*Incentives and Stability in Large Two-Sided Matching Markets*



**Citation:** Kojima, Fuhito, and Parag A. Pathak. 2009. "Incentives and Stability in Large Two-Sided Matching Markets." American Economic Review, 99(3): 608–27.© 2009 AEA

**As Published:** http://dx.doi.org/10.1257/aer.99.3.608

**Publisher:** American Economic Association

**Persistent URL:** http://hdl.handle.net/1721.1/61663

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

*American Economic Review 2009, 99:3, 608–627*
*http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.3.608*

# Incentives and Stability in Large Two-Sided Matching Markets

*By* Fuhito Kojima and Parag A. Pathak*

*A number of labor markets and student placement systems can be modeled as many-to-one matching markets. We analyze the scope for manipulation in many-to-one matching markets under the student-optimal stable mechanism when the number of participants is large. Under some regularity conditions, we show that the fraction of participants with incentives to misrepresent their preferences when others are truthful approaches zero as the market becomes large. With an additional condition, truthful reporting by every participant is an approximate equilibrium under the student-optimal stable mechanism in large markets.* (*JEL* C78)

In recent years, game theoretic ideas have been used to study the design of markets. Auctions have been employed to allocate radio spectrum, timber, electricity, and natural gas involving hundreds of billions of dollars worldwide (Paul R. Milgrom 2004). Matching procedures have found practical applications in centralized labor markets as well as school assignment systems in New York City and Boston.[1] Connections between auctions and matching have been explored and extended (Alexander Kelso and Vincent Crawford 1982; John Hatfield and Milgrom 2005; and Robert Day and Milgrom 2008).

The practical aspects of market design have led to an enrichment of the theory of market design. This paper investigates a theoretical problem motivated by the use of stable matching mechanisms in large markets, inspired by a practical issue first investigated by Roth and Peranson (1999).[2] A matching is stable if there is no individual agent who prefers to become unmatched or pair of agents who prefer to be assigned to each other to being assigned their allocation in the matching. In real-world applications, empirical studies have shown that stable mechanisms often succeed, whereas unstable ones often fail.[3]

* Kojima: Cowles Foundation, Yale University, New Haven, CT 06510 (e-mail: fuhitokojima1979@gmail.com); Pathak: Society of Fellows, Harvard University, Cambridge, MA 02138, and Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142 (e-mail: ppathak@mit.edu). We appreciate helpful discussions with Attila Ambrus, Eric Budish, Federico Echenique, Drew Fudenberg, Akihiko Matsui, Alvin E. Roth, Tayfun Sönmez, Satoru Takahashi, and seminar participants at Harvard, Hitotsubashi University, Kyoto University, Osaka University, Tokyo University, the 2006 North American Summer Meetings of the Econometric Society at Minneapolis, the 17th International Conference on Game Theory at SUNY Stony Brook, and 2006 INFORMS Annual Meeting at Pittsburgh. Discussions with Michihiro Kandori and Yohei Sekiguchi were especially important for the current version of the paper. Detailed comments by the coeditor and anonymous referees greatly improved the substance and exposition of the paper. For financial support, Pathak is grateful to the National Science Foundation, the Spencer Foundation, and the Division of Research at the Harvard Business School.

[1] For a survey of this theory, see Alvin E. Roth and Marilda A. Oliveira Sotomayor (1990). For applications to labor markets, see Roth (1984a) and Roth and Elliot Peranson (1999). For applications to student assignment, see for example Atila Abdulkadiroğlu and Tayfun Sönmez (2003), Abdulkadiroğlu et al. (2005), and Abdulkadiroğlu, Pathak, and Roth (2005).

[2] A recent paper by Jeremy Bulow and Jon Levin (2006) theoretically investigates a matching market with price competition, motivated by an antitrust case against the National Residency Matching Program. See also Muriel Niederle (2007).

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

Although stable mechanisms have a number of virtues, they are susceptible to various types of strategic behavior before and during the match. Roth (1982) shows that any stable mechanism is manipulable via preference lists: for some participants, reporting a true preference list (ordinal ranking over potential matches) may not be a best response to reported preferences of others. In many-to-one markets such as matching markets between colleges and students, where some colleges have more than one position, Sönmez (1997b, 1999) shows that there are additional strategic concerns. First, any stable mechanism is manipulable via capacities; that is, colleges may sometimes benefit by underreporting their quotas. Second, any stable mechanism is manipulable via prearranged matches; that is, 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 unstable mechanism: "Before you might have a situation where a school was going to take 100 new children for ninth grade, they might have declared only 40 seats, and then placed the other 60 outside of the process" (*New York Times,* November 19, 2004).[4] Roth and Uriel G. 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 United States.

The aim of this paper is to understand why, despite these negative results, many stable mechanisms appear to work well in practice. We focus on the student-optimal stable mechanism, which forms the basis of many matching mechanisms used in the field, such as in NYC and the NRMP. Lester E. Dubins and David A. Freedman (1981) and Roth (1982) show that students do not have incentives to manipulate markets with the student-optimal stable mechanism, though colleges do have incentives to do so. Our results show that the mechanism is immune to various kinds of manipulations by colleges when the market is large. In real-world two-sided matching markets, there are often a large number of applicants and institutions, and each applicant submits a preference list containing only a small fraction of institutions in the market. For instance, in the NRMP, the length of the applicant preference list is about 15, while the number of hospital programs is between 3,000 and 4,000 and the number of students is over 20,000 per year. In NYC, about 75 percent of students submit preference lists of fewer than 12 schools, and there are about 500 school programs and over 90,000 students per year.[5] Motivated by these features, we study how the scope for manipulations changes when the number of market 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 have incentives to manipulate the student-optimal stable mechanism when every other college is truth-telling converges to zero as the number of colleges approaches infinity. The key step of the proof involves showing that, when there are a

[4] A careful reader may recognize that this quote is not about a stable mechanism but about an unstable one. We present this quote just to suggest that strategic behavior may be a realistic problem in general. Indeed, concern about strategic behavior motivated New York City's recent adoption of a stable mechanism, and this paper suggests that scope for manipulations may be limited under the current stable mechanism despite theoretical possibility of manipulation.

[5] For data regarding the NRMP, see http://www.nrmp.org/2006advdata.pdf. For data regarding New York City high school match, see Abdulkadiroğlu et al. (2008).

large number of colleges, the chain reaction caused by a college's strategic rejection of a student is unlikely to make another, preferred student apply to that college.[6]

This result does not necessarily mean that agents report true preferences *in equilibrium*. Thus, we also conduct equilibrium analysis in the large market. We introduce an additional condition, called sufficient thickness, and show that truthful reporting is an approximate equilibrium in a large market that is sufficiently thick.

### A. *Related Literature*

Our paper is most closely related to Roth and Peranson (1999) and Nicole Immorlica and Mohammed Mahdian (2005). Roth and Peranson (1999) conduct a series of simulations on data from the NRMP and on randomly generated data. In their simulations, very few agents could have benefited by submitting false preference lists or by manipulating capacity in large markets when every other agent is truthful. These simulations lead them to conjecture that the fraction of participants with preference lists of limited length who can manipulate tends to zero as the size of the market grows.[7]

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 to the same student in all stable matchings approaches one. Since a college does not have incentives to manipulate via preference lists if and only if it is matched to the same student in all stable matchings (see David Gale and Sotomayor 1985 and Gabrielle Demange, Gale, and Sotomayor 1987), this result implies that most colleges cannot manipulate via preference lists when the market is large.

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 a college can sometimes manipulate via preference lists even if the college is matched to the same set of students in all stable matchings.[8] Moreover, in many-to-one markets there exists the additional possibilities of capacity manipulation and manipulation via pre-arrangement (unraveling) which are not present in a one-to-one market.[9] We introduce new techniques to overcome these complications.

Second, previous research mostly focuses on counting the average number of participants that can manipulate the student-optimal stable mechanism, assuming that others report their preferences truthfully. This leaves open the question of whether participants will behave truthfully *in equilibrium*. A substantial part of this paper investigates this question and shows that truthful reporting is an approximate equilibrium in a large market that is sufficiently thick.[10]

[6] Throughout the paper we focus on the SOSM. The college-optimal stable mechanism can be similarly defined by letting colleges propose to students. By making complementary assumptions about the model, we can derive results concerning incentives of students under the college-optimal stable mechanism similar to those in the current paper. However, additional assumptions may be needed for analyzing the incentives of colleges in this case, since truth-telling is not a dominant strategy for colleges with a quota larger than one in the college-optimal stable mechanism.

[7] The property that in large matching problems the size of the set of stable matchings has also been documented using data from Boston Public School's assignment system. Pathak and Sönmez (2008) report that in Boston, for the first two years of data from SOSM, in elementary school, the student-optimal stable matching and college-optimal stable matching coincide, while for middle school, at most three students are assigned to different schools in the two matchings.

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

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

[10] Immorlica and Mahdian (2005) claim that truth-telling is an approximate equilibrium in a one-to-one market even without sufficient thickness. In Section IV, we present an example to show that this is not the case, but truth-telling is

Incentive properties in large markets are studied in other areas of economics. For example, in the context of double auctions, Thomas Gresik and Mark Satterthwaite (1989), Aldo Rustichini, Sattherwaite, and Steven Williams (1994), Wolfgang Pesendorfer and Jeroen Swinkels (2000), Swinkels (2001), Drew Fudenberg, Markus Mobius, and Adam Szeidl (2005), and Martin Cripps and Swinkels (2006) show that the equilibrium behavior converges to truth-telling as the number of traders increases under various informational structures. Papers more closely related to ours are discussed in Section IV.

There is a literature that analyzes the consequences of manipulations via preference lists and capacities in complete information finite matching markets. See Roth (1984b), Roth (1985), and Sönmez (1997a) for games involving preference manipulation and Hideo Konishi and Utku Ünver (2005) and Kojima (2006) for games of capacity manipulation. Some of the papers most relevant to ours are discussed in Section IV.

The paper proceeds as follows. Section I presents the model and introduces a lemma which is key to our analysis. Section II defines a large market and presents our main result. Section III conducts equilibrium analysis. Section IV concludes. All proofs are in the online Appendix (available at http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.3.608).

## I. Model

### A. *Preliminary Definitions*

Let there be a set of students $S$ and a set of colleges $C$. Each student $s$ has a strict preference relation $P_s$ over the set of colleges and being unmatched (being unmatched is denoted by $s$). Each college $c$ has a strict preference relation $\succ_c$ 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 $cP_s s$. A market is tuple $\Gamma = (S, C, P_S, \succ_C)$ where $P_S = (P_s)_{s \in S}$, and $\succ_C = (\succ_c)_{c \in C}$.

Since only rankings of acceptable mates matter for our analysis, we often write only acceptable mates to denote preferences. For example,

$$P_s\!: c_1, c_2,$$

means that student $s$ 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 the ranking of a student is independent of her colleagues, and any set of students exceeding quota $q_c$ is unacceptable (see Roth 1985 for a formal definition).[11] We will assume that all preferences are responsive throughout the paper.

Let $P_c$ be the corresponding *preference list of college c*, which is the preference relation over individual students and $\emptyset$. Sometimes only the preference list structure and quotas are relevant for the analysis. Therefore at times we 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: $P_{-i} = (P_j)_{j \in S \cup C \setminus i}$ and $q_{-c} = (q_{c'})_{c' \in C \setminus c}$.

an approximate equilibrium under an additional assumption of sufficient thickness.

[11] Note that, given a responsive preference with quota $\succ_c$, we can always find a utility function $u_c : S \rightarrow \mathbb{R}$ with the property that for all $S', S'' \subseteq S$ such that $|S'|, |S''| \leq q_c$, $S' \succ_c S''$ if and only if $\sum_{s \in S'} u_c(s) > \sum_{s \in S''} u_c(s)$. In Section III, we will use a particular additive representation of responsive preferences.

A *matching* $\mu$ is a mapping from $C \cup S$ to itself 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 $s \in \mu(c)$. That is, a matching simply specifies the college where each student is assigned or if the student is unmatched, and the set of students assigned to each college, if any.

We say a matching $\mu$ is *blocked* by a pair of student $s$ and college $c$ if $s$ strictly prefers $c$ to $\mu(s)$ and either (i) $c$ strictly prefers $s$ to some $s' \in \mu(c)$ or (ii) $|\mu(c)| < q_c$ and $s$ is acceptable to $c$. In other words, the student $s$ in the pair prefers college $c$ over her assignment in $\mu$, and college $c$ prefers $s$ either because it has a vacant seat or $s$ is more preferred than another student assigned to $c$ under $\mu$. A matching $\mu$ is *individually rational* if for each student $s \in S$, $\mu(s) P_s \emptyset$ or $\mu(s) = \emptyset$, and for each $c \in C$, (i) $|\mu(c)| \leq q_c$ and (ii) $s \succ_c \emptyset$ for every $s \in \mu(c)$. A matching $\mu$ 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 with respect to reported preferences for every market.

We consider the SOSM, denoted by $\varphi$, which is analyzed by Gale and Lloyd Shapley (1962).[12]

- **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 the 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 (if any). 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 the 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. For two preference relations that induce the same pair of preference lists and quotas, the outcome of the algorithm is the same. Thus, we sometimes write the resulting matching by $\varphi(S, C, P, q)$. We denote by $\varphi(S, C, P, q)(i)$ the assignment given to $i \in S \cup C$ under matching $\varphi(S, C, P, q)$.

### B. *Manipulating the Student-Optimal Stable Mechanism*

We illustrate two ways that the SOSM can be manipulated through a simple example.

EXAMPLE 1: *Consider the following market with two colleges $c_1$ and $c_2$, and five students $s_1, \ldots, s_5$. Suppose $q_{c_1} = 3$ and $q_{c_2} = 1$, and the preference lists of colleges are*

$$P_{c_1} : s_1, s_2, s_3, s_4, s_5,$$

$$P_{c_2} : s_3, s_2, s_1, s_4, s_5,$$

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

and student preferences are

$$P_{s_1} : c_2, c_1,$$

$$P_{s_2} : c_1, c_2,$$

$$P_{s_3} : c_1, c_2,$$

$$P_{s_4} : c_1, c_2,$$

$$P_{s_5} : c_1, c_2.$$

The matching produced by the SOSM is

$$\mu = \begin{pmatrix} s_2 & s_3 & s_4 & s_1 \\ c_1 & c_1 & c_1 & c_2 \end{pmatrix},$$

which means that $c_1$ is matched to $s_2$, $s_3$, and $s_4$, $c_2$ is matched to $s_1$, and $s_5$ is unmatched.[13]

The first type of manipulation we will focus on is a *preference manipulation (manipulation via preference lists)*, first identified by Dubins and Freedman (1981). Suppose college $c_1$ submitted the following preference list:

$$P'_{c_1} : s_1, s_2, s_5, s_4, s_3,$$

while reporting its true quota, 3. Then the SOSM produces the following matching:

$$\mu' = \begin{pmatrix} s_1 & s_2 & s_5 & s_3 \\ c_1 & c_1 & c_1 & c_2 \end{pmatrix}.$$

If $c_1$ prefers $\{s_1, s_2, s_5\}$ to $\{s_2, s_3, s_4\}$, say because student $s_1$ is particularly desirable, then it can benefit by misreporting its preferences.[14]

Another way a stable mechanism can be manipulated, identified by Sönmez (1997b), is by *capacity manipulation (manipulation via capacities)*. Suppose that college $c_1$ states its quota as $q'_{c_1} = 1$, while reporting its preference list $P_{c_1}$ truthfully. Then the SOSM produces the following matching:

$$\mu'' = \begin{pmatrix} s_1 & s_3 \\ c_1 & c_2 \end{pmatrix}.$$

If $s_1$ is more desirable than $\{s_2, s_3, s_4\}$, then college $c_1$ can benefit by reporting $q'_{c_1} = 1$.[15]

These two cases illustrate how the SOSM can be manipulated. We will consider both preference and capacity manipulations and their combination, and refer to this as a manipulation:

[13] Similar notation is used throughout the paper.

[14] Note that relation $\{s_1, s_2, s_5\} \succ_{c_1} \{s_2, s_3, s_4\}$ is consistent with the assumption that $\succ_c$ is responsive.

[15] We note that, in this example, capacity manipulation may not benefit $c_1$ for some responsive preferences consistent with a fixed preference list $P_{c_1}$. Both $s_1 \succ_{c_1} \{s_2, s_3, s_4\}$ and $\{s_2, s_3, s_4\} \succ_{c_1} s_1$ are consistent with $P_{c_1}$, and capacity manipulation $q'_{c_1}$ benefits $c_1$ in the former but not in the latter. Konishi and Ünver (2006) and Kojima (2007b) characterize the subclass of responsive preferences under which capacity manipulations benefit manipulating colleges under stable mechanisms. For preference manipulations, on the contrary, there may exist manipulations that benefit the manipulating college for *all* responsive preferences that are consistent with a given preference list. In the current example, for instance, if $c_1$ declares $\tilde{P}_{c_1} : s_1, s_2, s_4, s_5$, then $c_1$ is matched to $\{s_1, s_2, s_4\}$. This is unambiguously preferred by $c_1$ to the match under truth-telling $\{s_2, s_3, s_4\}$ for any responsive preferences consistent with $P_{c_1}$.

DEFINITION 1: *A college c can* ***manipulate the SOSM*** *if there exists a pair of a preference list and a quota* $(P_c', q_c')$ *with* $q_c' \in \{1, \ldots, q_c\}$ *such that*

$$\varphi(S, C, (P_c', P_{-c}), (q_c', q_{-c}))(c) \succ_c \varphi(S, C, P, q)(c).$$

We assume $q_c' \leq q_c$ because it is easily seen that reporting a quota larger than its true quota is never profitable.

## C. *Dropping Strategies*

The previous section presented an example illustrating preference and capacity manipulations. To study how likely it is that manipulations are successful, one way to begin may be to consider all possible strategies of a particular college. For a college with a preference list of five students and a quota of three as in Example 1, this would involve considering all possible combinations of preference lists and quotas, which is 975 possible strategies, and verifying whether any of these strategies benefits the college.[16]

Fortunately, there is a general property of the manipulations we have discussed that allows us to greatly simplify the analysis and is one of the main building blocks of the analysis that follows. A reported pair of a preference list and a quota is said to be a *dropping strategy* if it simply declares some students who are acceptable under the true preference list as unacceptable. In particular, it does not misreport quotas, change the relative ordering of acceptable students, or declare unacceptable students as acceptable.[17]

Returning to our example before, the outcome of the previous preference manipulation $P_{c_1}'$ can be achieved by the dropping strategy $P_{c_1}''$: $s_1, s_2, s_5$, which simply drops students $s_3$ and $s_4$ from the true preference list. In the case of the previous capacity manipulation, if the college just dropped all students except $s_1$, then it would receive the same outcome as reducing the quota to one. Note that in both of these cases, the original ordering of students is unchanged and only students on the original preference list are dropped.

The observation above turns out to be general for any stable mechanism. The following lemma, which is formally stated and proved in Appendix B, shows that the outcome of any manipulation can be replicated or improved upon by some dropping strategy.

LEMMA 1 (Dropping strategies are exhaustive)*: Consider an arbitrary stable mechanism. Fix preferences of colleges other than c. Suppose the mechanism produces* $\mu$ *under some report of c. Then there exists a dropping strategy that produces a matching that c weakly prefers to* $\mu$ *according to its true preferences.*

Some intuition for this result can be seen in the example above. Consider a preference manipulation $P_{c_1}'$, which demoted $s_3$ to the bottom of the list of acceptable students. At one step of the SOSM, $s_3$ is rejected by $P_{c_1}'$ because three other students applied to $c_1$, and $s_3$ is declared to be the least desirable. Then $s_3$ applies to $c_2$, which rejects $s_1$. Then $s_1$ applies to $c_1$ which accepts her, which benefits $c_1$. A similar chain of rejection and acceptances can be initiated if $c_1$ declares just $s_3$ as unacceptable. Similarly, additional chains of rejections and acceptances produced by under-reporting of quotas can be replicated by declaring some students as unacceptable instead.

[16] All possible combinations are calculated by observing that there are $5! + 5 \times 4 \times 3 \times 2 + 5 \times 4 \times 3 + 5 \times 4 + 5$ ways to submit preference lists and 3 ways to report capacities.

[17] Let $(P_c, q_c)$ be a pair of the true preference list and true quota of college $c$. Formally, a dropping strategy is a report $(P_c', q_c)$ such that (i) $sP_cs'$ and $sP_c'\emptyset$ imply $sP_c's'$, and (ii) $\emptyset P_c s$ implies $\emptyset P_c' s$.

There are a few remarks about this lemma. First, while most of our analysis focuses on the SOSM, the lemma identifies a general property of any stable matching mechanism, which may be of independent interest. Second, a truncation strategy that drops students only from the end of its preference list can replicate any profitable manipulation in a simpler one-to-one matching market (Roth and John H. Vande Vate 1991), and our lemma offers its counterpart in a more general many-to-one matching market.[18] Finally, and most important in the context of our work, the lemma allows us to focus on a particular class of strategies to investigate manipulations. In the previous example, we need only consider $2^{|S|} = 2^5 = 32$ possible dropping strategies instead of all 975 strategies. In addition to being small in number, the class of dropping strategies turns out to be conceptually simple and analytically tractable.

## II. Large Markets

### A. *Regular Markets*

We have seen that a finite many-to-one matching market can be manipulated. To investigate how likely a college can manipulate the SOSM in large markets, we introduce the following random environment. A *random market* is a tuple $\tilde{\Gamma} = (C, S, \succ_C, k, \mathcal{D})$, 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 $c \in C$.[19] Each random market induces a market by randomly generating preferences of each student $s$ as follows (Immorlica and Mahdian 2005):

- **Step 1**: Select a college independently from distribution $\mathcal{D}$. List this college as the top ranked college of student $s$.

In general,

- **Step $t \leq k$**: Select a college independently from distribution $\mathcal{D}$ until a college is drawn that has not been previously drawn in steps 1 through $t-1$. List this college as the $t^{\text{th}}$ most preferred college of student $s$.

Student $s$ finds these $k$ colleges acceptable, and all other colleges unacceptable. For example, if $\mathcal{D}$ is the uniform distribution on $C$, then the preference list is drawn from the uniform distribution over the set of all preference lists of length $k$. For each realization of student preferences, a market with perfect information is obtained.

In the main text of the paper, we focus on the procedure above for distribution $\mathcal{D}$ to generate preferences of students for the sake of simplicity. For readers interested in how this can be generalized, we refer to the discussion in Section IV and Appendix A.5, where we describe additional results under weaker assumptions.

A *sequence of random markets* is denoted by $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \ldots)$, where $\tilde{\Gamma}^n = (C^n, S^n, \succ_{C^n}, k^n, \mathcal{D}^n)$ is a random market in which $|C^n| = n$ is the number of colleges.[20] Consider the following regularity conditions.

[18] Truncation strategies have found use in subsequent work, such as Roth and Peranson (1999), Roth and Rothblum (1999), Jinpeng Ma (2002), and Lars Ehlers (2004). Truncation strategies may not be exhaustive when colleges may have a quota larger than one. For instance, in Example 1 we have seen that $c_1$ can be matched with $\{s_1, s_2, s_5\}$ by a dropping strategy $P''_{c_1}$, but such a matching or anything better for $c_1$ cannot be attained by a truncation strategy.

[19] We impose this assumption to compare our analysis with existing literature. Our analysis remains unchanged when one allows for probabilities to be zero.

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

DEFINITION 2: *A sequence of random markets* $(\tilde{\Gamma}^1, \tilde{\Gamma}^2, \ldots)$ *is* ***regular*** *if there exist positive integers* $k$ *and* $\overline{q}$ *such that*

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

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

(3) $|S^n| \leq \overline{q}n$ for all $n$,[21] and

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

Condition (1) assumes that the length of students' preference lists does not grow when the number of market participants grows. Condition (2) requires that the number of positions of each college is bounded across colleges and markets. Condition (3) requires that the number of students does not grow much faster than that of colleges. Condition (4) requires colleges to find any student acceptable, but preferences are otherwise arbitrary.[22] This paper focuses on regular sequences of random markets and makes use of each condition in our arguments. In the last section, we will discuss directions in which these conditions can be weakened.

### B. *Main Result*

Consider the expected number of colleges that can manipulate the SOSM when others are truthful. Formally, such a number is defined as

$$\alpha(n) = E[\#\{c \in C \mid \varphi(S, C, (P'_c, P_{-c}), (q'_c, q_{-c}))(c) \succ_c \varphi(S, C, P, q)(c)$$
$$\text{for some } (P'_c, q'_c) \text{ in the induced market}\} \mid \tilde{\Gamma}^n].$$

The expectation is taken with respect to random student preferences, given the random market $\tilde{\Gamma}^n$. Note that we consider the possibility of manipulations under complete information, that is, we investigate for colleges' incentives to manipulate when they know the preferences of every agent. The randomness of student preferences is used only to assess the frequency of situations in which colleges have incentives to manipulate.

THEOREM 1: *Suppose that the sequence of random markets is regular. Then the expected proportion of colleges that can manipulate the SOSM when others are truthful,* $\alpha(n)/n$*, goes to zero as the number of colleges goes to infinity.*

This theorem suggests that manipulation of any sort within the matching mechanism becomes unprofitable to most colleges, as the number of participating colleges becomes large. In Appendix A, we show that a manipulation that involves unraveling outside the centralized mechanism,

[21] As mentioned later, this condition can be relaxed to state: there exists $\tilde{q}$ such that $|S^n| \leq \tilde{q}n$ for any $n$ (in other words, here we are assuming $\overline{q} = \tilde{q}$ just for expositional simplicity). In particular, the model allows situations in which there are more students than the total number of available positions in colleges.

[22] Condition (4) ensures that our economy has a large number of individually rational matchings, so potentially there is nontrivial scope for manipulations. It is possible to weaken this condition such that many, but not all, colleges find all students to be acceptable.

named manipulation via pre-arranged matches (Sönmez 1999), also becomes unprofitable as the market becomes large.

A limitation of Theorem 1 is that it does not imply that most colleges report true preferences *in equilibrium*. Equilibrium analysis is conducted in the next section.

The theorem has an implication about the structure of the set of stable matchings. It is well-known that if a college does not have incentives to manipulate the SOSM, then it is matched to the same set of students in all stable matchings. Therefore the following is an immediate corollary of Theorem 1.

COROLLARY 1: *Suppose that the sequence of random markets is regular. Then the expected proportion of colleges that are matched to the same set of students in all stable matchings goes to one as the number of colleges goes to infinity.*

Corollary 1 is referred to as a "core convergence" result by Roth and Peranson (1999). The main theorem of Immorlica and Mahdian (2005) shows Corollary 1 for one-to-one matching, in which each college has a quota of one.

The formal proof of Theorem 1 is in Appendix B. For the main text, we give an outline of the argument. We begin the proof by recalling that if a college can manipulate the SOSM, then it can do so by a dropping strategy. Therefore, when considering manipulations, we can restrict attention to a particular class of strategies that simply reject students who are acceptable under the true preferences.

The next step involves determining the outcome of dropping strategies. One approach might be to consider all possible dropping strategies and determine which ones are profitable for a college. While the set of dropping strategies is smaller than the set of all possible manipulations, this task is still daunting because the number of possible dropping strategies is large when there is a large number of students.

Thus, we consider an alternative approach. We start with the student-optimal stable matching under the true preferences, and examine whether a college might benefit by dropping some student assigned in the student-optimal matching. Specifically, we consider a process where beginning with the student-optimal matching under true preferences, we drop some students from a particular college's preference list and continue the SOSM procedure starting with the original students rejected by the college because of the dropping. We refer to this process as *rejection chains*.[23]

For instance, in Example 1, starting with the student-optimal stable matching, if college $c_1$ dropped students $\{s_3, s_4\}$, let us first consider what happens to the least preferred student in this set, student $s_4$. If we examine the continuation of the SOSM, when $s_4$ is rejected, this student will propose to $c_2$, but $c_2$ is assigned to $s_1$ whom it prefers, so $s_4$ will be left unassigned. Thus, college $c_1$ does not benefit by rejecting student $s_4$ in this process, because this rejection does not spur another more preferred student to propose to $c_1$. On the other hand, consider what happens when college $c_1$ rejects student $s_3$. This student will propose to college $c_2$, who will reject $s_1$, freeing $s_1$ up to propose to $c_1$. By dropping $s_3$, college $c_1$ has created a new proposal which did not take place in the original SOSM procedure.

Our rejection chains procedure results in two cases: (i) a rejection never leads to a new proposal at the manipulating college as in the case of $s_4$, and (ii) a rejection leads to a new student proposal at the college as in the case of $s_3$. The example above suggests that case (i) is never beneficial for the manipulating college, whereas case (ii) may benefit the college. In Lemma 3 in

[23] This process is formally defined in the Appendix B.

the Appendix, we make this intuition precise: if the rejection chain process does not lead to a new proposal (case (i)), then a college cannot benefit from a dropping strategy.[24] This result allows us to link the dropping strategy to the rejection chains algorithm. This connection gives us traction in large markets, as the number of cases to consider is bounded by a constant $2^{\bar{q}}$ for rejection chains, whereas there are $2^{|S^n|}$ potential dropping strategies in a market with $n$ colleges, and the latter number may approach infinity as $n$ approaches infinity.

The last step of the proof establishes that the probability that a rejection chain returns to the manipulating college is small when the market is large. To see the intuition for this step, suppose that there is a large number of colleges in the market. Then there is also a large number of colleges with vacant positions with high probability. We say that a college is popular if it is given a high probability in the distribution from which students preferences are drawn. Any student is much more likely to apply to one of those colleges with vacant positions rather than the manipulating college unless it is extremely popular in a large market, since there is a large number of colleges with vacant positions. Since every student is acceptable to any college by assumption, the rejection chains algorithm terminates without returning to the manipulating college if such an application happens. Thus, the probability that the algorithm returns to the manipulating college is very small unless the college is one of the small fraction of very popular colleges. 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, we conclude that the expected proportion of colleges that can successfully manipulate converges to zero when the number of colleges grows.

Both Roth and Peranson (1999) and Immorlica and Mahdian (2005) attribute the lack of manipulability to the "core-convergence" property as stated in Corollary 1. While this interpretation is valid in one-to-one markets, the fact that a college is matched to the same set of students in all stable matchings is necessary but not sufficient for the college to have incentives for truth-telling 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.

Roth and Peranson (1999) analyze the NRMP data and argue 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 conduct simulations using randomly generated data in one-to-one matching, and observe that the proportion of colleges that can successfully manipulate 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 for one-to-one matching. Theorem 1 improves upon their results and fully explains observations of Roth and Peranson (1999) in the following senses: (i) it studies manipulations via preference lists in many-to-one markets, and (ii) it studies manipulations via capacities. Neither of these points is 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 misreporting preference lists or misreporting capacities.

## III. Equilibrium Analysis

In the last section, we investigated individual colleges' incentives to manipulate the SOSM when all agents are truth-telling. As noted in the previous section, Theorem 1 does not necessarily

[24] We note that the manipulating college may not be made better off even when the rejection chain leads to a new proposal to the college (case (ii)).

mean that agents report true preferences *in equilibrium*. We now allow all participants to behave strategically and investigate equilibrium behavior in large markets. This section focuses on the simplest case to highlight the analysis of equilibrium behavior. Appendix A.5 presents the more general treatment 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 $2^S \to \mathbb{R}$ on the set of subsets of students. More specifically, we assume that

$$u_c(S') \begin{cases} = \sum_{s \in S'} u_c(s) \text{ if } |S'| \leq q_c, \\ < 0 \text{ otherwise,} \end{cases}$$

where $u_c(s) = u_c(\{s\})$. We assume that $sP_c s' \Leftrightarrow u_c(s) > u_c(s')$. If $s$ is acceptable to $c$, $u_c(s) > 0$. If $s$ is unacceptable, $u_c(s) < 0$. Further, suppose that utilities are bounded. Formally, $\sup u_c(s)$ is finite where the supremum is over the size of the market $n$, students $s \in S^n$, and colleges $c \in C^n$.

The normal-form game is specified by a random market $\tilde{\Gamma}$ coupled with utility functions of colleges, and is defined as follows. The set of players is $C$, with von Neumann–Morgenstern expected utility functions induced by the utility functions above. All the colleges move simultaneously. Each college submits a preference list and quota pair. 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 $\varphi$ 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.[25] 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 $\varphi$ (Dubins and Freedman 1981; Roth 1982).

First, let us define some additional notation. Let $P_C = (P_c)_{c \in C}$ and $P_{C-c} = (P_{c'})_{c' \in C, c' \neq c}$. Given $\varepsilon > 0$, a profile of preferences $(P_C^*, q_C^*) = (P_c^*, q_c^*)_{c \in C}$ is an *$\varepsilon$-Nash equilibrium* if there is no $c \in C$ and $(P_c', q_c')$ such that

$$E[u_c(\varphi(S, C, (P_S, P_c', P_{C-c}^*), (q_c', q_{-c}^*))(c))] > E[u_c(\varphi(S, C, (P_S, P_c^*), q^*)(c))] + \varepsilon,$$

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 example.[26]

EXAMPLE 2: *Consider the following market $\tilde{\Gamma}^n$ for any n. There are n colleges and students. Preference lists of $c_1$ and $c_2$ are given as follows:*[27]

$$P_{c_1} : s_2, s_1, \ldots,$$

$$P_{c_2} : s_1, s_2, \ldots.$$

[25] 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. See Appendix A.4.

[26] This example shows that Corollaries 3.1 and 3.3 in Immorlica and Mahdian (2005) are not correct.

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

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 $c \neq c_1, c_2$.[28] 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

$$P_{s_1} : c_1, c_2, \ldots,$$

$$P_{s_2} : c_2, c_1, \ldots.$$

Under the student-optimal matching $\mu$, we have $\mu(c_1) = s_1$ and $\mu(c_2) = s_2$. Now, suppose that $c_1$ submits preference list $P'_{c_1} : s_2$. Then, under the new matching $\mu'$, $c_1$ 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 $\varepsilon$-Nash equilibrium if $\varepsilon > 0$ is sufficiently small, as $c_1$ has an incentive to deviate.

The example above shows that, while the *proportion* of colleges that can manipulate via preference lists becomes small, *for an individual college* the opportunity for such manipulation may remain large. Note, on the other hand, that this is consistent with Theorem 1 since the scope for manipulation becomes small for any $c \neq c_1, c_2$.[29]

A natural question is under what conditions we can expect truthful play to be an $\varepsilon$-Nash equilibrium. One feature of Example 2 is that $c_1$ and $c_2$ are popular and remain so even as the market becomes large. This suggests that it is the colleges that are extremely popular and remain so in a large market that may be able to manipulate.

Consider a market where there is not too much concentration of popularity among a small set of colleges. This feature of the market will reduce the influence of the exceptionally popular colleges that we observed in Example 2. One way to formally define this situation is to consider a sequence of random markets with the following property: there exists a finite bound $T$ and fraction of $a$ colleges such that

$$p^n_1/p^n_{[an]} \leq T, \tag{1}$$

for large $n$, where $p^n_1$ is the popularity of the most popular college and $p^n_{[an]}$ is the popularity of the $a^{\text{th}}$-tile college in market $\tilde{\Gamma}^n$. The condition means that the ratio of the popularity of the most popular college to the popularity of the college at the $a^{\text{th}}$-tile does not grow without bound as the size of the market grows. This condition is satisfied if there is not a small number of colleges that are much more popular than all of the other colleges.

In Example 2, when $n$ is large, $p^n_{c_1}/p^n_{c_2} = 1$, but $p^n_{c_1}/p^n_c = n - 2$ when $c$ is not college $c_1$ or $c_2$, and this ratio grows without bound as $n \to \infty$. We will show later that inequality (1) is sufficient for truth-telling to be an $\varepsilon$-Nash equilibrium.

While this condition is easy to explain, it is not the most general condition that will ensure truthful play. Appendix A.3 presents additional examples with weaker assumptions. We can describe a more general condition with the help of additional notation. Let

$$V_T(n) = \{c \in C^n \mid \max_{c' \in C^n} \{p^n_{c'}\}/p^n_c \leq T, \#\{s \in S^n \mid cP_s s\} < q_c\}.$$

[28] We use superscript $n$ on $p_c$ to indicate the probability of college $c$ for market $\tilde{\Gamma}^n$. This notation is used in subsequent parts of the text.

[29] Capacity manipulation may remain profitable for some colleges in a large market as well. See the Appendix A.6 for an example illustrating this point.

In words, $V_T(n)$ is a random set that denotes the set of colleges sufficiently popular ex ante ($\max_{c'}\{p^n_{c'}\}/p^n_c \leq T$) but there are fewer potential applicants than the number of positions ($\#\{s \in S^n \mid cP_s s\} < q_c$) ex post.

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

$$E[|V_T(n)|] \to \infty,$$

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 vacant seats.[30] In the example above, it is easy to verify that the market is not sufficiently thick. However, many types of markets satisfy the condition of sufficient thickness. For instance, if all student preferences are drawn from the uniform distribution, the market will be sufficiently thick. This is the environment first analyzed by Roth and Peranson (1999). The sufficient thickness condition is satisfied in more general cases, which are described in Appendix A.3.

THEOREM 2: *Suppose that the sequence of random markets is regular and sufficiently thick. Then, for any* $\varepsilon > 0$*, there exists* $n_0$ *such that truth-telling by every college is an* $\varepsilon$*-Nash equilibrium for any market in the sequence with more than* $n_0$ *colleges.*

The proof of Theorem 2 (shown in Appendix B.3) is similar to that of Theorem 1, except for one point. With sufficient thickness, we can make sure that the rejection chain fails to return to any college with high probability, as opposed to only unpopular ones. This is because, with sufficient thickness, 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. The key difference is that, in this section, we have an upper bound of the probability of successful manipulation for *every* college, while in the previous section, we have an upper bound only for unpopular colleges.

We pursue generalizations of this result in the Appendix. First, we consider an incomplete information environment and show the conditions needed for truth-telling to be an $\varepsilon$-Bayesian equilibrium of the game (Appendix A.4). Second, we obtain a similar result when we allow for the possibility of pre-arrangement (Appendix A.6). Third, when we allow a coalition of colleges to manipulate, we extend the result to show the conditions under which coalitions of colleges will have little incentive to manipulate (Appendix A.7). The precise statements and details of each of these extensions is in the corresponding sections of the Appendix.

[30] 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. 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 definition above is finite as $n$ goes to infinity.

## IV. Discussion and Conclusion

Why do many stable matching mechanisms work in practice, even though existing theory suggests that they can be manipulated in many ways? This paper established that the fraction of participants who can profitably manipulate the student-optimal stable mechanism (SOSM) 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. These results suggest that a stable matching will be realized under the SOSM in a large market. Since a stable matching is efficient in general, the results suggest that the SOSM achieves a high level of efficiency in a large market.[31] Taken together, these results provide a strong case for the SOSM as a market design in large markets.

Based on the results in this paper, when do we expect certain matching mechanisms to be hard to manipulate? We will highlight three crucial assumptions here. First, our result is a limit result, and requires that the market is regular and large.[32] As discussed below, in a finite economy, the possibility of manipulation cannot be excluded without placing restrictive assumptions on preferences. Our results suggest that larger markets employing the SOSM will be less prone to manipulation than smaller markets, holding everything else the same.

Second, we have required that the length of the student preference list is bounded. As mentioned in the introduction, the primary motivation for this assumption is empirical. In the NRMP, the length of applicant preference lists is typically less than or equal to 15. In NYC, almost 75 percent of students ranked fewer than the maximum of 12 schools in 2003–04 and there were over 500 programs to choose from.[33] In Boston public schools, which recently adopted the SOSM, more than 90 percent of students ranked 5 or fewer schools at elementary school during the first and second year of the new mechanism, and there are about 30 different elementary schools in each zone.[34] The conclusions of our results are known to fail if the assumption of bounded preference lists is not satisfied, and instead students regard every college as acceptable (Donald Knuth, Rajeev Motwani, and Boris Pittel 1990). Roth and Peranson (1999) conduct simulations on random data illustrating this point.

One reason students do not submit long preference lists is that it may be costly for them to do so. For example, medical school students in the United States 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. Another reason may be that there is an exogenous restriction on the length of student preference lists as in NYC. However, there is an additional concern for manipulation under such exogenous constraints on preference lists, since students may not be able to report their true preferences under the restriction.[35] We submit as an open question how the large market

[31] In a somewhat different model of matching with price competition, Bulow and Levin (2006) show that stable matching mechanisms in their environment achieve high efficiency when there is a large number of market participants.

[32] It is easy to relax some of our conditions on the definition of a regular market. For instance, our assumption on bounded preference lists could be generalized to state: there is a sequence $(k^n)_{n=1}^{\infty}$ with $\lim_{n\to\infty}(k^n/\ln(n)) = 0$ such that, for any $n$ and any student in $S^n$, the length of her preference list for the student is at most $k^n$. Similarly, the condition on the number of students can be relaxed to state: there exists $\tilde{q}$ such that $|S^n| \leq \tilde{q}n$ for any $n$.

[33] In subsequent years, the fraction of students with preference lists of fewer than 12 schools ranged between 70 percent and 78 percent.

[34] There is a total of 84 different elementary schools in the entire city. A small handful of these programs are city-wide, and can be ranked by a student who lives in any zone. There is also a small fraction of students who can apply across zones because they live near the boundary of two zones.

[35] NYC is a two-sided matching market using the SOSM, which has an explicit restriction on the number of choices that can be ranked. Note, however, that even in NYC, in the first year of data from the new system about 75 percent of students rank fewer than the maximum number of 12 schools in the main round, so our analysis may still hold approxi-

argument may be affected when there is a restriction on the length of the preference list (see Guillaume Haeringer and Flip Klijn (2007) for analysis in this direction, though in the finite market setting).

Finally, our analysis relies on the distributional assumptions on student preferences. For any given number of colleges in the market, with enough data on the distribution of student preferences and college preferences, our methods can be adapted to obtain the bound on the potential gain from deviation from truth-telling. However, to verify whether a market is sufficiently thick requires making assumptions on college popularity as the market grows.

The main text has focused on a particular process for generating student preferences. Some distributions are excluded by this specification. For example, we assume that preferences of students are independently drawn from one another, and they are drawn from an identical distribution. The i.i.d. assumption and the particular process we analyzed in the main text excludes, for example, cases where students in a particular region are more likely to prefer colleges in that region than students in other regions.

In the Appendix A.5, however, we analyze a more general model that allows students to belong to one of several groups and draw their preferences from different distributions across groups. We believe that the model with group-specific student preferences may be important for studying a number of real-world markets, for participants may often have systematically different preferences according to their residential location, academic achievement, and other characteristics. For instance, in NYC, more than 80 percent of applicants rank a program that is in the borough of their residence as their top choice, for all four years we have data. We show that truthful reporting is an approximate equilibrium in large markets that are sufficiently thick when this sort of heterogeneity across groups is present.

Our result is known to fail when there is some form of interdependence of student preferences.[36] Obtaining the weakest possible condition on the distribution of preferences for our result to hold is a challenging problem for future work.

Another direction for generalization is to consider weakening the assumption of responsive preferences, or considering models of matching with transfers. With general substitutable preferences, however, telling the truth may not be a dominant strategy for students. Hatfield and Milgrom (2005) have shown that substitutable preferences, together with the "law of aggregate demand," is sufficient to restore the dominant strategy property for students. It would be interesting to investigate if a result similar to ours holds in this environment.

### A. *Large Markets and Matching Market Design*

Our results have established the virtues of the student-optimal stable mechanism in a large market. This 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 mechanism proceeds as follows:[37]

mately. The only other example we are aware of is Spanish college admissions studied by Antonio Romero-Medina (1998).

[36] Immorlica and Mahdian (2005) present a model in which preferences cannot be generated by our procedure, and the fraction of colleges that can manipulate does not go to zero even as the size of the market goes to infinity. Since their environment is a special case of ours, it shows that some assumptions similar to ours are needed to obtain our results.

[37] With slight abuse of terminology, we will refer to this mechanism where colleges rank students as the Boston mechanism, even though the Boston mechanism was introduced as a one-sided matching mechanism. That is, college preferences are given as an exogenously given priority structure.

- **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 when every unmatched student has been rejected by every acceptable college.[38]

Under the Boston mechanism, colleges have no incentive to manipulate either via preference lists or via capacity, even in a small market with an arbitrary preference profile (Haluk Ergin and Sönmez 2006). Nevertheless, 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 et al. 2006). An example in Appendix A.8 shows that students' incentives to manipulate the Boston mechanism remain large even when the number of colleges increases in a regular sequence of markets.

We view this paper as one of the first attempts in the matching literature to show how performance of matching mechanisms in large economies can be used to compare mechanisms. One traditional approach in the matching literature is to find restrictions on preferences to ensure that a mechanism produces a desirable outcome. José Alcalda and Salvador Barberà (1994) investigate this question focusing on preference manipulation, and Onur Kesten (2008), Kojima (2007b), and Konishi and Ünver (2006) focus on capacity manipulation and manipulation via pre-arranged matches. One message from these papers is that the conditions that prevent the possibility of manipulation are often quite restrictive. In this paper, we have developed a different approach, which uses large market arguments to obtain possibility results.

More generally, we think that the kind of large market arguments similar to that in the current paper will be fruitful in future work on matching and related allocation mechanisms. In the problem of allocating indivisible objects such as university housing, Kojima and Mihai Manea (2008) show that the probabilistic serial mechanism, which has desirable efficiency and fairness properties (Anna Bogomolnaia and Hervé Moulin 2001), also has a desirable form of incentive compatibility in a large market. In the kidney exchange problem, Roth, Sönmez, and Ünver (2004, 2007) show that efficiency can be achieved by conducting only kidney exchanges of small sizes when the number of incompatible patient-donor pairs is sufficiently large. Both of these are cases where large market arguments support a particular matching market design.

There are also open questions where large market analysis may yield new insights. Roth (2007) suggests that large market arguments may be useful to understand why a stable matching was always found in the NRMP, even though the existence of couples can make the set of stable matchings empty.[39] In the school choice setting with indifferences, Aytek Erdil and Ergin (2008) propose a new procedure to construct a student-optimal matching. While there are potentially large efficiency gains from their procedure (Abdulkadiroğlu et al., forthcoming), in their mechanism, it is not a dominant strategy for students to reveal their preferences truthfully. Evaluating

[38] Note the difference between this mechanism and the SOSM. At each step of the Boston mechanism, students who are not rejected are *guaranteed* positions; the matches of these students and colleges are permanent rather than temporary, unlike in the SOSM.

[39] To ensure the existence of stable matchings with couples, we need restrictive assumptions on preferences of couples (Bettina Klaus and Klijn 2005). Kojima (2007a) develops an algorithm to find stable matchings with couples without imposing assumptions on preferences that might be a first step to investigate this issue in large markets.

incentive properties of this new procedure in large markets may be an interesting direction for future research.

As market design tackles more complex environments, it will be harder to obtain finite market results on the properties of certain mechanisms. The current paper explores an alternative approach, which is based on a large market assumption. Since many markets of interest can be modeled as large markets, explicitly analyzing the limit properties will be a useful approach to guide policymakers and help evaluate designs in these environments.

## REFERENCES

**Abdulkadiroğlu, Atila, and Tayfun Sönmez.** 2003. "School Choice: A Mechanism Design Approach." *American Economic Review*, 93(3): 729–47.

**Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth.** Forthcoming. "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match." *American Economic Review*.

**Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth.** 2005. "The New York City High School Match." *American Economic Review*, 95(2): 364–67.

**Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez.** 2005. "The Boston Public School Match." *American Economic Review*, 95(2): 368–71.

**Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez.** 2006. "Changing the Boston School Choice Mechanism." National Bureau of Economic Research Working Paper 11965.

**Alcalde, José, and Salvador Barberà.** 1994. "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems." *Economic Theory*, 4(3): 417–35.

**Blum, Yosef, Alvin E. Roth, and Uriel Rothblum.** 1997. "Vacancy Chains and Equilibration in Senior-Level Labor Markets." *Journal of Economic Theory*, 76(2): 362–411.

**Bogomolnaia, Anna, and Hervé Moulin.** 2001. "A New Solution to the Random Assignment Problem." *Journal of Economic Theory*, 100(2): 295–328.

**Bulow, Jeremy, and Jonathan Levin.** 2006. "Matching and Price Competition." *American Economic Review*, 96(3): 652–68.

**Cantala, David.** 2004. "Restabilizing Matching Markets at Senior Level." *Games and Economic Behavior*, 48(1): 1–17.

**Cripps, Martin W., and Jeroen M. Swinkels.** 2006. "Efficiency of Large Double Auctions." *Econometrica*, 74(1): 47–92.

**Day, Robert, and Paul Milgrom.** 2008. "Core-Selecting Package Auctions." *International Journal of Game Theory*, 36(3–4): 393–407.

**Demange, Gabrielle, David Gale, and Marilda A. Oliveira Sotomayor.** 1987. "A Further Note on the Stable Matching Problem." *Discrete Applied Mathematics*, 16(3): 217–22.

**Dubins, Lester E., and David A. Freedman.** 1981. "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly*, 88(7): 485–94.

**Ehlers, Lars.** 2004. "In Search of Advice for Participants in Matching Markets Which Use the Deferred-Acceptance Algorithm." *Games and Economic Behavior*, 48(2): 249–70.

**Erdil, Aytek, and Haluk Ergin.** 2008. "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice." *American Economic Review*, 98(3): 669–89.

**Ergin, Haluk, and Tayfun Sönmez.** 2006. "Games of School Choice under the Boston Mechanism." *Journal of Public Economics*, 90(1–2): 215–37.

**Fudenberg, Drew, Markus Mobius, and Adam Szeidl.** 2007. "Existence of Equilibrium in Large Double Auctions." *Journal of Economic Theory*, 133(1): 550–67.

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

**Gale, David, and Marilda A. Oliveira Sotomayor.** 1985. "Some Remarks on the Stable Matching Problem." *Discrete Applied Mathematics*, 11(1985): 223–32.

**Gresik, Thomas A., and Mark A. Satterthwaite.** 1989. "The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms." *Journal of Economic Theory*, 48(1): 304–32.

**Haeringer, Guillaume, and Flip Klijn.** 2007. "Constrained School Choice." Unpublished.

**Hatfield, John W., and Paul R. Milgrom.** 2005. "Matching with Contracts." *American Economic Review*, 95(4): 913–35.

**Immorlica, Nicole, and Mohammad Mahdian.** 2005. "Marriage, Honesty, and Stability." In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 53–62. Philadelphia: Society for Industrial and Applied Mathematics.

**Kelso, Alexander S., Jr., and Vincent P. Crawford.** 1982. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, 50(6): 1483–1504.

**Kesten, Onur.** 2008. "On Two Kinds of Manipulation for School Choice Problems." Unpublished.

**Klaus, Bettina, and Flip Klijn.** 2005. "Stable Matchings and Preferences of Couples." *Journal of Economic Theory*, 121(1): 75–106.

**Knuth, Donald E.** 1976. *Mariages Stables et Leurs Relations avec d'Autres Problemes Combinatoires: Introduction a l'Analyse Mathematique des Algorithmes*. Montreal: Les Presse de l'Universite de Montreal.

**Knuth, Donald E., Rajeev Motwani, and Boris Pittel.** 1990. "Stable Husbands." *Random Structures and Algorithms*, 1(1): 1–14.

**Kojima, Fuhito.** 2006. "Mixed Strategies in Games of Capacity Manipulation in Hospital-Intern Markets." *Social Choice and Welfare*, 27(1): 25–28.

**Kojima, Fuhito.** 2007a. "Finding All Stable Matchings with Couples." Unpublished.

**Kojima, Fuhito.** 2007b. "When Can Manipulations Be Avoided in Two-Sided Matching Markets? Maximal Domain Results." *Contributions to Theoretical Economics*, 7(1): Article 32.

**Kojima, Fuhito, and Mihai Manea.** 2008. "Strategy-Proofness of the Probabilistic Serial Mechanism in Large Random Assignment Problems." Unpublished.

**Konishi, Hideo, and M. U. Ünver.** 2006. "Games of Capacity Manipulation in Hospital-Intern Markets." *Social Choice and Welfare*, 27(1): 3–24.

**Ma, Jinpeng.** 2002. "Stable Matchings and the Small Core in Nash Equilibrium in the College Admissions Problem." *Review of Economic Design*, 7(2): 117–34.

**McVitie, D. G., and L. B. Wilson.** 1970. "Stable Marriage Assignments for Unequal Sets." *BIT Numerical Mathematics*, 10(3): 295–309.

**Milgrom, Paul R.** 2004. *Putting Auction Theory to Work*. Churchill Lectures in Economics. Cambridge: Cambridge University Press.

**Niederle, Muriel.** 2007. "Competitive Wages in a Match with Ordered Contracts." *American Economic Review*, 97(5): 1957–69.

**Pathak, Parag A., and Tayfun Sönmez.** 2008. "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*, 98(4): 1636–52.

**Pesendorfer, Wolfgang, and Jeroen M. Swinkels.** 2000. "Efficiency and Information Aggregation in Auctions." *American Economic Review*, 90(3): 499–525.

**Romero-Medina, Antonio.** 1998. "Implementation of Stable Solutions in a Restricted Matching Market." *Review of Economic Design*, 3(2): 137–47.

**Roth, Alvin E.** 1982. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research*, 7(4): 617–28.

**Roth, Alvin E.** 1984a. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92(6): 991–1016.

**Roth, Alvin E.** 1984b. "Misrepresentation and Stability in the Marriage Problem." *Journal of Economic Theory*, 34(2): 383–87.

**Roth, Alvin E.** 1985. "The College Admissions Problem Is Not Equivalent to the Marriage Problem." *Journal of Economic Theory*, 36(2): 277–88.

**Roth, Alvin E.** 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica*, 70(4): 1341–78.

**Roth, Alvin E.** 2008. "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions." *International Journal of Game Theory*, 36(3–4): 537–69.

**Roth, Alvin E., and Elliot Peranson.** 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89(4): 748–80.

**Roth, Alvin E., and Uriel G. Rothblum.** 1999. "Truncation Strategies in Matching Markets--In Search of Advice for Participants." *Econometrica*, 67(1): 21–43.

**Roth, Alvin E., and Marilda A. Oliveira Sotomayor.** 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph 18. Cambridge: Cambridge University Press.

**Roth, Alvin E., and John H. Vande Vate.** 1991. "Incentives in Two-Sided Matching with Random Stable Mechanisms." *Economic Theory*, 1(1): 31–44.

**Roth, Alvin E., Tayfun Sönmez, and Utku Ünver.** 2004. "Kidney Exchange." *Quarterly Journal of Economics*, 119(2): 457–88.

**Roth, Alvin E., Tayfun Sönmez, and Utku Ünver.** 2007. "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences." *American Economic Review*, 97(3): 828–51.

**Rustichini, Aldo, Mark A. Satterthwaite, and Steven Williams.** 1994. "Convergence to Efficiency in a Simple Market with Incomplete Information." *Econometrica*, 62(5): 1041–63.

**Sönmez, Tayfun.** 1997a. "Games of Manipulation in Marriage Problems." *Games and Economic Behavior*, 20(2): 169–76.

**Sönmez, Tayfun.** 1997b. "Manipulation via Capacities in Two-Sided Matching Markets." *Journal of Economic Theory*, 77(1): 197–204.

**Sönmez, Tayfun.** 1999. "Can Pre-arranged Matches Be Avoided in Two-Sided Matching Markets?" *Journal of Economic Theory*, 86(1): 148–56.

**Swinkels, Jeroen M.** 2001. "Efficiency of Large Private Value Auctions." *Econometrica*, 69(1): 37–68.