Manipulation via Capacities Revisited*

Lars Ehlers†

March 2009

Abstract

This paper revisits manipulation via capacities in centralized two-sided matching markets. Sönmez (1997) showed that no stable mechanism is non-manipulable via capacities. We show that non-manipulability via capacities can be equivalently described by two types of non-manipulation via capacities: non-Type-I-manipulability meaning that no college with vacant positions can manipulate by dropping some of its empty positions; and non-Type-II-manipulability meaning that no college with no vacant positions can manipulate by dropping some of its filled positions. Our main result shows that the student-optimal stable mechanism is the unique stable mechanism which is non-Type-I-manipulable via capacities and independent of truncations. Our characterization supports the use of the student-optimal stable mechanism in these matching markets because of its limited manipulability via capacities by colleges.

Keywords: Two-Sided Matching, Stability, Manipulation, Capacities.

JEL Classification: C78, D78, I20.

---

*I am especially grateful to Fuhito Kojima and Al Roth for their detailed comments and suggestions. I acknowledge financial support from the SSHRC (Canada) and the FQRSC (Québec) and the hospitality of Harvard Business School where part of this paper was written.

†Département de Sciences Économiques and CIREQ, Université de Montréal, Montréal, Québec H3C 3J7, Canada, e-mail: `lars.ehlers@umontreal.ca`.

1 Introduction

Matching markets arise in many economic environments. There are two sides of the market, called students (or workers) and colleges (or hospitals or schools) and students need to be matched to colleges. Specifically we are interested in many-to-one matching markets where monetary compensations are fixed, i.e. salaries are prespecified in each match and colleges can admit one or possibly more students. In a number of those environments decentralized markets have failed and centralized clearinghouses (or mechanisms) have emerged. Leading examples are entry-level medical markets in Great Britain and the United States, college admissions, and school choice in American municipalities. In those markets the success of a centralized matching procedure has been proven to depend on the (in)stability of the mechanism. More precisely, it has been shown that stable mechanisms outperform unstable mechanisms in centralized matching markets.

This is surprising because any stable mechanism is susceptible to different kinds of manipulation. In centralized matching markets colleges' preferences are private information and need to be reported to the clearinghouse. A mechanism is manipulable via preferences if a college can gain (in terms of the true preference) by submitting a false preference instead of its true preference. Roth (1985) showed that any stable mechanism is manipulable via preferences (for colleges).[1] Similarly the colleges' capacities are private information and a mechanism is manipulable via capacities if a college can gain (in terms of the true preference) by underreporting its true capacity. Sönmez (1997) showed a counterpart of Roth's result: any stable mechanism is manipulable via capacities. Our first result links these two types of manipulation as follows: for any given problem, if a stable mechanism is manipulable via capacities, then this mechanism is manipulable via preferences. Hence, the impossibility result (for preference manipulation) by Roth (1985) is a direct consequence of the

[1] Roth (1982) showed that any stable mechanism is manipulable via preferences for colleges or students.

impossibility result (for capacity manipulation) by Sönmez (1997).

Motivated by the impossibility result by Sönmez (1997), one strand of recent literature (Konishi and Ünver, 2006; Kojima, 2006; Romero-Medina and Triossi, 2007) considered games of capacity manipulation induced by the student-optimal stable mechanism and the college-optimal stable mechanism. Unfortunately pure strategy Nash equilibria may not exist (Konishi and Ünver, 2006). Two other contributions (Kojima, 2007; Kesten, 2008) determine domains where no college can gain by manipulation via capacities. More precisely, Kojima (2007) shows that a college with a fixed preference relation and capacity of at least two can almost always be embedded in a problem such that the student-optimal stable mechanism or the college-optimal stable mechanism is manipulable via capacities by this college. Similarly, Kesten (2008) shows that the student-optimal stable mechanism for a given problem is non-manipulable via capacities if and only if the problem satisfies a strong "acyclicity" condition. All these results are disappointing regarding manipulation via capacities (of the student-optimal stable mechanism and/or the college-optimal stable mechanism).

Here we will not narrow ourselves to the student-optimal stable mechanism or the college-optimal stable mechanism and not restrict the domain of problems under consideration. We try to understand when and how mechanisms can be manipulated via capacities. We show that the non-manipulability via capacities of an arbitrary mechanism can be equivalently described by two types of non-manipulation via capacities: (i) non-Type-I-manipulability meaning that no college with vacant positions can manipulate by dropping some of its empty positions, and (ii) non-Type-II-manipulability meaning that no college with no vacant positions can manipulate by dropping some of its filled positions. We show that a mechanism is non-manipulable via capacities if and only if the mechanism is both non-Type-I-manipulable via capacities and non-Type-II-manipulable via capacities. Our main result shows that the student-optimal stable mechanism is the unique stable mechanism which is non-Type-I-manipulable via capacities and independent of truncations. Hence, the student-optimal stable mechanism is characterized in terms of non-manipulation properties of colleges and these properties imply that the mechanism is non-manipulable via preferences for students.

Our main result further supports the use of the student-optimal stable mechanism in applications. The NRMP changed the mechanism from the college-optimal stable mechanism to the student-optimal stable mechanism because of its non-manipulability via preferences for students. This change did not have only a positive effect on the students' side, but also on the colleges' side because a limited form of non-manipulability via capacities is guaranteed. Note that this is not guaranteed by the college-optimal stable mechanism because it is both Type-I- and Type-II-manipulable via capacities. Our main result also points out positive effects of the use of the student-optimal stable mechanism in school choice (like in Boston): here priorities of schools are fixed or known and a school may only manipulate via capacities. Again under the student-optimal stable mechanism schools with vacant seats cannot manipulate by dropping some of their empty seats.

A consequence of our main result is that any stable mechanism is Type-II-manipulable via capacities. Indeed avoiding manipulability via capacities by colleges with no vacant positions is difficult. Of course, any such solution will be unstable. For problems where each college has exactly one position no college can gain by underreporting its capacity. We show that one may use an iterative stable mechanism in order to avoid manipulations via capacities.[2] In determining which "no blocking" conditions an iterative stable mechanism may possess, recall that in school choice stability (of a matching) is equivalent to non-wastefulness (no empty positions are wasted) and fairness (no student justifiably envies another student at a college). We adopt weaker notions of these two "no blocking" conditions and establish in school choice a variant of the impossibility result by Sönmez (1997): (a) there exists no mechanism which is

[2]We provide details later.

non-wasteful, weakly fair, and non-manipulable via capacities; and (b) there exists no mechanism which is weakly non-wasteful, fair, and non-manipulable via capacities. If both "no blocking" conditions are weakened, then a possibility emerges. Namely, any iterative stable mechanism satisfies weak non-wastefulness, weak fairness and non-manipulability via capacities. In applications, if non-manipulation via capacities is more important than stability, then an iterative stable mechanism may provide a practical solution since it satisfies the weaker versions of the "no blocking" conditions non-wastefulness and fairness.

The paper is organized as follows. Section 2 introduces the two-sided matching market, stability and manipulation via capacities. Section 3 provides the direct link for stable mechanisms between manipulation via capacities and manipulation via preferences. Section 4 introduces two types of non-manipulation via capacities and gives our main result: the characterization of the student-optimal mechanism. Section 5 introduces iterative (stable) mechanisms and establishes the variant of Sönmez (1997) for school choice.

2 The Model

A college admissions problem is a quadruple (S, C, q, R) where (i) S denotes the finite set of students, (ii) C denotes the finite set of colleges, (iii) $q = (q_c)_{c \in C}$ is list of natural numbers where q_c is the capacity (or the number of available slots at college c), and (iv) $R = (R_v)_{v \in S \cup C}$ is a list of preference relations. Since we consider the case where S and C remain fixed, we write (q, R) instead of (S, C, q, R). Furthermore, for any $T \subseteq C$, let $q_T = (q_c)_{c \in T}$ and $q_{-T} = (q_c)_{c \in C \setminus T}$, and for any $c \in C$, let $R_{-c} = (R_v)_{v \in S \cup (C \setminus \{c\})}$.

For any $v \in S \cup C$, R_v is a complete and transitive preference relation. Let P_v denote the strict preference relation associated with R_v. For any $s \in S$, R_s is a strict preference relation on $C \cup \{\emptyset\}$ where $\emptyset$ stands for being unmatched. For any $c \in C$, R_c is a preference relation on 2^S such that R_c is strict on $S \cup \{\emptyset\}$ and R_c is responsive over 2^S (to $R_c|_{S\cup\emptyset}$[3]): for all $S' \subseteq S$ and all $s, s' \in S \backslash S'$,

$$\text{(i) } S' \cup \{s\} P_c S' \cup \{s'\} \Leftrightarrow s P_c s' \text{ and (ii) } S' \cup \{s\} P_c S' \Leftrightarrow s P_c \emptyset.^4$$

Let $\mathcal{R}_c$ denote the set of all responsive preferences over 2^S.

Given R_s, college c is *acceptable under* R_s if $cP_s\emptyset$. Similarly, for any R_c, student s is *acceptable under* R_c if $sP_c\emptyset$. Let $A(R_c) = \{s \in S : sP_c\emptyset\}$ denote the set of students who are acceptable under R_c.

A matching for a given capacity vector q is a function $\mu : S \cup C \to 2^{S \cup C}$ such that

(i) for all $s \in S$, $|\mu(s)| \leq 1$ and $\mu(s) \subseteq C$;

(ii) for all $c \in C$, $|\mu(c)| \leq q_c$ and $\mu(c) \subseteq S$; and

(iii) for all $s \in S$ and all $c \in C$, $\mu(s) = c$ if and only if $s \in \mu(c)$.

The main concept is stability of a matching: no student should be matched to an unacceptable college, no college should be matched to any unacceptable student, and no student-college pair blocks the matching because they mutually prefer each other. Given a problem (q, R), a matching μ for q is stable if

(a) (individual rationality for students) for all $s \in S$, $\mu(s)R_s\emptyset$;

(b) (individual rationality for colleges) for all $c \in C$, $\mu(c) \subseteq A(R_c)$;

(c) (no blocking pair) there exists no $s \in S$ and $c \in C$ such that $cP_s\mu(s)$ and either $[|\mu(c)| < q_c$ and $sP_c\emptyset]$ or $[sP_cs'$ for some $s' \in \mu(c)]$.

Gale and Shapley (1962) show that the set of stable matchings is non-empty for any problem (q, R). Furthermore, the set of stable matchings has a lattice structure and there exists a stable matching, called the student-optimal stable matching which is weakly preferred to any other stable matching by the students and which is worst

[3]Here $R_c|_{S\cup\emptyset}$ denotes the restriction of R_c to $S \cup \emptyset$.

[4]For convenience, we drop set brackets for singleton sets and write s instead of $\{s\}$.

for the colleges among all stable matchings. Similarly, there exists a college-optimal stable matching which is weakly preferred to any other stable matching by the colleges. The student-optimal stable matching can be calculated via the deferred-algorithm (DA) with students proposing and the college-optimal stable matching via the DA-algorithm with colleges proposing. Furthermore, at any two stable matchings, any college fills the same number of positions; and if a college does not fill all its positions at a stable matching, then this college is matched to the same set of students under all stable matchings.[5]

A mechanism (or mechanism) φ associates with any problem (q, R) a matching $\varphi(q, R)$ for the capacity vector q. A mechanism φ is stable if for any problem (q, R), $\varphi(q, R)$ is stable. Let DA_S denote student-optimal stable mechanism choosing for each problem (q, R) its student-optimal stable matching (determined for each problem via the DA-algorithm with students proposing). Let DA_C denote the college-optimal stable mechanism.

In many situations capacities are private information and a college may attempt to manipulate a mechanism via underreporting its capacity.

Definition 1 *Let R be a profile, q be a capacity vector, and φ be a mechanism. Then φ is manipulable via capacities at (q, R) if there exists $c \in C$ and $q'_c \in \{0, 1, \ldots, q_c\}$ such that*

$$\varphi(q'_c, q_{-c}, R)(c) P_c \varphi(q, R)(c).$$

We say that φ is non-manipulable via capacities if for any problem (q, R), φ is not manipulable via capacities at (q, R).

The principal result of Sönmez (1997) is the following.

Theorem 1 (Sönmez, 1997, Theorem 1) *Suppose there are at least three students and two colleges. Then there exists no mechanism that is stable and non-manipulable via capacities.*

[5] All these properties of stable matchings are stated in the illuminating introduction to two-sided matching by Roth and Sotomayor (1990).

3 Manipulation via Preferences

An important concern of clearinghouses in centralized markets is whether a mechanism can be profitably manipulated via misreporting preferences. If this is the case, then the outcome of the mechanism may not be based on the true information of the participants.

Definition 2 *Let R be a profile, q be a capacity vector, and φ be a mechanism. Then φ is manipulable via preferences at (q, R) if there exists $c \in C$ and $R'_c \in \mathcal{R}_c$ such that*

$$\varphi(q, R'_c, R_{-c})(c) P_c \varphi(q, R)(c).$$

We say that φ is non-manipulable via preferences if for any problem (q, R), φ is not manipulable via preferences at (q, R).

Note that the above definition focusses only on manipulation via preferences by colleges. The following establishes for stable mechanisms an important link between manipulation via capacities and manipulation via preferences.[6]

Theorem 2 *Let φ be stable mechanism, q be a capacity vector, and R be a profile. If φ is manipulable via capacities at (q, R), then φ is manipulable via preferences at (q, R).*

Proof. If φ is manipulable via capacities at (q, R), then there exists some $c \in C$ and $q'_c \in \{1, \ldots, q_c\}$ such that $\varphi(q'_c, q_{-c}, R)(c) P_c \varphi(q, R)(c)$. Let $R'_c \in \mathcal{R}_c$ be such that $A(R'_c) = \varphi(q'_c, q_{-c}, R)(c)$. Obviously, $\varphi(q'_c, q_{-c}, R)$ is stable under (q, R'_c, R_{-c}). Since φ is stable, c fills the same number of positions under all stable matchings, and $A(R'_c) = \varphi(q'_c, q_{-c}, R)(c)$, we must have $\varphi(q, R'_c, R_{-c})(c) = \varphi(q'_c, q_{-c}, R)(c)$. Hence, $\varphi(q, R'_c, R_{-c})(c) P_c \varphi(q, R)(c)$ and φ is manipulable via preferences at (q, R), the desired conclusion. □

[6] This result is related to the "Dropping Strategies Lemma" by Kojima and Pathak (2008).

Remark 1 By Theorem 1, we know that there exists no stable mechanism that is non-manipulable via capacities. Thus, for any stable mechanism there exist problems at which the mechanism is manipulable via capacities. Now by Theorem 2, for any stable mechanism there exist problems at which the mechanism is manipulable via preferences. Hence, there exists no stable mechanism that is non-manipulable via preferences (for colleges) and Roth (1985, Proposition 2) is a corollary from Sönmez (1997, Theorem 1).

4 Two Types of Capacity Manipulation

In the sequel we will distinguish between two types of capacity manipulation.

Definition 3 *Let q be a capacity vector, R be a profile, and φ be a mechanism. For all $c \in C$, let $|\varphi(q, R)(c)| = f_c$ denote the number of filled positions at college c under $\varphi(q, R)$.*

(i) Then φ is Type-I-manipulable via capacities at (q, R) if there exists $c \in C$ and $q'_c \in \{f_c, f_c + 1, \ldots, q_c\}$ such that

$$\varphi(q'_c, q_{-c}, R)(c) P_c \varphi(q, R)(c).$$

We say that φ is non-Type-I-manipulable via capacities if for any problem (q, R), φ is not Type-I-manipulable via capacities at (q, R).

(ii) Then φ is Type-II-manipulable via capacities at (q, R) if there exists $c \in C$ with $f_c = q_c$ and $q'_c \in \{1, \ldots, f_c - 1\}$ such that

$$\varphi(q'_c, q_{-c}, R)(c) P_c \varphi(q, R)(c).$$

We say that φ is non-Type-II-manipulable via capacities if for any problem (q, R), φ is not Type-II-manipulable via capacities at (q, R).

These two types of manipulation via capacities have the following interpretations. Type-I-manipulability means that a college with vacant positions gains from giving up some of its unfilled positions. Type-II-manipulability means that a college with no vacant positions gains from giving up some of its filled positions. Then a college is ready to forego some of its students (leaving these positions empty) in order to exchange some of the other students for better students. Of course, a college could simultaneously give up some of its filled positions *and* forego some of its students, giving rise to a third type of manipulability via capacities. By the lemma below, such a third type of non-manipulability is unnecessary for determining the non-manipulability via capacities of a mechanism.

Lemma 1 *Let φ be a mechanism. Then φ is non-manipulable via capacities if and only if φ is both non-Type-I-manipulable via capacities and non-Type-II-manipulable via capacities.*

Proof. (Only if) It is straightforward that if φ is non-manipulable via capacities, then φ is both non-Type-I-manipulable via capacities and non-Type-II-manipulable via capacities.

(If) Suppose that φ is both non-Type-I-manipulable via capacities and non-Type-II-manipulable via capacities. Let q be a capacity vector and R be a profile. Let $q'_c \in \{1, \ldots, q_c\}$. Now if $|\varphi(q, R)(c)| = q_c$, then from non-Type-II-manipulability we obtain $\varphi(q, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$; and if $|\varphi(q, R)(c)| < q_c$ and $|\varphi(q, R)(c)| \leq q'_c$, then from non-Type-I-manipulability we obtain $\varphi(q, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$. Otherwise, if both $|\varphi(q, R)(c)| < q_c$ and $q'_c < |\varphi(q, R)(c)| = f_c$, then from non-Type-I-manipulability we obtain

$$\varphi(q, R)(c) R_c \varphi(f_c, q_{-c}, R)(c). \tag{1}$$

Now if $|\varphi(f_c, q_{-c}, R)(c)| = f_c$, then from non-Type-II-manipulability we obtain that $\varphi(f_c, q_{-c}, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$. By transitivity of R_c and (1), $\varphi(q, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$, the desired conclusion. If both $|\varphi(f_c, q_{-c}, R)(c)| < f_c$ and $|\varphi(f_c, q_{-c}, R)(c)| \leq q'_c$, then from non-Type-I-manipulability we obtain $\varphi(f_c, q_{-c}, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$. Thus, by transitivity of R_c and (1), $\varphi(q, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$, the desired conclusion. If both $|\varphi(f_c, q_{-c}, R)(c)| < f_c$ and $q'_c < |\varphi(f_c, q_{-c}, R)(c)| = f'_c$, then from non-Type-I-manipulability we obtain

$$\varphi(f_c, q_{-c}, R)(c) R_c \varphi(f'_c, q_{-c}, R)(c).$$

By using similar arguments and the transitivity of R_c, this and (1) yield in a finite number of steps (since q_c is finite) $\varphi(q, R)(c) R_c \varphi(q'_c, q_{-c}, R)(c)$, the desired conclusion. □

Before stating the main result, we introduce an invariance property for a mechanism: it says that for any given problem, if a college truncates its preference by leaving unchanged its ranking over students and restricting its set of acceptable students without dropping any of the students it is matched to, then for the problem with the truncated preference the college should be matched to the same set of students.

Let $R_c, R'_c \in \mathcal{R}_c$. Then we call R'_c a truncation of R_c if (i) $R'_c|_S = R_c|_S$ and $A(R'_c) \subseteq A(R_c)$ and (ii) for all $S', S'' \subseteq A(R'_c)$, we have $S' R'_c S'' \Leftrightarrow S' R_c S''$.

We say that a mechanism φ is *independent of truncations* if for any problem (q, R), for any $c \in C$, and for any truncation R'_c of R_c such that $\varphi(q, R)(c) \subseteq A(R'_c)$, we have $\varphi(q, R'_c, R_{-c})(c) = \varphi(q, R)(c)$.

Independence of truncations is a weak invariance property which is satisfied by many stable mechanisms: DA_S and DA_C, or strictly order all matchings according to $>$ and choose for any problem (q, R) the $>$-greatest matching which is stable under (q, R). Furthermore, as Ehlers (2008) shows, all mechanisms which are used in British entry-level medical markets satisfy this property.[7]

The result below characterizes DA_S in terms of stability and axioms of non-manipulability *for colleges*. Furthermore, the properties of stability, independence of

[7] According to the author's knowledge, any mechanism, which is used in a real life market, satisfies this property.

truncations and non-Type-I-manipulability via capacities imply that the mechanism is non-manipulable via preferences *for students* since the student-optimal mechanism DA_S satisfies this property (Dubins and Freedman, 1981; Roth, 1982).

Theorem 3 *The student-optimal stable mechanism DA_S is the unique stable mechanism which is independent of truncations and non-Type-I-manipulable via capacities.*

Proof. First, we show that DA_S is a stable mechanism which is independent of truncations and non-Type-I-manipulable via capacities. By definition, DA_S is stable and by Ehlers (2008), independent of truncations. In order to show that DA_S is non-Type-I-manipulable via capacities, let q be a capacity vector and R be a profile. Let $q'_c \in \{|DA_S(q,R)(c)|, |DA_S(q,R)(c)|+1,\ldots,q_c\}$. Obviously, $DA_S(q,R)$ is stable under (q'_c, q_{-c}, R). Since $DA_S(q'_c, q_{-c}, R)$ is the worst stable matching for the colleges, it follows $DA_S(q,R)(c) R_c DA_S(q'_c, q_{-c}, R)(c)$, the desired conclusion.

Second, let φ be a stable mechanism which is independent of truncations and non-Type-I-manipulable via capacities. Suppose that $\varphi \neq DA_S$. Then for some problem (q,R) and some $c \in C$ we have $\varphi(q,R)(c) \neq DA_S(q,R)(c)$. Since both $\varphi(q,R)$ and $DA_S(q,R)$ are stable and $\varphi(q,R)(c) \neq DA_S(q,R)(c)$, college c must fill all its positions at all stable matchings, i.e. $|DA_S(q,R)(c)| = q_c$. Since $\varphi(q,R)$ is stable and $DA_S(q,R)$ is the stable matching which is worst for the colleges, we obtain

$$\varphi(q,R)(c) P_c DA_S(q,R)(c). \tag{2}$$

Let s' be the R_c-worst student in $DA_S(q,R)(c)$ and R'_c be a truncation of R_c such that s' is the R'_c-worst acceptable student. Note that $R'_c|_S = R_c|_S$. Let $R' = (R'_c, R_{-c})$. Since both φ and DA_S are independent of truncations, we obtain both $\varphi(q,R') = \varphi(q,R)$ and $DA_S(q,R') = DA_S(q,R)$. Thus, by (2),

$$\varphi(q,R')(c) P_c DA_S(q,R')(c). \tag{3}$$

Consider the problem (q_c+1, q_{-c}, R'). We show that $DA_S(q,R')$ is stable under (q_c+1, q_{-c}, R'): if not, then some pair $(\hat{s}, \hat{c})$ blocks $DA_S(q,R')$ under (q_c+1, q_{-c}, R');

obviously then we must have $\hat{c} = c$ and $\hat{s}P'_c\emptyset$. Since s' is the R'_c-worst acceptable student, we must have $\hat{s}P'_c s'$, which means that $DA_S(q, R')$ is not stable under (q, R'), a contradiction. Since $DA_S(q, R')$ is stable under $(q_c + 1, q_{-c}, R')$ and and $|DA_S(q, R')(c)| = q_c < q_c + 1$, c is matched to the same set of students at all matchings which are stable under $(q_c + 1, q_{-c}, R')$. Thus, by stability of φ, we have $\varphi(q_c + 1, q_{-c}, R')(c) = DA_S(q, R')(c)$. Now by (3), we obtain $\varphi(q, R')(c)P_c\varphi(q_c+1, q_{-c}, R')(c)$. Since $|\varphi(q_c + 1, q_{-c}, R')(c)| = q_c$, this means that φ is Type-I-manipulable via capacities, a contradiction. □

Note that in Theorem 3 non-Type-I-manipulability via capacities is a weak condition because it requires only that colleges with vacant positions cannot profitably manipulate by dropping some of its empty positions.

The independence of the properties in Theorem 3 is easily established: (i) DA_C is a stable mechanism which is independent of truncations (but DA_C violates non-Type-I-manipulability via capacities); (ii) the mechanism leaving for all problems all students unmatched and all colleges having all positions empty is independent of truncations and non-Type-I-manipulable via capacities (but violates stability); and (iii) the mechanism choosing for all problems the same matching as DA_S except for the problems (q, R) where $q_c = 1$ for all $c \in C$, all students are acceptable for all colleges (i.e. $A(R_c) = S$ for all $c \in C$), all colleges are acceptable for all students (i.e. $cP_s\emptyset$ for all $s \in S$ and all $c \in C$), and there are more students than colleges (i.e. $|S| > |C|$)). For those problems the mechanism chooses the college-optimal stable matching. It is straightforward to verify that this mechanism is stable and non-Type-I-manipulable via capacities but violates independence of truncations.

Remark 2 The feature of the statement of Theorem 3 that optimal stable mechanism for the students, DA_S, is characterized in terms of non-manipulation properties of the other side (by colleges) has appeared in iterative elimination of dominated strategies. More precisely, for one-to-one matching markets Alcalde (1996) showed that the student-optimal mechanism is dominance solvable and for any problem it uniquely implements (in terms of dominance solvability) the college-optimal stable matching.

By Theorem 3, DA_S is non-Type-I-manipulable via capacities. Now from Theorem 1 by Sönmez (1997) we obtain the following result.

Theorem 4 *Suppose there are at least three students and at least two colleges. Then there exists no mechanism that is stable and non-Type-II-manipulable via capacities.*

Theorem 4 points out why manipulation via capacities is problematic: colleges with no vacant positions may gain from manipulation.

5 Iterative Mechanisms

We know that any stable mechanism is susceptible to (Type-II-)manipulations via capacities. Colleges may underreport their capacities and fewer positions (than the true numbers) may be revealed which may result in more unmatched students. In applications it may be important to deter such manipulations in order to avoid unemployment (in entry-level labor markets) or unassigned students not attending any college (in education).

We will propose an iterative procedure which will be non-manipulable via capacities. We call (q, R) a base problem if for all $c \in C$, $q_c \in \{0, 1\}$. A base mechanism ϕ associates with any base problem a matching. Below we provide a heuristic way to extend any base mechanism from the set of base problems to the set of all problems such that non-manipulability via capacities is guaranteed.

For any set $C' \subseteq C$, let $1_{C'}$ denote the vector of capacities such that all colleges in C' have capacity 1 and all other colleges have capacity 0. Let $q_{\max} = \max_{c \in C} q_c$ be the maximal capacity in q.

Iterative ϕ-Mechanism: Let (q, R) be a problem. For any natural number l, let $C_l = \{c \in C : q_c \geq l\}$ and $S_l = \cup_{c \in C} \phi(1_{C_l}, R_{-(S_1 \cup \cdots \cup S_{l-1})})(c)$ (with the convention $S_0 = \emptyset$). The iterative ϕ-mechanism, denoted by $I(\phi)$, is defined as follows. For all $c \in C$, let

$$I(\phi)(q, R)(c) = \cup_{l=1}^{q_{\max}} \phi(1_{C_l}, R_{-(S_1 \cup \cdots \cup S_{l-1})})(c).$$

In other words, $I(\phi)$ gives first each college capacity 1 and calculates $\phi(1_{C_1}, R)$. Then all assigned students are removed and it determines the colleges which have at least capacity 2 and gives all those ones again capacity 1 and determines again ϕ for the reduced problem, and so on. Each college is assigned to the union of all the students it is matched to at all steps in the iterative ϕ-mechanism.

Examples of iterative mechanisms are $I(DA_S)$ and $I(DA_C)$, or iterative Boston mechanism, iterative priority mechanisms, iterative top-trading cycles algorithm, etc..[8] An iterative stable mechanism is an iterative mechanism where the base mechanism chooses for any base problem a stable matching.

It turns out that any iterative mechanism is (coalitionally) non-manipulable via capacities.

Definition 4 *Let R be a profile, q be a capacity vector, and φ be a mechanism. We say that φ is coalitionally manipulable via capacities at (q, R) if there exists $\emptyset \neq T \subseteq C$ and $q'_T = (q'_c)_{c \in T}$ with $q'_c \in \{0, 1, \ldots, q_c\}$ for any $c \in T$, such that $\varphi(q'_T, q_{-T}, R)(c) P_c \varphi(q, R)(c)$ for all $c \in T$. We say that φ is coalitionally non-manipulable via capacities if for any problem (q, R), φ is not coalitionally non-manipulable via capacities at (q, R).*

Proposition 1 *Let ϕ be a base mechanism. The iterative ϕ-mechanism $I(\phi)$ is coalitionally non-manipulable via capacities.*

Proof. Let (q, R) be a problem, $\emptyset \neq T \subseteq C$ and $q'_T = (q'_c)_{c \in T}$ be a capacity vector such that $q'_c \leq q_c$ for all $c \in T$. Suppose that for all $c \in T$ we have

[8] We refer the interested reader to Ehlers (2008) and Abdulkadiroğlu and Sönmez (2003) for a detailed description of these mechanisms.

$I(\phi)(q'_T, q_{-T}, R)(c) P_c I(\phi)(q, R)(c)$. Let $q' = (q'_T, q_{-T})$, $C'_l = \{c \in C : q'_c \geq l\}$, and $S'_l = \cup_{c \in C} \phi(1_{C'_l}, R_{-(S'_1 \cup \cdots S'_{l-1})})(c)$.

First, suppose that for some $c \in T$, we have $I(\phi)(q, R)(c) \backslash I(\phi)(q'_T, q_{-T}, R)(c) \neq \emptyset$. Choose the minimal index $k \in \{1, \ldots, q_{\max}\}$ such that for some $c' \in T$ we have $\phi(1_{C_k}, R_{-(S_1 \cup \cdots S_{k-1})})(c') \notin I(\phi)(q'_T, q_{-T}, R)(c')$. Suppose that $T \subseteq C'_k$. Then from the fact that $N \backslash T$ did not change their capacities, by our choice of k and $T \subseteq C'_k$, we have for all $l \in \{1, \ldots, k-1\}$, $1_{C_l} = 1_{C'_l}$, $S_l = S'_l$, $\phi(1_{C_l}, R_{-(S_1 \cup \cdots S_{l-1})}) = \phi(1_{C'_l}, R_{-(S_1 \cup \cdots S_{l-1})})$, and $\phi(1_{C_l}, R_{-(S_1 \cup \cdots S_{l-1})})(c) \in I(\phi)(q'_T, q_{-T}, R_{-(S_1 \cup \cdots S_{l-1})})(c)$ for all $c \in C$. By $T \subseteq C'_k$, we then have $1_{C_k} = 1_{C'_k}$ and $\phi(1_{C_k}, R_{-(S_1 \cup \cdots S_{k-1})}) = \phi(1_{C'_k}, R_{-(S_1 \cup \cdots S_{k-1})})$, which contradicts the fact $\phi(1_{C_k}, R_{-(S_1 \cup \cdots S_{k-1})})(c') \notin I(\phi)(q'_T, q_{-T}, R)$.

If $T \subsetneq C'_k$, then there exists $c \in T$ such that $q'_c < k$. By our choice of k and $I(\phi)(q'_T, q_{-T}, R)(c) P_c I(\phi)(q, R)(c)$, we have $I(\phi)(q, R)(c) \subsetneq I(\phi)(q'_T, q_{-T}, R)(c)$. Then the argument below can be used to show that this case cannot occur.

For all $c \in T$, let $I(\phi)(q, R)(c) \subseteq I(\phi)(q'_T, q_{-T}, R)(c)$. If $I(\phi)$ were coalitionally manipulable at (q, R), then for all $c \in T$ we have $I(\phi)(q'_T, q_{-T}, R)(c) P_c I(\phi)(q, R)(c)$ and $I(\phi)(q, R)(c) \subsetneq I(\phi)(q'_T, q_{-T}, R)(c)$. Now for all $c \in T$ we have

$$|I(\phi)(q, R)(c)| < q'_c \leq q_c. \tag{4}$$

Choose $c' \in T$ such that $q'_{c'} = k \leq q'_c$ for all $c \in T$. Then from the fact that $N \backslash T$ did not change their capacities and by our choice of c' and $q'_{c'} = k$, we have for all $l \in \{1, \ldots, |I(\phi)(q, R)(c')|\}$, $1_{C_l} = 1_{C'_l}$, $S_l = S'_l$, and $\phi(1_{C_l}, R_{-(S_1 \cup \cdots S_{l-1})})(c) = \phi(1_{C'_l}, R_{-(S_1 \cup \cdots S_{l-1})})(c)$. By (4) and our choice of c' and k, we obtain that for all $l \in \{|I(\phi)(q, R)(c')| + 1, \ldots, q'_{c'}\}$, $1_{C_l} = 1_{C'_l}$, $S_l = S'_l$, and $\phi(1_{C_l}, R_{-(S_1 \cup \cdots S_{l-1})}) = \phi(1_{C'_l}, R_{-(S_1 \cup \cdots S_{l-1})})$. Thus, $I(\phi)(q, R)(c') = I(\phi)(q'_T, q_{-T}, R)(c')$, a contradiction to $I(\phi)(q, R)(c') \subsetneq I(\phi)(q'_T, q_{-T}, R)(c')$.

Hence, $I(\phi)$ is not coalitionally manipulable via capacities at (q, R), the desired conclusion. □

Remark 3 A weaker form of coalitional manipulation via capacities is where in Definition 4 we have $\varphi(q'_T, q_{-T}, R)(c) R_c \varphi(q, R)(c)$ for all $c \in T$ with strict preference holding for at least one $c \in T$. We say that φ strongly coalitionally non-manipulable via capacities if for any problem, there is no weaker form of coalitional manipulation via capacities. It can be checked that the iterative DA_S-mechanism $I(DA_S)$ is strongly coalitionally non-manipulable via capacities[9] whereas the iterative DA_C-mechanism $I(DA_C)$ does not satisfy this property.

Now one may wonder which properties may be inherited by the iterative ϕ-mechanism from the base mechanism ϕ. Recall that in school choice stability is divided into two "no blocking" properties: non-wastefulness and fairness. We will be interested whether iterative mechanism can satisfy any kind of these two "no blocking" conditions.

Non-wastefulness means that there is no student-college pair (s, c) such that s prefers c to her current assignment and c has a vacant position which it prefers to fill with s instead of having the position empty.

Non-Wastefulness: For all problems (q, R), there exists no student-college pair (s, c) such that $cP_s\varphi(q, R)(s)$, $|\varphi(q, R)(c)| < q_c$, and $sP_c\emptyset$.

Fairness means that there is no student-college pair (s, c) such that s prefers c to her current assignment and c prefers s to one of its assigned students.

Fairness: For any problem (q, R), there exists no student-college pair (s, c) such that $cP_s\varphi(q, R)(s)$ and sP_cs' for some $s' \in \varphi(q, R)(c)$.

We also introduce weaker notions of these two "no blocking" conditions. Weak fairness requires that no student-college pair mutually prefers each other to the match-

[9] The proof follows closely the proof of Proposition 1 and uses in the second part the fact that for any $C' \subsetneq C'' \subseteq C$ and any profile R, if $DA_S(1_{C''}, R)(c) = \emptyset$ for all $c \in C''\backslash C'$, then $DA_S(1_{C'}, R) = DA_S(1_{C''}, R)$.

ing such that the college strictly prefers the student to *all* its assigned students.

Weak Fairness: For all (q, R), there exists no student-college pair (s, c) such that $cP_s\varphi(q, R)(s)$ and sP_cs' for all $s' \in \varphi(q, R)(c)$.

A weaker notion of non-wastefulness is the following: there is no student-college pair (s, c) such that both s is unmatched and s prefers c to her current assignment and c has a vacant position which it prefers to fill with s instead of having the position empty.

Weak Non-Wastefulness: For all problems (q, R), there exists no student-college pair (s, c) such that $\varphi(q, R)(s) = \emptyset$, $cP_s\varphi(q, R)(c)$, $|\varphi(q, R)(c)| < q_c$, and $sP_c\emptyset$.

For school choice we establish the following variant of Sönmez's impossibility result for college admissions.

Theorem 5 *Suppose there are at least four students and at least two colleges. Then*

(a) there exists no mechanism that is non-wasteful, weakly fair, and non-manipulable via capacities; and

(b) there exists no mechanism that is weakly non-wasteful, fair and non-manipulable via capacities.

Proof. We prove both (a) and (b) via the same example. Consider the following problem: let $S = \{s_1, s_2, s_3, s_4\}$, $C = \{c_1, c_2\}$, and R be a profile such that R_{s_1} : $c_1c_2\emptyset$[10], $R_{s_2} : c_2c_1\emptyset$, $R_{s_3} : c_1c_2\emptyset$, $R_{s_4} : c_1c_2\emptyset$, both $R_{c_1} : s_2s_3s_1s_4\emptyset$ and $s_2P_{c_1}\{s_1, s_3, s_4\}$, and $R_{c_2} : s_1s_2s_3s_4\emptyset$. Let[11]

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

[10] This means $c_1P_{s_1}c_2P_{s_1}\emptyset$.

[11] If a student is not indicated in the matching, then the student is unmatched under this matching.

First, we show (a). Suppose to the contrary that there exists a mechanism φ satisfying the properties of (a) in Theorem 5. By weak fairness of φ (which implies fairness for base problems) and non-wastefulness, we have $\varphi(1,1,R) = \mu$. By non-wastefulness, $\varphi(4,1,R) = \mu'$. Now by $s_2 P_{c_1} \{s_1, s_3, s_4\}$, $\varphi(1,1,R)(c_1) P_{c_1} \varphi(4,1,R)(c_1)$, which means that φ is manipulable via capacities at $(4,1,R)$, a contradiction.

Second, we show (b). Suppose to the contrary that there exists a mechanism φ satisfying the properties of (b) in Theorem 5. By fairness and weak non-wastefulness we have $\varphi(1,1,R) = \mu$. Consider the problem $(3,1,R)$. By weak non-wastefulness and our construction, for all $s \in S$, $\varphi(3,1,R)(s) \neq s$. Now if $\varphi(3,1,R)(s_4) = c_2$, then $c_2 P_{s_2} \varphi(3,1,R)(s_2)$ and $s_2 P_{c_2} s_4$, which means that φ violates fairness, a contradiction. Thus, $\varphi(3,1,R)(s_4) = c_1$. Similarly, it follows that $\varphi(3,1,R)(s_3) = c_1$.

Now from $s_4 \in \varphi(3,1,R)(c_1)$, $s_1 P_{c_1} s_4$, $s_1 P_{c_1} s_4$ and fairness we obtain $\varphi(3,1,R) = \mu'$. By $s_2 P_{c_1} \{s_1, s_3, s_4\}$, we have $\varphi(1,1,R)(c_1) P_{c_1} \varphi(3,1,R)(c_1)$, which means that φ is manipulable via capacities at $(3,1,R)$, a contradiction. $\square$

Since stability implies (weak) non-wastefulness and (weak) fairness, Theorem 1 follows from Theorem 5 (when there are more than four students). Reformulated for school choice, Sönmez (1997, Theorem 1) shows that there exists no mechanism which is non-wasteful, fair, and non-manipulable via capacities. Weakening either non-wastefulness or fairness as above still results in an impossibility regarding non-manipulability via capacities.

When both "no blocking" properties are weakened, a possibility emerges.

Proposition 2 *Let ϕ be a stable base mechanism. Then the iterative ϕ-mechanism $I(\phi)$ is weakly non-wasteful, weakly fair, and non-manipulable via capacities.*

Proof. By Proposition 1, $I(\phi)$ is non-manipulable via capacities. Let (q,R) be a problem.

In showing weak non-wastefulness, suppose that there exists (s,c) such that $I(\phi)(q,R)(s) = \emptyset$, $cP_s I(\phi)(q,R)(s)$, $|I(\phi)(q,R)(c)| < q_c$, and $sP_c\emptyset$. But then for all

$l \in \{1, \ldots, q_{\max}\}$, $s \notin S_l$, and for some $k \in \{1, \ldots, q_c\}$, $\phi(1_{C_k}, R_{-(S_1 \cup \cdots \cup S_{k-1})})(c) = \emptyset$. This means that $\phi(1_{C_k}, R_{-(S_1 \cup \cdots \cup S_{k-1})})$ is not stable because $cP_s\emptyset$, $sP_c\emptyset$, and $s \notin S_1 \cup \cdots \cup S_{k-1}$. Thus, ϕ is not a stable base mechanism, a contradiction.

In showing weak fairness, suppose that there exists (s, c) such that $cP_sI(\phi)(q, R)(s)$ and $s'P_cs$ for all $s' \in I(\phi)(q, R)(c)$. Since ϕ is stable for base problems, we have $\phi(1_{C_1}, R)(c) \neq \emptyset$. Let $s' = \phi(1_{C_1}, R)(c)$. But then by stability of $\phi(1_{C_1}, R)$ and $s' \neq s$, $s'P_cs$, a contradiction. □

In applications, if non-manipulation via capacities is more important than stability, then an iterative stable mechanism may provide a practical solution since it satisfies the weaker versions of the two "no blocking" conditions non-wastefulness and fairness. If stability is more important than non-manipulation via capacities, then the student-optimal stable mechanism is a good solution because of its limited manipulability via capacities.

References

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

Alcalde, J. (1996): "Implementation of Stable Solutions to Marriage Problems," *Journal of Economic Theory* 69:240–254.

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

Ehlers, L. (2008): "Truncation Strategies in Matching Markets," *Mathematics of Operations Research* 33:327–335.

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

Kesten, O. (2008): "On Two Kinds of Manipulation for School Choice Problems," Working Paper.

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

Kojima, F. (2007): "When can Manipulations be Avoided in Two-Sided Matching Markets? – Maximal Domain Results," *The B.E. Journal of Theoretical Economics*, Vol. 7: Iss. 1 (Contributions), Article 32.

Kojima, F., and P. Pathak (2008): "Incentives and Stability in Large Two-Sided Matching Markets," *American Economic Review*, forthcoming.

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

Romero-Medina, A., and M. Triossi (2007): "Games of Capacities: A (Close) Look at Nash Equilibria," Working Paper.

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

Roth, A.E. (1985): "The College Admissions Problem is not equivalent to the Marriage Problem," *Journal of Economic Theory* 36:277–288.

Roth, A.E., and M. Sotomayor, Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis, Cambridge Univ. Press, London/New York, 1990.

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