Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Economic Theory 125 (2005) 189–193

JOURNAL OF
Economic
Theory

www.elsevier.com/locate/jet

Notes, Comments, and Letters to the Editor

Pairwise kidney exchange: Comment

John William Hatfield

Graduate School of Business, Stanford University, CA, USA

Received 13 May 2005; final version received 10 September 2005

Abstract

In their recent paper, Roth et al. [Pairwise kidney exchange, J. Econ. Theory 125 (2005) 151–188] consider pairwise kidney exchanges, and show within this subset of feasible exchanges that a priority mechanism is strategy-proof. We show that this result can be broadened to allow much more general mechanisms and restrictions on the feasible set of allocations, including allowing three-way exchanges, regional specifications, and others. The key requirement is that the choice mechanism be consistent, i.e., if an allocation is chosen from some set of feasible allocations, it is also chosen from any subset of that set.

JEL classification: D02

Keywords: Kidney exchange; Strategy-proofness; Matching markets

---

1. Introduction

How to construct efficient kidney exchanges is a new and important application of economic theory. Thousands of patients have died while waiting for a donor kidney, and instituting a procedure which allows trading of donor kidneys could greatly reduce these numbers in the future. Yet when designing these procedures, it is important that we do not give agents incentives to lie about medical details or their available donors in order to improve their chances of getting a match.

Roth et al. [4] show that for pairwise exchanges, a priority mechanism ensures that it is a dominant strategy for patients to truthfully reveal both the set of donors they can receive kidneys from and the set of patients that their donor can donate a kidney to. However, many patients cannot be accommodated using only a pairwise exchange; for instance, if exchanges involving three patient–donor pairs were allowed, more patients could receive donated kidneys. In recent work, Saidman et al. [5] estimate that a matching mechanism that included exchanges with three patient–donor pairs would allow approximately 10% more of the patients in the pool with

---

E-mail address: hatfield@stanford.edu.

doi:10.1016/j.jet.2005.09.002

incompatible donors to receive kidneys. In this paper, we find necessary and sufficient conditions for the exchange mechanism and restrictions on the set of feasible exchanges such that the agents will wish to truthfully reveal; for instance, a priority mechanism allowing exchanges with up three patient–donor pairs is indeed strategy-proof. One may also wish to allow mechanisms that give a greater weight to some patients, even at the cost of the overall size of the allocation, if these agents are the most likely to die before another exchange is organized; these too will be strategy-proof. The key condition is that the mechanism must be consistent; that is, if a certain allocation is chosen from some set of feasible allocations over another allocation, then there cannot exist a set of feasible allocations that include both of these allocations such that the latter is chosen.

In the prior literature, Roth et al. [3] discuss a different model of kidney exchange, where any number of kidneys can be exchanged simultaneously. This is based on the housing assignment problem, first introduced by Shapley and Scarf [7], and is known to be a strategy-proof mechanism. This model, however, only considers unrestricted exchanges. Our goal in this comment is to completely characterize the set of one item exchange mechanisms, for binary preferences, that are strategy-proof.

2. Model

Let $\mathbb{N} = \{1, 2, \ldots, N\}$ be the set of patients, each of whom has one or more associated incompatible donors. We shall refer to each patient and his associated donors as an agent. Each agent has very simple preferences; he prefers to receive a kidney rather than not.

Let $\mathbb{P}$ denote the set of ordered pairs (i, j) such that agent i can donate to agent j. A *feasible allocation* F is a set of ordered pairs (i, j) such that

1. if $(i, j) \in F$, then $(i, j) \in \mathbb{P}$,
2. if $(i, j) \in F$, then there exists a $k \in \mathbb{N}$ such that $(k, i) \in F$ and
3. if $(i, j) \in F$, then there does not exist a $k \neq j$ such that $(i, k) \in F$.

Let the set of feasible allocations be denoted $\mathcal{F}(\mathbb{P})$. The first restriction simply states that each particular donation is feasible. The second requires that any patient–donor pair which donates a kidney also receives one. The third condition ensures that each agent is donating only one kidney. Hence we can think of allocations as a set of cycles of agents, where within each cycle each agent donates and receives a kidney.

Roth et al. [4] require that allocations contain only pairwise exchanges, i.e., if $(i, j) \in F$, then $(j, i) \in F$. We may, however, wish to introduce other restrictions. For instance, it may be possible to do three-way exchanges as well. Or it may be that three-way exchanges are allowed, but only if at least two members of the exchange are in the same region of the country. For our purposes, the nature of the restriction can be completely arbitrary. We define a *restriction* $\rho(\mathbb{P})$ as taking the intersection of the set of feasible allocations $\mathcal{F}(\mathbb{P})$ and some other predetermined set of allocations, $\mathcal{F}_\rho$, that is, $\rho(\mathbb{P}) = \mathcal{F}(\mathbb{P}) \cap \mathcal{F}_\rho$. We might have $\rho(\mathbb{P}) = \mathcal{F}(\mathbb{P})$, if we wish to allow any feasible allocation, or perhaps $\rho(\mathbb{P}) = \{F \in \mathcal{F}(\mathbb{P}) \mid (i, j) \in F \Rightarrow (j, i) \in F\}$ if we wish to restrict attention to pairwise exchanges.

A *mechanism* M is a function from the set of ρ-feasible allocations $\rho(\mathbb{P})$ to a particular ρ-feasible allocation F.[1]

---

[1] Of course, this assumes that the information available to the mechanism is correct. If the mechanism chooses an allocation which is not in fact feasible due to misinformation, we shall assume that we simply implement the cycles within that allocation that are actually feasible.

We can think of mechanisms as choice mechanisms, and we shall say a mechanism is *consistent* if $F = M(\mathcal{A})$ and $F \in \mathcal{A}' \subset \mathcal{A}$ imply that $F = M\left(\mathcal{A}'\right)$.[2] This requirement simply states that if the mechanism chooses F from a set $\mathcal{A}$, it still chooses F from the subset $\mathcal{A}'$, as long as F is in $\mathcal{A}'$. For instance, consider the *priority* mechanism introduced in Roth et al. [4]. This mechanism ranks allocations in the following way: if $|F| > \left|F'\right|$, then F is chosen over F'. For two allocations of the same size, let i be the lowest numbered agent who is a member of one of the allocations but not both. Then the priority mechanism ranks F over F' if $(i, j) \in F$ for some j. Since we can rank all allocations against each other, the priority mechanism is consistent; an allocation preferred from a set $\mathcal{A}$ will still outrank all the other allocations in the subset $\mathcal{A}'$.

Roth et al. [4] differentiate between the mechanism used to pick over feasible exchanges and the restriction on the set of feasible exchanges, as we have above. However, this distinction is in some sense superfluous. We can instead think of a *procedure* $K(\mathbb{P})$ which maps the set of feasible donations directly to an allocation. For any mechanism M and restriction ρ, we can define the corresponding procedure $K(\mathbb{P}) \equiv M(\rho(\mathbb{P}))$, encapsulating both the restrictions on allocations and our choice among allocations. Considering procedures may give us more flexibility: for instance, we may wish to allow exchanges involving three agents only if it greatly increases the number of agents who receive a kidney.

A procedure K is *consistent* if $F = K(\mathbb{P})$, $\mathbb{P}' \subseteq \mathbb{P}$ and $F \in \mathcal{F}\left(\mathbb{P}'\right)$ imply that $F = K\left(\mathbb{P}'\right)$. We reuse the term consistent as, for any consistent mechanism, and for any restriction, the corresponding procedure is consistent.

Theorem 1. *For any consistent mechanism M, and any restriction ρ, the procedure $K(\mathbb{P}) \equiv M(\rho(\mathbb{P}))$ is consistent.*

Proof. Suppose $K(\mathbb{P}) = M(\rho(\mathbb{P})) = F$. Now consider $\mathbb{P}' \subseteq \mathbb{P}$ such that $F \in \mathcal{F}\left(\mathbb{P}'\right)$. Since $F = M(\rho(\mathbb{P}))$, $F \in \rho(\mathbb{P})$, so $F \in \mathcal{F}_\rho$, and hence $F \in \mathcal{F}_\rho \cap \mathcal{F}\left(\mathbb{P}'\right) = \rho\left(\mathbb{P}'\right)$, so F can still be chosen by M. Since $\rho\left(\mathbb{P}'\right) = \mathcal{F}_\rho \cap \mathcal{F}\left(\mathbb{P}'\right)$, and $\mathcal{F}\left(\mathbb{P}'\right) \subseteq \mathcal{F}(\mathbb{P})$, $\rho\left(\mathbb{P}'\right) \subseteq \rho(\mathbb{P})$. Hence, by the consistency of M, $M\left(\rho\left(\mathbb{P}'\right)\right) = F$. $\square$

Note that while consistency is an inclusive condition, not all reasonable procedures will satisfy it. For instance, we may wish to give a higher priority (within a priority mechanism, as defined above) to agents who bring in more or better donors. For instance, consider a pairwise priority mechanism (as in [4]) with the following twist: priority is given to agents who bring in "good donors"; that is, the more people you can donate to, the higher your priority. Hence, if an agent adds to the number of agents he can donate to, he may receive a kidney when he did not before, even if none of those new donation possibilities are used, and so this mechanism will not be consistent.

Consider the case where Alice can donate to Bob or Carol, Bob can donate to Alice or David, Carol can donate to Alice, and finally, David and Ellen can donate to no one. Then the mechanism will choose to have Alice and Bob donate to each other. However, if Carol lies and brings in a "shill donor" who can donate to David and Ellen, now the mechanism chooses to have Alice and Carol donate to each other, so Carol is better off. (Note that the allocation chosen by the mechanism is still feasible, even though it received incorrect data.) Hence, the mechanism will

---

[2] This condition is known as condition α in social choice theory [6]. If we think of the mechanism as an agent, it is equivalent to stating that the agent's preferences satisfy the weak axiom of revealed preference, as defined in Mas-Colell et al. [1].

not be strategy-proof, as an agent might then wish to bring in fake donors, who would increase their priority, but who had no real intention to donate, if actually called upon to do so.

3. Strategy-proofness

Roth et al. [4] consider the case where we restrict attention to the priority mechanism and restrict the set of feasible allocations to just those of pairwise exchanges. They show that it is a dominant strategy for agents to truthfully reveal both the full set of kidneys they can accept and the full set of agents to whom they can donate kidneys. Our goal is to show a much more general result, allowing the use of more general restrictions, as well as possibly different mechanisms.

A procedure is *strategy-proof* if it is a dominant strategy for an agent to reveal both the set of agents from which he can accept kidneys and the set of agents to whom he can donate kidneys. In reality, an agent will not be able to manipulate his report to the mechanism quite so precisely. However, if the mechanism is strategy-proof in our strong sense, an agent will not wish to perform any manipulations of the sets of agents he can give a kidney to and receive a kidney from.

Theorem 2. *Any consistent procedure K is strategy-proof.*

Proof. Consider agent $i \in \mathbb{N}$. Let the set of agents from which he can receive kidneys be denoted $\mathbb{R}_i$, and the set of agents to whom he can donate kidneys be denoted $\mathbb{D}_i$. For the mechanism to be strategy-proof, it must be that i does not want to lie about the set of agents in these two sets.

If $(i, j) \in K(\mathbb{P})$, then agent i is happy with the outcome and has no incentive to lie. If not, suppose by way of contradiction that there exist sets $\bar{\mathbb{R}}_i$ and $\bar{\mathbb{D}}_i$, which generate a new set of ordered pairs $\bar{\mathbb{P}}$ such that $(j, i), (i, k) \in K\left(\bar{\mathbb{P}}\right)$ for some $j \in \mathbb{R}_i$ and $k \in \mathbb{D}_i$. (Otherwise, the cycle of $K\left(\bar{\mathbb{P}}\right)$ which includes i will not be feasible and so i has gained nothing by lying.) Then, since K is consistent, if agent i reports that the set of agents from which he can receive kidneys is $\{j\}$ and the of agents to whom he can donate kidneys is $\{k\}$, then K must still choose $K\left(\bar{\mathbb{P}}\right)$. Denote the set of feasible donations when i reports these singleton sets as $\hat{\mathbb{P}}$.

However, if i now reveals that he can receive kidneys from all the agents in $\mathbb{R}_i$, and donate kidneys to all the agents in $\mathbb{D}_i$, the mechanism must now choose an allocation in which i receives (and donates) a kidney, as the only allocations that are newly feasible are those which involve i; hence, by consistency, either $K(\hat{\mathbb{P}})$ is chosen or a new allocation involving i is chosen (as otherwise $K(\mathbb{P})$ is chosen from the larger set $\mathbb{P}$ but not from the subset $\hat{\mathbb{P}}$).

Since i is indeed a member of the allocation $K(\mathbb{P})$, this contradicts the assumption that i obtained a kidney by lying and we are done. □

Since the priority mechanism advocated in Roth et al. [4] is consistent, we have that

Corollary 3. *For any restriction ρ on the set of feasible allocations, a priority mechanism is strategy-proof.*

Hence, we see that we can use a mechanism that allows exchanges involving up to some maximal number of agents, or exchanges that depend on location, and so on.

In fact, consistency is the key condition to ensure strategy-proofness for this problem. A procedure is *nonbossy* if there does not exist a connection (i, j) and set of connections $\mathbb{P}$ such that (1) $(i, j) \notin K(\mathbb{P} \cup \{(i, j)\})$, (2) $K(\mathbb{P} \cup \{(i, j)\}) \neq K(\mathbb{P})$ and (3) whether or not i or j obtains a kidney does not change between $K\left(\mathbb{P}\cup\{(i,j)\}\right)$ and $K\left(\mathbb{P}\right)$. Any procedure that is strategy-proof and nonbossy is consistent.

Theorem 4. *Any nonbossy, strategy-proof procedure is consistent.*

Proof. Suppose not. Then there exists an inconsistent, nonbossy, strategy-proof procedure K. Since K is inconsistent, there exists sets $\mathbb{P}'$ and $\mathbb{P}''$, $\mathbb{P}'\subseteq\mathbb{P}''$, such that $F\neq K\left(\mathbb{P}'\right)$ and $F\in\mathcal{F}\left(\mathbb{P}'\right)$, but $F=K\left(\mathbb{P}''\right)$. Then, there exists a set $\mathbb{P}$ and connection (i,j) such that $F\neq K\left(\mathbb{P}\right)$ and $F\in\mathcal{F}\left(\mathbb{P}\right)$, but $F=K\left(\mathbb{P}\cup(i,j)\right)$, that is, when we add (i,j) to the set of feasible donations $\mathbb{P}$ the procedure chooses a new allocation that does not involve (i,j). If this new allocation changes the state of i or j, then it is not strategy-proof. If it does not change the state of i or j, then the mechanism is bossy. □

Hence, consistency is the key requirement we should impose on a mechanism if we are concerned with agent's motivations to truthfully reveal information relating to their ability to donate and accept kidneys. In some sense, we need the mechanism itself to have complete, transitive preferences over allocations in order to ensure strategy-proofness for the agents.

4. Conclusion

We have shown that strategy-proofness can be assured for a wide variety of restrictions on the set of feasible allocations, as well as for a variety of ranking procedures of various allocations. Roth et al. [4] restrict attention to pairwise exchanges, while noting it has been feasible in some cases to organize three-way exchanges, and some four-way exchanges have been organized in Romania [2]. Furthermore, large gains in welfare can be achieved when exchanges involving three patient–donor pairs are considered [5]. By extending their results to allow for more complex restrictions, the efficiency of these kidney exchange programs may be improved.

Acknowledgments

We are grateful to Paul Milgrom, Al Roth, Azeem Shaikh and Tayfun Sönmez, as well as two anonymous referees, for their helpful comments. This research has been supported by the B.F. Haley and E.S. Shaw Fund at Stanford University.

References

[1] A. Mas-Colell, M. Whinston, J. Green, Microeconomic Theory, Oxford University Press, New York, NY, 1995.
[2] M. Lucan, P. Rotariu, D. Neculoiu, G. Iacob, Kidney exchange program: a viable alternative in countries with a low rate of cadaver harvesting, Transplantation Proc. 35 (2003) 933–934.
[3] A. Roth, T. Sönmez, M. Ünver, Kidney exchange, Quart. J. Econ. 119 (2004) 457–488.
[4] A. Roth, T. Sönmez, M. Ünver, Pairwise kidney exchange, J. Econ. Theory 125 (2005) 151–188.
[5] S. Saidman, A. Roth, T. Sönmez, M. Ünver, F. Delmonico, Increasing the opportunity of live kidney donation by matching for two and three way exchanges, Working paper.
[6] A. Sen, Choice functions and revealed preference, Rev. Econ. Stud. 38 (1971) 307–317.
[7] L. Shapley, H. Scarf, On cores and indivisibility, J. Math. Econ. 1 (1974) 23–28.