Games and Economic Behavior 91 (2015) 284–296

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

# Mix and match: A strategyproof mechanism for multi-hospital kidney exchange ☆

Itai Ashlagi [a,1], Felix Fischer [b,2], Ian A. Kash [c], Ariel D. Procaccia [d,*]

[a] Department of Operations Management, Sloan School of Management, Massachusetts Institute of Technology, United States
[b] Statistical Laboratory, University of Cambridge, United Kingdom
[c] Microsoft Research, Cambridge, United Kingdom
[d] Computer Science Department, Carnegie Mellon University, United States

ARTICLE INFO

*Article history:*
Received 15 October 2011
Available online 31 May 2013

*JEL classification:*
C72
C78
D47
D82

*Keywords:*
Approximate mechanisms without money
Kidney exchange

ABSTRACT

As kidney exchange programs are growing, manipulation by hospitals becomes more of an issue. Assuming that hospitals wish to maximize the number of their own patients who receive a kidney, they may have an incentive to withhold some of their incompatible donor–patient pairs and match them internally, thus harming social welfare. We study mechanisms for two-way exchanges that are strategyproof, i.e., make it a dominant strategy for hospitals to report all their incompatible pairs. We establish lower bounds on the welfare loss of strategyproof mechanisms, both deterministic and randomized, and propose a randomized mechanism that guarantees at least half of the maximum social welfare in the worst case. Simulations using realistic distributions for blood types and other parameters suggest that in practice our mechanism performs much closer to optimal.



## 1. Introduction

Transplantation of a healthy kidney is the best treatment today for severe kidney disease. Since humans normally have two kidneys and need only one to lead a healthy life, many patients have a family member or friend willing to donate them a kidney. However, not all potential donors are compatible with their desired recipient. This raises the possibility of *kidney exchange*, in which two or more incompatible donor–patient pairs exchange kidneys such that each patient receives a compatible kidney from the donor of another patient.[3]

Incentives of donor–patient pairs and efficiency in kidney exchange programs have respectively been studied by Roth et al. (2004, 2005, 2007a). As kidney exchange programs grow, however, manipulation by hospitals also becomes an issue. In particular, a hospital may choose to withhold some of its incompatible donor–patient pairs and match them internally, in order to maximize the number of its own patients who receive a kidney. This kind of strategic behavior has a negative

☆ We have benefited from valuable discussions with Moshe Tennenholtz and David Parkes. An earlier version of this paper appeared in the Proceedings of the 11th ACM Conference on Electronic Commerce.

* Corresponding author.
*E-mail addresses:* iashlagi@mit.edu (I. Ashlagi), fischerf@statslab.cam.ac.uk (F. Fischer), iankash@microsoft.com (I.A. Kash), arielpro@cs.cmu.edu (A.D. Procaccia).

[1] The author thanks the NSF for financial support.
[2] Support from the Deutsche Forschungsgemeinschaft under grant FI 1664/1-1 is gratefully acknowledged.

[3] These cyclic exchanges can also be combined with chains, starting with a deceased donor or an "undirected" donor without a particular intended recipient and ending with a patient who has a high priority on the deceased-donor waiting list or with a donor who will donate at some point in the future.

0899-8256/$ – see front matter 
http://dx.doi.org/10.1016/j.geb.2013.05.008

effect on social welfare and runs counter to the whole idea of having a large exchange. It is therefore an interesting question how hospitals can be incentivized to fully participate in an exchange by submitting all of their incompatible donor–patient pairs.

This problem can be modeled formally as a matching problem on a graph in which each vertex corresponds to an incompatible donor–patient pair and an edge between two such pairs indicates that the donor of each pair is compatible with the recipient of the respective other pair. Moreover disjoint sets of vertices are controlled by self-interested agents, in the sense that their existence is private information of the agent controlling them. Agents then reveal subsets of their vertices, and matches are determined based on the induced subgraph. An agent can seek to manipulate by hiding some of its vertices and then proceeding to benefit both from the inter-agent matches and matches on its hidden and unmatched vertices. We assume that each agent seeks to maximize the number of its own vertices that end up being matched.[4]

The above model was first used by Roth et al. (2007b) and later by Ashlagi and Roth (2011) in order to study the incentives of the hospitals in an exchange. Roth et al. observed that no efficient and strategyproof mechanisms exist for this problem. Ashlagi and Roth showed that no deterministic strategyproof mechanism can guarantee more than half the size of an efficient matching, whereas a nearly efficient incentive compatible mechanism exists in a Bayesian setting. A more detailed discussion of these results can be found in Section 2.

In this paper we take a prior-free approach to the nonexistence of efficient and strategyproof mechanisms and relax efficiency rather than strategyproofness. We say that a mechanism is an *$\alpha$-approximation* mechanism if the size of the maximum cardinality matching is always at most $\alpha$ times that of the matching returned by the mechanism.[5] Our goal is to design mechanisms that are strategyproof and at the same time provide a good approximation ratio. This approach is interesting for at least two reasons. First, strategyproof mechanisms are more robust in the worst case against information hospitals might have about each others' patients. Interestingly, we will see that their efficiency loss on average (in simulations) is still very small. Second, together with the results of Ashlagi and Roth (2011), our results provide insights into the tradeoff between different degrees of incentive compatibility on the one hand and social welfare on the other.

We begin in Section 4 by establishing lower bounds on the approximation ratio achievable by strategyproof mechanisms. To this end, we refine an example used by Roth et al. (2007b) to illustrate that no efficient mechanism can be strategyproof, and observe that no deterministic strategyproof mechanism can provide an approximation ratio better than 2 and no randomized strategyproof mechanism can provide an approximation ratio better than 8/7.[6]

In Section 5 we then introduce a mechanism, termed $\textsc{Match}_{\Pi}$, that is parameterized by a bipartition $\Pi = (\Pi_1, \Pi_2)$ of the set of agents. Roughly speaking, for any given graph, the mechanism returns a matching that has maximum cardinality among all the matchings that (i) contain no edges between the vertex sets of two agents on the same side of the bipartition, and (ii) are a maximum cardinality matching when restricted to the vertex set of each individual agent. We show that $\textsc{Match}_{\Pi}$ is strategyproof for any bipartition of the set of agents and can be executed in polynomial time. Unfortunately, for any fixed bipartition $\Pi$, $\textsc{Match}_{\Pi}$ does not generally provide a bounded approximation ratio. We observe, however, that $\textsc{Match}_{\Pi}$ yields a 2-approximation in the two-agent case when used with the obvious bipartition that places the two agents on opposite sides. This mechanism is in fact the optimal deterministic strategyproof mechanism for two agents, since the deterministic lower bound of 2 holds even in this case. In Section 6 we finally construct a randomized mechanism, termed Mix-and-Match, that first *mixes* the agents by choosing a random bipartition $\Pi$, then *matches* the vertices by applying $\textsc{Match}_{\Pi}$. We show that Mix-and-Match is strategyproof and provides a 2-approximation.

An average-case analysis of Mix-and-Match, using simulations with realistic values for parameters like the structure and frequency of blood types, is given in Section 7. These simulations suggest a practical performance that is much closer to optimal than the theoretical worst-case bounds. Section 8 concludes with a discussion of our results and possible directions for future work.

## 2. Related work

Most closely related to our work is that of Roth et al. (2007b), Ashlagi and Roth (2011), and Toulis and Parkes (2011), who consider mechanisms for multi-hospital kidney exchange. Following the negative result of Roth et al. (2007b), Ashlagi and Roth (2011) and Toulis and Parkes (2011) studied mechanisms in the Bayesian setting.

Ashlagi and Roth (2011) show that under reasonable prior information, there exists an individually rational mechanism that is $\epsilon$-Bayesian incentive compatible and almost efficient. Their analytical results are obtained for large markets and require a regularity condition that roughly means that hospitals are not too big. The mechanism of Ashlagi and Roth (2011) finds a maximum set of donor–patient pairs for each hospital that it can match internally, and finds a maximum matching in the graph that guarantees that this set of pairs will be matched (not necessarily to each other). This is made possible by the large market assumption and perfect-matching results for Erdös–Rényi graphs. Using the prior information about the population (blood types and tissue-type compatibilities) the authors further identify a set of nodes that should be

[4] This model more generally applies to settings where information about clients and potential trades among clients is partitioned among a set of agents. What distinguishes kidney exchanges from other such settings is the absence of monetary transfers: in most countries, payments in return for organs are both illegal and considered immoral, so we are interested in mechanisms without payments.

[5] Since the social welfare of a matching is exactly twice its cardinality, approximating the two is equivalent.

[6] The preliminary version of this paper incorrectly stated the bound as 4/3. Ashlagi and Roth (2011) show this result in a slightly different setting.

given an "extra" chance in the match in order to achieve Bayesian incentive compatibility. In their work they consider also three-way exchanges. Toulis and Parkes (2011) establish similar results in the Bayesian setting assuming that each hospital is sufficiently large.[7]

By contrast, our mechanism is strategyproof and does not require any assumptions about the structure of the market. Nevertheless, to complement results about the Bayesian setting and make use of information about the population, we also study the performance of our mechanism on inputs drawn from the distribution used by Ashlagi and Roth. On a technical level, our mechanism ensures that the outcome contains a maximum cardinality internal matching (though not a specific one) for each hospital, and imposes some restrictions on exchanges across hospitals. Mechanisms currently in use also give a somewhat higher priority to exchanges among donors and patients of the same hospital. This is done mainly to minimize geographical distance between donors and recipients, and is usually not enough to incentivize hospitals to fully reveal their information to the mechanism (also see Ashlagi and Roth, 2011).

While it is possible to exchange kidneys among more than two donor–patient pairs at a time, finding an efficient set of exchanges becomes computationally hard in this case. In this paper, we therefore restrict our attention to two-way exchanges. We note, however, that there exist algorithms that allow multi-way exchanges and have good performance in practice (Abraham et al., 2007; Biró et al., 2009).

Kidney exchange has also been studied in dynamic environments where patients arrive and depart, but not from the perspective of incentives. Ünver (2010) provides an elegant characterization of optimality under the assumption that there are no tissue-type incompatibilities. Awasthi and Sandholm (2009) and Dickerson et al. (2012) design and analyze stochastic optimization algorithms for a dynamic environment.

Finally, our work is part of a line of research that seeks to approximate optimal outcomes in mechanism design settings without monetary transfers, which was initiated by Procaccia and Tennenholtz (2009). This approach is particularly intriguing in the context of problems that are computationally feasible: while there is no need to approximate the optimal solution for strictly computational reasons, there might be a need for that to maintain strategyproofness (when the optimal solution is not strategyproof).

## 3. Preliminaries

Let $N = \{1, \ldots, n\}$ be a set of agents. For each $i \in N$, let $V_i$ be a set of private vertices of agent $i$. Let $G = (V, E)$ with $V = \bigcup_{i \in N} V_i$ be an undirected labeled graph, that is, each vertex is labeled by its agent. We slightly abuse terminology by simply referring to such labeled graphs as "graphs."

A *matching* $M \subseteq E$ on $G$ is a subset of edges such that each vertex is incident to at most one edge of $M$. For $i, j \in N$ we denote

$$M_{ij} = \big\{(u, v) \in M\colon\ u \in V_i \wedge v \in V_j\big\}.$$

Given $i \in N$, we refer to edges in $M_{ii}$ as *internal edges* and to edges in $M_{ij}$, where $j \in N \setminus \{i\}$, as *external edges*.

Given a graph $G$ and a matching $M$ on $G$, the utility of agent $i$ for this matching is

$$u_i(M) = \big|\big\{u \in V_i\colon\ \exists v \in V \text{ s.t. } (u, v) \in M\big\}\big|,$$

that is, it is equal to the number of vertices of $V_i$ that are matched under $M$.

We now turn to the definition of a mechanism, without being too formal. For a fixed number $n$ of agents, a *deterministic mechanism* is a function that maps any (labeled) graph for $n$ agents to a matching of this graph. A *randomized mechanism* maps any graph to a probability distribution over matchings, that is, it can select a matching randomly. For conciseness, we treat deterministic mechanisms as a special case of randomized mechanisms in the rest of this section.

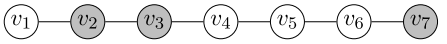
For a randomized mechanism $f$ and a (possibly random) graph $G$, define

$$u_i\big(f(G)\big) = \mathbb{E}_{M \sim f(G)}\big[u_i(M)\big],$$

where the expectation is taken over the distribution on matchings returned by the mechanism. In other words, the utility of an agent simply equals the expected number of its vertices being matched.

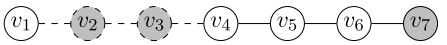
We are concerned with situations where an agent "hides" a subset of its vertices and then internally matches them among themselves or with vertices not matched by the mechanism. To make this formal we need some notation. We however feel that the idea is rather intuitive, and will avoid the rather cumbersome formalism in the rest of the paper. For any subset $V' \subseteq V$, let $G[V']$ be the subgraph of $G$ induced by $V'$. For a graph $G$, an agent $i \in N$, and a matching $M$, let $X_i(M)$ be the set of vertices in $V_i$ that are not matched in $M$; if $M$ is chosen randomly, then $X_i(M)$ is a random variable. Furthermore, let $f^*$ be a mechanism that maps each graph $G$ to a maximum cardinality matching of $G$. We say

[7] Both Ashlagi and Roth and Toulis and Parkes use realistic values for parameters like the structure and frequency of blood types.




(a) Graph $G$. Vertices of $V_1$ are shown in white, vertices of $V_2$ in gray.


(b) Graph $G'$ obtained from $G$ when agent 1 hides the vertices $v_5$ and $v_6$. The hidden vertices are not part of the graph, but are shown to give the complete picture.




(c) Graph $G''$ obtained from $G$ when agent 2 hides vertices $v_2$ and $v_3$.

**Fig. 1.** Construction used in the proof of Theorem 4.1.

that a mechanism $f$ is *strategyproof* if for every graph $G = (V, E)$ with $V = \bigcup_{i \in N} V_i$, for every $i \in N$, and for every $V_i' \subseteq V_i$ it holds that

$$u_i\big(f(G)\big) \geqslant u_i\big(f\big(G\big[V \setminus V_i'\big]\big)\big) + u_i\big(f^*\big(G\big[V_i' \cup X_i\big(f\big(G\big[V \setminus V_i'\big]\big)\big)\big]\big)\big).$$

In words, a mechanism is strategyproof if an agent can never benefit by hiding some of its vertices. The agent's utility after hiding a subset $V_i'$ of its vertices equals the (expected) number of its vertices that the mechanism matches given the subgraph induced by all vertices but those in $V_i'$, plus the (expected) number of vertices in a maximum cardinality matching of the subgraph induced by $V_i'$ and the vertices not matched by the mechanism. In our model, individual rationality requires that an agent cannot benefit from the special case when $V_i' = V_i$, and is therefore implied by strategyproofness.

We are interested in mechanisms that, while being strategyproof, produce matchings that maximize social welfare, i.e., the sum of agent utilities. For any matching $M$, $\sum_{i \in N} u_i(M) = 2|M|$, so what we are looking for are matchings that are as large as possible. We say that a randomized mechanism $f$ provides an $\alpha$-approximation if for every graph $G$,

$$\frac{|f^*(G)|}{\mathbb{E}[|f(G)|]} \leqslant \alpha, \tag{1}$$

where once again $f^*(G)$ is a maximum cardinality matching of $G$. For deterministic mechanisms, the expectation in (1) can simply be dropped.

## 4. Lower bounds

It may not be immediately apparent that the optimal mechanism is not strategyproof. Given a graph, the optimal mechanism simply returns a maximum cardinality matching (while employing a consistent tie-breaking rule to decide between different maximum cardinality matchings).

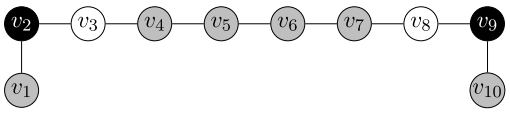
To see how this can fail to be strategyproof, consider graph $G$ in Fig. 1(a). This graph has an odd number of vertices, so every matching leaves some vertex unmatched. However, each agent has a pair of vertices such that removing these vertices from the graph results in a graph with a unique maximum cardinality matching in which all of that agent's vertices are matched (Figs. 1(b) and 1(c)). Thus, one of the agents must have an unmatched vertex in $G$, and this agent can hide two of his vertices to increase his utility. This simple example, which is due to Roth et al. (2007b), can be used to derive lower bounds that will later turn out to be, at least in one case, tight (see also Ashlagi and Roth, 2011, for similar bounds in a slightly different setting).

**Theorem 4.1.** *If there are at least two agents,*

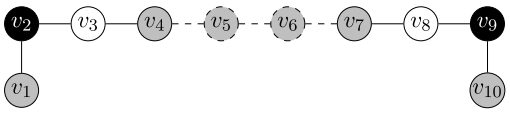
1. *no deterministic strategyproof mechanism can provide an $\alpha$-approximation for $\alpha < 2$, and*
2. *no randomized strategyproof mechanism can provide an $\alpha$-approximation for $\alpha < 8/7$.*

**Proof.** For the first part of the theorem, we consider the case where $N = \{1, 2\}$; the proof can easily be extended to the case where $n > 2$ by adding agents with vertices that are not incident to any edges. Let $f$ be a deterministic mechanism, and consider graph $G$ given in Fig. 1(a). Since $G$ has an odd number of vertices, it does not have a perfect matching, and so $f(G)$ must leave some $v \in V_1$ or some $v \in V_2$ unmatched. Thus, either $u_1(f(G)) \leqslant 3$ or $u_2(f(G)) \leqslant 2$.

We first deal with the case where $u_1(f(G)) \leqslant 3$. Consider the graph $G'$ that is obtained when agent 1 hides vertices $v_5$ and $v_6$ (see Fig. 1(b)). The unique maximum cardinality matching of this graph is $\{(v_1, v_2), (v_3, v_4)\}$, a matching of cardinality 2. However, agent 1 could internally match the pair $(v_5, v_6)$ and obtain a utility of 4, contradicting strategyproofness. Therefore, $f(G')$ must have cardinality at most 1, meaning that its approximation ratio on $G'$ cannot be smaller than 2.




(a) The original graph $G$, where the vertices of $V_1$ are white, the vertices of $V_2$ are gray, and the vertices of $V_3$ are black.




(b) The graph $G'$, agent 2 hides vertices $v_5$ and $v_6$.

**Fig. 2.** The naïve three-agent mechanism is not strategyproof.

The case where $u_2(f(G)) \leqslant 2$ can be handled similarly. Consider the graph $G''$ obtained when agent 2 hides vertices $v_2$ and $v_3$ (see Fig. 1(c)). Once again there is a unique maximum matching of cardinality 2, but $f$ cannot return this matching since it would yield a utility of 3 to agent 2, in contradiction to strategyproofness. As before the mechanism is forced to select a matching of cardinality at most 1.

The second part of the theorem can be derived using the same construction. Let $f$ be a randomized strategyproof mechanism. Since $G$ does not have a perfect matching, it must be that $u_1(f(G)) + u_2(f(G)) \leqslant 6$. Therefore, either $u_1(f(G)) \leqslant 7/2$ or $u_2(f(G)) \leqslant 5/2$.

We now proceed as before. If $u_1(f(G)) \leqslant 7/2$, we consider graph $G'$; by strategyproofness $f$ can only match both of agent 1's pairs with probability at most $3/4$, for a maximum of $7/4$ pairs in expectation, but the optimum is 2. If $u_2(f(G)) \leqslant 5/2$, we use graph $G''$ to show that $f$ can only match $7/4$ pairs in expectation, while the optimum is 2. □

## 5. Deterministic mechanisms

Let us now focus on deterministic mechanisms. We begin by designing a deterministic mechanism that is strategyproof for any number of agents, but may not provide a bounded approximation ratio. We then leverage this mechanism to obtain an optimal deterministic strategyproof mechanism for two agents. The more powerful application of our deterministic mechanism will only appear in the next section, when we discuss randomized mechanisms.

Let us first address the issue of designing strategyproof deterministic mechanisms without worrying, for now, about approximate optimality or computational tractability. Consider the following mechanism for two agents. Given a graph $G$, the mechanism computes the set of all matchings on $G$ that have maximum cardinality on $V_1$ and $V_2$, and among these selects a matching with maximum overall cardinality. Since every matching that this mechanism considers has maximum cardinality on $V_1$ and $V_2$, it clearly is individually rational. We will show momentarily that it is also strategyproof.

But let us first consider what this mechanism does when applied to the graph of Fig. 1(a). Any matching that is a maximum cardinality matching on $V_2$ would have to match $(v_2, v_3)$, and there are two maximum cardinality matchings on $V_1$: one can either match $(v_4, v_5)$ or $(v_5, v_6)$. If we match $(v_5, v_6)$, no additional edges can be added. Hence, the unique matching of cardinality 3 that maximizes the number of internal edges is $\{(v_2, v_3), (v_4, v_5), (v_6, v_7)\}$. The only unmatched vertex in this matching is $v_1$. With the proof of Theorem 4.1 in mind, let us verify that agent 1 cannot benefit by hiding $v_5$ and $v_6$. Given graph $G'$ in Fig. 1(b), the mechanism would simply return the matching $(v_2, v_3)$, since this is the unique matching that is a maximum cardinality matching on $V_2$.

The two-agent mechanism suggested above seems promising from the perspective of strategyproofness. Let us extend it to an $n$-agent mechanism in the natural way, and consider the mechanism that selects a matching of maximum cardinality among the matchings that have maximum cardinality on each $V_i$, $i = 1, \ldots, n$. In addition, let us break ties *serially*: among all the matchings that meet the above criteria, we select a matching that maximizes the utility of agent 1; if there are several such matchings, we choose one that maximizes the utility of agent 2, and so on.

Interestingly enough, this $n$-agent mechanism is not strategyproof, even when $n = 3$. Consider graph $G$ given in Fig. 2(a). Any matching that has maximum cardinality on $V_2$ must match $(v_4, v_5)$ and $(v_6, v_7)$; by the tie-breaking rule the mechanism then returns the matching $\{(v_2, v_3), (v_4, v_5), (v_6, v_7), (v_8, v_9)\}$. When agent 2 hides $v_5$ and $v_6$ we obtain graph $G'$ given in Fig. 2(b). On this graph the mechanism returns a perfect matching $\{(v_1, v_2), (v_3, v_4), (v_7, v_8), (v_9, v_{10})\}$. After internally matching $(v_5, v_6)$ agent 1 gains two additional matched vertices compared to the matching on $G$. Clearly this example can be modified to work if ties are broken in a different order.

The deeper reason why the above mechanism fails to be strategyproof is rather subtle, and has to do with the following observation: if one takes the union of the matchings generated on the graphs of Figs. 2(a) and 2(b), and contracts each $V_i$ to one vertex, one obtains an odd-length cycle among $V_1$, $V_2$, and $V_3$, as the matching on $G$ has an edge between $V_1$ and $V_3$, and the matching on $G'$ has edges between $V_1$ and $V_2$, and $V_2$ and $V_3$. We proceed to refine the above mechanism in order to avoid such odd cycles; this turns out to be sufficient to guarantee strategyproofness. The following is in fact a family of mechanisms, parameterized by a fixed bipartition $\Pi = (\Pi_1, \Pi_2)$ of the set of agents.

MATCH$_{\Pi}$

1. Given a graph $G$, consider all the matchings that have maximum cardinality on each $V_i$ and do not have any edges between $V_i$ and $V_j$ when $i, j \in \Pi_l$ for some $l \in \{1, 2\}$, i.e., those that maximize the number of internal edges and do not have any edges between sets on the same side of the bipartition.
2. Among these matchings select one of maximum cardinality, breaking ties serially in favor of agents in $\Pi_1$ and then agents in $\Pi_2$.

By letting $N = \{1, 2\}$, $\Pi_1 = \{1\}$, and $\Pi_2 = \{2\}$, we obtain the two-agent mechanism described above. The naïve generalization of this mechanism to three agents, on the other hand, is not an instance of MATCH$_{\Pi}$: for the example of Fig. 2 showing that the mechanism is not strategyproof, the sets $M_{12}$, $M_{13}$, and $M_{23}$ are all non-empty.

We proceed to show that MATCH$_{\Pi}$ is strategyproof for any bipartition of the set of agents. The main idea behind the proof of this theorem is again rather subtle. It relies on the fact that if one takes the union of the two matchings produced by the mechanism before and after an agent hides some of its vertices, then this union cannot contain a cycle that visits the vertex sets of an odd number of agents. This property holds because the mechanism does not match vertices of agents on the same side of the bipartition.

**Theorem 5.1.** *For any number of agents, and for any bipartition $\Pi$ of the set of agents,* MATCH$_{\Pi}$ *is strategyproof.*

**Proof.** Fix some bipartition $\Pi = (\Pi_1, \Pi_2)$ of $N$. Consider a graph $G$, and let $M = \text{MATCH}_{\Pi}(G)$. Assume that agent $i \in N$ hides a subset of vertices, inducing a subgraph $G'$, and let $M'$ be the matching that results from applying the mechanism to $G'$, along with the internal matching of agent 1 on its hidden and unmatched vertices, that is,

$$M' = \text{MATCH}_{\Pi}\left(G'\right) \cup \hat{M},$$

where $\hat{M}$ is a maximum cardinality matching of agent $i$ on its hidden and unmatched vertices.

The symmetric difference

$$M \Delta M' = M \cup M' \setminus \left(M \cap M'\right)$$

then consists of vertex-disjoint paths (some of which may be cycles) with alternating edges of $M$ and $M'$. For example, consider the two-agent version of MATCH$_{\Pi}$ applied to graphs $G$ and $G'$ given in Fig. 1(a) and Fig. 2(a). It holds that

$$M = \text{MATCH}_{\Pi}(G) = \left\{(v_2, v_3), (v_4, v_5), (v_6, v_7)\right\},$$

whereas, say, $M' = \{(v_2, v_3), (v_5, v_6)\}$. Then, $M \Delta M'$ is the single path $\{(v_4, v_5), (v_5, v_6), (v_6, v_7)\}$ where the first and last edge are in $M$ and the middle edge is in $M'$.

In order to simplify notation, we henceforth assume that $M \Delta M'$ consists of just one path. This assumption is made without loss of generality, because we show that *each* such path satisfies one of the following properties: either $M$ matches at least as many vertices of $V_i$ as $M'$ for every $i \in N$, or one can derive a contradiction to the way $M$ or $M'$ were selected by switching between some (or all) of their edges on the path. Since the contradiction can be derived for each path *separately*, it follows that the first property holds on every path, that is, the overall utility of agent $i$ for $M$ is at least as large as its utility for $M'$.

If the path in $M \Delta M'$ is a cycle, then this cycle must be of even length, because otherwise there would be a vertex that is incident to two edges of the same matching. It follows that both $M$ and $M'$ match all the vertices on the cycle, hence agent $i$ is indifferent between the two matchings. We may therefore assume that $M \Delta M'$ is not a cycle.

It will prove useful to arbitrarily fix a direction over the (undirected) edges of the single path in $M \Delta M'$. Since the path is not a cycle, this direction pinpoints two specific vertices as the start and the end of the path. We further say that the (directed) edge $(u, v)$ *enters* $V_j$ if $u \notin V_j$ and $v \in V_j$, and *exits* $V_j$ if $u \in V_j$ and $v \notin V_j$.

We consider two cases.

*Case 1:* $|M_{ii}| > |M'_{ii}|$. We claim that

$$\sum_{j \in N \setminus \{i\}} |M_{ij}| \geqslant \sum_{j \in N \setminus \{i\}} \left|M'_{ij}\right| - 2. \tag{2}$$

Since both $M$ and $M'$ are maximum cardinality matchings on $V_j$ for all $j \neq i$, it must hold that every subpath of $M \Delta M'$ on $V_j$ has even length (see Fig. 3); otherwise we would have, say, more edges of $M$ than $M'$ on the subpath, and by switching from $M'$ to $M$ on the subpath we would be able to increase the size of $M'$ on $V_j$. This implies that for any $j \in N \setminus \{i\}$, any subpath entering $V_j$ with an edge of $M'$ must exit $V_j$ with an edge of $M$, and any subpath entering $V_j$ with and edge of $M$ must exit $V_j$ with an edge of $M'$.

The next part of the proof is crucial, and uses the main idea behind mechanism MATCH$_{\Pi}$. We argue that it also holds that when the path exits $V_i$ with an edge of $M'$ it can only enter $V_i$ again, the first time after exiting, with an edge of $M$.

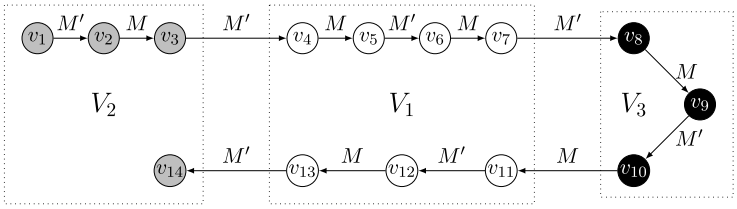



**Fig. 3.** Illustration of Case 1 of the proof of Theorem 5.1, with $i = 1$ as the manipulator, and $\Pi = (\{1\}, \{2, 3\})$. $M \Delta M'$ is shown as a single directed path with alternating edges of $M$ and $M'$. It holds that $3 = |M_{11}| > |M'_{11}| = 2$. Every subpath inside $V_2$ and $V_3$ has even length (those from $v_1$ to $v_3$ and from $v_8$ to $v_{10}$), but subpaths inside $V_1$ may not have (like that from $v_4$ and $v_7$). The subpath of $M \Delta M' \setminus (M_{11} \cup M'_{11})$ from $v_1$ to $v_4$ enters $V_1$ but does not exit it, while the subpath from $v_{13}$ to $v_{14}$ exits $V_1$ but does not enter it. This example satisfies (2) with equality.

Assume without loss of generality that $i \in \Pi_1$. By the above argument the subpath that exits $V_i$ immediately enters $V_{j_1}$, for some $j_1 \in \Pi_2$, with an edge of $M'$, and therefore next exits it with an edge of $M$, thus entering $V_{j_2}$ for some $j_2 \in \Pi_1$. If $j_2 \neq i$, and the subpath exits $V_{j_2}$, then it does so with an edge of $M'$, and by the same arguments returns to the vertex set of an agent in $\Pi_1$ with an edge of $M$. If eventually the subpath enters $V_i$ again, it must be with an edge of $M$. Analogously, if the subpath exits $V_i$ with an edge of $M$, it can only enter $V_i$ with an edge of $M'$. See Fig. 3 for an illustration.

Now consider $(M \Delta M') \setminus (M_{ii} \cup M'_{ii})$, which again is a collection of vertex-disjoint subpaths. Some start and end in $V_i$, and it follows by the discussion above that such subpaths have exactly one edge in $M_{ij}$ and one edge in $M'_{ik}$, for $k, j \in N \setminus \{i\}$. There can only be one subpath that starts in $V_i$ but does not end in $V_i$, and at most one subpath that ends in $V_i$ but does not start in $V_i$. Eq. (2) directly follows.

We now have that

$$\begin{aligned} u_i(M) &= 2|M_{ii}| + \sum_{j \in N \setminus \{i\}} |M_{ij}| \\ &\geqslant 2\left(\left|M'_{ii}\right| + 1\right) + \left( \sum_{j \in N \setminus \{i\}} \left|M'_{ij}\right| - 2 \right) = u_i\left(M'\right), \end{aligned}$$

where the inequality follows from the fact that $|M_{ii}| > |M'_{ii}|$ and from (2).

*Case 2:* $|M'_{ii}| = |M_{ii}|$. Note that it holds that $|M_{jj}| = |M'_{jj}|$ for all $j \in N$, that is, $M \Delta M'$ has to be of even length inside every $V_j$. This includes $M_{ii}$ and $M'_{ii}$, because the total number of internal edges for $i$ is even. If some subpath of $i$'s internal edges has odd length with more edges from $M$ there must be another subpath with more internal edges from $M'$. Swapping the edges of $M$ for those of $M'$ in the second subpath results in a matching $M''$ such that $|M''_{ii}| > |M_{ii}|$, contradicting the construction of $M$ to have maximum cardinality on each $V_i$. It follows that $|M| \geqslant |M'|$, since $M$ is a maximum cardinality matching under the constraint that it has maximum cardinality inside each $V_i$.

We claim that if $|M| > |M'|$ then $\sum_j |M_{ij}| \geqslant \sum_j |M'_{ij}|$. Together with the assumption that $|M'_{ii}| = |M_{ii}|$ this implies that agent $i$ cannot benefit. Indeed, in this case $M \Delta M'$ is a path of odd length that starts and ends with an edge of $M$. Recall that every subpath of $M \Delta M'$ consisting of $i$'s internal edges has even length. This means that when the path enters $V_i$ with an edge of $M'$ it cannot end inside $V_i$, as otherwise it would end with an edge of $M'$. In other words, every time the path enters $V_i$ with an edge of $M'$ it must exit $V_i$ with an edge of $M$. Similarly, every time the path exits $V_i$ with an edge of $M'$ it must have entered $V_i$ with an edge of $M$, otherwise the path must start in $V_i$ with an edge of $M'$. This proves our claim, so we can assume that $|M| = |M'|$.

Suppose that $|M| = |M'|$. Therefore $M \Delta M'$ has even length, and moreover we know it has even length inside each $V_k$. Note that all the vertices on the path are matched under both $M$ and $M'$, except for the start and the end vertices. Hence, if agent $i$ gains from the manipulation, it must be the case (when fixing a specific direction on the edges) that the start vertex is a vertex of $V_i$ and the first edge is an edge of $M'$, whereas the end vertex is in $V_j$, for some $j \in N \setminus \{i\}$, and the last edge is an edge of $M$.

Now, if tie-breaking favors $i$ over $j$, then by switching the edges of $M$ with those of $M'$ we get a matching of equal size that has maximum cardinality on each $V_k$ and is better for $i$, in contradiction to the tie-breaking rule. We will therefore assume that tie-breaking favors $j$ over $i$. Consider the subpath $\rho$ of $M \Delta M'$ that starts with the last edge that exits $V_i$ and ends with the last edge in $M \Delta M'$. We argue that $\rho$ must start with an edge of $M'$. To see why, note that $M \Delta M'$ starts in $V_i$ with an edge of $M'$. Since the subpaths of $M \Delta M'$ in $V_i$ have even length, it exits with an edge of $M'$. By the same argument as in Case 1, the bipartition ensures that, if $M \Delta M'$ re-enters $V_i$, it does so with an edge from $M$. Since all subpaths of $M \Delta M'$ in $V_i$ are of even length, the path always exits $V_i$ with an edge of $M'$.

By replacing all the edges of $M'$ with the edges of $M$ on $\rho$, we obtain a matching $M''$ that is identical to $M'$ inside $V_i$, has maximum cardinality on $V_k$ for each $k \in N$, is as large as $M'$ overall, and satisfies $u_j(M'') = u_j(M') + 1$, $u_i(M'') = u_i(M') - 1$, and $u_k(M'') = u_k(M')$ for all $k \in N \setminus \{i, j\}$. This is a contradiction, since $M''$ should have been chosen over $M'$ due to the way the mechanism breaks ties. See Fig. 4 for an illustration. □

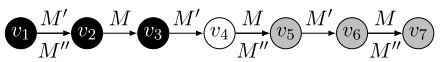



**Fig. 4.** An illustration of the last argument in Case 2 of the proof of Theorem 5.1 with $i = 3$ and $j = 2$. The vertices of $V_1$ are white, the vertices of $V_2$ are gray, and the vertices of $V_3$ are black. By switching from $M'$ to $M''$ we increase the utility of agent 2 and decrease the utility of agent 3, thereby obtaining a legal matching that contradicts the choice of $M'$.

We next show that MATCH$_{\Pi}$ can be executed in polynomial time by a reduction to the maximum weighted matching problem (for a polynomial time algorithm for the latter see Gabow, 1990).

**Theorem 5.2.** MATCH$_{\Pi}$ *can be executed in polynomial time.*

**Proof.** Assume without loss of generality that $|E| > 1$, and let $\epsilon_i = 1/|E|^{i+1}$. We assign weights to edges as follows. An (internal) edge $(u, v)$ such that $u, v \in V_i$ for some $i \in N$ receives weight $|E| + 3$. An (external) edge $(u, v)$ such that $u \in V_i$ and $v \in V_j$ with $i \in \Pi_1$ and $j \in \Pi_2$ receives weight $1 + \epsilon_i + \epsilon_j/|E|^{n+1}$. An (external) edge $(u, v)$ such that $u \in V_i$ and $v \in V_j$ with $i \neq j$ but $i, j \in \Pi_1$ or $i, j \in \Pi_2$ receives weight 0.

The sum of the weights of all external edges is at most $|E|(1 + 1/|E|^2 + 1/|E|^{n+3}) < |E| + 3$, which is less than the weight of a single internal edge. Thus a maximum weight matching of this graph maximizes the number of internal edges. All edges between sets on the same side of the bipartition have weight zero, so no such edges will be included.

To complete the proof we need to verify that the maximum weight matching has maximum cardinality among those with a maximum number of internal edges and no edges across the bipartition, and that ties are broken appropriately. Each edge across the bipartition has weight at least 1 and at most $1 + 1/|E|^2 + 1/|E|^{n+3}$. Thus, given two matchings $M$ and $M'$ satisfying the above constraints such that $|M| > |M'|$, the difference in their weights is at least

$$\begin{aligned} 1 - \left|M'\right|\left(1/|E|^2 + 1/|E|^{n+3}\right) &\geqslant 1 - |E|\left(1/|E|^2 + 1/|E|^{n+3}\right) \\ &= 1 - 1/|E| - 1/|E|^{n+2} > 0. \end{aligned}$$

The maximum weight matching thus has maximum cardinality subject to the constraints. For tie-breaking, observe that $\epsilon_i \geqslant |E|\epsilon_j$ if $i < j$, meaning that among agents on the same side of the bipartition those with smaller indices have higher priority. The factor of $1/|E|^{n+1}$ finally ensures that agents in $\Pi_1$ have priority over agents in $\Pi_2$. □

Recall that by Theorem 4.1 no deterministic strategyproof mechanism can have an approximation ratio smaller than 2, even when there are only two agents. We will see momentarily that MATCH$_{\Pi}$ provides an approximation ratio of 2 when $N = \{1, 2\}$ and $\Pi = (\{1\}, \{2\})$, i.e., it is the best possible deterministic strategyproof mechanism for the case of two agents. Indeed, consider a graph $G$, let $M^*$ be an optimal matching of $G$, and $M$ the matching returned by MATCH$_{(\{1\},\{2\})}$. $M$ is inclusion-maximal. Therefore, for every $(u, v) \in M^*$, either $u$ is matched by $M$ or $v$ is matched by $M$. We conclude that $|M| \geqslant |M^*|/2$. Strategyproofness is obtained from Theorem 5.1.

**Corollary 5.3.** *Let* $N = \{1, 2\}$. *Then,* MATCH$_{(\{1\},\{2\})}$ *is strategyproof and provides a* 2*-approximation.*

Unfortunately, when $n \geqslant 3$, MATCH$_{\Pi}$ does not provide a finite approximation ratio for any fixed bipartition. To see this, let $\Pi = (\Pi_1, \Pi_2)$ be a bipartition of the set of agents. Then there must be two distinct agents $i, j \in N$ such that $i, j \in \Pi_l$ for some $l \in \{1, 2\}$. Now consider a graph where the only edge is an external edge between $V_i$ and $V_j$; given this graph MATCH$_{\Pi}$ returns an empty matching, whereas the optimum is a matching of cardinality 1.

## 6. Randomized mechanisms

We have seen above that MATCH$_{\Pi}$ does not provide a bounded approximation ratio for any fixed bipartition $\Pi$. The natural next step is to choose the bipartition uniformly at random. This leads to the eponymous MIX-AND-MATCH mechanism.

MIX-AND-MATCH

1. **Mix:** Construct a random bipartition $\Pi = (\Pi_1, \Pi_2)$ of the agents by independently flipping a fair coin for each agent to determine whether the agent is in $\Pi_1$ or in $\Pi_2$.
2. **Match:** Apply MATCH$_{\Pi}$ to the given graph, where $\Pi$ is the bipartition constructed in Step 1.

It immediately follows from Theorem 5.1 that MIX-AND-MATCH is strategyproof, and in fact in a stronger sense than the one defined in Section 3, namely *universal strategyproofness*. A randomized mechanism is called universally strategyproof if agents cannot gain by lying regardless of the random choices made by the mechanism, i.e., if the mechanism is a distribution over strategyproof deterministic mechanisms.

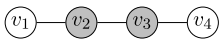



**Fig. 5.** Graph illustrating that Mix-and-Match cannot provide an approximation ratio smaller than two. $V_1$ is shown in white, $V_2$ is shown in gray. Mix-and-Match returns the matching $(v_2, v_3)$.

A naïve analysis of Mix-and-Match would yield a rather unimpressive approximation ratio. Indeed, the reason why $\textsc{Match}_{(\{1\},\{2\})}$ does not provide a better approximation ratio than two is that it may have to sacrifice two external edges for one internal edge. The fact that Mix-and-Match will not be able to match many of the edges in the graph because they are not between the two elements of the constructed bipartition would seem to cause the approximation ratio to deteriorate further. Fortunately, these two problems effectively cancel out: sacrificing two external edges for an internal edge is less of a problem when each of those external edges is allowed to be part of the matching for only half of the bipartitions. Formally, we prove the following result.

**Theorem 6.1.** *For any number of agents,* Mix-and-Match *is* (*universally*) *strategyproof and provides a* 2*-approximation.*

**Proof.** We prove the theorem by taking a maximum cardinality matching $M^*$ and constructing a matching $M'$ that, when restricted to a random bipartition $\Pi$ (by removing edges between agents on the same side of the bipartition), has at least half the size of $M^*$ in expectation. We then show that the matching produced by $\textsc{Match}_{\Pi}$ is always at least as large as $M'$ restricted to $\Pi$.

Consider a graph $G$, and let $M^*$ be a maximum cardinality matching of $G$. For each $i \in N$, let $M_i^{**}$ be a maximum cardinality matching on $V_i$, and let $M^{**} = \bigcup_{i\in N} M_i^{**}$.

We construct a matching $M'$ as follows. Consider the symmetric difference $M^* \Delta M^{**}$ which, as in Theorem 5.1, consists of a set of paths with alternating edges of $M^*$ and $M^{**}$. For each alternating path in the symmetric difference, if there are more internal edges among the edges from $M^{**}$, we add these edges to $M'$. Otherwise, we add the edges from $M^*$ to $M'$ (note that this is where external edges can be added to $M'$).

Since $M^{**}$ has maximum cardinality on each $V_j$ and $M'$ has at least as many internal edges from each path as $M^{**}$, $M'$ has maximum cardinality on each $V_j$. Furthermore, since $M^*$ is a maximum cardinality matching, each path has either the same number of edges from $M^*$ and $M^{**}$ or one extra edge from $M^*$. In any given path, all external edges are from $M^*$, so if the edges from $M^{**}$ on the path are chosen to be in $M'$ then the number of internal edges gained relative to $M^*$ is at least the number of external edges lost minus one. In the worst case $M'$ has two fewer external edges for each extra internal edge relative to $M^*$. Thus $M'$ satisfies

$$\sum_{i\in N}\left(\left|M'_{ii}\right| - \left|M^*_{ii}\right|\right) \geqslant \frac{1}{2}\sum_{i\in N}\sum_{j>i}\left(\left|M^*_{ij}\right| - \left|M'_{ij}\right|\right),$$

where we sum over $j > i$ so as not to count the same edges twice. Rearranging, we get

$$\sum_{i\in N}\left|M'_{ii}\right| + \frac{1}{2}\sum_{i\in N}\sum_{j>i}\left|M'_{ij}\right| \geqslant \sum_{i\in N}\left|M^*_{ii}\right| + \frac{1}{2}\sum_{i\in N}\sum_{j>i}\left|M^*_{ij}\right|. \tag{3}$$

Now let $M^{\Pi}$ be the matching produced by $\textsc{Match}_{\Pi}$ for the fixed bipartition $\Pi$. Since $M^{\Pi}$ has maximum cardinality under the constraints, we have

$$\left|M^{\Pi}\right| = \sum_{i\in N}\left|M^{\Pi}_{ii}\right| + \sum_{i\in\Pi_1}\sum_{j\in\Pi_2}\left|M^{\Pi}_{ij}\right| \geqslant \sum_{i\in N}\left|M'_{ii}\right| + \sum_{i\in\Pi_1}\sum_{j\in\Pi_2}\left|M'_{ij}\right|.$$

Since each pair of agents appears on opposite sides in exactly half of the bipartitions, the expected size of the matching produced by Mix-and-Match is

$$\begin{aligned}\sum_{\Pi}\left(\frac{1}{2^n}\cdot\left|M^{\Pi}\right|\right) &\geqslant \sum_{i\in N}\left|M'_{ii}\right| + \frac{1}{2}\sum_{i\in N}\sum_{j>i}\left|M'_{ij}\right| \\ &\geqslant \sum_{i\in N}\left|M^*_{ii}\right| + \frac{1}{2}\sum_{i\in N}\sum_{j>i}\left|M^*_{ij}\right| \geqslant \frac{1}{2}\cdot\left|M^*\right|,\end{aligned}$$

where the second inequality follows from (3). □

The graph in Fig. 5 shows that the analysis of Mix-and-Match is tight even for $n = 2$. Still one might hope to do better, given that Theorem 4.1 only provides a randomized lower bound of 8/7, and indeed Caragiannis et al. (2011) recently were able to provide an upper bound of 3/2 for the case where $n = 2$ using the following mechanism.

WEIGHT-AND-MATCH

1. Given a graph $G$, assign a weight of 1 to internal edges and a weight of $1/2$ to external edges.
2. Flip a fair coin.
3. If the outcome is heads, return a maximum cardinality matching among all maximum weight matchings.
4. If the outcome is tails, return a minimum cardinality matching among all maximum weight matchings.

Despite the improvement over MIX-AND-MATCH for the case of two agents, this mechanism still leaves a small gap between the randomized upper and lower bounds.

## 7. Average-case performance

In the worst case, MIX-AND-MATCH achieves only half of the maximum social welfare, and in fact no strategyproof mechanism can do much better. It is therefore natural to ask how MIX-AND-MATCH performs in practice, where the occurrence of a worst-case instance may be very unlikely. To answer this question, we simulate the practical performance of MIX-AND-MATCH using incompatible donor–patient pairs drawn at random according to realistic parameters, and compare it to the optimal outcome without manipulation as well as to the outcome obtained when hospitals match their donor–patient pairs internally and reveal only the unmatched ones.

The comparison is done for a varying number $n$ of hospitals, each with $k$ patients.[8] For given values of $n$ and $k$, we generate 400 graphs, each of which is obtained by generating patients and donors until the desired number $nk$ of incompatible pairs is reached. Each patient is assigned a blood type and a number $[0, 1]$ representing the likeliness of a tissue-type incompatibility with a random donor. Both are drawn from realistic distributions: we use probabilities of 48%, 34%, 14%, and 4% for blood types O, A, B, and AB; the probability for tissue-type incompatibility is set to 0.05 with probability 70%, to 0.45 with probability 20%, and to 0.9 with probability 10% (cf. Roth et al., 2007a). For each donor, we draw a blood type and a uniform random number in $[0, 1]$. If a patient and its corresponding donor have incompatible blood types or if the number of the donor is smaller than that of the patient (corresponding to a negative outcome of the so-called Panel Reactive Antibody test), they are added to the pool of incompatible pairs. Otherwise they are considered compatible and are discarded. Compatibility between any pair of vertices—each corresponding to an incompatible donor–patient pair—is determined analogously, resulting in a random graph with $nk$ vertices. Finally, the vertices are partitioned into $n$ sets, each of which corresponds to a hospital.

Results for the three outcomes, averaged across all 400 graphs, are shown in Table 1. Since the performance of MIX-AND-MATCH depends on the bipartition, it is run 200 times on each graph, each time with a different random partition. We also repeated the experiment for a harder-to-match population in which a patient is added only after being incompatible with between 1 and 4 potential donors, with the number of such donors chosen uniformly at random. The results for this type of population did not show a significant difference, and are therefore omitted.

The columns labeled “opt”, “mm”, and “selfish” respectively report the number of patients matched for the optimal matching, MIX-AND-MATCH, and hospitals “selfishly” matching as many patients as possible internally before submitting the rest to the pool. The next two columns give the performance of MIX-AND-MATCH relative to the other two outcomes. While MIX-AND-MATCH is consistently worse than the optimal outcome, it is also significantly better than the worst-case bound of 0.5. In general, it is within a few percent of the optimum for $n = 2$, and within 15% of the optimum for larger values of $n$ when $k$ is sufficiently large. Of course, we would not expect the outcome of a mechanism that simply finds a maximum matching to be as good as this optimal outcome in practice. The column labeled “1h-opt” lists the fraction of instances in which a particular hospital would gain in such a mechanism by withholding a maximum internal matching. Observe that as the hospital size grows this percentage becomes very high. This suggests that most hospitals will have an incentive to deviate, leading to an outcome closer to the “selfish” one.

The last two columns report standard errors of performance across the 400 graphs. As MIX-AND-MATCH uses randomization internally, we also calculated standard deviations on each graph. We do not give a full table, but note that standard deviation increase with the size of the pool with averages of 2.24, 5.88, 8.69, and 9.7 when $n$ is 4 and $k$ is 5, 50, 100, and 150, respectively. Other choices of $n$ and $k$ that yield the same overall pool size lead to similar standard deviations.

For $n = 2$, MIX-AND-MATCH performs better than the selfish outcome (in fact, it is easy to see that this must be the case), especially when $k$ is large. This result suggests that strategyproofness may have a positive effect on social welfare in certain practical settings, and is particularly relevant for mergers between exchange programs, each of which represents the patients of a large number of hospitals.

The conclusion in the opposite direction, that strategyproofness is undesirable for settings with more than two hospitals, cannot easily be drawn, since the selfish outcome is not an equilibrium. To further investigate this matter we conducted another computational experiment with a small change to MIX-AND-MATCH, so that it no longer prevents patients of hospitals on two sides of a bipartition from being matched. While the resulting mechanism is not strategyproof, it makes manipulation more difficult in practice. We refer to the mechanism as MAX-IR, because it returns a maximum matching subject to individual rationality.

[8] Values of $k$ were chosen to limit the overall size of the graph $nk$ since each experiment involves repeatedly solving many graphs.

**Table 1**
Performance of Mix-and-Match, the optimal outcome, and heuristic strategic behavior.

| $n$ | $k$ | opt | mm | selfish | 1h-opt | se-opt | se-mm |
|---|---|---|---|---|---|---|---|
| 2 | 5 | 1.88 | 1.84 | 1.82 | 0.01 | 0.09 | 0.08 |
| 2 | 10 | 4.78 | 4.70 | 4.59 | 0.10 | 0.13 | 0.13 |
| 2 | 20 | 12.19 | 11.92 | 11.49 | 0.34 | 0.21 | 0.20 |
| 2 | 30 | 20.81 | 20.52 | 19.68 | 0.42 | 0.26 | 0.26 |
| 2 | 50 | 39.74 | 38.83 | 37.34 | 0.56 | 0.38 | 0.37 |
| 2 | 100 | 91.44 | 88.99 | 85.75 | 0.72 | 0.51 | 0.49 |
| 2 | 150 | 146.51 | 142.54 | 138.10 | 0.78 | 0.66 | 0.65 |
| 2 | 200 | 200.54 | 194.67 | 189.04 | 0.82 | 0.72 | 0.69 |
| 4 | 5 | 4.85 | 3.65 | 4.52 | 0.02 | 0.15 | 0.11 |
| 4 | 10 | 12.32 | 10.09 | 11.39 | 0.16 | 0.20 | 0.17 |
| 4 | 15 | 21.02 | 17.28 | 19.39 | 0.30 | 0.26 | 0.22 |
| 4 | 20 | 30.11 | 24.98 | 27.86 | 0.40 | 0.36 | 0.30 |
| 4 | 30 | 49.64 | 42.23 | 45.79 | 0.58 | 0.39 | 0.36 |
| 4 | 50 | 91.05 | 79.26 | 83.48 | 0.73 | 0.50 | 0.45 |
| 4 | 100 | 201.61 | 180.46 | 183.50 | 0.95 | 0.80 | 0.71 |
| 4 | 150 | 312.45 | 283.28 | 283.69 | 0.89 | 0.95 | 0.84 |
| 10 | 5 | 16.27 | 12.73 | 15.35 | 0.09 | 0.23 | 0.19 |
| 10 | 10 | 41.11 | 33.17 | 37.94 | 0.24 | 0.33 | 0.28 |
| 10 | 15 | 65.05 | 53.28 | 60.03 | 0.40 | 0.44 | 0.35 |
| 10 | 20 | 91.28 | 76.14 | 83.77 | 0.51 | 0.53 | 0.44 |
| 10 | 30 | 146.10 | 124.90 | 133.60 | 0.72 | 0.63 | 0.54 |
| 20 | 5 | 39.31 | 31.20 | 37.11 | 0.06 | 0.34 | 0.26 |
| 20 | 10 | 91.08 | 74.76 | 84.46 | 0.27 | 0.54 | 0.43 |
| 20 | 15 | 146.40 | 123.74 | 135.12 | 0.44 | 0.68 | 0.55 |
| 20 | 20 | 201.65 | 173.16 | 185.05 | 0.54 | 0.73 | 0.60 |
| 30 | 5 | 64.59 | 52.32 | 61.03 | 0.10 | 0.43 | 0.34 |
| 30 | 10 | 145.62 | 122.82 | 135.95 | 0.32 | 0.66 | 0.55 |
| 30 | 15 | 229.16 | 197.66 | 212.11 | 0.46 | 0.82 | 0.69 |

Results for Max-IR are given in Table 2. The rightmost column shows the fraction of instances in which a particular hospital would gain by withholding a maximum internal matching. This fraction is significantly smaller compared to the optimal outcome. The third and fourth columns compare the performance of Max-IR to the optimal and selfish outcomes. We observe that the cost of using Max-IR compared to the optimal outcome is small (always less than 5%), while Max-IR provides more than a 3% improvement compared to a situation where hospitals withhold donor–patient pairs. We emphasize though that we are making the strong assumption that hospitals report all their donor patient–pairs under Max-IR; this is of course a plausible assumption to make under Mix-and-Match, which is provably strategyproof, but it is difficult to predict how hospitals would behave when faced with the Max-IR mechanism.

## 8. Discussion and future work

We have seen that Mix-and-Match provides near-optimal worst-case guarantees: the outcome it achieves is always within a factor of two of the optimal matching, which matches the lower bound for deterministic mechanisms and is close to the lower bound for randomized mechanisms. While a factor of two might not be acceptable in practice, in particular in the context of kidney exchanges, simulations suggest a practical performance that is much closer to optimal and sometimes better than that of mechanisms that incentivize agents to hide donors and patients and match them internally. More importantly, what distinguishes Mix-and-Match from mechanisms that are not strategyproof[9] is that it is robust against information asymmetries, has zero deliberation cost, and zero ex-post regret. Arguably, all of these properties are important in the context of kidney exchanges.

An aspect of Mix-and-Match that might be problematic in practice is that it prevents vertices of agents on the same side of the bipartition to be matched: it may be hard to convince hospitals that they best serve their patients by refusing to match them with patients of roughly half of the other hospitals, despite the fact that this would not have a negative impact on social welfare, neither in the worst case nor on average assuming there are sufficiently many patients. One might therefore ask to what extent this characteristic of Mix-and-Match is necessary to guarantee strategyproofness and large social welfare, or one could more generally try to characterize the set of strategyproof mechanisms. Our results suggest that there probably is no simple characterization: quite a few straightforward mechanisms are instances of Match$_{\Pi}$, like the one

[9] This includes mechanisms that are not incentive compatible, but also mechanisms satisfying weaker notions of incentive compatibility like the one proposed by Ashlagi and Roth (2011).

**Table 2**
Comparison of Max-IR and the optimal outcome.

| $n$ | $k$ | mi/o | mi/s | 1h-mi |
|---|---|---|---|---|
| 2 | 5 | 0.97 | 1.02 | 0.00 |
| 2 | 10 | 0.99 | 1.03 | 0.00 |
| 2 | 20 | 0.98 | 1.04 | 0.00 |
| 2 | 30 | 0.98 | 1.04 | 0.00 |
| 2 | 50 | 0.98 | 1.04 | 0.00 |
| 2 | 100 | 0.97 | 1.03 | 0.00 |
| 2 | 150 | 0.97 | 1.03 | 0.00 |
| 2 | 200 | 0.97 | 1.03 | 0.00 |
| 4 | 5 | 0.97 | 1.02 | 0.00 |
| 4 | 10 | 0.96 | 1.03 | 0.04 |
| 4 | 15 | 0.97 | 1.05 | 0.06 |
| 4 | 20 | 0.96 | 1.04 | 0.07 |
| 4 | 30 | 0.96 | 1.04 | 0.10 |
| 4 | 50 | 0.96 | 1.04 | 0.13 |
| 4 | 100 | 0.95 | 1.04 | 0.16 |
| 4 | 150 | 0.95 | 1.04 | 0.15 |
| 10 | 5 | 0.95 | 1.01 | 0.01 |
| 10 | 10 | 0.95 | 1.03 | 0.03 |
| 10 | 15 | 0.95 | 1.04 | 0.10 |
| 10 | 20 | 0.96 | 1.04 | 0.12 |
| 10 | 30 | 0.95 | 1.05 | 0.22 |
| 20 | 5 | 0.96 | 1.02 | 0.00 |
| 20 | 10 | 0.96 | 1.03 | 0.03 |
| 20 | 15 | 0.95 | 1.04 | 0.11 |
| 20 | 20 | 0.95 | 1.04 | 0.15 |
| 30 | 5 | 0.96 | 1.01 | 0.00 |
| 30 | 10 | 0.96 | 1.03 | 0.04 |
| 30 | 15 | 0.96 | 1.04 | 0.12 |

that only allows edges inside hospitals, but a mechanism that selects two agents and runs the two-agent mechanism on these agents is not.[10]

Several gaps still remain between our upper and lower bounds, the most enigmatic one of which concerns deterministic mechanisms for three or more agents. While Theorem 4.1 provides a deterministic lower bound of 2, we were unable to design a deterministic strategyproof mechanism with a constant approximation ratio, and indeed we conjecture that such a mechanism does not exist when there are more than two agents. For randomized mechanisms, there is a gap between the lower bound of 8/7 and the upper bound of 2 provided by Mix-and-Match. For the two-agent case, Caragiannis et al. (2011) recently provided a strategyproof 3/2-approximate mechanism, but it is unknown whether this improved upper bound is tight.

An interesting direction for future work would be to incorporate weights into the model. In practice, different exchanges involving the same vertex may be valued differently, either by an agent or by society, or one vertex may be more important than another. Another direction would be to allow exchanges of length greater than two. This is important, as the number of matched vertices can be increased substantially already through three-way exchanges (Roth et al., 2007a). Finally, one could ask for the stronger requirement of group-strategyproofness to prevent groups of agents to deviate in a coordinated fashion, or consider solution concepts like the core to ensure that no group of agents would want to leave and form a smaller pool.

## References

Abraham, D., Blum, A., Sandholm, T., 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In: Proceedings of the 8th ACM Conference on Electronic Commerce, pp. 295–304.

Ashlagi, I., Roth, A.E., 2011. Free riding and participation in large scale, multi-hospital kidney exchange. Working paper.

Awasthi, P., Sandholm, T., 2009. Online stochastic optimization in the large: Application to kidney exchange. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, pp. 405–411.

Biró, P., Manlove, D.F., Rizzi, R., 2009. Maximum weight cycle packing in directed graphs, with application to kidney exchange programs. Discrete Math. Algorithms Appl. 1 (4), 499–517.

Caragiannis, I., Filos-Ratsikas, A., Procaccia, A.D., 2011. An improved 2-agent kidney exchange mechanism. In: Proceedings of the 7th International Workshop on Internet and Network Economics, pp. 37–48.

Dickerson, J., Procaccia, A.D., Sandholm, T., 2012. Dynamic matching via weighted myopia with application to kidney exchange. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence, pp. 1340–1346.

[10] While these examples are fairly close to Match$_{\Pi}$, we are also aware of a (relatively complex) strategyproof mechanism that works quite differently.

Gabow, H.N., 1990. Data structures for weighted matching and nearest common ancestors with linking. In: Proceedings of the 1st Annual ACM–SIAM Symposium on Discrete Algorithms, pp. 434–443.
Procaccia, A.D., Tennenholtz, M., 2009. Approximate mechanism design without money. In: Proceedings of the 10th ACM Conference on Electronic Commerce, pp. 177–186.
Roth, A.E., Sönmez, T., Ünver, M.U., 2004. Kidney exchange. Quart. J. Econ. 119, 457–488.
Roth, A.E., Sönmez, T., Ünver, M.U., 2005. Pairwise kidney exchange. J. Econ. Theory 125, 151–188.
Roth, A., Sönmez, T., Ünver, M., 2007a. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. Amer. Econ. Rev. 97, 828–851.
Roth, A., Sönmez, T., Ünver, M., 2007b. Notes on forming large markets from small ones: Participation incentives in multi-center kidney exchange. Personal communication.
Toulis, P., Parkes, D.C., 2011. A random graph model of kidney exchanges: Efficiency, individual-rationality and incentives. In: Proceedings of the 12th ACM Conference on Electronic Commerce, pp. 323–332.
Ünver, M.U., 2010. Dynamic kidney exchange. Rev. Econ. Stud. 77 (1), 372–414.