S, C, P, q'_c, q'_{-c})(c)) > u(\phi(S, C, P, q_c, q_{-c})(c))\right] < \frac{\varepsilon}{\sup_{\Gamma_n, s \in S} \hat{u}(s)}.$$

Therefore q_c is an ε-dominant strategy. Note that n_0 can be chosen independent of $c \in C$.

Proof of Theorem 6

Consider an arbitrary $c \in C$. As in the proof of Theorem 3, the probability that c profitably manipulates via pre-arrangement is at most π_c. By Lemma 2 we have $\pi_c \to 0$ as $n \to \infty$, completing the proof.

References

Abdulkadiroğlu, A., P. A. Pathak, and A. E. Roth (2005): "The Boston Public School Match," *American Economic Review Papers and Proceedings*, 95, 368–372.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2005): "Changing the Boston School Choice Mechanism," Harvard University, Unpublished mimeo.

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

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 (2005): "Depth and Efficiency of Large Double Auctions," forthcoming, *Econometrica*.

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

Ergin, H., and T. Sonmez (2005): "Games of school choice under the Boston mechanism," forthcoming, Journal of Public Economics.

Fudenberg, D., M. Mobius, and A. Szeidl (2003): "Existence of Equilibria in Large Double Auctions," Unpublished, Harvard University.

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., 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.

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 J. VANDE VATE (1990): "Random path to stability in two-sided matching," *Econometrica*, 58(6), 1475–1480.

——— (1991): "Incentives in two-sided matching with random stable mechanisms," *Economic Theory*, 1, 31–44.

RUSTICHINI, A., M. SATTERTHWAITE, AND S. WILLIAMS (1994): "Converge 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.

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