JOURNAL OF ECONOMIC THEORY 77, 197–204 (1997)
ARTICLE NO. ET972316

Manipulation via Capacities in Two-Sided Matching Markets

Tayfun Sönmez*

Department of Economics, University of Michigan, Ann Arbor, Michigan 48109-1220

Received March 7, 1996; revised March 5, 1997

We study manipulation of solutions by hospitals via underreporting their capacities in the context of centralized two-sided matching markets. We show that the solution that is used to match medical interns and hospitals in United States is manipulable in this way. Our main result is that there is no solution that is *stable* and *non-manipulable via capacities*. *Journal of Economic Literature* Classification Numbers: C71, C78, D71 D78. © 1997 Academic Press

1. INTRODUCTION

Recently there has been a growing interest on manipulation and implementation in economic domains. One such domain is the domain of *two-sided matching problems* [3].[1] Here there are two finite and disjoint sets of agents, say medical interns and hospitals; each hospital has a capacity that limits the maximum number of interns it can employ, each intern has a preference relation over the set of hospitals and being unemployed, and each hospital has a preference relation over the set of groups of interns. An allocation is a matching of interns and hospitals such that no hospital is assigned more interns than its capacity and no intern is assigned more than one hospital. A matching is *stable* if (i) no hospital prefers keeping a position vacant rather than filling it with one of its assignments, (ii) no intern prefers remaining unemployed to his/her assignment, and (iii) there is no unmatched hospital-intern pair such that the intern prefers the hospital to his/her assignment and the hospital prefers the intern to one of its assignments or keeping a vacant position (in the case it has one).

* I thank Ted Bergstrom, Alvin Roth, William Thomson, an associate editor, and seminar participants at the University of Michigan and University of Pittsburgh for their comments and suggestions. All errors are my own responsibility.

[1] See Roth and Sotomayor [14] for an extensive analysis of two-sided matching problems.

0022-0531/97 $25.00

The stability criterion has been very central to studies concerning two-sided matching problems as well as to its real life applications. For example Roth [11] shows that the National Resident Matching Program (NRMP) has been using a stable solution to match the medical interns and hospitals in United States since 1950. Unfortunately this solution employs some strategic opportunities: agents can manipulate it via their preferences. Nevertheless this is not the fault of this particular solution: Roth [10] shows that there is no solution that is *stable* and *non-manipulable via preferences*.[2] Some of the recent papers concerning manipulation and implementation in two-sided matching problems include Alcalde [1], Alcalde and Barberà [2], Kara and Sönmez [4, 5], Ma [6, 7], Shin and Suh [16], and Sönmez [17, 18].

Recently there has been increased activity in the medical community concerning a possible change in the solution that is used by the NRMP: The American Medical Students Association (AMSA) has been urging changes in the current solution anal the Board of Directors of the NRMP retained Alvin Roth (University of Pittsburgh) to direct a study concerning the effects of possible changes. (See Public Citizen's Health Research Group and the AMSA Report on Hospital Bias in the NRMP [9] and Roth [13] respectively.) Among other things the AMSA and the Board of Directors of the NRMP are particularly interested in the strategic implications of possible changes. Naturally both these parties as well as the earlier papers focus their attention to the manipulation possibilities via the preferences.

In this paper we depart from this line and study the manipulation opportunities of the hospitals by underrepresenting their capacities. Obviously in many situations the capacities are private information and hence such an attempt may be plausable if it is profitable. This is in the same spirit with Postlewaite [8] who studies manipulation via endowments in the context of exchange economies.[3] (See also Sertel [15] and Thomson [20, 21, 22].) Our main result is a counterpart to Roth [10]: there is no solution that is *stable* and *non-manipulable via capacities*. We find this result surprising as unlike the manipulation possibilities via preferences, the manipulation possibilities via capacities are very limited. For example if the capacity of a hospital is two then the only possibly profitable manipulation is reporting a capacity of one. Nevertheless, it turns out that even such a limited opportunity may be good enough to manipulate any stable solution.

[2] This property is widely known as *strategy-proofness*.

[3] Postlewaite [8] studies both *manipulation via witholding the endowments* as well as *manipulation via destroying the endowments*. As hospitals have no use for empty slots, our notion of *manipulation via capacities* is analogous to *manipulation via destroying the endowments*. He shows that there is no solution that is *Pareto efficient*, *individually rational*, and *non-manipulable via witholding the endowments*. On the other hand he constructs a class of solutions that is *Pareto efficient*, *individually rational*, and *non-manipulable via destroying the endowments*.

2. THE MODEL

A (*many-to-one*) *matching problem* is a four-tuple (H, I, R, q). The first two components are non-empty, finite, and disjoint sets of hospitals and interns $H = \{h_1, ..., h_n\}$ and $I = \{i_1, ..., i_m\}$. The third component $R = (R_k)_{k \in H \cup I}$ is a list of preference relations of hospitals and interns. Let P_k denote the strict relation associated with the preference relation R_k for all $k \in H \cup I$. The last component is a vector of positive natural numbers $q = (q_{h_1}, ..., q_{h_n})$, where q_{h_i} is the capacity of hospital $h_i \in H$. We consider the case where H and I are fixed and hence each matching problem is defined by a preference profile and a capacity vector.

The preference relation R_i of each intern $i \in I$ is a binary relation on $\Sigma_i = \{\{h_1\}, ..., \{h_n\}, \varnothing\}$ which is *reflexive* (for all $\sigma \in \Sigma_i$ we have $\sigma R_i \sigma$), *transitive* (for all $\sigma, \sigma', \sigma'' \in \Sigma_i$ if $\sigma R_i \sigma'$ and $\sigma' R_i \sigma''$ then $\sigma R_i \sigma''$), and *total* (for all $\sigma, \sigma' \in \Sigma_i$ with $\sigma \neq \sigma'$ we either have $\sigma R_i \sigma'$ or $\sigma' R_i \sigma$ buth not both). Such preference relations are referred to as *linear orders* (or strict preferences). Let $\mathscr{R}_i$ be the class of all such preference relations for intern $i \in I$. The preference relation R_h of hospital $h \in H$ is a linear order on $\Sigma_h = 2^I$ and it is *responsive* (Roth [12]): For all $J \subset I$,

1. for all $i \in I \backslash J$, $J \cup \{i\}\ P_h J$ if and only if $\{i\}\ P_h \varnothing$,
2. for all $i, i' \in I \backslash J$, $J \cup \{i\}\ P_h J \cup \{i'\}$ if and only if $\{i\}\ P_h \{i'\}$.

Let $\mathscr{R}_h$ be the class of all such preferences for hospital $h \in H$. Let $\mathscr{E} = \mathbb{N}_+^n \times \Pi_{k \in H \cup I} \mathscr{R}_k$. That is, $\mathscr{E}$ is the class of all matching problems for given H and I.

The *choice* of a hospital h from a group of interns $J \subseteq I$ under the preference R_h and capacity q_h is defined as

$$Ch_h(R_h, q_h, J) = \{J' \subseteq J: |J'| \leqslant q_h, J' R_h J'' \text{ for all } J'' \subseteq J \text{ such that } |J''| \leqslant q_h\}.$$

A *matching* μ for a given capacity vector q is a function from the set $H \cup I$ into $2^{H \cup I}$ such that:

1. for all $i \in I$, $|\mu(i)| \leqslant 1$ and $\mu(i) \subseteq H$;
2. for all $h \in H$, $|\mu(h)| \leqslant q_h$ and $\mu(h) \subseteq I$;
3. for all $(h, i) \in H \times I$, $\mu(i) = \{h\}$ if and only if $i \in \mu(h)$.

We denote the set of all matchings for a given q by $\mathscr{M}(q)$ and the set of all matchings by $\mathscr{M}$. Given a preference relation R_h of a firm $h \in H$, initially defined over Σ_h, we extend it to the set of matchings $\mathscr{M}$, in the following natural way: h prefers the matching μ to the matching μ' if and only if it

prefers $\mu(h)$ to $\mu'(h)$. We slightly abuse the notation and also use R_h to denote this extension. We do the same for each intern $i \in I$.

A matching μ is *blocked by an intern* $i \in I$ if $\varnothing P_i \mu(i)$. A matching μ is *blocked by a hospital* $h \in H$ if $\mu(h) \neq Ch_h(R_h, q_h, \mu(h))$. Note that whenever the preferences are responsive this statement is equivalent to the following: A matching μ is blocked by a hospital $h \in H$ if there is an intern $i \in \mu(h)$ such that $\varnothing P_h \{i\}$. A matching μ is *individually rational* if it is not blocked by an intern or a hospital. A matching μ is *blocked by a hospital–intern pair* $(h, i) \in H \times I$ if $\{h\}\ P_i \mu(i)$ and $\mu(h) \neq Ch_h(R_h, q_h, \mu(h) \cup \{i\})$. A matching μ is *stable* if it is not blocked by an intern, a hospital, or a hospital-intern pair. We denote the set of stable matchings under (R, q) by $\mathscr{S}(R, q)$. Roth [11] shows that for any matching problem $(R, q) \in \mathscr{E}$ there exists a matching $\mu_H(R, q) \in \mathscr{S}(R, q)$ such that

$$\text{for all } h \in H, \text{ for all } \mu \in \mathscr{S}(R, q); \mu_H(R, q)(h)\ R_h \mu(h).$$

We refer to this matching as the *hospital-optimal stable matching* for the matching problem $(R, q) \in \mathscr{E}$. There is an analogous matching which favors the interns and we refer to it as the *intern-optimal stable matching*.

A *matching rule* is a function $\varphi: \mathscr{E} \to \mathscr{M}$ such that, for all $(R, q) \in \mathscr{E}$ we have $\varphi(R, q) \in \mathscr{M}(q)$. An example of a matching rule is the one which selects the hospital-optimal stable matching for each problem. We denote this rule by μ_H and refer to it as the *hospital-optimal stable rule*.

A matching rule φ is *stable* if $\varphi(R, q) \in \mathscr{S}(R, q)$ for all $(R, q) \in \mathscr{E}$. A matching rule φ is *non-manipulable via capacities* if

$$\text{for all } (R, q) \in \mathscr{E}, \text{ for all } h \in H, \text{ for all } q'_h \leqslant q_h,$$

$$\varphi(R, q)(h)\ R_h \varphi(R, q_{-h}, q'_h)(h).$$

That is, a matching rule is *non-manipulable via capacities* if no hospital can ever benefit by underreporting its capacity.

3. MANIPULATION VIA CAPACITIES

The NRMP uses the hospital-optimal stable rule to match medical interns and hospitals in United States. We first show that this matching rule is not immune to *manipulation via capacities* as long as there are at least two hospitals and two interns.

PROPOSITION 1. *Suppose there are at least two hospitals and two interns. Then the hospital-optimal stable rule is not immune to manipulation via capacities.*

Proof. We first prove the proposition for two hospitals and two interns. Let $H = \{h_1, h_2\}$, $I = \{i_1, i_2\}$,

$$\{i_1, i_2\}\ P_{h_1}\{i_1\}\ P_{h_1}\{i_2\}\ P_{h_1}\varnothing,$$
$$\{i_1, i_2\}\ P_{h_2}\{i_2\}\ P_{h_2}\{i_1\}\ P_{h_2}\varnothing,$$
$$\{h_2\}\ P_{i_1}\{h_1\}\ P_{i_1}\varnothing,$$
$$\{h_1\}\ P_{i_2}\{h_2\}\ P_{i_2}\varnothing,$$

$q_{h_1} = q'_{h_2} = 1$ and $q_{h_2} = 2$.

We have $\mathscr{S}(R, q_{h_1}, q_{h_2}) = \{\mu_1\}$, $\mathscr{S}(R, q_{h_1}, q'_{h_2}) = \{\mu_1, \mu_2\}$, where

$$\mu_1 = \begin{pmatrix} h_1 & h_2 \\ \{i_2\} & \{i_1\} \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} h_1 & h_2 \\ \{i_1\} & \{i_2\} \end{pmatrix},^4$$

and therefore $\mu_H(R, q_{h_1}, q_{h_2}) = \mu_1$ and $\mu_H(q_{h_1}, q'_{h_2}) = \mu_2$. Hence we have

$$\mu_H(R, q_{h_1}, q'_{h_2})(h_2)\ P_{h_2}\mu_H(R, q_{h_1}, q_{h2})(h_2)$$

completing the proof for the case of two interns and two hospitals. Finally we can include hospitals whose top choice is keeping all its positions vacant and interns whose top choice is staying unemployed to generalize this proof to situations with more than two interns and two hospitals.

Q.E.D.

Our main theorem shows that the hospital-optimal stable rule is not the only matching rule that suffers from *manipulation via capacities.*

THEOREM 1. *Suppose there are at least two hospitals and three interns. Then these exists no matching rule that is stable and non-manipulable via capacities.*

Proof. We first prove the theorem for two hospitals and three interns. Let $\varphi: \mathscr{E} \to \mathscr{M}$ be *stable*, $H = \{h_1, h_2\}$, $I = \{i_1, i_2, i_3\}$,

$$\{i_1, i_2, i_3\}\ P_{h_1}\{i_1, i_2\}\ P_{h_1}\{i_1, i_3\}\ P_{h_1}\{i_1\}\ P_{h_1}\{i_2, i_3\}\ P_{h_1}\{i_2\}\ P_{h_1}\{i_3\}\ P_{h_1}\varnothing,$$
$$\{i_1, i_2, i_3\}\ P_{h_2}\{i_2, i_3\}\ P_{h_2}\{i_1, i_3\}\ P_{h_2}\{i_3\}\ P_{h_2}\{i_1, i_2\}\ P_{h_2}\{i_2\}\ P_{h_2}\{i_1\}\ P_{h_2}\varnothing,$$
$$\{h_2\}\ P_{i_1}\{h_1\}\ P_{i_1}\varnothing,$$
$$\{h_1\}\ P_{i_2}\{h_2\}\ P_{i_2}\varnothing,$$
$$\{h_1\}\ P_{i_3}\{h_2\}\ P_{i_3}\varnothing,$$

$q_{h_1} = q_{h_2} = 2$ and $q'_{h_1} = q'_{h_2} = 1$.

[4] We read this as $\mu_1(h_1) = \{i_2\}$ and $\mu_1(h_2) = \{i_1\}$. Likewise for μ_2.

We have $\mathscr{S}(R, q_{h_1}, q_{h_2}) = \{\mu_1\}$, $\mathscr{S}(R, q_{h_1}, q'_{h_2}) = \{\mu_1, \mu_2\}$ and $\mathscr{S}(R, q'_{h_1}, q'_{h_2}) = \{\mu_3\}$ where

$$\mu_1 = \begin{pmatrix} h_1 & h_2 \\ \{i_2, i_3\} & \{i_1\} \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} h_1 & h_2 \\ \{i_1, i_2\} & \{i_3\} \end{pmatrix}, \qquad \mu_3 = \begin{pmatrix} h_1 & h_2 \\ \{i_1\} & \{i_3\} \end{pmatrix}.$$

Therefore $\varphi(R, q_{h_1}, q_{h_2}) = \mu_1$, $\varphi(R, q'_{h_1}, q'_{h_2}) = \mu_3$, and $\varphi(R, q_{h_1}, q'_{h_2}) \in \{\mu_1, \mu_2\}$. If $\varphi(R, q_{h_1}, q'_{h_2}) = \mu_1$ then $\varphi(R, q'_{h_1}, q'_{h_2})(h_1) = \mu_3(h_1) = \{i_1\}$ and $\varphi(R, q_{h_1}, q'_{h_2})(h_1) = \mu_1(h_1) = \{i_2, i_3\}$ and hence

$$\varphi(R, q'_{h_1}, q'_{h_2})(h_1)\ P_{h_1}\varphi(R, q_{h_1}, q'_{h_2})(h_1)$$

which implies hospital 1 can *manipulate* φ *via capacities* when its capacity is $q_{h_1} = 2$ and hospital 2's capacity is $q'_{h_2} = 1$ by underreporting its capacity as $q'_{h_1} = 1$. Otherwise $\varphi(R, q_{h_1}, q'_{h_2}) = \mu_2$ and therefore $\varphi(R, q_{h_1}, q'_{h_2})(h_2) = \mu_2(h_2) = \{i_3\}$, $\varphi(R, q_{h_1}, q_{h_2})(h_2) = \mu_1(h_2) = \{i_1\}$. Hence

$$\varphi(R, q_{h_1}, q'_{h_2})(h_2)\ P_{h_2}\varphi(R, q_{h_1}, q_{h_2})(h_2)$$

which implies hospital 2 can *manipulate* φ *via capacities* when its capacity is $q_{h_2} = 2$ and hospital 1's capacity is $q_{h_1} = 2$ by underreporting its capacity as $q'_{h_2} = 1$. Hence φ is *manipulable via capacities* completing the proof for the case of two hospitals and three interns. Finally we can include hospitals whose top choice is keeping all its positions vacant and interns whose top choice is staying unemployed to generalize this proof to situations with more than three interns and two hospitals. Q.E.D.

Remark 1. Theorem 1 does not hold if (i) there is only one hospital, (ii) there is only one intern, and (iii) there are two hospitals and two interns. In the first two cases there is a single stable matching for each problem and the unique stable matching rule is *non-manipulable via capacities*. In the last case the intern-optimal stable rule is *non-manipulable via capacities*.

Remark 2. Agents can also manipulate matching rules by pre-arranging matches before the centralized procedure. In a related impossibility theorem Sönmez [19] shows that there is no matching rule that is *stable* and *non-manipulable via prearranged matches*.

Remark 3. Alcalde and Barberà [2] improve upon Roth [10] and show that there is no matching rule that is *Pareto efficient individually rational*, and *non-manipulable via preferences*. One cannot obtain a counterpart to this result by replacing *non-manipulability via capacities* with *non-manipulability*

via preferences. An example of a matching rule that is *Pareto-efficient*, *individually rational*, and *non-manipulable via capacities* is

$$\varphi(h_1) = Ch_{h_1}(R_{h_1}, q_{h_1}, \{i \in I: \{h_1\}\ P_i \varnothing\})$$

$$\varphi(h_k) = Ch_{h_k}\left(R_{h_k}, q_{h_k}, \{i \in I: \{h_k\}\ P_i \varnothing\} \setminus \bigcup_{l=1}^{k-1} \varphi(h_l)\right) \qquad k = 2, ..., n.$$

Remark 4. Alcalde and Barber'a [2] show that when the preferences of the hospitals are responsive and the class of the preferences of hospitals satisfy the *top dominance condition*,[5] the intern-optimal stable rule is *non-manipulable via preferences*. The preferences of the hospitals in the proof of Theorem 1 are consistent with this requirement and therefore the analogue of Alcalde and Barberà's positive result does not hold when *manipulation* is *via capacities*.

REFERENCES

1. J. Alcalde, Implementation of stable solutions to the marriage problem, *J. Econ. Theory* **69** (1996), 240–254.
2. J. Alcalde and S. Barberà, Top dominance and the possibility of strategy-proof stable solutions to matching problems, *Econ. Theory* **4** (1994), 417–435.
3. D. Gale and L. Shapley, College admissions and the stability of marriage, *Amer. Math. Monthly* **69** (1962), 9–15.
4. T. Kara and T. Sönmez, Implementation of college admission rules, mimeo, University of Rochester, 1994. [*Econ. Theory*, in press]
5. T. Kara and T. Sönmez, Nash implementation of matching rules, *J. Econ. Theory* **68** (1996), 425–439.
6. J. Ma, Manipulation and stability in a college admissions problems, mimeo, Rutgers University, 1994.
7. J. Ma, Stable matchings and rematching-proof equilibria in a two-sided matching market, *J. Econ. Theory* **66** (1995), 352–369.
8. A. Postlewaite, Manipulation via endowments, *Rev. Econ. Stud.* **46** (1979), 255–262.
9. Public Citizen's Health Research Group and the Americal Medical Student Association report on hospital bias in the NRMP, web page: http://pubweb.acns.nwu.edu/@lan/nrmp2.html, 1995.
10. A. E. Roth, The economics of matching: stability and incentives, *Math. Oper. Res.* **7** (1982), 617–628.
11. A. E. Roth, The evolution of the labor market for medical interns and residents: A case study in game theory, *J. Polit. Econ.* **92** (1984), 991–1016.
12. A. E. Roth, The college admissions problem is not equivalent to the marriage problem, *J. Econ. Theory* **36** (1985), 277–288.

[5] A class of preferences $\mathscr{R} \subseteq \mathscr{R}_h$ satisfies the *top dominance condition* if and only if for any pair of preferences $R_h, R'_h \in \mathscr{R}$, and any $i, i' \in I$ if we have $\{i\}\ P_h \varnothing$, $\{i'\}\ P'_h \varnothing$, $\{i\}\ P_h \{i'\}$ and $\{i'\}\ P'_h \{i\}$ then there is no $i'' \in I$ with $\{i''\}\ P_h \{i\}$ and $\{i''\}\ P'_h \{i'\}$.

13. A. E. Roth, Proposed research program: Evaluation of changes to be considered in the NRMP algorithm, consultant's report and mimeo, University of Pittsburgh, web page: http://www.pitt.edu/@lroth/nrmp.html, 1995.
14. A. E. Roth and M. Sotomayor, "Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis," Cambridge Univ. Press, London/New York, 1990.
15. M. Sertel, Manipulating Lindahl equilibrium via endowments, *Econ. Lett.* **46** (1994), 167–171.
16. S. Shin and S-C. Suh, A mechanism implementing the stable rule in marriage problems, *Econ. Lett.* **51** (1996), 185–189.
17. T. Sönmez, Strategy-proofness in many-to-one matching problems, *Econ. Design* **1** (1996), 365–380.
18. T. Sönmez, Games of manipulation in marriage problems, mimeo, University of Michigan, 1996. [*Games Econ. Behav.*, in press]
19. T. Sönmez, Can pre-arranged matches be avoided in two-sided matching markets?, mimeo, University of Michigan, 1996.
20. W. Thomson, Monotonic allocation mechanisms, mimeo, University of Rochester, 1987.
21. W. Thomson, Monotonic allocation mechanisms in economies with public goods, mimeo, University of Rochester, 1987.
22. W. Thomson, Endowment monotonicity in economies with single-peaked preferences, mimeo, University of Rochester, 1995.