Free riding and participation in large scale, multi-hospital kidney exchange

ITAI ASHLAGI
Sloan School of Management, MIT

ALVIN E. ROTH
Department of Economics, Stanford University

As multi-hospital kidney exchange has grown, the set of players has grown from patients and surgeons to include hospitals. Hospitals can choose to enroll only their hard-to-match patient–donor pairs, while conducting easily arranged exchanges internally. This behavior has already been observed.

We show that as the population of hospitals and patients grows, the cost of making it individually rational for hospitals to participate fully becomes low in almost every large exchange pool (although the worst-case cost is very high), while the cost of failing to guarantee individual rationality is high—in lost transplants. We identify a mechanism that gives hospitals incentives to reveal all patient–donor pairs. We observe that if such a mechanism were to be implemented and hospitals enrolled all their pairs, the resulting patient pools would allow efficient matchings that could be implemented with two- and three-way exchanges.

KEYWORDS. Market design, kidney exchange.

JEL CLASSIFICATION. D47, D82.

1. INTRODUCTION

A marketplace is similar to a public good: it provides opportunities to every potential trader. As the market becomes larger, the trades it offers become more numerous and varied. However, when it is costly to bring goods to market, and when some goods are easier to trade than others, a kind of free riding can occur. Traders may be tempted to bring only their hard-to-trade goods to market and to trade their easy-to-trade goods elsewhere (e.g., nearer home). When this leads to loss of efficiency, the task of the market designer is to make the marketplace attractive enough for even the easy-to-trade goods.[1]

[1]This was the case in the labor market for gastroenterologists. What had been a national marketplace collapsed into a set of small local marketplaces in which gastroenterology fellowship positions were increasingly filled by local candidates, often from the same hospital (see Niederle and Roth 2003). To repair

We are today seeing this in kidney exchange, in a way that allows the free riding and the consequent loss of efficiency to be clearly understood. When kidney exchange was just beginning, most exchanges were conducted in single hospitals or in closely connected networks of hospitals like the 14 New England transplant centers organized by the New England Program for Kidney Exchange (Roth et al. 2005b). But today exchanges often involve multiple hospitals that may have relatively little repeated interaction outside of kidney exchange. The present paper establishes a theoretical framework to study the kinds of problems that have developed as the United States moves toward nationally organized exchange, as it has begun to do since the passage of facilitating legislation in 2007.[2]

We study the growing problem of giving hospitals incentives to participate fully, to achieve the gains that kidney exchange on a large scale makes possible. We characterize the efficient exchanges that would arise in large markets if hospitals participated fully and how this efficiency can be lost if hospitals withhold easy-to-match pairs. Our results suggest that if care is taken in how kidney exchange mechanisms are organized, the problems of participation may be less troubling in large exchange programs than they are starting to be in multi-hospital exchanges as presently organized. We propose a "bonus mechanism," similar in spirit to frequent flyer programs, and show that it provides incentives for hospitals to enroll their easy- as well as their hard-to-match patient–donor pairs.

1.1 *Background*

Kidney transplantation is the treatment of choice for end stage renal disease, but there are many more people in need of kidneys than there are kidneys available. Kidneys for transplantation can come from deceased donors or from live donors (since healthy people have two kidneys and can remain healthy with one). However, not everyone who is healthy enough to donate a kidney and wishes to do so can donate a kidney to his or her intended recipient, since a successful transplant requires that donor and recipient be compatible in blood and tissue types. This raises the possibility of *kidney exchange*, in which two or more incompatible patient–donor pairs exchange kidneys, with each patient in the exchange receiving a compatible kidney from another patient's donor.[3]

Note that it is illegal for organs for transplantation to be bought or sold in the United States and throughout much of the world (see Roth 2007 and Leider and Roth 2010).

---

this market and reestablish a national marketplace, it was necessary to make changes in the market rules; see McKinney et al. (2005), Niederle and Roth (2005), Niederle et al. (2006, 2008). While the efficiency losses could not be quantified in those markets, one of the contributions of the present paper is that we can show the size of the efficiency losses from withholding easy transactions in the markets we consider here. (Related issues arose in the market for college football championship games, in which a variety of market designs were deployed to create a thicker market; see Roth and Xing 1994 and Fréchette et al. 2007.)

[2]110th Congress, Public Law 110–144, Charlie W. Norwood Living Organ Donation Act, Dec. 21, 2007.

[3]In addition to such cyclic exchanges, chains are also possible, which involve not only incompatible patient–donor pairs, and begin with a deceased donor or an undirected donor (one without a particular intended recipient) and end with a patient with high priority on the deceased donor waiting list or with a donor who will donate at a future time.

Kidney exchange thus represents an attempt to organize a barter economy on a large scale, with the aid of a computer-assisted clearinghouse.[4]

The first kidney exchange in the United States was carried out in 2000 at the Rhode Island Hospital, between two of the hospital's own incompatible patient–donor pairs.[5] Roth et al. (2004) made an initial proposal for organizing kidney exchange on a large scale, which included the ability to integrate cycles and chains, and considered the incentives that well designed allocation mechanisms would give to participating patients and their surgeons to reveal relevant information about patients. The surgical infrastructure available in 2004 meant that only pairwise exchanges (between exactly two incompatible patient–donor pairs) could initially be considered, and Roth et al. (2005a) proposed a mechanism for accomplishing this, again paying close attention to the incentives for patients and their surgeons to participate straightforwardly. As kidney exchanges organized around these principles gained experience, Saidman et al. (2006) and Roth et al. (2007) showed that efficiency gains could be achieved by incorporating chains and larger exchanges that required only relatively modest additional surgical infrastructure, and today there is growing use of larger exchanges and longer chains, particularly following the publication of Rees et al. (2009).

Roth et al. (2005b) describe the formation of the New England Program for Kidney Exchange (NEPKE) under the direction of Dr. Frank Delmonico, which initially organized the 14 transplant centers in New England. Those proposals were also instrumental in helping to organize the Alliance for Paired Donation (APD) under the direction of Dr. Mike Rees.[6] In 2010, a National Kidney Paired Donation Pilot Program organized by the United Network for Organ Sharing (UNOS) became operational, still on a very small scale.[7]

Kidney exchange is growing fast, but it still accomplishes well under a thousand transplants a year.[8] 54 hospitals participate (actively) in the privately organized National Kidney Registry (NKR), for example, and 49 hospitals participate in APD. In the last year, the number of incompatible pairs that join these programs is between 30–40 pairs per month (these numbers are growing). 20% of the centers provide more than 50% of the pairs and are roughly the same size. These large centers currently enter just a few pairs every month.

During the initial startup period, attention to the incentives of patients and their surgeons to reveal information was important. But as infrastructure has developed,

---

[4]Recall that Jevons (1876) proposed that precisely the difficulties of organizing barter economies—in particular, the difficulty of satisfying the "double coincidence" of wants involved in simultaneous exchange without money—had led to the invention of money.

[5]For an account of this and other early events in kidney exchange, see Roth (2010).

[6]Today, in addition to those two large kidney exchange clearinghouses, kidney exchange is practiced by a growing number of hospitals and formal and informal consortia (see Roth 2008). Computer scientists have become involved, and an algorithm of Abraham et al. (2007) designed to handle large populations is used in the national pilot program.

[7]The national pilot program ran two initial pilot matches in October and December of 2010. Under its initial guidelines, only exchanges were considered, not chains. In December 2011 NEPKE formally ceased operation to merge its efforts with the national pilot program.

[8]Massie et al. (2013) report that 93 kidney exchange transplants were conducted in 2006, and between 500 and 600 in each of 2010 and 2011.

the information contained in blood tests has come to be conducted and reported in a more standard manner (sometimes at a centralized testing facility), reducing some of the choice about what information to report and with what accuracy. So some strategic issues have become less important over time (and indeed current practice does not deal with the provision of information that derives from blood tests as an incentive issue).

However, as kidney exchange has become more widespread, and as multi-hospital exchange consortia have been formed and a national exchange is being explored, the "players" are not just (and perhaps not even) patients and their surgeons, but hospitals (or directors of transplant centers). And as kidney exchange is practiced on a wider scale, a new phenomenon has emerged. Free riding has become possible, with hospitals having the option of participating in one or more kidney exchange networks but also of withholding some of their patient–donor pairs or some of their nondirected donors, and enrolling those of their patient–donor pairs who are hardest to match, while conducting more easily arranged exchanges internally. Some of this behavior is already observable.

The present paper considers the "kidney exchange game," with hospitals as the players, to clarify the issues currently facing hospitals in existing multi-hospital exchange consortia and those that would face hospitals in a large-scale national kidney exchange program.[9]

While we concentrate on the incentives created by the matching algorithms, the fact that presently used algorithms do not make it individually rational for hospitals to fully participate in kidney exchange is not the only reason that hospitals withhold patients. Other reasons include lack of standardization in compatibility tests, and bureaucratic and other difficulties in registering pairs to the various kidney exchange systems. For example, not all hospitals collect all the medical data that some programs require. Financing is another obstacle; for example, hospitals may have difficulty recovering the costs of testing a donor who will eventually donate to a patient at another hospital (Rees et al. 2012).

1.2 *Free riding*

Hospitals participate in a multi-center exchange by reporting a list of incompatible patient–donor pairs to a central clearinghouse, and a matching mechanism chooses which exchanges to carry out. At the same time, some hospitals conduct exchanges only internally among their own patients, and even hospitals participating in multi-center exchange programs may conduct some internal exchanges and may participate in more than one exchange program.

---

[9]One referee asks why, if hospital participation is a problem, kidney exchange cannot be designed without hospitals, with patients registering directly (or through dialysis centers). A second referee proposes that the problem of hospital participation could be "simply" solved by legislation requiring hospitals to participate. It seems to us that these suggestions make more sense in the abstract than in connection with practical market design. It is difficult to pass legislation mandating how hospitals treat patients, since hospitals need to exercise a good deal of discretion about individual cases. And patients in need of transplants presently get most of their advice from surgeons associated with hospitals, and it would be difficult and not obviously desirable to bypass this process. And each transplant adds revenues to hospitals and subtracts it from dialysis centers, so it is far from clear that dialysis centers are natural partners for promoting kidney exchange.

Centralized kidney exchange programs substantially increase the number of matches found and also the chance for highly sensitized patients to be matched compared to decentralized matching within individual hospitals. The efficiency gains from centralization grow as the number of (moderate-sized) hospitals increases.[10]

However, when kidney exchange clearinghouses try to maximize the (weighted) number of transplants without attention to whether those transplants are internal to a hospital, it may not be individually rational for a hospital to contribute those pairs it can match internally (cf. Roth 2008).[11] For example, consider a hospital a with two pairs, $a1$ and $a2$, that it can match internally. Suppose it enters those two pairs in a centralized exchange. It may be that the weighted number of transplants is maximized by including $a1$ in an exchange but not $a2$, in which case only one of hospital a's patients will be transplanted, when it could have performed two transplants on its own.

This is becoming a first-order problem, as membership in a kidney exchange network does not mean that a hospital does not also do some internal exchanges.[12] Mike Rees, the director of the APD, writes (personal communication)

> …competing matches at home centers is becoming a real problem. Unless it is mandated, I'm not sure we will be able to create a national system. I think we need to model this concept to convince people of the value of playing together.

This paper attempts to understand the problem raised by the APD director. We will see that when the number of hospitals and incompatible pairs is small, it may be costly (in terms of lost transplants) for a centralized clearinghouse to guarantee hospitals individual rationality, compared to how many transplants could be accomplished if all pairs were submitted to a centralized exchange despite no guarantee of individual rationality. However, in large markets we will show that this cost becomes very low. In the market we study, the number of hospitals grows large and each hospital satisfies a regularity assumption, which implicitly requires that its number of patient–donor pairs is not "too big," yet not zero. In particular, we show that there is an individually rational allocation that is almost efficient. We further begin to explore incentive compatible mechanisms for achieving full participation by hospitals as efficiently as possible; We introduce an (almost) efficient mechanism under which full participation (not withholding pairs) is an approximated Bayes Nash equilibrium under a slightly stronger regularity assumption.

1.3 *Related literature*

Roth et al. (2007) studied efficiency in large markets without considering incentives or directly modeling tissue-type incompatibilities. They showed that exchanges of size

---

[10]See Toulis and Parkes (2011), who quantify the benefit from a centralized clearinghouse for organizing two-way exchange.

[11]Some weighted matching algorithms currently in use put some weight on internal exchanges, but this does not solve the problem, since it neither guarantees a hospital the exchanges it could conduct internally nor does it guarantee that the pairs that could be internally exchanged will be used efficiently if submitted to the central clearinghouse.

[12]The national pilot program has to date completed very few transplants, in part because of this problem.

more than 4 are not needed for efficiency. In this paper, we model tissue-type incompatibilities using a random graph framework (and show that even four-way exchanges are not needed for efficiency), but more importantly, we study the hospitals' incentives. In unpublished notes from 2007, Roth, Sönmez, and Ünver introduced the problem of withholding internal matches by hospitals and showed that there is no efficient strategyproof mechanism for kidney exchange.[13] Our work extends that negative result for small markets to show that efficient mechanisms cannot even be individually rational, but, more importantly, we provide positive results in large random markets.

Toulis and Parkes (2011) also adopt a random graph model to study mechanisms for kidney exchange and provide useful quantitative welfare results. Their results are close to ours, but with a very different model of how the market grows large. While we model a growing number of "small" hospitals, they let a fixed number of hospitals each become very large. We further discuss the differences throughout the paper. In another paper, Ashlagi et al. (forthcoming(b)) also analyze hospitals' incentives, but worst case rather than in a Bayesian (random graph) environment. Finally, Ünver (2010) analyzes an efficient algorithm for a dynamic environment in which full participation is assumed.

2. Kidney exchange and individual rationality

2.1 *Exchange pools*

An exchange pool consists of a set of patient–donor pairs. A patient p and a donor d are *compatible* if patient p can receive the kidney of donor d and are *incompatible* otherwise. It is assumed that every pair in the pool is incompatible.[14] Thus, a pair is a tuple $v = (p, d)$ in which donor d is willing to donate his kidney to patient p, but p and d are incompatible. We assume for simplicity that each donor and each patient belong to a single pair.

An exchange pool V induces a *compatibility graph* $D(V) = D(V, E(V))$ that captures the compatibilities between donors and patients as follows: the set of nodes is V, and for every pair of nodes $u, v \in V$, (u, v) is an edge in the graph if and only if the donor of node u is compatible with the patient of node v. We will use the terms nodes and pairs interchangeably.[15]

An exchange can now be described through a cycle in the graph. Thus, an *exchange* in V is a cycle in $D(V)$, i.e., a list $v_1, v_2, \ldots, v_k$ for some $k \geq 2$ such that for every i, $1 \leq i < k$, $(v_i, v_{i+1}) \in E(V)$ and $(v_k, v_1) \in E(V)$. The size of an exchange is the number of nodes in the cycle. An *allocation* in V is a set of distinct exchanges in $D(V)$ such that each node belongs to at most one exchange. Since, in practice, the size of an exchange

---

[13]This result appears in the survey by Sönmez and Ünver (2013).

[14]Pairs that are compatible would presently go directly to transplantation and not join the exchange pool (but see, e.g., Roth et al. 2005b and Sönmez and Ünver 2014 on the advantages of changing this policy).

[15]In practice, a patient may have more than one incompatible donor. In this case, the model can be naturally extended by adding outgoing edges from the patient's node to pairs such that at least one of her incompatible donors is compatible with the patient at that node.

is limited (mostly due to logistical constraints), we assume there is an exogenous maximum size limit $k > 0$ for any exchange. Thus, if $k = 3$, only exchanges of size 2 and 3 can be conducted.[16]

Let M be an allocation in V. We say that node v is *matched* by M if there exists an exchange in M that includes v. For any set of nodes $V' \subseteq V$, let $M(V')$ be the set of all nodes in V' that are matched (or "covered") by M.

We will be interested in finding efficient allocations that have as many transplants as possible. Two types of efficiency will be considered. M is called *k-efficient* if it matches the maximum number of transplants possible for exchanges of size no more than k, i.e., there exists no other allocation M' consisting of exchanges of size no more than k such that $|M'(V)| > |M(V)|$.[17] M is called *k-maximal* if there exists no such allocation M' such that $M'(V) \supsetneq M(V)$. A matching will be called *efficient* (or *maximal*) if it is k-efficient (or k-maximal) for unbounded k, i.e., for no limit on how many transplants can be included in an exchange. Note that every k-efficient allocation is also k-maximal. The converse is not true. However, for $k = 2$, both types of efficiency coincide, since the collection of sets of simultaneously matched nodes in allocations forms a matroid (see Edmonds 1971).

A *kidney exchange program* (or simply a kidney exchange) consists of a set of n hospitals $H_n = \{h_1, \dots, h_n\}$ and a set of incompatible pairs V_h for each hospital $h \in H_n$. We let $V_{H_n} = \bigcup_{h \in H_n} V_h$. The compatibility graph induced by V_{H_n} is called the *underlying graph*. We will take the hospitals (e.g., the director of transplantation at each hospital) as the active decision makers in the kidney exchange, whose choices are which incompatible pairs to reveal to the exchange. We will approximate the preferences of hospitals as being concerned only with their own patients. Mostly we will assume hospitals are concerned only with the *number* of their patients who receive transplants, although we do not rule out hospitals having preferences over which of their patients are transplanted.

An exchange that matches only pairs from the same hospital is called *internal*. Hospital h can match a set of pairs $B_h \subseteq V_h$ *internally* if there exists an allocation in V_h such that all nodes in B_h are matched.

2.2 *Participation constraints: Individual rationality for hospitals*

The kidney exchange setting invites discussions of various types of individual rationality (IR). In this paper, an allocation is not individually rational if some hospital can internally match more pairs than the number of its pairs matched in the allocation. Formally, an allocation M in V_{H_n} is not *individually rational* if there exists a hospital h and an allocation M_h in V_h such that $|M(V_h)| < |M_h(V_h)|$.

To illustrate this, consider the compatibility graph in Figure 1, where nodes a_1 and a_2 belong to hospital a, and b_1 and b_2 belong to hospital b. The only individually rational allocation is the one that matches a_1 and a_2.

---

[16]In the APD and NEPKE, k was originally set to 2, was increased to 3, and now optimization is conducted over even larger exchanges and chains, and the pilot national program considers exchanges up to size 3. Exchanges are generally conducted simultaneously, so an exchange of size k requires $2k$ operating rooms and surgical teams for the k nephrectomies (kidney removals) and k transplants.

[17]In graph theory, a 2-efficient allocation is referred to as a maximum matching.
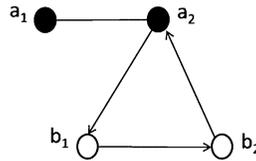
FIGURE 1. No 3-efficient allocation is individually rational.

REMARK. Throughout this paper, undirected edges represent two directed edges, one in each direction.

Other formulations of individual rationality may sometimes be appropriate, such as requiring not merely that a hospital be allocated the same *number* of transplants it can achieve on its own, but that it be guaranteed a *set* of transplants that includes all the individuals it could match on its own. It is worth mentioning that all our positive results hold even with this stronger individual rationality.

In the next section, we study worst-case efficiency loss from choosing IR allocations.

3. IR AND EFFICIENCY: WORST-CASE RESULTS FOR COMPATIBILITY GRAPHS

By choosing the individually rational allocation in Figure 1, we obtain two transplants whereas the efficient allocation provides three. The next result, proved in the Appendix, shows that a maximum individually rational allocation can be very costly in the worst case.[18]

THEOREM 1. *Let $k \geq 3$. In every compatibility graph the size of a k-maximal allocation is at least $1/(k-1)$ times the size of a k-efficient allocation. This bound is tight: there exists a compatibility graph such that no k-maximal allocation that is also individually rational matches more than $1/(k-1)$ of the number of nodes matched by a k-efficient allocation.*[19]

Thus there is a very high potential cost of individual rationality, but it gives a worst-case result. However, it appears that the *expected* efficiency loss from requiring individual rationality can be very small. Indeed, our simulations show that *if all incompatible pairs are in the same exchange pool, the average number of patients who do not get a kidney due to requiring IR is less than* 1 (see Table 1). But as we shall see in Section 8, the cost of *failing* to guarantee individual rationality could be large if that causes hospitals to match their own internal pairs. In Section 9, we explain how we conduct the simulations and provide further simulation results.

In the next sections, we will prove that the efficiency loss from choosing an IR allocation of maximum size is small in large compatibility graphs, supporting the simulation results.

---

[18]Note that in every compatibility graph, one can find a k-maximal allocation that is also individually rational: first choose a k-efficient allocation in V_h for every hospital h and then repeatedly search for an allocation that increases the total number of matched pairs without unmatching any pair that was previously matched (although possibly rematching such pairs using different edges).

[19]If hospitals can conduct only two-way exchanges, then there is always a 2-efficient allocation that is individually rational since every 2-maximal allocation is 2-efficient.

No. of hospitals	IR, $k = 3$	Efficient, $k = 3$
2	6.91	7.07
4	17.02	17.42
6	27.31	27.92
8	39.35	40.04
10	51.72	52.44
12	63.44	64.19
14	75.89	76.72
16	88.08	88.81

TABLE 1. Number of transplants achieved using maximum-size individually rational allocations versus using efficient (and not necessarily individually rational) allocations. Each hospital has, on average, 10 pairs.

4. RANDOM EXCHANGE POOLS

To discuss the Bayesian setting, it is useful to consider random compatibility graphs. Each person in the population has one of four blood types A, B, AB, and O, according to whether her blood contains the proteins A, B, both A and B, or neither. The probability that a random person's blood type is X is given by $\mu_X > 0$. We will assume that $\mu_O > \mu_A > \mu_B > \mu_{AB}$ (as in the U.S. population).[20] For any two blood types X and Y, a donor of blood type Y and a patient with blood type X are blood-type compatible if X includes whatever blood proteins A and B are contained in Y.[21]

A patient–donor pair has pair type (or just type, whenever it is clear from the context) X-Y if the patient has blood type X and the donor has blood type Y. The set of pair types will be denoted by $\mathcal{P}$. For a donor and a patient to be compatible, they need to be *both* blood-type compatible and tissue-type compatible. To test tissue type compatibility, a *cross-match* test is performed. Each patient has a level of *percentage reactive antibodies* (PRA) that determines the likelihood that the patient will be compatible with a random donor. The lower the PRA of a patient, the more likely the patient is compatible with a random donor. For simplicity, we assume that there exist two levels of PRA, γ_L and γ_H ($\gamma_L < \gamma_H$); the probability that a patient p with PRA γ and a donor are tissue-type *in*compatible is given by γ. Furthermore, the probability that a random patient has PRA γ_L is given by $v > 0$. Let $\bar{\gamma}$ denote the expected PRA level of a random patient, that is, $\bar{\gamma} = v\gamma_L + (1 - v)\gamma_H$.[22]

---

[20]In practice, $\mu_O = 0.48$, $\mu_A = 0.34$, $\mu_B = 0.14$, and $\mu_{AB} = 0.04$.

[21]Thus type O patients can receive kidneys only from type O donors, while type O donors can give kidneys to patients of any blood type. Note that since only *incompatible* pairs are present in the kidney exchange pool, donors of blood type O will be underrepresented, since most such donors will be compatible with their intended recipients; the only incompatible pairs with an O donor will be tissue-type incompatible. (Roth et al. 2005b showed that a significant increase in the number of kidney exchanges could be achieved by allowing compatible pairs to participate, but this has not become common practice.)

[22]Our results hold for any number of different PRA levels as long as the probabilities for compatibility are constant.

DEFINITION 1 (Random compatibility graph). A random (directed) compatibility graph of order m, denoted $D(m)$, consists of m incompatible patient–donor pairs and is generated as follows:

Nodes. A patient p and a potential donor d are generated using the blood type and PRA distribution, and (p, d) forms a new node if and only if they are incompatible with each other.

Edges. Between every two pairs v_1 and v_2, a directed edge is generated if the donor of v_1 is compatible with the patient of v_2.

We will often denote a random compatibility graph by $D(H_n)$; thus $D(H_n) = D(m)$, where m is the total number of pairs in all hospitals belonging to H_n. We also denote by $\mu_{X\text{-}Y}$ the posterior probability that an incompatible pair (p, d) is of type X-Y.

We will derive results for large random compatibility graphs (with many hospitals), and use results and methods from random graph theory. We adopt the following formalism from this literature: *if the probability that a given property Q is satisfied in a random graph G tends to* 1 *when m tends to* ∞, *we say that Q holds in* almost every (large) G.

The relative number of pairs of various types will be useful in studying large compatibility graphs.

LEMMA 1. *In almost every large $D(m)$, the following statements hold:*

1. *For all $X \in \{A, B, AB\}$, the number of* O-X *pairs is larger than the number of X-O pairs.*

2. *For all $X \in \{A, B\}$, the number of X-AB pairs is larger than the number of AB-X pairs.*

3. *The absolute difference between the number of* A-B *pairs and* B-A *pairs is $o(m)$. Consequently, this difference is smaller than the number of pairs of any other pair type.*[23]

[Toulis and Parkes (2011)](#) prove a similar lemma, and use the blood-type and tissue-type distributions to characterize the size of each set of pairs in the graph.

[Lemma 1](#), whose proof appears in the [Appendix](#), motivates the following partition of patient–donor pair types $\mathcal{P}$ (see also [Roth et al. 2007](#) and [Ünver 2010](#)). Let

$$\mathcal{O} = \{\text{A-O, B-O, AB-O, AB-A, AB-B}\}$$

be the set of *overdemanded* types.

Let

$$\mathcal{U} = \{\text{O-A, O-B, O-AB, A-AB, B-AB}\}$$

be the set of *underdemanded* types.

---

[23][Terasaki et al. (1998)](#) claim that the frequency of A-B pairs (0.05) is larger than B-A pairs (0.03), but they do not give any data or other explanation to support their claim. Our result just asserts that the absolute difference is small.

Let

$$\mathcal{S} = \{\text{O-O, A-A, B-B, AB-AB}\}$$

be the set of *self-demanded* types and, finally, let $\mathcal{R}$ be the set of *reciprocally demanded* types that consists of types A-B and B-A.

Intuitively, an overdemanded pair is offering a kidney in greater demand than the one being sought. For example, a patient whose blood type is A and a donor whose blood type is O form an overdemanded pair. Underdemanded types have the reverse property: they are seeking a kidney that is in greater demand than the one they are offering in exchange. A donor and patient in a pair with a self-demanded type have the same blood type.

The following notations will be useful in later sections and proofs. For any type $t \in \mathcal{P}$ and set of pairs S, we denote by $\tau(S, t)$ the set of pairs with type t in S and for a set of types $T \subseteq \mathcal{P}$, let $\tau(S, T) = \bigcup_{t \in T} \tau(S, t)$ and let $\mu_T = \sum_{t \in T} \mu_t$. For any set of pairs V, let M_T^V be a (random) allocation in the graph induced by the set of pairs V that maximizes the number of matches with type belonging to T.

In the next section, we study efficiency in large random compatibility graphs. We let γ_L and γ_H (the probability of tissue type incompatibility for patients with low or high PRA) be nondecreasing functions of m, with the important special case in which both are constants.

5. Efficient allocations in large random compatibility graphs

We construct here an efficient allocation in a large random compatibility graph. We make the following assumptions, which are compatible with blood-type frequencies and with observed tissue-type sensitivity frequencies. Zenios et al. (2001) reported that for nonrelated blood-type donors and recipients, $\bar{\gamma} = 0.11$.

ASSUMPTION A (Non-highly-sensitized patients). $\bar{\gamma} < \frac{1}{2}$.[24]

ASSUMPTION B (Blood type frequencies). $\mu_O < 1.5\mu_A$.[25]

PROPOSITION 1. *Almost every large $D(m)$ has an efficient allocation that requires exchanges of no more than size 3 with the following properties:*

1. *Every self-demanded pair X-X is matched in a two-way or a three-way exchange with other self-demanded pairs (no more than one three-way exchange is needed, in the case of an odd number of X-X pairs).*

---

[24]This assumption is also used for avoiding case-by-case analysis; one can provide similar results for the opposite inequality. However, the limit results we obtain here for large compatibility graphs are less of a good approximation to the situation facing *very* high PRA patients in the finite graphs we see in practical applications than they are for the situation facing the large majority of patients who are not extremely highly sensitized. We will return to this, and the open questions it raises, in the conclusion.

[25]We will use this assumption to construct the efficient allocation. However, even if this assumption does not hold, using a similar method of proof, one can construct a very similar allocation. The details of the efficient allocation would slightly change, but not our results about individually rational allocations.
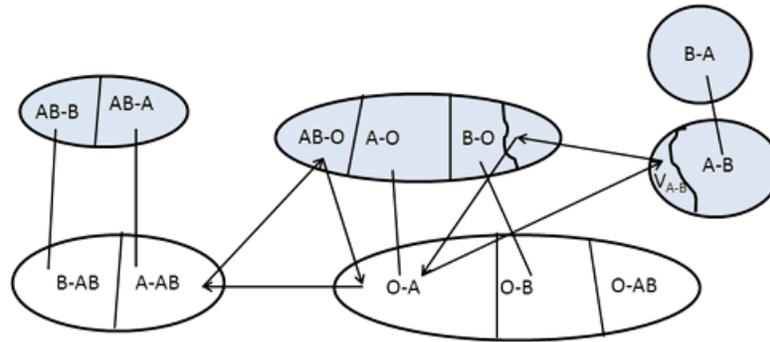
FIGURE 2. The structure of an efficient allocation in the graph $D(m)$ (excluding all self-demanded pairs). The shaded region is the set of overdemanded pairs, none of which remains unmatched after an efficient matching. All B-A pairs are matched to A-B (assuming there are more B-A than A-B); the remainder of the A-B pairs ($V_{A\text{-}B}$) are matched in three-way exchanges using O-A's and B-O's. AB-O are matched in three ways each using two overdemanded pairs, and every other overdemanded pair is matched to a corresponding underdemanded pair.

2. *Either every* B-A *pair is matched in a two-way exchange with an* A-B *pair or every* A-B *pair is matched in a two-way exchange with a* B-A *pair.*

3. *Let* $X, Y \in \{A, B\}$ *and* $X \neq Y$. *If there are more* Y-X *than* X-Y, *then every* Y-X *pair that is not matched to an* X-Y *pair is matched in a three-way exchange with an* O-Y *pair and an* X-O *pair.*

4. *Every* AB-O *pair is matched in a three-way exchange with an* O-A *pair and an* A-AB *pair.*

5. *Every overdemanded pair* X-Y *that is not matched as above is matched to an under-demanded* Y-X *pair.*

The structure of the efficient allocation described in the proposition is given Figure 2. The proof of the proposition is deferred to the Appendix.

Roth et al. (2007) showed that exchanges of size at most 4 are sufficient for efficiency and assumed compatibilities are determined merely by blood types. Interestingly, they used the four-way exchanges (AB-O, O-A, A-B, B-AB) whenever there were many more A-B pairs than B-A pairs. Our random model assures that this difference is small enough to avoid the need for such four-way exchanges (note from Figure 2 that such a four-way exchange becomes inefficient, since it uses an AB-O pair that could instead have been used in a three-way exchange and an A-B pair that could have been used in either a two-way or a three-way exchange, for a total of more than four transplants). Toulis and Parkes (2011) prove a similar result to ours.

Similarly to Proposition 1, one can show that the size of a 2-efficient allocation is at most $\mu_{\text{AB-O}}m + o(m)$ smaller than the size of an efficient allocation.[26] One possibly

---

[26]In particular, AB-O pairs can be matched to O-AB pairs using two-way allocations rather than being matched in a three-way allocation as described in Proposition 1, and the three-way exchanges that use A-B pairs (or B-A pairs) can be ignored.

undesirable feature of the efficient allocation is that underdemanded pairs of type O-AB will all be left unmatched. While it is inevitable that many underdemanded pairs will be left unmatched, there is sometimes discomfort in medical settings having a priori identifiable pairs seemingly singled out. A natural outcome would be that hospitals would seek to match such pairs internally, a point to which we will return later when we observe that precisely these internal matches account for most of the efficiency cost of individual rationality.

Until this point, nothing has been said about individual rationality in the Bayesian setting. In the next section, we study the efficiency cost of requiring an allocation to be individually rational in large exchange pools.

6. Individual rationality is not very costly in large random compatibility graphs

One way in which individual rationality might conflict with efficiency is if hospitals' internal exchanges make inefficient use of overdemanded pairs, e.g., if an overdemanded A-O pair were matched internally in a two-way exchange with a B-A, an A-A, or an AB-A pair, in each case resulting in two transplants instead of four. In Section 3, we proved tight worst-case bounds on the efficiency loss from having to honor hospitals' internal exchanges to guarantee individual rationality. We derive here a much smaller upper bound on this loss for large random compatibility graphs.

One way to bound the efficiency loss is by attempting to construct an efficient allocation as in Proposition 1 that matches the pairs each hospital can internally match. Unfortunately such an allocation is not always feasible.

Consider, for example, the following two *unbalanced* three-way exchanges (B-O, O-A, A-B) and (A-O, O-B, B-A). Too many three-way internal exchanges of the second type, for example, as well as other internal exchanges that include O-B pairs but not B-O pairs, could lead to a situation in which, to satisfy individual rationality, more O-B pairs would potentially need to be matched than the total number of B-O pairs. This can harm efficiency since, as Theorem 1 suggests, more transplants are obtained by choosing the two two-way exchanges rather than the three-way exchange in Figure 3.

Individual rationality, however, does not require the clearinghouse to match a specific maximum set of pairs that each hospital can internally match, but only to guarantee to match at least the *number* of pairs each hospital can internally match. For example, if a hospital has an internal unbalanced exchange A-O, O-B, B-A and an internally unmatched O-A pair, then to satisfy individual rationality, it is sufficient to match the A-O, B-A, and O-A pairs.

As the above discussion suggests, individually rational allocations may contain (many) more underdemanded pairs of a specific type than its reciprocally overdemanded type. However, if each hospital is not "too big," this is very unlikely. We will use the following definition that implicitly bounds the size of a hospital by not letting it match internally too many underdemanded pairs.
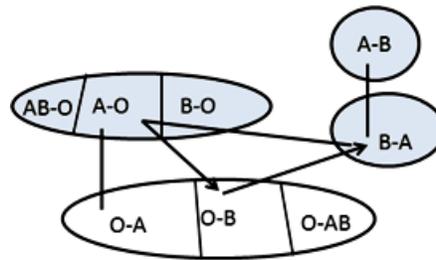
FIGURE 3. A three-way (internal) exchange that matches an O-B pair for which there may not be a corresponding B-O pair.

We say that a size c of a hospital is *regular* if in a random internal allocation that maximizes the number of matched underdemanded pairs and any underdemanded type X-Y, the expected number of X-Y matched pairs is less than the expected number of overdemanded Y-X pairs in its pool.

A formal definition is given in Appendix A.4. We further motivate this definition. Note that if each hospital is large enough, it can internally match with high probability the same set of pairs that are matched in the efficient allocation described in Proposition 1 and illustrated in Figure 2 (this can be shown using the Erdos–Renyi theorem (see Theorem 4) and using the fact that the probability for each edge is a constant). Therefore, if hospitals were large enough, centralized kidney exchange would not yield more matches than a decentralized system. And, in fact, American transplant centers have grown in numbers more than in size. Today there are over 200 centers that perform kidney transplants, and the largest do fewer transplants they did when there were only a handful of centers (see Massie et al. 2013, who compile data from 207 American transplant centers).

Using simulations with distributions from clinical data (see Ashlagi et al. 2011a, 2011b), we find that hospitals of size up to at least $c = 70$ are regular.[27]

This allows us to state our first main result.

THEOREM 2. *Suppose every hospital size is regular and bounded by some $\bar{c} > 0$, and let $\epsilon > 0$. In almost every large graph $D(H_n)$, there exists an individually rational allocation using exchanges of size at most 3, which is at most $\mu_{\text{AB-O}}m + \epsilon\mu_{\text{A-B}}m$ smaller than the efficient allocation, where m is the number of pairs in the graph.*

As suggested in the theorem (and as shown in the proof), most of the efficiency loss comes from matching of (otherwise unmatched) underdemanded O-AB pairs in two-way exchanges to AB-O pairs. This means that the *efficiency loss is only about* 1%, which is the (simulated) frequency of the AB-O pairs. Note also that, as remarked earlier, it is hard to regret this small decrease in the total number of matched pairs, since no

---

[27]In practice, hospitals indeed withhold, but in a given month, no hospital has ever enrolled more than 10 pairs to either the National Kidney Registry or the Alliance for Paired Donation.

O-AB pairs would have been matched had the goal been to maximize the number of transplants.[28,29]

Theorem 2 is a limit theorem, but Table 1 showed simulation results that demonstrate that the cost of individual rationality is very low even for sizes of exchange pools observed in present-day clinical settings.

The proof is deferred to the Appendix. The key step is to match for each hospital as many underdemanded pairs it can internally match (since these are the pairs that compete to be matched). In particular, we show that with high probability there exists a *satisfiable* set of underdemanded pairs that can be matched, where a satisfiable set is a set in which (i) for each hospital, the set contains at least the number of underdemanded pairs the hospital can internally match and (ii) for each underdemanded type X-Y, the number of pairs in the set is the same number as the total number of overdemanded Y-X pairs in the entire pool.

REMARKS. 1. Toulis and Parkes (2011) provide an algorithm that finds an efficient individually rational allocation. However, there is a major difference in their model; in contrast to our regularity assumption, they assume each hospital is *large* enough so that it contains an efficient allocation with the structure provided in Figure 5, which they term a *canonical allocation*. Note that for a hospital to contain a canonical allocation, it cannot be regular by definition (recall that the condition that our simulations provide regularity is only violated for hospitals of size more than 70, much larger than hospitals enroll in a "reasonable" time period). Also, Ashlagi and Roth (2012) observed that there are many very highly sensitized patients, making it very unlikely that hospitals contain a canonical internal allocation.

   Finally, as we mentioned, if every hospital has an internal allocation with a canonical structure, there is almost no need for a centralized mechanism.

2. The proof of Theorem 2 is by construction and thus defines an algorithm that finds an individually rational allocation (in almost every large graph). Interestingly, our algorithm runs in polynomial time. The algorithm first finds within each hospital internal allocations, a step that runs in linear time (in the number of hospitals) since each hospital is of a constant size. This step identifies the set of pairs S that will be matched in the final allocation.

   In the second step, we identify an allocation that matches all pairs in S, and even when using only two-way exchanges in this step, we achieve the same bound for the efficiency loss. As our proof shows in almost every large graph, such an allocation exists. Using this fact, our final step runs in polynomial time; it is equivalent to the following simplified problem: we are given a graph and a set of nodes S in the graph that can be matched in a maximum matching, and the task is to find a maximum

---

[28]We conjecture that the requirement that every hospital size be regular can be relaxed (to a weaker definition of regular) or eliminated entirely.

[29]The allocation constructed in the proof can fail to be individually rational with low probability. However, using similar ideas as in the proof of Proposition 1, one can construct in *every* graph an individually rational allocation and show that in almost every graph the construction will be within the indicated efficiency bound.

matching that indeed matches all nodes in S. Since finding a maximum matching in a graph can be done using linear programming in polynomial time, one can add a linear constraint so that each node in S is indeed matched.[30,31]

7. Kidney exchange mechanisms

We have seen that a mechanism that is individually rational for hospitals need not be costly in terms of lost transplants, and individual rationality can be seen as a necessary condition for full participation in a world in which a hospital can withdraw participation after seeing the allocation proposed by the centralized mechanism. But a mechanism that makes it individually rational for hospitals to participate may still not be sufficient to elicit full participation if it does not also make it a dominant strategy, or a Bayesian equilibrium, for hospitals to reveal all their patient–donor pairs. We next begin the exploration of the incentive properties of exchange mechanisms, starting (as in the case of individual rationality) with some negative worst-case results.

A kidney exchange *mechanism*, φ, maps a profile of incompatible pairs $V = (V_1, V_2, \ldots, V_n)$ to an allocation, denoted by $\varphi((V_h)_{h \in H_n})$. A mechanism φ is IR if for every profile V, $\varphi(V)$ is IR. Efficient and maximal mechanisms are defined similarly.

Every kidney exchange mechanism φ induces a game of incomplete information $\Gamma(\varphi)$ in which the players are the hospitals. The type of each hospital h is its set of incompatible pairs. The realized type will be denoted by V_h and at this point we assume no prior over the set of types. At strategy σ_h, hospital h reports a subset of its incompatible pairs $\sigma_h(V_h)$. For any strategy profile σ, let $\sigma(V) = (\sigma_1(V_1), \ldots, \sigma_n(V_n))$ be the profile of subsets of pairs each hospital submits under σ given V. Therefore, for any profile $V = (V_1, \ldots, V_n)$, at strategy profile σ, mechanism φ chooses the allocation $\varphi(\sigma(V))$.

A kidney exchange mechanism does not necessarily match all pairs in $V_{H_n} = \bigcup_{h \in H_n} V_h$, either because it did not match all reported pairs or because hospitals did not report all pairs. Therefore, we assume that each hospital also chooses an allocation in the set of its pairs that are not matched by the mechanism. Formally, let φ be a kidney exchange mechanism, and let σ be a strategy profile and V_h be the type of each hospital. After the mechanism chooses $\varphi(\sigma(V))$, h finds an allocation in $V_h \setminus \varphi(\sigma(V))(V_h)$, where $\varphi(\sigma(V))(V_h)$ is the set of all pairs in V_h that are matched by the allocation $\varphi(\sigma(V))$. In particular, every hospital $h \in H_n$ has an allocation function φ_h that maps any set of pairs X_h to an allocation $\varphi_h(X_h)$.

Since each hospital wishes to maximize the number of its own matched pairs, the utility of hospital h, u_h, at profile V and strategy profile σ, is defined by the number of pairs in V_h that are matched by the centralized match plus the number of its remaining pairs that h can match using internal exchanges:

$$u_h(\sigma_h(V_h), \sigma_{-h}(V_{-h})) = |\varphi(\sigma(V))(V_h)| + \left|\varphi_h\big(V_h \setminus \varphi(\sigma(V))(V_h)\big)(V_h)\right|. \tag{1}$$

---

[30]If the second step uses also three-way exchanges, finding a maximum allocation is well known to be an NP-hard problem as Abraham et al. (2007) showed.

[31]There is also a computational difference with the algorithm Toulis and Parkes (2011) provide; their algorithm selects an efficient allocation repeatedly from the compatibility graph until they find one that is individually rational, which is computationally inefficient.

In the next section we study incentives of hospitals in the games induced by kidney exchange mechanisms.

8. Incentives

Loosely speaking, most of the kidney exchange mechanisms presently employed choose an efficient allocation in the (reported) exchange pool.[32] As already emphasized, maximizing the number (or the weighted number) of transplants in the pool of patient–donor pairs reported by hospitals is not the same as maximizing the number of transplants in the whole pool unless the whole pool is reported. We next consider the tensions between achieving efficiency and making reporting of the whole pool a dominant strategy for each hospital.

8.1 *Strategyproofness: Negative results for compatibility graphs*

Section 3 showed that for any largest feasible exchange size $k > 2$, no individually rational mechanism can be efficient and obtained discouraging worst-case bounds (although efficiency can be achieved for $k = 2$). Here we show that for $k \geq 2$, no mechanism that always produces a k-maximal allocation (even if not efficient) can be individually rational and strategyproof, again with discouraging worst-case bounds.

A mechanism φ is strategyproof if it makes it a dominant strategy for every hospital to report all of its incompatible pairs in the game $\Gamma(\varphi)$. Formally, φ is *strategyproof* if for every hospital h, every V_h, every strategy σ'_h, and every V_{-h},

$$u_h(\varphi(V_h, V_{-h})) \geq u_h(\varphi(\sigma'_h(V_h), V_{-h})).$$

In unpublished notes from 2007, Roth, Sönmez, and Ünver showed that (even for a maximum exchange size $k = 2$) the following proposition holds.

PROPOSITION 2 (Roth, Sönmez, and Ünver). *No IR mechanism is both maximal and strategyproof.*

Strategyproof mechanisms do exist, e.g., a mechanism that chooses allocations that maximize the number of matched nodes using only internal exchanges. By allowing randomization between allocations (in particular, allowing inefficient allocations to be chosen with positive probability), one can hope to obtain outcomes that are close to efficient in expectation. Unfortunately, building on the proof of Proposition 2, both deterministic and randomized mechanisms are not close to being efficient (again even for $k = 2$).[33]

PROPOSITION 3. *For $k \geq 2$, (i) no IR strategyproof mechanism can always guarantee more than $\frac{1}{2}$ of the efficient allocation and (ii) no IR strategyproof (in expectation) randomized mechanism can always guarantee more than $\frac{7}{8}$ of the efficient allocation.*

---

[32]The mechanisms often maximize a *weighted* sum of transplants, rather than a simple sum, to implement priorities, such as for children and for how difficult it is to match a patient (due to high PRA levels).

[33]Strategyproofness in the randomized case means that for any reports of other hospitals, no hospital h is better off in expectation by reporting anything other than its type V_h.

Ashlagi et al. (forthcoming(b)) study dominant strategy mechanisms for $k = 2$ and provide a strategyproof (in expectation) randomized mechanism that guarantees 0.5 of the 2-efficient allocation.[34] But it remains an open question whether the bounds established in this section can be achieved.

Strategyproofness is independent of any probability distribution of the underlying compatibility graphs. As in the case for individual rationality, using information about the (approximate) distribution of compatibility graphs might be useful for finding mechanisms that can achieve (almost) efficient allocations as Bayesian equilibria.[35] We proceed by studying the Bayesian setting in a large random kidney exchange program, in the spirit of recent advances in the study of two-sided matching in large markets (cf. Immorlica and Mahdian 2005, Kojima and Pathak 2009, Kojima et al. 2013, and Ashlagi et al. forthcoming(a)).

8.2 *The Bayesian setting*

To study hospitals' incentives in a given mechanism, we consider a Bayesian game in which hospitals strategically report a subset of their set of incompatible pairs and the mechanism chooses an allocation. Thus a *kidney exchange game* is now a Bayesian game $\Gamma(\varphi) = (H, (T_h)_{h \in H}, (u_h)_{h \in H})$, where H is the set of hospitals, u_h is the utility function for hospital h, and T_h is the set of possible private types for each hospital, drawn independently from a known distribution. The type for each hospital is the subgraph induced by its pairs in the random compatibility graph. In particular, the random compatibility graph is drawn and then the nodes of the graph are partitioned randomly among the hospitals. The set of feasible partitions is determined by the possible hospitals' sizes.

The expected utility for hospital h at strategy profile σ given V_h is

$$E_{V_{-h}}[u_h(\sigma_h(V_h), \sigma_{-h}(V_{-h}))],$$

where the utility function u_h is defined as in (1). In words, the hospitals wishes to maximize the expected number of its own matched pairs either by the mechanism or by itself. Let σ be a strategy profile and let $\epsilon > 0$. Strategy σ_h is an ϵ-best response against σ_{-h} if for every σ_h' and every V_h,

$$E_{V_{-h}}[u(\sigma_h(V_h), \sigma_{-h}(V_{-h}))] \geq E_{V_{-h}}[u(\sigma_h'(V_h), \sigma_{-h}(V_{-h}))] - \epsilon.$$

σ is an ϵ-Bayes Nash equilibrium if for every hospital h, σ_h is an ϵ-best response against σ_{-h}. For $\epsilon = 0$, σ is the standard Bayes Nash equilibrium.

A particular strategy of interest is the *truth-telling* strategy: a hospital always reports its entire set of incompatible pairs. To analyze mechanisms for large random exchange pools, it will be useful to consider a sequence of random kidney exchange games $(\Gamma^1(\varphi), \Gamma^2(\varphi), \ldots)$, where $\Gamma^n(\varphi) = (H_n, (T_h)_{h \in H_n}, (u_h)_{h \in H_n})$ denotes a random kidney exchange game with $|H_n| = n$ hospitals.

---

[34]The model in Ashlagi et al. (forthcoming(b)) does not allow hospitals to choose an internal allocation after the mechanism has chosen an allocation. However, their algorithm works in our model.

[35]An efficiency approximation gap between the Bayesian approach and prior free approach has been shown, for example, by Babaioff et al. (2010) in an online supply problem.
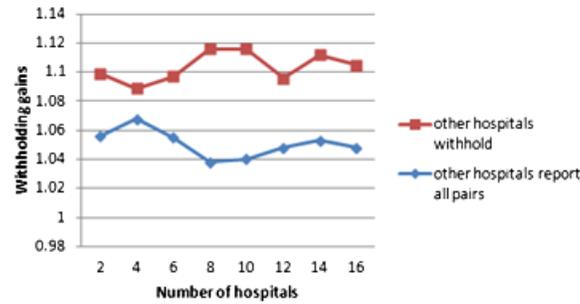
FIGURE 4. Withholding internal matches versus reporting truthfully in the status quo mechanism ($k = 3$). Hospitals have, on average, 10 pairs.

No. of hospitals	Reporting truthfully	Withholding
10	52.44	44.79
12	64.19	55.26
14	76.72	66.18
16	88.81	76.89

TABLE 2. Number of transplants achieved in the status quo (random efficient) mechanism under two different strategies: (i) each hospital withholds an efficient internal allocation and (ii) each hospital reports truthfully. Each hospital has, on average, 10 pairs. k is set to 3.

8.2.1 *Toward a new mechanism*   A stylized version of status quo kidney exchange mechanisms is to choose (randomly) an efficient allocation. We observed that such a mechanism can violate individual rationality. Moreover, it is often the case that a hospital will benefit (nonnegligibly) from withholding pairs. For example, if a hospital has two pairs, A-O and O-A, it can internally match, it is better off withholding them, since only a fraction of the underdemanded O-A pairs in the pool will be matched.

We simulated such a (status quo) mechanism and examined two types of behavior for hospitals: *truth-telling*, in which a hospital reports all its incompatible pairs, and a naive strategy called *withhold internal matches*, in which a hospital withholds a maximum set of pairs it can match internally. As Figure 4 shows, withholding provides more transplants, on average, than truth-telling for an arbitrary hospital given that all other hospitals are truth-telling. The benefit from withholding becomes even larger when all other hospitals also withhold internal matches.

If all hospitals use the withhold internal matches strategy under the status quo mechanism, the total number of transplants achieved (by the mechanism and the internal matches) results in more than 12% efficiency loss as shown in Table 2. See Section 9 for further simulations.

The underdemanded pairs are the ones that "compete" to be matched; an attempt to solve this problem is by guaranteeing each hospital to match at least as many underdemanded pairs as it can internally match (and randomly choose maximum allocations under this constraint). Unfortunately, hospitals will often still benefit from withholding.
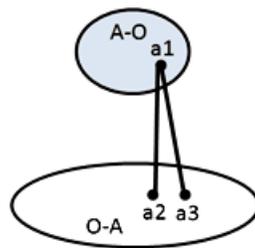
FIGURE 5. Hospital a has one overdemanded A-O pair and two O-A pairs, and it can internally match either one.

To see this, suppose all hospitals but a report truthfully and suppose a has the compatibility graph in Figure 5. Using the fact that any O-A pair in the graph is likely to be chosen with probability $p < \frac{1}{2}$, one can show that withholding only the overdemanded A-O pair a_1 makes a strictly better off, since it can then match one O-A pair internally if either of them is not matched by the mechanism. That is, hospitals will sometimes have an incentive to hold in reserve their overdemanded pairs to be matched ex post to underdemanded pairs left unmatched by the centralized mechanism.

8.2.2 *A new mechanism*   One way to prevent hospitals from withholding overdemanded pairs is to give priority to underdemanded pairs from hospitals that contribute overdemanded pairs that could be part of internal matches. To do this, we propose an "underdemanded lottery" that will determine which underdemanded pairs will be matched. We first give a sketch of the lottery for a setting with only A-O and O-A pairs, and illustrate it with a simple example.

The underdemanded lottery consists of the following two main steps that output a set of O-A pairs S, which contains the same number of O-A pairs as there are reported A-O pairs:

1. For each hospital, select randomly a maximum set of O-A pairs it can internally match and add them to S.

2. Consider a bin that contains for each hospital the same number of balls as the number of its reported underdemanded O-A pairs. Until S reaches the target size,[36] iteratively draw balls without replacement and after each draw, if the ball belongs to hospital h, add one of h's O-A pairs that has not yet been chosen (if any still exist) to S.

Note that in the beginning of the second step, each hospital begins with the same number of balls as its O-A pairs even if some of its O-A pairs have *already been chosen* in the first step.

EXAMPLE 1.   Consider a hospital h that has three pairs $a1$, $a2$, and $a3$, and its compatibility graph can internally match $a1$ to $a2$ and to $a3$ as in Figure 6(a). For simplicity, assume that no other two pairs belong to the same hospital. We show that the O-A pairs are

---

[36]One should think of this size as the number of reported A-O pairs.

(a) Hospital a reports $a1$ (b) Hospital a withholds $a1$
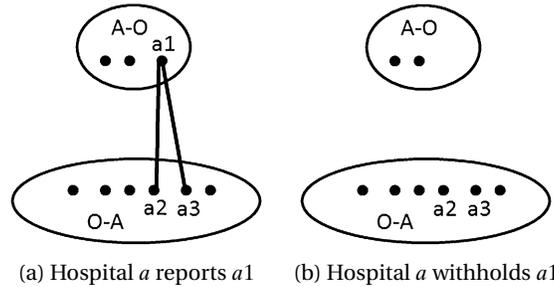
FIGURE 6. Hospital a has one overdemanded pair, $a1$, and two underdemanded pairs, $a2$ and $a3$, and can match internally $a1$ to $a2$ and to $a3$.

chosen by the underdemanded lottery and reporting $a1$ (Figure 6(a)) results in a higher utility for h than withholding $a1$ (Figure 6(b)).[37] In the former case, the lottery will select three O-A pairs (to be matched to the A-O pairs) and in the latter case, the lottery will select two O-A pairs.

If h withholds $a1$, the underdemanded lottery will choose randomly without replacement two O-A pairs (so as to match them to the two A-O) pairs. In this case, the probability that none of a's O-A pairs will be chosen is $\binom{4}{2}/\binom{6}{2} = \frac{12}{30}$; the probability that one of its O-A pairs will be chosen is $\frac{2}{6} \cdot \frac{4}{5} + \frac{4}{6} \cdot \frac{2}{5} = \frac{16}{30}$ (choosing one of a_2 or a_3 in the first or second round of the second step of the lottery); the probability that both of its O-A pairs will be chosen is $\frac{2}{6} \cdot 15 = \frac{2}{30}$. Since h can internally match $a1$ to one of its O-A pairs if it has not been selected by the lottery, its expected utility (expected total number of transplants) is $2 \cdot \frac{12}{30} + 3 \cdot \frac{16}{30} + 2 \cdot \frac{2}{30} = 2\frac{8}{15}$.

If h reports $a1$, either $a2$ or $a3$ is chosen in the first step of the lottery, say $a2$. The probability that $a3$ will not be chosen in the second step of the lottery is $\frac{4}{6} \cdot \frac{3}{5} = \frac{12}{30}$ and, therefore, the probability that $a3$ will be chosen in that step is $\frac{18}{30}$. Therefore, a's expected utility in this case is $2 \cdot \frac{12}{30} + 3 \cdot \frac{18}{30} = 2\frac{9}{15}$. Note the incentive to report $a1$: it is that, after $a2$ is chosen, the probability that $a3$ will also be chosen is the same as the probability that either one of $a2$ or $a3$ would have been chosen if $a1$ had not been reported.   ◊

In this section, we present a mechanism for kidney exchange that uses this kind of lottery to make truth-telling an approximate Bayes Nash equilibrium, assuming that hospitals satisfy a stronger regularity condition. This stronger regularity condition, which now deals with each underdemanded type and its reciprocal overdemanded type separately, will allow us to separate the reporting problem for each type of overdemanded pair. This will allow a mechanism in which there is no incentive to withhold an overdemanded pair of some type so as to influence the match probability of an underdemanded pair that is not of its reciprocal type.

We say that a size $c > 0$ of a hospital is *strongly regular* if for every underdemanded type X-Y, the expected maximum number of X-Y pairs with which it can be internally matched is less than the expected number of overdemanded Y-X pairs in its pool.

---

[37]An implicit assumption in this example is that there exists a perfect matching between the A-O pairs and the chosen O-A pairs.

A formal definition is given in Appendix A.6. Using simulations, we find that hospitals of size up to at least $c = 30$ are strongly regular.

Throughout this section, we assume hospitals' sizes are strongly regular and bounded. The mechanism we introduce provides an allocation that uses only two-way exchanges with similar properties to the one constructed in the proof of Theorem 2. The following is a high level description of the new mechanism:

1. Find a maximum allocation in the graph induced by all self-demanded pairs.

2. Find a maximum allocation in the graph induced by all A-B and B-A pairs.

3. Choose which underdemanded X-Y pairs to match and match them to overdemanded Y-X pairs.

The missing key part is how to choose the underdemanded pairs that will be matched in part 3. We will use a lottery like the one described in the example, called the *underdemanded lottery*, to determine for each underdemanded type X-$Y \in \mathcal{U}$ a *set* of X-Y pairs, denoted by $S_h(X\text{-}Y)$, that will be matched for each hospital h (ideally we want to match all overdemanded pairs, so the total number of X-Y pairs that will be matched equals the total number of Y-X pairs in the pool).

We now formally describe the underdemanded lottery for a given underdemanded type X-$Y \in \mathcal{U}$. For each h, $S_h(X\text{-}Y)$ will be initialized to be a set of X-Y pairs with maximum cardinality that h can match internally, and the lottery will output for each hospital a set of pairs that are chosen to be matched. We need the following notation: let $\mathcal{M}_T^V$ be the set of allocations in V that maximize the number of matched pairs in V whose type belongs to T (whenever $T = \{t\}$ singleton, we will just write $\mathcal{M}_t^V$).

UNDERDEMANDED LOTTERY.   1. *Input*. A set of hospitals H_n, a profile of subsets of pairs $(B_1, \dots, B_n)$, an underdemanded type X-Y, and an integer $0 < \theta < |\tau(B_{H_n}, X\text{-}Y)|$, which is interpreted as the number of X-Y pairs that we want to choose in total.[38]

2. *Initialization*. For each hospital h, let $Q_h(X\text{-}Y) = |\tau(B_h, X\text{-}Y)|$ and let $S_h(X\text{-}Y)$ be an arbitrary maximum set of X-Y pairs h can internally match in B_h.[39]

3. *Main Step*. Let J be a bin containing $Q_h(X\text{-}Y)$ balls for each hospital h. As long as $\sum_{h \in H_n} |S_h(X\text{-}Y)| < \theta$, the following choice holds:

    (a) Choose uniformly at random a ball from J without replacement. If the ball belongs to hospital h, then add an arbitrary X-Y pair to $S_h(X\text{-}Y)$ from $B_h \setminus S_h(X\text{-}Y)$ if such exists.

In Example 1, $S_a(\text{O-A})$ is initialized to be either $\{a_2\}$ or $\{a_3\}$, say $S_a(\text{O-A}) = \{a_2\}$. However, *two* balls are initially placed in the bin J for hospital a, and if either one of them is

---

[38]The parameter θ is not set here to be the number of Y-X pairs in B_{H_n} since, as we shall see later, the mechanism will apply the lottery twice, each time with a different set of $\frac{1}{2}n$ hospitals and θ will be the number of Y-X pairs in the set of other $\frac{1}{2}n$ hospitals. This will be further discussed below.

[39]Formally, $S_h(X\text{-}Y) = \tau(M_{X\text{-}Y}^{B_h}(B_h), X\text{-}Y)$ for some $M_{X\text{-}Y}^{B_h} \in \mathcal{M}_{X\text{-}Y}^{B_h}$.

drawn, a_3 is added to S_a(O-A). Therefore, the fact that the hospital can internally match one of its underdemanded pairs increases the probability that another of its underdemanded pairs of the same type will be matched.

We are now ready to state the mechanism formally. For simplicity of exposition, we assume throughout this section that n is even. All results hold when n is odd (see also footnote 41 below).

THE BONUS MECHANISM. 1. *Input.* A set of hospitals $H_n = \{1, \dots, n\}$ and a profile of incompatible pairs $(B_1, B_2, \dots, B_n)$, each of a strongly regular size.

2. *Step 1* (Match self-demanded pairs). Find a maximum allocation, M_S, in the graph induced by all self-demanded pairs B_{H_n}.

3. *Step 2* (Match A-B and B-A pairs). For each hospital h, choose randomly an allocation $M_h \in \mathcal{M}_{\mathcal{R}}^{B_h}$.[40] Find a maximum allocation, M_R, in the graph induced by A-B and B-A pairs among those that maximize the number of matched pairs in $\bigcup_{h \in H_n} \tau(M_h(B_h), \mathcal{R})$.

4. *Step 3* (Match overdemanded and underdemanded pairs). Partition the set of hospitals into two sets $H_n^1 = \{1, \dots, \frac{1}{2}n\}$ and $H_n^2 = \{\frac{1}{2}n + 1, \dots, n\}$. For each underdemanded type, X-$Y \in \mathcal{U}$ and for each $j = 1, 2$, perform the following substeps.[41]

    Step 3(a) Set $\theta_j(Y\text{-}X) = |\tau(B_{H_n^{3-j}}, Y\text{-}X)|$ to be the number of Y-X pairs in the set $B_{H_n}^{3-j}$. Then, using the underdemanded lottery procedure with the inputs $(B_h)_{h \in H_{n^j}}$, $\theta_j(Y\text{-}X)$, and X-Y, construct one subset $S_h(X\text{-}Y)$ for each hospital in $h \in H_n^j$.

    Step 3(b) Find a maximum allocation $M_{X\text{-}Y}^j$ in the subgraph induced by the sets of pairs $\bigcup_{h \in H_n^j} S_h(X\text{-}Y)$ and $\tau(B_{H_n^{3-j}}, Y\text{-}X)$.[42]

5. *Step 4* (Output). Let $M_U = \bigcup_{j=1,2} \bigcup_{X\text{-}Y \in \mathcal{U}} M_{X\text{-}Y}^j$. Output $M_S \cup M_R \cup M_U$.

We can now state our second main result.

THEOREM 3. *Let H_n be a set of hospitals. If every hospital size is strongly regular, the truth-telling strategy profile is an $\epsilon(n)$ Bayes Nash equilibrium in the game induced by the Bonus Mechanism, where $\epsilon(n) = o(1)$. Furthermore, for any $\epsilon > 0$, the efficiency loss*

---

[40]Recall that $\mathcal{R} = \{$A-B, B-A$\}$.

[41]So as to choose the sets of underdemanded pairs of each type that will be matched, we partition the set of hospitals into sets, each with $\frac{1}{2}n$ hospitals (if n is odd, one set will have one more hospital than the other). For each set in the partition, we will match the overdemanded pairs of the hospitals in one set to the chosen underdemanded pairs of the hospitals in the other set so as to avoid lack of independence (see also the proof of Theorem 2).

[42]The size of $|\bigcup_{h \in H_n^j} S_h(X\text{-}Y)|$ will equal $|\tau(B_{H_n^{3-j}}, Y\text{-}X)|$ with high probability and, therefore, the maximum allocation in this subgraph will match with high probability all pairs in $\bigcup_{h \in H_n^j} S_h(X\text{-}Y)$.

*under the truth-telling strategy profile in almost every $D(H_n)$ is at most $\mu_{\text{AB-O}}m + \epsilon\mu_{\text{A-B}}m$,
where m is the number of pairs in the pool.*

We conjecture that, here too, the strong regularity assumption can be relaxed and even entirely eliminated. In the next section, we provide simulations that demonstrate the effectiveness of our mechanism.

REMARKS. 1. Toulis and Parkes (2011) show a similar result only under their assumption that each hospital has a canonical structure (see the first remark at the end of Section 6). Another important difference is that they do not allow hospitals to withhold a single pair (only internal allocations). As Example 1 illustrates, a hospital will have an incentive to withhold only an overdemanded pair (and these pairs are exactly the ones we wish to incentivize hospitals to report).

2. The Bonus Mechanism algorithm can be adapted so that in each step it allows the output allocation to use only two-way exchanges and Theorem 3 will still hold (the internal allocations in the algorithm still use two- and three-way exchanges). This implies that with the adaptation, the Bonus Mechanism algorithm runs in polynomial time (assuming it indeed finds an allocation with the desired properties, which we show to exist in almost every graph). The arguments are similar to the arguments for the complexity of the construction of an almost efficient individually rational allocation (see the second remark in Section 6).

3. When the compatibility graph is not too large, there is often the knowledge about which pairs are "hard" and which pairs are "easy" to match (see, e.g., Ashlagi et al. 2012). The idea of the underdemanded lottery, which is the key part in our mechanism, can be adapted so that hospitals will indeed be incentivized to enroll their easy-to-match pairs.

9. SIMULATIONS

Simulations are useful to evaluate whether the conclusions of limit theorems apply even in relatively small finite settings. We first explain the Monte Carlo simulations we have conducted. To generate incompatible pairs, we follow our definition of a random compatibility graph, which is also consistent with the method used in Saidman et al. (2006). First we create a patient and donor with blood types drawn from the national distributions as reported by Roth et al. (2007). Each patient is also assigned a percentage reactive antibody (PRA) level also drawn from a distribution as described in Roth et al. (2007). The patient PRA is interpreted as the probability of a positive cross-match (tissue-type incompatibility) with a random donor. If the generated pair is compatible, i.e., if they are both blood-type compatible and have a negative cross-match, they are discarded (this captures the fact that compatible pairs go directly to transplantation). Otherwise, the population generation continues until each hospital accumulates a certain number of incompatible pairs. In all simulations, we have bounded the number of pairs in total by
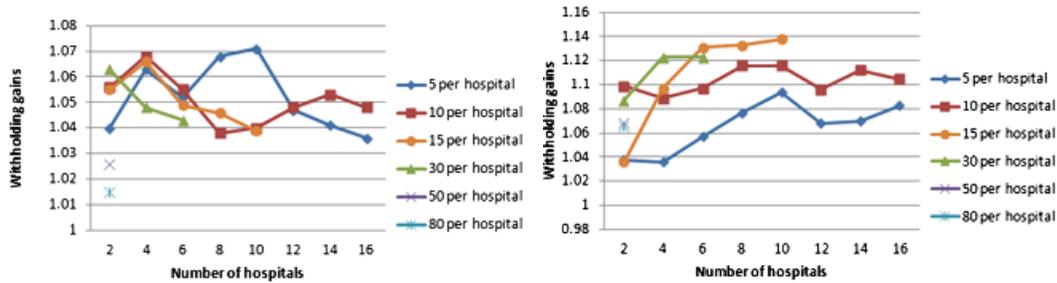
FIGURE 7. Withholding internal matches versus reporting truthfully in the status quo mechanism ($k = 3$).
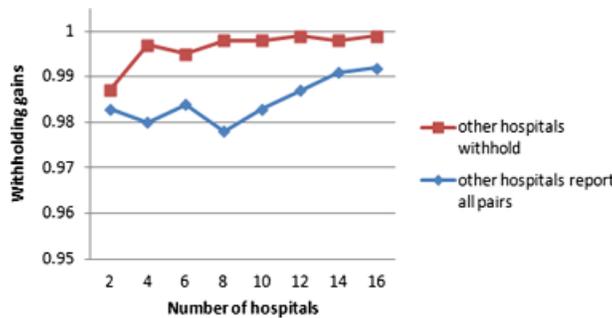
FIGURE 8. Withholding versus not withholding under the Bonus Mechanism. Each hospital has an average of 10 pairs.

180 per iteration (thus, for some scenarios there are fewer hospitals than others).[43] For each random population, we ran 500 trials.

Figure 7 provides, for various average-sized hospitals, the gain from withholding under a current status quo-like mechanism that randomly chooses an allocation that maximizes the number of transplants.

We also simulated the gains from withholding under the Bonus Mechanism. Figure 8 provides the results for the gains when hospitals average 10 incompatible pairs. Notice that hospitals never gain from withholding. We obtained very similar results for different average-sized hospitals.

To further test the Bonus Mechanism, we let the average size of hospitals vary. Figure 9 shows, for different average-sized hospitals, that withholding is not beneficial under the Bonus Mechanism. These results not only support Theorem 3, but also our conjecture that the theorem holds without the strong regularity assumption.

We further simulated the efficiency gains under the Bonus Mechanism by comparing to a status quo-like mechanism, and assuming that in the Bonus Mechanism, hospitals

---

[43]In practice, the current popular exchange programs receive fewer than 50 pairs per period. When allowing three-way exchanges, finding an allocation that maximizes the number of matches is an NP-hard problem (see Abraham et al. 2007 and Biro et al. 2009). The compatibility graph is generally sparse enough, however, that the problem is tractable in reasonably sized populations.
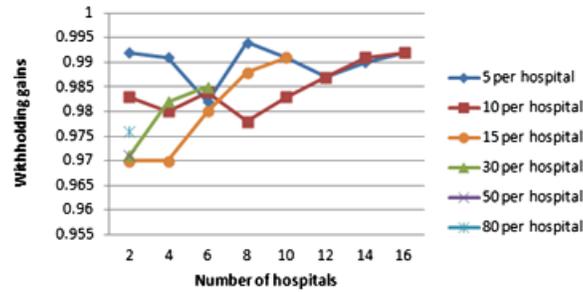
FIGURE 9. Withholding versus not withholding under the Bonus Mechanism.

(a) Percentage of lost matches

(b) Percentage of lost highly sensitized matches
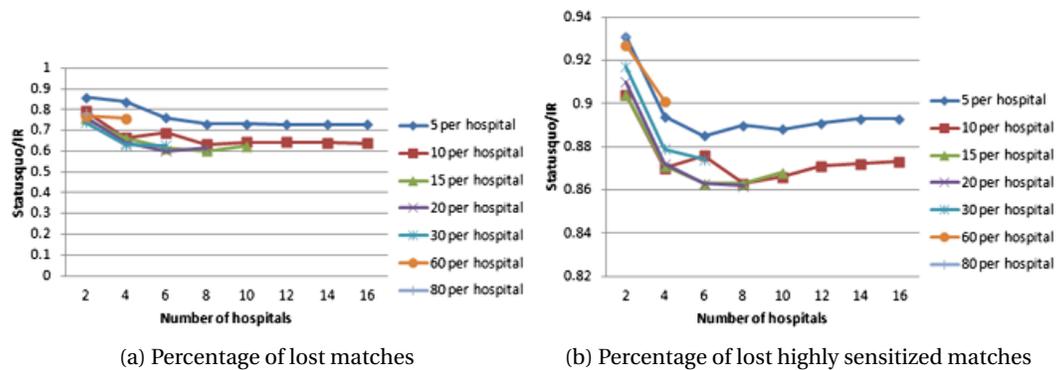
FIGURE 10. Ratio between the number of matches achieved under an "efficient" mechanism to the number of matches achieved under the Bonus Mechanism.

report truthfully, and under the status quo-like mechanism, hospitals withhold an internal maximum allocation. Figures 10(a) and 10(b) provide the percentage of number of lost matches and the percentage of lost high PRA (highly sensitized) matches.

Finally, we compared the number of matches obtained under the Bonus Mechanism to the number of transplants obtained under a mechanism that randomly chooses the maximum number of matches, assuming that under both mechanisms hospitals report truthfully. The results are give in Figure 11. The results also support our conjecture that even with "larger-sized" hospitals, i.e., without the regularity assumption, there is an individually rational allocation that is almost efficient.

10. CONCLUSIONS AND OPEN QUESTIONS

There are a number of ways in which barter may be inefficient. Jevons (1876) famously pointed to the double coincidences needed to make pairwise exchanges (a difficulty that is only partially eased by allowing larger exchanges and further relieved when chains are possible). A second difficulty is that profitable but inefficient transactions may take place that prevent efficient ones from occurring (cf. Roth and Postlewaite 1977).[44] A well

---

[44]Roth and Postlewaite look at the model proposed by Shapley and Scarf (1974) in which traders each have only a single indivisible good, and observe that there are inefficient transactions of this sort in the
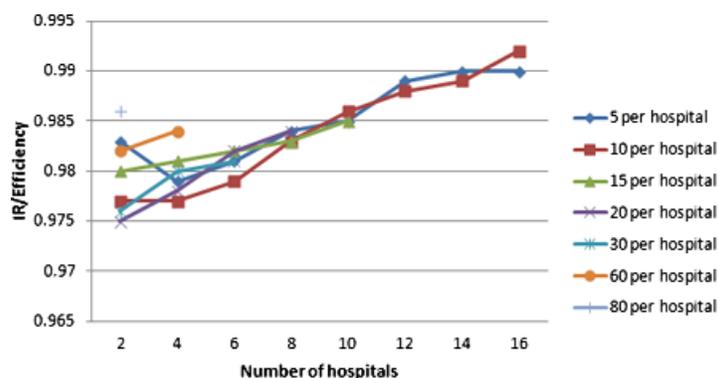
FIGURE 11. The cost from using IR and not an efficient mechanism assuming hospitals fully participate.

designed centralized clearinghouse can address not only the first problem (by making a thick market), but also the second, by guaranteeing hospitals that they will not suffer by foregoing potentially inefficient internal exchanges and, instead, reporting all their patient–donor pairs.

The problem of inefficient exchanges has come to the fore as kidney exchange in the United States has grown from being carried out rarely in only a few hospitals to being carried out regularly in a variety of kidney exchange networks of hospitals, and is presently being explored at the national level. The National Kidney Paired Donation Pilot Program was approved by the OPTN/UNOS Board of Directors in June 2008, and ran its first two match runs in October and December 2010, with 43 patient–donor pairs in October and 62 in December, registered by kidney exchange consortia representing 77 transplant programs. For the purposes of the present paper, it is notable that only a small fraction of the patient–donor pairs registered in the participating hospitals were enrolled in the national pilot program.[45] So the problem of full participation by hospitals is both real and timely. It has also begun to be observed in the active kidney exchange networks that are fully operational.

---

core of the game that are not supported by any market-clearing prices. Consider three traders {1, 2, 3}, with endowments $w = (w1, w2, w3)$ and preferences such that each trader can get his first choice via a three-way exchange that yields the allocation $x = (w3, w1, w2)$. There can, nevertheless, be a profitable two-way exchange that yields, e.g., $y = (w2, w1, w3)$ via a trade between 1 and 2, and that gives 1 his second choice and 2 his first choice. This is in the core of the game when the initial endowments are w, but not after the trade has taken place and the (new) endowment is y (since, from y, 1 and 3 could trade $w2$ and $w3$). Kidneys of course cannot be reexchanged after being transplanted. But a centralized clearinghouse can take into account the potential trade between 1 and 2, and make it rational for them to enter the centralized mechanism, knowing that it will produce an allocation x that must be at least as good for them as y (cf. Roth 1982).

[45] We hasten to note that there are many reasons other than the incentive problems discussed here that contribute to this initial very low participation rate. These include the new bureaucratic procedures for enrolling patients, the novelty and lack of track record of the national program, the desire to start small and see what happens, the exclusion of chains and nondirected donors, etc. See http://optn.transplant.hrsa.gov/resources/KPDPP.asp.

One way to solve this problem is by forcing hospitals to disclose all their pairs and, thus, have incompatible living donors be a national resource as happens with cadaver organs. This will be very difficult with private hospitals that act independently from one another.[46]

The present paper observes that one contributory cause of the lack of full participation is that the matching algorithms currently employed in practice do not make it individually rational for hospitals to always contribute all their patient–donor pairs. We show that, in worst cases, this could be very costly, but we prove that in large markets, it is possible to redesign the matching mechanisms to guarantee individually rational allocations to hospitals at very modest cost in terms of "lost" transplants. Note that these lost transplants are not really lost if, instead, hospitals would have withheld patient–donor pairs; on the contrary, we show that individually rational allocations produce a big gain in transplants compared to having hospitals withhold pairs.

To obtain analytical results about large markets, we approximate them as large random graphs whose properties we can study with limit theorems based on the classical results of Erdos and Renyi. But we also show by simulation with clinically relevant distributions of patients and donors that these main results apply on the scale of exchange we are presently seeing. The fast convergence we see in simulations suggests that these limit theorems from random graph theory may have much wider application than if convergence were slow.

The highly interconnected compatibility graphs that we see in the limit theorems do not approximate well the much sparser compatibility graphs we see in practice, which contain many very highly sensitized patients. One of several causes of the high percentage of highly sensitized patients is that many transplant centers are withholding their easy-to-match patient–donor pairs and only enrolling their hard-to-match pairs. This raises a number of open questions that are likely to arise in practice regarding this most vulnerable class of patients.

The first of these questions is how to model the situation facing highly sensitized patients, who will be only sparsely connected in the compatibility graph, because they may be compatible with a very small number of donors, even in a large graph of finite size. This is closely related to the second question, which is how to effectively integrate nondirected donors and chains with the cyclic exchanges that have been used initially in the national pilot program and that are the subject of the present paper. In addition to cycles of length k, there has been growing use of various kinds of chains in kidney exchange, and it remains an open question how the relative importance of chains and cyclic exchanges will change as the size of the pool (and the number of nondirected donors) grows large. It seems likely that even in large markets, chains will be especially helpful to the most highly sensitized patients (Ashlagi et al. 2012). A related question is how the composition of the patient pool changes dynamically, as easier-to-match pairs are matched and removed. Like the withholding of easy-to-match pairs studied here,

---

[46]Even if such a law exists, hospitals can still enroll all their pairs while specifying that only the internal matches are acceptable by, for example, specifying the acceptable travel distance for pairs involved in internal matches to be 0.

this will also impact the composition of kidney exchange pairs in ways that make the compatibility graphs relatively sparse.

A fourth open question is under what conditions individually rational and incentive compatible mechanisms exist that are as efficient as we have shown them to be under regularity conditions on the size of hospitals. We conjecture that these regularity conditions can be relaxed. In any case, such mechanisms could be useful in eliciting full participation in a full scale national exchange, as it appears from simulations that hospitals are, in fact, of regular size (although the largest hospitals may not be strongly regular). However, our results suggest that the benefits of a national exchange could also be realized if there was sufficient regulatory power to require transplant centers to either participate fully or not at all, since that would reduce the strategy space so that individual rationality would be the primary consideration.[47]

The final open question we raise here is how these strategic concerns would be different in a world in which the players are not only hospitals and a (single) centralized exchange, but in which there are multiple kidney exchange networks, some with strategic concerns of their own. This is, of course, the situation that is currently in place.

In conclusion, as kidney exchange has grown, the strategy sets, the strategic players, and, hence, the incentive constraints have changed. The new incentive issues, concerning full participation by hospitals, arise out of the growth of kidney exchange and are potential obstacles to further growth. These are problems shared with barter exchange generally and by marketplaces with money as long as there are both easy- and hard-to-trade goods (such as the markets for gastroenterologists mentioned in the Introduction). However, the results of this paper strongly suggest that these new barriers can also be overcome.

<p align="center">Appendix</p>

A.1 *Preliminaries*

We briefly describe here some results that will provide intuition and be building blocks in our proofs. A random graph $G(m, p)$ is a graph with m nodes and between each two different nodes, an *undirected* edge exists with probability p (p is a nonincreasing function of m). A bipartite random graph $G(m, m, p)$ consists of two disjoint sets of nodes V and W, each of size m, and an *undirected* edge between any two nodes $v \in V$ and $w \in W$ exists with probability p (no two nodes within the same set V or W have an edge between them). It will be useful to think of an undirected edge as two directed edges, one in each direction.

Throughout the paper, by saying just a "random graph," we will not refer to a specific type, but a graph that is generated by any of the graph generating processes defined in this paper (e.g., $D(m)$, $G(m, p)$, and $G(m, m, p)$).

---

[47]It might be possible for a single kidney exchange network to institute such a rule as a condition of participation, since the fact that hospitals could exercise discretion over whether or not to participate at all might not present too great a challenge to the discretion that hospitals and surgeons demand over treatment decisions.

For any graph theoretic property Q, there is a probability that a random graph G satisfies Q, denoted by $\Pr(G \models Q)$.

A *matching* in an undirected graph is a set of edges for which no two edges have a node in common. A matching is *nearly perfect* if it matches (contains) all but at most one node in the graph and is *perfect* if it matches all nodes.

Erdos and Renyi provided a threshold function $f(m) = (\ln m)/m$ such that for $p(m) \gg f(m)$, a perfect matching exist in $G(m, p(m))$. We state here a corollary of their result (see, e.g., Janson et al. 2000).

THEOREM 4 (Erdos–Renyi theorem). *Let Q be the property that there exists a nearly perfect matching. For any constant p, the following equalities hold.*

1. $\Pr(G(m, p) \models Q) = 1 - o(1)$.[48,49]

2. $\Pr(G(m, m, p) \models Q) = 1 - o(1)$.

REMARK ON THE CONVERGENCE RATE. The probability of a perfect matching in $G(m, p)$ and $G(m, m, p)$ converges to 1 at an exponential rate for any constant p. More precisely, as shown in Janson et al. (2000),

$$\Pr(G(m, m, p) \models Q) = 1 - O(me^{-mp}) = 1 - o(2^{-mp}),$$

and clearly a perfect matching in $G(m, p)$ exists with at least the same convergence rate. From now on, whenever we write $1 - o(1)$, it can be replaced with a rate of $1 - o(2^{-\alpha mp})$ for some constant α, where α will be the linear coefficient of the least probable pairs of blood types in the compatibility graph. So the convergence rate in all our large graph results is exponential.

A.2 *Proof of Theorem 1*

Let V be a set of nodes, and let M be a k-efficient allocation and M' be a k-maximal individually rational allocation in V. Since M' is k-maximal, every exchange in M must intersect an exchange in M' (otherwise a disjoint exchange could be added to M', contradicting maximality). Fix an exchange c with size $2 \le l \le k$ in M'. The maximum number of nodes that might be covered by M and not M' would be achieved if for each such exchange c, M contains $l - 1$ exchanges each of size k, which each intersect exactly one node of c (and M'). (Note that if all l nodes of c were in such exchanges, then M' would not be maximal.) For each such exchange c, M matches $(l - 1)k$ nodes and M' matches l nodes, so the ratio is $l/(l - 1)k$, which is minimized at $1/(k - 1)$ when $l = k$, giving the desired bound.

---

[48]For any two functions f and g, we write $f = o(g)$ if the limit of the ratio $f(n)/g(n)$ tends to zero when n tends to infinity.

[49]The Erdos–Renyi theorem showed stronger results, which assert that $r(m) = (\ln m)/m$ is a threshold function for the existence of a perfect matching; that is, if $p = p(m)$ is such that $r(m) = o(p(m))$, then the probability a nearly perfect matching exists converges to 1, and if $p(m) = o(r(m))$, the probability a nearly perfect matching exists converges to 0.
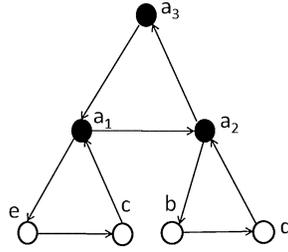
FIGURE 12. Worst-case efficiency loss from choosing an individually rational allocation ($k = 3$).

To see that the bound is tight, observe that the construction used to find the bound achieves it: fix some hospital a with k vertices and suppose that a has a single internal exchange consisting of all of its pairs (see Figure 12 for an illustration for $k = 3$). The bound $1/(k-1)$ is obtained by letting the k-efficient allocation in the underlying graph consist of exactly $k - 1$ exchanges, each of size k, at which a single pair of a is part of each such exchange. That is, the efficient allocation matches all but one of hospital a's pairs, each in exchanges of size k with $k - 1$ pairs from other hospitals.

A.2.1 *Proof of Lemma 1*   For each pair type X-Y, let $Z_{X\text{-}Y}(m)$ be the random variable that indicates the number of X-Y pairs in $D(m)$.

CLAIM 1. *Let $0 < \delta < 1$ and $D(m)$ be a random compatibility graph and consider the event*

$$B_\delta(m) = \{\forall X\text{-}Y \in \mathcal{P}, (1-\delta)\mu_{X\text{-}Y} m < Z_{X\text{-}Y}(m) < (1+\delta)\mu_{X\text{-}Y} m\}. \tag{5}$$

Then $\Pr[B_\delta(m)] = 1 - o(m^{-1})$.

PROOF. Let $D(m)$ be a random compatibility graph and let $\delta > 0$. By Hoeffding's bound (see, e.g., Alon and Spencer 2008), for every type X-Y,

$$\Pr[Z_{X\text{-}Y}(m) \notin ((1-\delta)m\mu_{X\text{-}Y}, (1+\delta)m\mu_{X\text{-}Y})] < e^{-m\mu_{X\text{-}Y}\delta^2/4} + e^{-m\mu_{X\text{-}Y}\delta^2/2} = o(m^{-1}).$$

Therefore,

$$\begin{aligned}
\Pr[B_\delta(m)] &= 1 - \Pr[\text{for some } X\text{-}Y \in \mathcal{P} : Z_{X\text{-}Y}(m) \notin ((1-\delta)m\mu_{X\text{-}Y}, (1+\delta)m\mu_{X\text{-}Y})] \\
&\geq 1 - \sum_{X\text{-}Y \in \mathcal{P}} \Pr[Z_{X\text{-}Y}(m) \notin ((1-\delta)m\mu_{X\text{-}Y}, (1+\delta)m\mu_{X\text{-}Y})] \\
&= 1 - o(m^{-1}),
\end{aligned}$$

where the last inequality follows since there are a finite number of pair types.   □

CLAIM 2. *Let $0 < \delta < \frac{1}{2}$ and let $D(m)$ be a random compatibility graph, and consider the event*

$$S_\delta(m) = \{|Z_{\text{A-B}}(m) - Z_{\text{B-A}}(m)| < m^{1/2+\delta}\}. \tag{6}$$

Then $\Pr[S_\delta(m)] = 1 - o(m^{-1})$.

Proof. By Hoeffding's bound,

$$\Pr\big(Z_{\text{A-B}}(m) \geq E[Z_{\text{A-B}}(m)] + m^{1/2+\delta}\big) \leq e^{-m^{2\delta}/2}.$$

Applying the same argument for B-A pairs, we obtain the result.   □

Proof of Lemma 1. Let $S_\delta(m)$ and $B_\delta(m)$ be as in Claims 2 and 1. By these claims, we obtain that the probability that either $S_\delta(m)$ or $B_\delta(m)$ does not hold is $o(m^{-1})$.   □

A.2.2 *Bounded directed random graphs: Definitions and Erdos–Renyi extensions*   In a random compatibility graph, the number of pairs of each type is not fixed. We will need Erdos–Renyi type results for random graphs in which the number of nodes as well as the number of edges are random.

We start by defining a vector that will represent bounds on the number of nodes of each pair type in a given subset of the compatibility graphs. For example, to represent the subgraph induced by all A-O pairs and all O-A pairs, by Lemma 1 and the event $B_\delta(m)$, we can use the vector $((1-\delta)\mu_{\text{A-O}}, (1+\delta)\mu_{\text{A-O}}, (1-\delta)\mu_{\text{O-A}}, (1+\delta)\mu_{\text{O-A}})$ for some $\delta < 1$; in particular, this vector is a tuple of coefficients for bounding from below and above the number of A-O pairs and the number of O-A pairs in this subgraph.

For any $r > 0$, a *quasi-ordered* vector is a vector $\bar\alpha_r = (\alpha_{0,1}, \alpha_{0,2}, \alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{r-1,1}, \alpha_{r-1,2})$, where $\alpha_{j,1} \leq \alpha_{j,2}$ for all $0 \leq j < r$ and $\alpha_{0,1} \leq \alpha_{1,1} \leq \cdots \leq \alpha_{r-1,1}$.[50]

The vector $\bar\alpha_r$ is called *feasible* if at most one pair type could have zero number of nodes, that is, $\alpha_{0,2} > 0$ and for every $j \geq 1$, $\alpha_{j,1} > 0$. Let $\bar\alpha_r$ be a feasible vector. We say that a tuple of r sets of nodes $(W_0, \ldots, W_{r-1})$ are $(\bar\alpha_r, m)$-*feasible* if for each $0 \leq j < r$, the interval $[\alpha_{j,1}m, \alpha_{j,2}m]$ contains at least one integer and if the sizes of these sets are drawn from some distribution over all possible r-tuples of integers that belong to $[\alpha_{0,1}m, \alpha_{0,2}m] \times \cdots \times [\alpha_{r-1,1}m, \alpha_{r-1,2}m]$. Note that for every sufficiently large m, the interval $[\alpha_{j,1}m, \alpha_{j,2}m]$ contains at least one integer if an only if $\alpha_{j,1} < \alpha_{j,2}$ or $\alpha_{j,1} = \alpha_{j,2}$ is an integer.

Definition 2 (Bounded directed random graphs). A graph is called a *bounded directed random graph*, denoted by $D(\bar\alpha_1, m, p)$, if it is generated as follows. A $(\bar\alpha_1, m)$-feasible set of nodes is generated and between each two nodes v, w, a directed edge is generated from v to w with probability *at least* p.[51]

A graph is called a *r-bounded directed random graph*, denoted by $D(\bar\alpha_r, m, p)$, if it is generated as follows: first, $r \geq 2$ distinct sets of nodes $W_0, W_1, \ldots, W_{r-1}$ that are $(\bar\alpha_r, m)$-feasible are generated. Then for each $i = 0, 1, \ldots, r-1$ and for each two nodes $v \in W_i$, $w \in W_{i+1}$ (i is taken modulo r), a directed edge is generated from v to w with probability at least p.

The definition of a bipartite graph can naturally be extended to an r-partite graph that contains r sets of nodes, each of size exactly m, and edges are generated as in Definition 2. Whenever there is no confusion, we will refer also to an r-bounded directed

---

[50]The vector is called quasi-ordered since only the lower bounds are ordered.

[51]Note that for $\alpha_{0,1} = \alpha_{0,2} = 1$, the number of nodes is m.

random graph as an r-partite graph. Note that in any r-partite graph, only exchanges of size $k = qr$ for positive integers q are feasible. (When $r = 1$, we think about subgraphs induced by self-demanded pairs of some given type. When $r = 2$, we think about subgraphs with potential two-way exchanges such as O-A, A-O, and when $r = 3$, we think about subgraphs with potential three-way exchanges such as AB-O, O-A, A-AB.)

LEMMA 2. *Let* $0 < p < 1$.

1. *For any feasible vector* $\bar{\alpha}_1$, *almost every large* $D(\bar{\alpha}_1, m, p)$ *has a nearly perfect allocation using exchanges of size* 2 *(i.e., an allocation that matches all nodes but at most one) and a perfect allocation for any* $k \geq 3$ *(i.e., an allocation that matches all nodes).*

2. *Let* $\bar{\alpha}_r$ *be a feasible vector with* $r > 1$. *Almost every large* $D(\bar{\alpha}_r, m, p)$ *contains a perfect allocation, i.e., an allocation that matches all nodes in some set* W_i. *Consequently, if* $j' \leq r - 1$ *is the least index for which* $\alpha_{j',2} < \alpha_{j'+1,1}$, *then every perfect allocation matches all nodes in some set* W_i *for some* $i \leq j'$.

PROOF. Observe that it is sufficient to prove the lemma for exact p, since by increasing p for some edges can only increase the probability for the existence of a (nearly) perfect allocation. Throughout the proof, we denote by 1_r the positive vector with $2r$ 1's $(1, 1, \ldots, 1)$.

We begin with the first part. Denote by Q the nearly perfect allocation property. Fix some feasible vector $\bar{\alpha}_1$. The proof for both $k = 2$ and $k \geq 3$ will follow from applying the Erdos–Renyi theorem to nondirected random graphs.

First consider $k = 2$. Let p_m be the probability that a nearly perfect allocation exists in the nondirected random graph $G(m, p^2)$ (recall that this graph has exactly m nodes and each edge is generated with probability p^2). That is,

$$p_m = \Pr[G(m, p^2) \models Q].$$

Consider the graph $D(1_1, m, p)$. Since a cycle of length 2 has probability p^2,

$$\Pr[D(1_1, m, p) \models Q] = p_m.$$

Let $m(\bar{\alpha}_1)$ be such that $[\alpha_{0,1}m, \alpha_{1,1}m]$ contains an integer for every $m \geq m(\bar{\alpha}_1)$. We define a sequence $(x_m)_{m \geq m(\bar{\alpha}_1)}$ by choosing arbitrarily the integer

$$x_m \in \arg\min_{x \in N \cap [\alpha_{0,1}m, \alpha_{1,1}m]} \Pr[D(1_1, x, p) \models Q].$$

Note that the minimum is attained at some value, since it is taken over a finite set that includes an integer. Therefore,

$$\Pr[D(\bar{\alpha}_1, m, p) \models Q] \geq \Pr[D(1_1, x_m, p) \models Q] = p_{x_m}.$$

By the Erdos–Renyi theorem, since p is a constant, $p_{x_m} \to 1$ as $m \to \infty$, completing the proof for $k = 2$.

We proceed with $k \geq 3$. If m is even, a perfect allocation exists using only two-way exchanges with probability $1 - o(1)$. If m is odd, we pick arbitrarily $m - 1$ nodes. In

the graph induced by these nodes, we find a perfect allocation, say M, using two-way exchanges (again, this can be found with probability $1 - o(1)$). Given that such M exists, it is sufficient to find a couple of nodes v, w that are matched to each other in M so that the single unmatched node can form a three-way exchange with v and w. Such two nodes v and w cannot be found with probability at most $(1 - p^2)^{m/2}$, completing the first part.

We sketch the second part of the proof, which follows by applying multiple times the Erdos–Renyi theorem. We sketch the proof for $r = 3$ (the proof for $r > 3$ is similar). Consider the 3-partite graph with realized sets of nodes W_0, W_1, W_2 and assume without loss of generality (w.l.o.g.) that W_0 is the smallest of those sets. Consider the directed graph induced by the nodes in W_0 and W_1. By the Erdos–Renyi theorem, with high probability there exists a disjoint set of edges $\bar{E}$ that covers all nodes in W_0 (since one can change every edge to a nondirected one and apply directly the theorem for nondirected bipartite graphs).

We next construct a bipartite directed graph $\bar{G}$ as follows. Let W_0 be the set of nodes on one side and let W_2 be the set of nodes on the other side. We construct the set of edges as follows. For any edge $e = (w_0, w_1) \in \bar{E}$, where $w_0 \in W_0$ and $w_1 \in W_1$, and any edge (w_1, w_2), where $w_2 \in W_2$ in the original graph, we construct an edge between w_0 and w_2 in $\bar{G}$. In addition, for any two nodes $w_2 \in W_2$ and $w_0 \in W_0$, an edge $\bar{G}$ exists if and only if it exists in the original graph. By the Erdos–Renyi theorem, there exists a perfect allocation $\bar{G}$ (this is just a 2-bipartite directed random graph). Finally, observe that by construction, a perfect allocation in $\bar{G}$ implies the existence of a perfect allocation in the original graph, which completes the proof.   □

A.3 *Proof of Proposition 1*

The proof is by construction. Let $D(m)$ be a random compatibility graph. We need to show that an allocation with the properties described in the proposition exists in $D(m)$ with probability $1 - o(1)$. Let δ be a constant such that $0 < \delta < \min\{(1 - 2.5\bar{\gamma})/(1 + 2.5\bar{\gamma}), 0.01, \frac{1}{100}\bar{\gamma}\}$.

Let $B_\delta(m)$ and $S_\delta(m)$ be the events defined in (5) and (6), respectively. Since $\Pr[B_\delta(m)] = 1 - o(m^{-1})$, we will assume throughout the proof that the events $B_\delta(m)$ and $S_\delta(m)$ occur (we will assume the probability that either one of these events does not occur toward nonexistence of a desired allocation). Let V be the set of realized pairs in $D(m)$. While we assume that the type of pair is realized, we assume that the edges are yet to be realized.

CLAIM 3. *1. With probability $1 - o(1)$, there exists a perfect allocation using only two-way or three-way exchanges in the subgraph induced by only self-demanded pairs.*

  2. With probability $1 - o(1)$, there exists a perfect allocation in the subgraph induced by only A-B and B-A pairs. In particular, either all A-B pairs or all B-A pairs are matched under such an allocation.

PROOF. Since $B_\delta(m)$ occurs, for every self-demanded type X-X, the subgraph induced by only X-X pairs is a bounded directed graph, $D(((1-\delta)\mu_{X\text{-}X}, (1+\delta)\mu_{X\text{-}X}), m, \gamma_H)$. Therefore, the first part follows by the first part of Lemma 2.

Similarly the graph induced by only A-B and B-A pairs is a 2-bounded directed graph, $D(((1-\delta)\mu_{A\text{-}B}, (1+\delta)\mu_{A\text{-}B}, (1-\delta)\mu_{B\text{-}A}, (1+\delta)\mu_{B\text{-}A}), m, \gamma_H)$. Hence the second part follows by the second part of Lemma 2. □

Let M_1 be an allocation in V that satisfies both parts of Claim 3. We will assume that such M_1 exists, and count the low probability that it does not toward failure of the desired allocation to exist. Further assume that M_1 matches all B-A pairs and, in particular, $Z_{A\text{-}B}(m) \geq Z_{B\text{-}A}(m)$ (the proof proceeds similarly if all B-A pairs are matched).

Let V' be the set of pairs that are not matched by M_1 in V. In particular, V' contains all overdemanded pairs, underdemanded pairs, and the A-B pairs that are not matched by M_1. The next claim shows that all A-B pairs that are not matched by M_1 can be matched as in the hypothesis. Recall that for a set of pairs S and type t, $\tau(S, t)$ denotes the set of pairs in S whose type is t.

CLAIM 4. *With probability $1 - o(1)$ there exists a perfect allocation in the subgraph induced by the sets of pairs $\tau(V', A\text{-}B)$, $\tau(V', B\text{-}O)$, and $\tau(V', O\text{-}A)$, which matches all pairs in $\tau(V', A\text{-}B)$.*

PROOF. Let $\bar\alpha_3 = (0, 2\delta\mu_{A\text{-}B}, (1-\delta)\mu_{B\text{-}O}, (1+\delta)\mu_{B\text{-}O}, (1-\delta)\mu_{O\text{-}A}, (1+\delta)\mu_{O\text{-}A})$. Since both $B_\delta(m)$ and $S_\delta(m)$ occur, the subgraph induced by the pairs in the statement is a 3-bounded directed random graph $D(\bar\alpha_3, m, \gamma_H)$, and the result follows by the second part of Lemma 2. □

Let M_2 be a perfect allocation as in Claim 4 (again assuming it exists). By Lemma 1, the size of this allocation is $o(m)$.

As the hypothesis suggests we wish to match every AB-O pair in a three-way exchange using one O-A pair and one A-AB pair (see Figure 3). Furthermore, we need to match every other overdemanded pair X-Y in a two-way exchange to a Y-X pair. Although we have already used some O-A pairs in M_2, the following claim shows that there are sufficiently many O-A pairs that are not matched by M_2 that can be used so as to match all A-O and AB-O pairs as we have just described. Similarly, there are sufficiently many A-AB pairs to match all AB-A and AB-O pairs.

CLAIM 5.   *1. $Z_{O\text{-}A}(m) \geq (1+\delta)m(\mu_{A\text{-}O} + \mu_{AB\text{-}O}) + \lambda m$ for some $\lambda > 0$.*

2. $Z_{A\text{-}AB}(m) \geq (1+\delta)m(\mu_{AB\text{-}A} + \mu_{AB\text{-}O})$.

PROOF. Let $1/\rho$ be the probability that a random patient and a random donor are incompatible.[52] Since $B_\delta(m)$ occurs,

$$Z_{O\text{-}A}(m) \geq \mu_{O\text{-}A}(1-\delta)m = \rho\mu_O\mu_A(1-\delta)m > \rho\mu_O\bar\gamma(\mu_A + \mu_{AB})(1+\delta)m,$$

---

[52]Thus if Y and X are blood types such that a donor of blood type Y is blood type compatible with a patient of blood type X, then $\mu_{X\text{-}Y} = \rho\mu_X\mu_Y\bar\gamma$ and otherwise $\mu_{X\text{-}Y} = \rho\mu_X\mu_Y$.

where the last inequality follows since $\mu_{\text{AB}} < \mu_{\text{A}}$ and $\delta < (1 - 2.5\bar{\gamma})/(1 + 2.5\bar{\gamma}) < (1 - 2\bar{\gamma})/(1 + 2\bar{\gamma})$, completing the first part. To see that the second part follows, note that

$$Z_{\text{A-AB}}(m) \geq \mu_{\text{A-AB}}(1 - \delta)m = \rho\mu_{\text{A}}\mu_{\text{AB}}(1 - \delta)m > \rho\mu_{\text{AB}}\bar{\gamma}(\mu_{\text{O}} + \mu_{\text{A}})(1 + \delta)m,$$

where the last inequality follows because $\mu_{\text{O}} + \mu_{\text{A}} < 2.5\mu_{\text{A}}$ (see Assumption B and footnote 21) and $\delta < (1 - 2.5\bar{\gamma})/(1 + 2.5\bar{\gamma})$. □

Let $M' = M_1 \cup M_2$ and let V'' be the set of all pairs that are not matched by M'. Consider the subgraph induced by the sets of pairs $\tau(V', \text{AB-O})$, $\tau(V', \text{O-A})$, and $\tau(V', \text{A-AB})$. Observe that this graph is a 3-bounded directed random graph; indeed by Claim 5, there exist constants c_1 and c_2 such that the number of pairs in $\tau(V', \text{A-AB})$ and $\tau(V', \text{AB-O})$ is at least $c_1 m$ and $c_2 m$. Therefore, by Lemma 2, with high probability there exists a perfect allocation that all AB-O pairs will be matched.

To complete the construction, it remains to show that for every overdemanded type X-Y except AB-O, the graph induced by all X-Y and Y-X pairs that are not yet matched contains a perfect allocation exchanges of size 2. This follows from similar arguments as above.

It remains to show that one cannot obtain more transplants by allowing exchanges of size more than 3. Let e be an exchange of any size and let $\tau(e, X$-$Y)$ be the set of pairs in e whose type is X-Y. It is enough to show that

$$\sum_{t \in \mathcal{U}} |\tau(e, t)| \leq 2|\tau(e, \text{AB-O})| + \sum_{t \in \mathcal{O} \setminus \{\text{AB-O}\}} |\tau(e, t)|.$$

We say that a pair v *helps* pair y if the there is either a directed edge from v to w or there is a directed path $v, z_1, z_2, \ldots, z_r, w$, where each z_i, $i \geq r$ is a self-demanded pair. Observe that every underdemanded O-X pair must be helped by some overdemanded Y-O pair. Similarly, any underdemanded pair X-AB must help an overdemanded AB-Y pair. Finally, since an O-X underdemanded pair can help an underdemanded pair Y-AB but not vice versa, we obtain the bound.

A.4 *Individual rationality and the proof of Theorem 2*

Before we prove Theorem 2, we need some preliminaries. First, it will be useful to write Claims 1 and 2 with respect to $D(H_n)$ rather than $D(m)$. We will need to rewrite the events (5) and (6) accordingly.

LEMMA 3. *Let $0 < \delta < \frac{1}{2}$ and let $H_n = \{1, \ldots, n\}$. Moreover, let χ_{H_n} be a random variable that denotes the size of all hospitals, that is, $\chi_{H_n} = \sum_{h \in H_n} |V_h|$. Consider the events*

$$W_\delta(H_n) = \{\forall X\text{-}Y \in \mathcal{P}, (1 - \delta)\mu_{X\text{-}Y}\chi_{H_n} < |\tau(V_{H_n}, X\text{-}Y)| < (1 + \delta)\mu_{X\text{-}Y}\chi_{H_n}\} \quad (7)$$

and

$$S_\delta(H_n) = \{||\tau(V_{H_n}, \text{A-B})| - |\tau(V_{H_n}, \text{B-A})|| = o(n)\}. \quad (8)$$

If every hospital $h \in H_n$ is of a positive and bounded size, then

$$\Pr[W_\delta(H_n), S_\delta(H_n)] = 1 - o(1). \tag{9}$$

DEFINITION 3 (Regularity). We say that $c > 0$ is a regular size if for every underdemanded type $X\text{-}Y \in \mathcal{U}$,

$$E_V\left[\int |\tau(M_{\mathcal{U}}^V(V), X\text{-}Y)|\, dF(V)| \#V = c\right] < \mu_{Y\text{-}X} c, \tag{10}$$

where $F(V)$ is any distribution over all allocations that maximize the number of matched underdemanded pairs in a given set of pairs V.

A.4.1 *Proof of Theorem 2*   Let $D(H_n)$ be a random compatibility graph with the set of hospitals H_n. We will prove the theorem for the case in which each hospital has the same regular size $c \leq \bar{c}$. The proof for the general case is similar (using the fact that the size of each hospital is bounded).

Let RHS(10) and LHS(10) be the right hand side and left hand side of inequality (10), respectively (see Definition 4).   Fix $\delta > 0$ such that $\delta < \min(\text{RHS}(10) - \text{LHS}(10), 0.01, \frac{1}{100}\bar{\gamma})$.

We assume that both events $W_\delta(H_n)$ and $S_\delta(H_n)$ as defined in (7) and (8), respectively, occur with low probability and count toward failure for the existence of an allocation with the properties described in the theorem.

The next lemma is a key step. Before we proceed, we require some definitions first. For every $h \in H_n$, let V_h be the set of pairs of hospital h. For a hospital $h \in H_n$ and a set of pairs $S \subseteq V_{H_n}$, denote by $\alpha(S, h) = V_h \cap S$ the set of pairs in S belonging to h. Note that $\tau(M_{\mathcal{U}}^{V_h}(V_h), \mathcal{U})$ is a maximum set of underdemanded pairs h can internally match. We let $U_{H_n} = \tau(V_{H_n}, \mathcal{U})$ and $O_{H_n} = \tau(V_{H_n}, \mathcal{O})$ be the set of all underdemanded and overdemanded pairs in H_n, respectively.

DEFINITION 4.   A set of underdemanded pairs $S \subseteq \tau(V_{H_n}, \mathcal{U})$ is called a *satisfiable set* if

1. $|\alpha(S, h)| \geq |\tau(M_{\mathcal{U}}^{V_h}(V_h), \mathcal{U})|$ for all $h \in H_n$

2. $|\tau(S, X\text{-}Y)| = |\tau(V_{H_n}, Y\text{-}X)|$ for all $X\text{-}Y \in \mathcal{U}$.

Note that the first part can be thought of as individual rationality with respect to underdemanded pairs.[53]

LEMMA 4 (Underdemanded rationality lemma). *Suppose every hospital size is regular and bounded by some $\bar{c} > 0$. With probability $1 - o(1)$, there exists a satisfiable set S_n in $D(H_n)$ and a perfect allocation in the bipartite subgraph induced by S_n and $\tau(V_{H_n}, \mathcal{O})$.*

---

[53]Even if a hospital can internally match more pairs using fewer underdemanded pairs, it is reasonable to consider this condition since pairs of other types will be "easy" to match as suggested by Proposition 1.

PROOF. One way to construct a satisfiable set S_n would be to first (i) choose randomly for each hospital a maximum set of underdemanded pairs it can internally match (by regularity and the law of large numbers this will satisfy the first property of Definition 4), and (ii) add arbitrary pairs of each underdemanded type so that the second property of Definition 4 is satisfied.

Suppose S_n is constructed as above. We want to show that with high probability for each underdemanded type $X\text{-}Y \in \mathcal{U}$, a perfect allocation exists in the subgraph induced by $\tau(S_n, X\text{-}Y)$ and the overdemanded pairs in $\tau(V_{H_n}, Y\text{-}X)$. Unfortunately, Lemma 2 cannot be directly applied since these graphs are not 2-bounded directed random graphs due to lack of independence of each edge in the graph (recall that we already have partial information on internal edges after phase (i) of the process above). Although it is true that with high probability such a perfect allocation exists, we use a slightly more subtle construction for a satisfiable set.

Instead, we will partition the set of hospitals into two sets H_n^1 and H_n^2, each with $\frac{1}{2}n$ hospitals, and find a satisfiable set S_n such that (i) the number of underdemanded pairs of each type $X\text{-}Y$ in S_n belonging to H_n^1 (H_n^2) equals the number of overdemanded pairs $Y\text{-}X$ belonging to H_n^2 (H_n^1). Then we will match overdemanded pairs of type $Y\text{-}X$, H_n^1 (H_n^2), to $X\text{-}Y$ underdemanded pairs in S_n belonging to H_n^2 (H_n^1), using the observation that these subgraphs are 2-uniform directed random graphs.

For every hospital $h \in H_n$, let M_h be a random allocation that maximizes the number of underdemanded pairs in the subgraph induced by its set of pairs V_h. For simplicity, we will assume throughout the proof that n is even. We partition the set of hospitals into two sets $H_n^1 = \{1, \ldots, \frac{1}{2}n\}$ and $H_n^2 = \{\frac{1}{2}n + 1, \ldots, n\}$. Define for each $j = 1, 2$,

$$S_n^j = \bigcup_{h \in H_n^j} \tau(M_h(V_h), \mathcal{U}),$$

and let $S = S_n^1 \cup S_n^2$. By construction, S satisfies the first property in Definition 4. Consider the following events for $j = 1, 2$:

$$Q_n^j = \left\{ \forall X\text{-}Y \in \mathcal{U}, |\tau(S_n^j, X\text{-}Y)| < (1 - \delta)\mu_{Y\text{-}X}\tfrac{1}{2}nc \right\}.$$

By the regularity assumption and the law of large numbers, $\Pr[Q_n^j] = 1 - o(1)$ for both $j = 1, 2$, and, therefore, $\Pr[Q_n^1, Q_n^2] = 1 - o(1)$.

Consider the events $W_\delta(H_n^j)$ for each $j = 1, 2$, where $W_\delta(H_n^j)$ is defined as in (7). Since the size of each H_n^j is $\frac{1}{2}n$, from Lemma 9 and the fact that there are only two sets in the partition with probability $1 - o(1)$, both $W_\delta(H_n^1)$ and $W_\delta(H_n^2)$ occur.

Therefore, with probability $1 - o(1)$ for each $j = 1, 2$,

$$|\tau(S_n^j, X\text{-}Y)| < |\tau(V_{H_n^{3-j}}, Y\text{-}X)|. \tag{11}$$

Finally, for each $j = 1, 2$, we add to S_n^j arbitrary underdemanded pairs belonging to H_n^j such that (11) becomes an equality for every $X\text{-}Y \in \mathcal{U}$. Observe that this is feasible by applying Lemma 1 for $\frac{1}{2}n$ hospitals. By construction, $S_n = S_n^1 \cup S_n^2$ is a satisfiable set.

Let X-$Y \in \mathcal{U}$ be an arbitrary type, and consider the subgraph induced by the sets of pairs $\tau(S_n^1, X$-$Y)$ and $\tau(V_{H_n^2}, X$-$Y)$. Note that this is a 2-bounded directed random graph (the realization of each edge is independent of the internal allocations M_h for each h since all potential edges in this graph are not internal). Therefore, there is perfect matching in this graph with probability $1 - o(1)$. Similarly, a perfect allocation exists with high probability in the graph induced by the sets of pairs $\tau(S_n^2, X$-$Y)$ and $\tau(V_{H_n^1}, X$-$Y)$. Finally, since there are a finite number of types, the proof follows.   □

We continue with the proof of the theorem. Let M_1 be a perfect allocation as in Lemma 4. We assume that such M_1 exists, again assuming that with the failure probability, no allocation with the desired properties exists.

So far, M_1 matches twice the number of overdemanded pairs in the graph, including for each hospital h the number of underdemanded pairs each h can internally match. As in the proof of Proposition 1, there exists a perfect allocation in the subgraph induced by all self-demanded pairs with probability $1 - o(1)$, say M_2.

Finally, we will show that there exists a perfect allocation in the subgraph induced by all A-B and B-A pairs that matches for each hospital at least the same number of A-B and B-A pairs it can internally match.

For each hospital, there exist probabilities $\epsilon_{\text{A-B}} > 0$ and $\epsilon_{\text{B-A}} > 0$ not depending on n for not matching all their A-B and B-A pairs, respectively. Therefore, there exists $\epsilon > 0$ not depending on n such that with probability $1 - o(1)$, the number of A-B pairs that cannot be internally matched is at least ϵn and the expected number of B-A pairs that cannot be internally matched is at least ϵn, i.e., linear in n.

However, by Lemma 1, the difference between the number of A-B and B-A pairs is sublinear with high probability, that is, with probability $1 - o(1)$,

$$\big||\tau(V_{H_n}, \text{A-B})| - |\tau(V_{H_n}, \text{B-A})|\big| = o(n). \tag{12}$$

Suppose that $|\tau(V_{H_n}, \text{A-B})| > |\tau(V_{H_n}, \text{B-A})|$ (the proof proceeds similarly if the converse inequality holds). By (12), with probability $1 - o(1)$ there exists $W \subseteq \tau(V_{H_n}, \text{A-B})$ such that (i) $|W| = |\tau(V_{H_n}, \text{B-A})|$ and (ii) for each hospital h, W contains at least the number of A-B pairs it can internally match.

Using similar arguments as in the proof of Lemma 4, there exists with high probability a perfect allocation in the graph induced by the sets of pairs W and $\tau(V_{H_n}, \text{B-A})$, say M_3.

It remains to bound the efficiency loss, which will follow from Proposition 1. We consider an efficient allocation M' as in Proposition 1 and let $M = M_1 \cup M_2 \cup M_3$. In both M and M', all self-demanded pairs are matched. M matches each AB-O pair in a two-way exchange to an O-AB pair rather than carrying out a three-way exchange as in M'. In both allocations M and M', after excluding all exchanges of which an AB-O pair is a part, all overdemanded pairs are matched and the same number of underdemanded pairs are matched. Finally, by (12), M leaves $o(n)$ A-B or B-A pairs unmatched, whereas M' matches all A-B and B-A pairs.
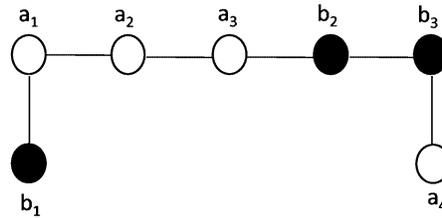
FIGURE 13. No 2-way mechanism is strategyproof.

A.5 *Proofs of Propositions 2 and 3*

We begin by proving Proposition 2. Consider a setting with two hospitals $H_2 = \{a, b\}$ such that $V_a = \{a_1, a_2, a_3, a_4\}$ and $V_b = \{b_1, b_2, b_3\}$. Further assume the compatibility graph induced by V_{H_2} is given in Figure 13.

Note that every maximal allocation leaves exactly one node unmatched. Suppose φ is both maximal and IR. We show that if a and b submit V_a and V_b, respectively, at least one hospital strictly benefits from withholding a subset of its nodes. Let $v \in V_{H_2}$ be unmatched in $\varphi(V_a, V_b)$. If $v \in V_a$, then $u_a(\varphi(V_a, V_b)) = 3$. However, by withholding a_1 and a_2, a's utility is 4, since the maximal allocation in $V \setminus \{a_1, a_2\}$ matches both a_3 and a_4, and a can match both a_1 and a_2 via an internal exchange. If $v \in V_b$, then by a symmetric argument, hospital b would benefit by withholding b_2 and b_3.

We continue with the proof for the first part of Proposition 3. Consider the same setting as in the proof of Proposition 2 (see Figure 13) and suppose φ is an IR strategyproof mechanism that always guarantees more than $\frac{1}{2}$ of the efficient allocation. Note that either $u_a(\varphi(V_a, V_b)) \leq 3$ or $u_b(\varphi(V_a, V_b)) \leq 2$. Suppose $u_a(\varphi(V_a, V_b)) \leq 3$. As in the proof of Proposition 2, for it not to be beneficial for a to withhold a_1 and a_2, the mechanism cannot match all pairs in $\{a_3, a_4\} \cup V_b$. Thus φ can choose at most a single exchange of size 2 in $\{a_3, a_4\} \cup V_b$, which is only half of the maximum (efficient) number, and not more, as required by assumption. The case in which $u_b(\varphi(V_a, V_b)) \leq 2$ is similar.

The proof of the second part of Proposition 3 is similar: Consider again the same setting as in the proof of Proposition 2 (see Figure 13) and assume there exists a randomized IR strategyproof mechanism φ that guarantees more than $\frac{7}{8}$ of the efficient allocation in every possible V. Any allocation leaves at least one node unmatched. Therefore, either $E[u_a(\varphi(V_a, V_b))] \leq 3.5$ or $E[u_b(\varphi(V_a, V_b))] \leq 2.5$. Suppose $E[u_a(\varphi(V_a, V_b))] \leq 3.5$. We argue that under the mechanism φ, hospital a benefits from withholding a_1 and a_2. Since φ guarantees more than $\frac{7}{8}$ of the efficient allocation in $\{a_3, a_4, b_1, b_2, b_3\}$, φ will choose the allocation containing exchanges a_3, b_2 and b_3, a_4 with probability more than $\frac{3}{4}$. Therefore, a's expected utility from reserving two transplants to do internally will be $2 + c$ for some $c > 1.5$. A similar argument holds if $E[u_b(\varphi(V_a, V_b))] \leq 2.5$.

A.6 *Proof of Theorem 3*

We first provide a formal definition for a strongly regular size.

DEFINITION 5. We say that $c > 0$ is a *strongly regular* size if for every underdemanded type $X\text{-}Y \in \mathcal{U}$,

$$E_V\big[\#\tau(M_{X\text{-}Y}^V(V), X\text{-}Y)|\#V = c\big] < \mu_{Y\text{-}X}c, \tag{13}$$

where $M_{X\text{-}Y}^V$ is an arbitrary allocation in $\mathcal{M}_{X\text{-}Y}^V$.

Let H_n be a set of bounded and strongly regular-sized hospitals, and let H_n^1 and H_n^2 be as in the theorem, i.e., a partition of H_n to two sets of hospitals each of size $\frac{1}{2}n$. For simplicity, we will assume that all hospitals have the same size $c > 0$.[54] Fix some hospital $\bar{h} \in H_n$ and fix $V_{\bar{h}}$ to be the set of pairs (type) of hospital $\bar{h}$. Without loss of generality, assume that $\bar{h} \in H_n^1$. We assume that all hospitals but $\bar{h}$ report truthfully their set of incompatible pairs.

Denote by φ the Bonus Mechanism. We need to show that for any subset of pairs $B_{\bar{h}} \subseteq V_{\bar{h}}$,

$$E_{V_{-h}}[u(\varphi(V_{\bar{h}}, V_{-h}))] \geq E_{V_{-h}}[u(\varphi(B_{\bar{h}}, V_{-h}))] - o(1). \tag{14}$$

Let RHS(13) and LHS(13) be the right hand side and left hand side of inequality (13), respectively (see Definition 5). Fix $\delta > 0$ such that $\delta < \min(\text{RHS}(13) - \text{LHS}(13), 0.01)$. We assume that the events $W_\delta(H_n^1)$, $W_\delta(H_n^2)$, $W_\delta(H_n)$, and $S_\delta(H_n)$ as defined in (7) and (8) occur, and as usual count the low probability they do not toward failure of the existence of an allocation as constructed in the Bonus Mechanism.[55]

The following claim will imply that the strategic problem of each hospital roughly comes down to maximizing its expected number of matched underdemanded pairs.

CLAIM 6. *If $\bar{h}$ reports truthfully $V_{\bar{h}}$, all its non-underdemanded pairs that can be internally matched will be matched by φ with probability $1 - o(1)$.*

PROOF. We first claim that in Step 1, the mechanism φ will find a perfect allocation in the graph induced by the set of self-demanded pairs with probability $1 - o(1)$. By the first part of Lemma 2 and its proof, in almost every subgraph induced by all self-demanded pairs except pairs of $\bar{h}$, there exists a perfect allocation M using two-way exchanges and at most one three-way exchange. Let v be a self-demanded pair belonging to $\bar{h}$. Using a similar argument to the proof of part 1 of Lemma 2, with high probability, v can form a three-way exchange with one of the two-way exchanges in M. Since $\bar{h}$ is bounded by a constant size, repeating this argument for each node of $\bar{h}$ proves the claim.

Similarly, as well as using the same arguments as in the proof of Theorem 2 to match A-B and B-A pairs, we obtain in Step 2 of the Bonus Mechanism that, with probability $1 - o(1)$, a perfect allocation will be found in the graph induced by A-B and B-A pairs that matches all A-B and B-A that can be internally matched. Finally, similarly to Lemma 4,

---

[54]Again, since all hospitals are of bounded size, a similar proof follows (one needs ca neglect sizes appear a finite number of times).

[55]Note that the internal graph of hospital $\bar{h}$ is not a random variable since it is fixed. However, Lemma 9 still holds since the size of $\bar{h}$ is bounded and does not affect the number of pairs in the limit. We skip here the formal details.

all overdemanded pairs will be matched in Step 3 (to underdemanded pairs) with probability $1 - o(1)$. Since there are only three steps and they are all independent of one another, the result follows.   □

For any $B_{\bar{h}} \subseteq V_{\bar{h}}$ and any underdemanded type $X\text{-}Y \in \mathcal{U}$, denote by $\psi_{X\text{-}Y}(B_{\bar{h}})$ the expected number of $X\text{-}Y$ pairs in $V_{\bar{h}}$ that will be matched when $\bar{h}$ reports $B_{\bar{h}}$ (both by the mechanism φ and, in the second stage, by $\bar{h}$).

Fix an arbitrary subset $B_{\bar{h}} \subseteq V_{\bar{h}}$ and an arbitrary underdemanded type $X\text{-}Y \in \mathcal{U}$. To see that (14) holds, by Claim 6 it is sufficient to show that

$$\psi_{X\text{-}Y}(B_{\bar{h}}) \leq \psi_{X\text{-}Y}(V_{\bar{h}}) + o(1). \tag{15}$$

The following lemma allow us to assume that all $X\text{-}Y$ pairs belonging to $\bar{h}$ that are chosen in the underdemanded lottery will be matched.

CLAIM 7. *All $X\text{-}Y$ pairs chosen by the underdemanded lottery will be matched by φ with probability $1 - o(1)$, regardless of whether $B_{\bar{h}}$ or $V_{\bar{h}}$ is reported.*

PROOF. Suppose $\bar{h}$ reports $B_{\bar{h}}$ (since $B_{\bar{h}}$ is arbitrary, all arguments in the proof hold also if $\bar{h}$ reports $V_{\bar{h}}$). Recall that $S_h(X\text{-}Y)$ is the set of $X\text{-}Y$ pairs belonging to h that are chosen in the underdemanded lottery and recall that $\theta_j(Y\text{-}X) = |\tau(B_{H_n^{3-j}}, Y\text{-}X)|$ for each $j = 1, 2$ (see Step 3(a) in the Bonus Mechanism).

By our assumption, every hospital is strongly regular (see Definition 4). Therefore, by the law of large numbers and since $\bar{h}$ is of bounded size, with probability $1 - o(1)$ for each $j = 1, 2$,

$$\sum_{h \in H_n^j} |S_h(X\text{-}Y)| < \theta_j(Y\text{-}X).^{56}$$

Therefore, with high probability, the underdemanded lottery will enter the Main Step of the underdemanded lottery.[57]

Since $W_\delta(H_n^1)$ and $W_\delta(H_n^2)$ occur, $\theta_j(Y\text{-}X) < |\tau(B_{H_n^{3-j}}, X\text{-}Y)|$ and $\theta_{3-j}(Y\text{-}X) < |\tau(B_{H_n^{3-j}}, X\text{-}Y)|$ for each $j = 1, 2$. Hence, for each $j = 1, 2$, the size of $\bigcup_{h \in H_n^j} S_h(X\text{-}Y)$ at the end of the underdemanded lottery is the same size as the number of reported $Y\text{-}X$ pairs by all hospitals in H_n^{3-j}.

In particular, each of the two subgraphs containing $X\text{-}Y$ and $Y\text{-}X$ pairs considered in Step 3(b) of the Bonus Mechanism is a 2-bounded directed random graph (here we used that nodes on each side of a graph cannot belong to the same hospital and, therefore, we still have independence of each edge). Therefore, by Lemma 2, both these subgraphs contain a perfect allocation with probability $1 - o(1)$ and by construction, all $X\text{-}Y$ pairs in these graph will be matched with probability $1 - o(1)$.   □

---

[56]We do not know if $|B_{\bar{h}}|$ is a strongly regular size, but since it is only one bounded hospital, the inequality holds.

[57]Again, we neglect formalizing that hospital $\bar{h}$'s set is fixed and not a random variable.

From this point on we will assume that all X-Y pairs chosen by the underdemanded lottery all end up matched (again counting the failure probability toward failing to match all underdemanded pairs of hospital $\bar{h}$ that are chosen in the underdemanded lottery).

By the Main Step of the underdemanded lottery, adding imaginary X-Y pairs to $B_{\bar{h}}$ (i.e., not from $V_{\bar{h}} \setminus B_{\bar{h}}$) can only increase $\psi_{X\text{-}Y}(B_{\bar{h}})$. We will add g new X-Y pairs to the set $B_{\bar{h}}$, assuming that each of these new pairs cannot be internally matched by $\bar{h}$, where

$$g = |\tau(V_{\bar{h}}, X\text{-}Y)| - |\tau(B_{\bar{h}}, X\text{-}Y)|.$$

Note that $g \geq 0$ and, with a slight abuse of notation, we refer from now on to $B_{\bar{h}}$ as the extended set containing the imaginary pairs. We need to show that (15) holds.

Let q and $\tilde{q}$ be the number of X-Y pairs $\bar{h}$ can match internally in $V_{\bar{h}}$ and $B_{\bar{h}}$, respectively. Observe that $\tilde{q} \leq q \leq |\tau(V_{\bar{h}}, X\text{-}Y)|$. We will assume that $q < |\tau(V_{\bar{h}}, X\text{-}Y)|$; otherwise (15) is satisfied since all pairs in $\tau(V_{\bar{h}}, X\text{-}Y)$ will be matched by φ.

Consider the Main Step in the underdemanded lottery. When $\bar{h}$ reports $V_{\bar{h}}$, each ball in J belonging to $\bar{h}$ is drawn with some identical probability $p > 0$. Similarly, when $\bar{h}$ reports $B_{\bar{h}}$, each ball in J belonging to $\bar{h}$ is drawn with some identical probability $\tilde{p} > 0$. Since the number of X-Y pairs and Y-X belonging to $\bar{h}$ is bounded, and the total number of X-Y and Y-X pairs in the pool approaches infinity,

$$\tilde{p} = p + o(1). \tag{16}$$

We set $z = |\tau(V_{\bar{h}}, X\text{-}Y)|$ and consider the case that $\bar{h}$ reports $V_{\bar{h}}$. In the initialization step of the underdemanded lottery, $S_{\bar{h}}(X\text{-}Y)$ is initialized to contain exactly q X-Y pairs of $\bar{h}$, and in the Main step of the lottery, for each one of $\bar{h}$'s that is drawn, an additional X-Y pair belonging to $\bar{h}$ is added to $S_{\bar{h}}(X\text{-}Y)$ as long as there are remaining X-Y pairs in $V_{\bar{h}}$. Therefore, since $\bar{h}$ has at most $z - q$ additional X-Y pairs (to the initial q ones),

$$\psi_{X\text{-}Y}(V_{\bar{h}}) = q + \sum_{j=1}^{z-q-1} j \binom{z}{j} p^j (1-p)^{z-j} + (z-q) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j}. \tag{17}$$

Consider now the case that $\bar{h}$ reports $B_{\bar{h}}$. Again, the initialized set $S_{\bar{h}}(X\text{-}Y)$ contains $\tilde{q}$ X-Y pairs, and for each of $\bar{h}$'s balls that is drawn in the Main Step, an additional X-Y pair is added to $S_{\bar{h}}(X\text{-}Y)$ (as long as it has such remaining in $B_{\bar{h}}$). Recall that we assumed that all pairs $S_{\bar{h}}(X\text{-}Y)$ at the end of the lottery will be matched by the mechanism φ.

Since $\bar{h}$ has not reported all its pairs, it can still use pairs in $V_{\bar{h}} \setminus B_{\bar{h}}$ in exchanges to match X-Y pairs in $\tau(V_{\bar{h}}, X\text{-}Y) \setminus S_{\bar{h}}(X\text{-}Y)$. By definition of q and the initialization of $S_{\bar{h}}(X\text{-}Y)$, $\bar{h}$ cannot match more than an additional $q - \tilde{q}$ X-Y pairs that the mechanism has not matched. Therefore,

$$\psi_{X\text{-}Y}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-\tilde{q}-1} \min(j+q-\tilde{q}, z-\tilde{q}) \binom{z}{j} \tilde{p}^j (1-\tilde{p})^{z-j}$$

$$+ (z-\tilde{q}) \sum_{j=z-\tilde{q}}^{z} \binom{z}{j} \tilde{p}^j (1-\tilde{p})^{z-j}, \tag{18}$$

where the second term on the right hand side follows since if j balls are drawn from J, $\bar{h}$ can match at most an additional $q - \tilde{q}$ X-Y pairs and altogether not more than $z - \tilde{q}$ additional X-Y pairs to the first $\tilde{q}$ pairs.

Since z, p and $\tilde{q}$ are all bounded, by (16) we can replace $\tilde{p}$ with p in the right hand side of (18) and add $o(1)$. Therefore,

$$\psi_{X\text{-}Y}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-\tilde{q}-1} \binom{z}{j} p^j (1-p)^{z-j} \min(j + q - \tilde{q}, z - \tilde{q})$$

$$+ (z - \tilde{q}) \sum_{j=z-\tilde{q}}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1).$$

Since $z - \tilde{q} \geq z - q$,

$$\psi_{X\text{-}Y}(B_{\bar{h}}) \leq \tilde{q} + \sum_{j=1}^{z-q-1} \binom{z}{j} p^j (1-p)^{z-j} (j + q - \tilde{q})$$

$$+ (z - \tilde{q}) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1)$$

$$= \tilde{q} + \sum_{j=1}^{z-q-1} j \binom{z}{j} p^j (1-p)^{z-j} + (q - \tilde{q}) \sum_{j=1}^{z-q-1} \binom{z}{j} p^j (1-p)^{z-j}$$

$$+ (z - q + q - \tilde{q}) \sum_{j=z-q}^{z} \binom{z}{j} p^j (1-p)^{z-j} + o(1)$$

$$\leq \psi_{X\text{-}Y}(V_{\bar{h}}) + o(1),$$

where the last inequality follows by (17) and since $(q - \tilde{q}) \sum_{j=1}^{z} \binom{z}{j} p^j (1-p)^{z-j} \leq q - \tilde{q}$. We have shown that inequality (15) is satisfied.

To see that the bound on the efficiency loss holds under the truth-telling strategy profile, note that the allocation constructed by φ has the same size/properties as the one constructed in the proof of Theorem 2, implying the result.

References

Abraham, David J., Avrim Blum, and Tuomas Sandholm (2007), "Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges." In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC)*, 295–304, ACM, New York. [819, 832, 841]

Alon, Noga and Joel Spencer (2008), *The Probabilistic Method*. Wiley, New York. [847]

Ashlagi, Itai, Mark Braverman, and Avinatan Hassidim (forthcoming(a)), "Stability in large matching markets with complementarities." *Operations Research*. [834]

Ashlagi, Itai, Felix Fischer, Ian Kash, and Ariel D. Procaccia (forthcoming(b)), "Mix and match." *Games and Economic Behavior*. [822, 834]

Ashlagi, Itai, David Gamarnik, Michael A. Rees, and Alvin E. Roth (2012), "The need for (long) chains in kidney exchange." Working paper. [840, 844]

Ashlagi, Itai, Duncan S. Gilchrist, Alvin E. Roth, and Michael A. Rees (2011a), "Nead chains in transplantation." *American Journal of Transplantation*, 11, 2780–2781. [830]

Ashlagi, Itai, Duncan S. Gilchrist, Alvin E. Roth, and Michael A. Rees (2011b), "Nonsimultaneous chains and dominos in kidney paired donation—Revisited." *American Journal of Transplantation*, 11, 984–994. [830]

Ashlagi, Itai and Alvin E. Roth (2012), "New challenges in multi-hospital kidney exchange." *American Economic Review: Papers and Proceedings*, 102, 354–359. [831]

Babaioff, Moshe, Liad Blumrosen, and Aaron Roth (2010), "Auctions with online supply." In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*, ACM, New York. [834]

Biro, Péter, David F. Manlove, and Romeo Rizzi (2009), "Maximum weight cycle packing in directed graphs, with application to kidney exchange programs." *Discrete Mathematics, Algorithms and Applications*, 1, 499–517. [841]

Edmonds, Jack (1971), "Matroids and the greedy algorithm." *Mathematical Programming*, 1, 127–136. [823]

Fréchette, Guillaume, Alvin E. Roth, and M. Utku Ünver (2007), "Unraveling yields inefficient matchings: Evidence from post-season college football bowls." *RAND Journal of Economics*, 38, 967–982. [818]

Immorlica, Nicol and Mohammad Mahdian (2005), "Marriage, honesty, and stability." In *Proceedings of the Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms*, 53–62, ACM, New York. [834]

Janson, Svante, Tomasz Luczak, and Andrzej Rucinski (2000), *Random Graphs*. Wiley, New York. [846]

Jevons, W. Stanley (1876), *Money and the Mechanism of Exchange*. D. Appleton and Co., New York. Available at http://mises.org/books/money_mechanism_exchange_jevons.pdf. [819, 842]

Kojima, Fuhito and Parag A. Pathak (2009), "Incentives and stability in large two-sided matching markets." *American Economic Review*, 99, 608–627. [834]

Kojima, Fuhito, Parag Pathak, and Alvin E. Roth (2013), "Matching with couples: Stability and incentives in large markets." *Quarterly Journal of Economics*, 128, 1585–1632. [834]

Leider, Stephen and Alvin E. Roth (2010), "Kidneys for sale: Who disapproves, and why?" *American Journal of Transplantation*, 10, 1221–1227. [818]

Massie, Allan B., Sommer E. Gentry, Robert A. Montgomery, Adam A. Bingaman, and Dorry L. Segev (2013), "Center-level utilization of kidney paired donation." *American Journal of Transplantation*, 13, 1317–1322. [819, 830]

McKinney, Nicholas C., Muriel Niederle, and Alvin E. Roth (2005), "The collapse of a medical labor clearinghouse (and why such failures are rare)." *American Economic Review*, 95, 878–889. [818]

Niederle, Muriel, Deborah D. Proctor, and Alvin E. Roth (2006), "What will be needed for the new gastroenterology fellowship match to succeed?" *Gastroenterology*, 130, 218–224. [818]

Niederle, Muriel, Deborah D. Proctor, and Alvin E. Roth (2008), "The gastroenterology fellowship match: The first two years." *Gastroenterology*, 135, 344–346. [818]

Niederle, Muriel and Alvin E. Roth (2003), "Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match." *Journal of Political Economy*, 111, 1342–1352. [817]

Niederle, Muriel and Alvin E. Roth (2005), "The gastroenterology fellowship market: Should there be a match?" *American Economic Review: Papers and Proceedings*, 95, 372–375. [818]

Rees, Michael A., Jonathan E. Kopke, Ronald P. Pelletier, Dorry L. Segev, Matthew E. Rutter, Alfredo J. Fabrega, Jeffrey Rogers, Oleh G. Pankewycz, Janet Hiller, Alvin E. Roth, Tuomas Sandholm, M. Utku Ünver, and Robert A. Montgomery (2009), "A non-simultaneous extended altruistic donor chain." *New England Journal of Medicine*, 360, 1096–1101. [819]

Rees, Michael A., Mark A. Schnitzler, Edward Zavala, James A. Cutler, Alvin E. Roth, Frank D. Irwin, Stephen W. Crawford, and Alan B. Leichtman (2012), "Call to develop a standard acquisition charge model for kidney paired donation." *American Journal of Transplantation*, 12, 1392–1397. [820]

Roth, Alvin E. (1982), "Incentive compatibility in a market with indivisble goods." *Economics Letters*, 9, 127–132. [843]

Roth, Alvin E. (2007), "Repugnance as a constraint on markets." *Journal of Economic Perspectives*, 21, 37–58. [818]

Roth, Alvin E. (2008), "What have we learned from market design?" *Economic Journal*, 118, 285–310. [819, 821]

Roth, Alvin E. (2010), "The first kidney exchange in the U.S., and other accounts of early progress." Available at http://marketdesigner.blogspot.com/2010/04/first-kidney-exchange-in-us-and-other.html. [819]

Roth, Alvin E. and Andrew Postlewaite (1977), "Weak versus strong domination in a market with indivisble goods." *Journal of Mathematical Economics*, 4, 131–137. [842]

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2004), "Kidney exchange." *Quarterly Journal of Economics*, 119, 457–488. [819]

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005a), "Pairwise kidney exchange." *Journal of Economic Theory*, 125, 151–188. [819]

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005b), "A kidney exchange clearinghouse in New England." *American Economic Review: Papers and Proceedings*, 95, 376–380. [818, 819, 822, 825]

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2007), "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences." *American Economic Review*, 97, 828–851. [819, 821, 826, 828, 840]

Roth, Alvin E. and Xiaolin Xing (1994), "Jumping the gun: Imperfections and institutions related to the timing of market transactions." *American Economic Review*, 84, 992–1044. [818]

Saidman, Susan L., Alvin E. Roth, Tayfun Sönmez, M. Utku Ünver, and Francis L. Delmonico (2006), "Increasing the opportunity of live kidney donation by matching for two and three way exchanges." *Transplantation*, 81, 773–782. [819, 840]

Shapley, Lloyd and Herbret Scarf (1974), "On the core and indivisibility." *Journal of Mathematical Economics*, 1, 23–37. [842]

Sönmez, Tayfun and M. Utku Ünver (2013), "Market design for kidney exchange." In *The Handbook of Market Design* (Nir Vulkan, Alvin E. Roth, and Zvika Neeman, eds.), 93–137, Oxford University Press, Oxford. [822]

Sönmez, Tayfun and M. Utku Ünver (2014), "Altruistically unbalanced kidney exchange." *Journal of Economic Theory*, 152, 105–129. [822]

Terasaki, Paul I., David W. Gjertson, and Michael J. Cecka (1998), "Paired kidney exchange is not a solution to ABO incompatibility." *Transplantation*, 65, 291. [826]

Toulis, Panos and David C. Parkes (2011), "A random graph model of kidney exchanges: Optimality and incentives." In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*, 323–332, ACM, New York. [821, 822, 826, 828, 831, 832, 840]

Ünver, M. Utku (2010), "Dynamic kidney exchange." *Review of Economic Studies*, 77, 372–414. [822, 826]

Zenios, Stefanos, Steve E. Woodle, and Lainie F. Ross (2001), "Primum non nocere: Avoiding harm to vulnerable candidates in an indirect kidney exchange." *Transplantation*, 72, 648–654. [827]