Altruistically Unbalanced Kidney Exchange*

Tayfun Sönmez† M. Utku Ünver‡

Boston College *Boston College*

This Draft: October 2011

Abstract

Although a pilot national live-donor kidney exchange program was recently launched in the US, the kidney shortage is increasing faster than ever. A new solution paradigm is able to incorporate *compatible pairs* in exchange. In this paper, we consider an exchange framework that has both compatible and incompatible pairs, and patients are indifferent over compatible pairs. Only two-way exchanges are permitted due to institutional constraints. We explore the structure of Pareto-efficient matchings in this framework. The mathematical structure of this model turns out to be quite novel. We show that under Pareto-efficient matchings, the same number of patients receive transplants, and it is possible to construct Pareto-efficient matchings that match the same incompatible pairs while matching the least number of compatible pairs. We extend the celebrated Gallai-Edmonds Decomposition in the combinatorial optimization literature to our new framework. We also conduct comparative static exercises on how this decomposition changes as new compatible pairs join the pool.

Keywords: Kidney Exchange, Market Design, Matching

JEL Classification Numbers: C78, D78, D02, D63

*We would like to thank Marek Pycia, Alvin E. Roth, Itai Sher, and Özgür Yilmaz for comments. Tayfun Sönmez and M. Utku Ünver acknowledge the research support of NSF. We would also like to thank participants at Villa La Pietra Applied Microeconomics Conference '11 at NYU Florence, CoED'11 at Montreal; PET'11 at Indiana; Johns Hopkins; BQGT '10 at Newport Beach, CA; Market Design Conference at Yonsei University, Korea; SAET '09 at Ischia, Italy; GAMES '08 at Northwestern; and Market Design Conference '08 at MIT. Initial draft of this paper was circulated under the title "Altruistic Kidney Exchange."

†Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467. E-mail: sonmezt@bc.edu. Web-site: www2.bc.edu/~sonmezt.

‡Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467. E-mail: unver@bc.edu. Web-site: www2.bc.edu/~unver.

1 Introduction

In the last decade, market design found an unexpected practical application in kidney exchange. What started as a scholarly interaction between economists, transplant surgeons, and immunology experts led to the establishment of the New England Program for Kidney Exchange (NEPKE) in 2004, the first kidney exchange program to utilize a formal mechanism. NEPKE's example was followed in 2005 by the Johns Hopkins Kidney Exchange Program and Alliance for Paired Donation (APD). These developments resulted in an amendment to the National Organ Transplant Act (NOTA) in 2007 clarifying that kidney exchanges are not in violation of this federal law, and subsequently a pilot national kidney exchange program was created in the U.S. in 2010.

To put the contribution of the current paper into perspective, it is helpful to describe how the collaboration between economists and transplantation community evolved over the years. In the early 2000s, economists observed that the two main types of kidney exchanges conducted in the U.S. corresponded to the most basic forms of exchanges in a house allocation model [Abdulkadiroğlu and Sönmez, 1999]. Building on this setup, they formulated a kidney exchange model and proposed a *top trading cycles and chains* (TTCC) mechanism [Roth, Sönmez, and Ünver, henceforth, RSÜ, 2004]. In their simulations RSÜ [2004] have shown that, in contrast to the 45 percent of the patients with willing donors who cannot receive a transplant in the absence of kidney exchanges, fewer than 10 percent would remain without a transplant under the TTCC mechanism. When economists shared their findings with Dr. Francis Delmonico, then president-elect of UNOS (the federal entity that now runs the national pilot kidney exchange program in the U.S.), he expressed two reservations about the proposed kidney exchange model. First of all, RSÜ [2004] allowed for potentially large exchanges that would be logistically hard to implement since all transplants in an exchange need to be carried out simultaneously. The second concern was that RSÜ [2004] assumed strict preferences between compatible kidneys, which is contrary to the general tendency in the U.S. where doctors assume that two compatible living-donor kidneys have essentially the same survival rates, regardless of the "genetic distance" between the patient and the donor [Gjertson and Cecka, 2000, Delmonico, 2004]. To address these concerns, RSÜ [2005a] proposed a second model that restricted the size of kidney exchanges to two patient-donor pairs and assumed that patients are indifferent between compatible kidneys. RSÜ [2005a] observed that their *pairwise kidney exchange* model is an application of a well-analyzed problem in discrete-optimization literature,[1] some of the techniques of which was recently imported to economic theory by Bogolomania and Moulin [2004] for two-sided matching markets.[2] The optimal-matching methodology proposed by RSÜ [2005a] became the basis of practical kidney exchange throughout the world including at NEPKE, APD, the National Matching Scheme for Paired and Pooled Donation in the UK, and most recently the National Kidney Paired Donation Pilot Program in the U.S., although all of these programs allow for three-way exchanges based on findings of RSÜ [2007] and Saidman et al. [2006].

An earlier, abstract version of the RSÜ [2005a] model was extensively analyzed in the 1960s. One of the most important contributions to this literature was that of Gallai [1963, 1964] and Edmonds [1965], who characterized the set of Pareto-efficient matchings. This result is known as the *Gallai-Edmonds Decomposition* (GED) Theorem, and it plays a central role in our current paper. One of the corollaries of the GED Theorem has a very plausible implication for pairwise kidney exchange: The same number of patients are matched at every Pareto-efficient matching. This means that, regardless of how the kidney exchange programs determine patient priorities, the same number of patients are matched. Hence a program never matches a high-priority patient at the expense of multiple patients under the *pairwise priority mechanisms* offered in RSÜ [2005a]. This result does not hold for the TTCC mechanism. Hence, it gives pairwise priority mechanisms an edge from a medical ethics perspective. However the elegance of the structure of Pareto-efficient matchings come at a very high cost to efficiency: In contrast to the TTCC mechanism, the number of patients who remain without a transplant more than triples under the pairwise priority mechanisms. To explain this large difference, we need to describe the basic mechanics for kidney transplantation.

A patient with a healthy and willing live donor might not be able to receive his kidney either because of blood-type incompatibility or because of tissue-type incom-

[1]See Lovasz and Plummer [1986] and Korte and Vygen [2002] for comprehensive surveys of this literature.

[2]See Yilmaz [2011a] for an application of this two-sided matching approach in kidney exchange. In his model, Yilmaz assumes any size of exchange is feasible and considers "list" exchanges as well as regular exchanges. He comes up with an egalitarian matching mechanism by treating patients and kidney donors as two sides of the market.

patibility. There are four blood types, A, B, AB, O, where 44 percent of the U.S. population have O blood type, 42 percent have A blood type, 10 percent have B blood type, and 4 percent have AB blood type. Furthermore:

- an O blood-type donor is blood-type compatible with all patients,
- an A blood-type donor is blood-type compatible with only A and AB blood-type patients,
- a B blood-type donor is blood-type compatible with only B and AB blood-type patients,
- and an AB blood-type donor is blood-type compatible with only AB blood-type patients.

This very important asymmetry in blood-type compatibility relation makes O blood-type donors highly sought after and O blood-type patients highly vulnerable. Based on the U.S. blood-type distribution given above, the odds for blood-type incompatibility are about 35 percent between a patient and a random donor.

A donor might also be tissue-type incompatible with his paired patient. Zenios, Woodle, and Ross [2001] report that the odds for tissue-type incompatibility are about 11 percent between a patient and a random donor. Consistent with figures for random pairs, a large majority of incompatible pairs are blood-type incompatible.

The key observation here is the following: With the exception of A blood-type patients with B blood-type donors and B blood-type patients with A blood-type donors, a blood-type-incompatible pair cannot be mutually compatible with any blood-type-incompatible pair. Hence any such pair has to receive a kidney from a blood-type-compatible pair. In a regime where patients are assumed to be indifferent between all compatible pairs, the only blood-type compatible pairs available for exchange are those that are tissue-type incompatible. In contrast, in a regime where patients have strict preferences over compatible pairs, essentially all pairs are available for exchange. This is by far the most important reason for the large efficiency gap between the RSÜ [2004] TTCC mechanism and RSÜ [2005a] pairwise priority mechanism. Blood-type O patients with blood-type A, B, or AB donors, and blood-type A or B patients with blood-type AB donors face much stronger competition for a fraction of tissue-type incompatible pairs in a program that excludes compatible pairs from the kidney exchange pool. This highly vulnerable group makes up more than 25 percent of *all* pairs.

Once it became clear that pairwise exchange among incompatible pairs will leave about half of these incompatible pairs without a transplant, economists were able to convince the transplantation community to be more flexible about the size of acceptable exchanges. RSÜ [2007] and Saidman et al. [2006] have shown that the percentage of incompatible pairs who receive transplants increases to 60 percent if three-way exchanges are allowed in addition to two-way exchanges, although larger exchanges, and especially those larger than four-way exchanges, essentially have minimal impact on efficiency. Based on these results, all major kidney exchange programs, including the pilot national kidney exchange program in the U.S., adopted mechanisms that allow three-way exchanges. One negative implication of this flexibility is the loss of the GED-type structure of Pareto-efficient matchings. In particular, the number of patients receiving transplants can differ between two Pareto-efficient matchings, and hence the priority mechanism used by NEPKE might no longer maximize the number of patients receiving transplants. That is perhaps a small price to pay in comparison to the gains from three-way exchanges, but there is an alternative that not only dramatically increases the efficiency gains from kidney exchange but also preserves the GED structure of Pareto-efficient matchings.

Inclusion of three-way exchanges was not the only suggestion economists made to increase the number of patients who can benefit from kidney exchange. RSÜ [2005b] proposed the inclusion of compatible pairs in the kidney exchange pool even if they do not strictly benefit from an exchange. They reported that the inclusion of compatible pairs in the kidney exchange pool would produce the largest efficiency gains in comparison to a number of other design modifications that also improve the efficiency. Assuming a pool of 100 pairs, they have shown that the percentage of patients who remain without a transplant can be reduced to less than 10 percent if compatible pairs are included in the exchange pool. This dramatically improved efficiency is due to the elimination of the above-discussed asymmetry, essentially solving the root of the problem. The idea of a kidney exchange between an incompatible pair and a compatible pair was not new; it was first introduced by Ross and Woodle [2000] as an *altruistically unbalanced kidney exchange*. Ironically Ross and Woodle [2000] themselves condemned this type of exchange as morally inappropriate on the grounds of potential coercion, even though they had not fully closed the door on its implementation. They concluded:

> Empirical data of the attitudes of potential and former organ donors and living kidney recipients would be useful to prove or disprove our concerns.

This strong objection resulted in altruistically unbalanced kidney exchange receiving no attention until RSÜ [2005b] strongly advocated for the inclusion of compatible pairs in exchange pools. This message has reached the transplantation community, and a number of recent papers in the transplantation literature make a convincing case for altruistically unbalanced kidney exchange.[3] Gentry et al. [2007] verify the large efficiency gains from the inclusion of compatible pairs in the exchange pool and advocate for this paradigm change in kidney exchange. Ratner et al. [2010] report a survey of 52 patients with compatible donors who were asked whether they would be willing to participate in an exchange. Less than 20 percent were opposed to the idea. This study presents a stark contrast to the long-held mainstream belief in the transplantation community regarding compatible pairs' attitudes toward altruistically unbalanced kidney exchange. Ratner et al. [2010] also report three altruistically unbalanced exchanges conducted at Columbia University as a proof of concept involving four compatible pairs. Thanks to these compatible pairs, five additional patients received transplants. Columbia University currently has an altruistically unbalanced kidney exchange program, the first one that we are aware of. As the attitude toward altruistically unbalanced kidney exchange has improved, some medical ethicists have started questioning the grounds on which the medical community has been opposed to these types of exchanges in the first place. Steinberg [2011] states:

> Despite their utilitarian value transplant ethicists have condemned this type of organ exchange as morally inappropriate. An opposing analysis concludes that these exchanges are examples of moral excellence that should be encouraged.

Motivated by this paradigm change, in this paper, we consider a pairwise kidney exchange model in which both compatible and incompatible pairs are available for exchange. Our main focus is understanding the structure of Pareto-efficient matchings

[3] Other economists also became interested in this paradigm. Nicolò and Rodriguez-Álvarez [2011] introduce a model that incorporates compatible pairs to kidney exchange under the assumption that patients have strict preferences over the ages of compatible donors. They study Pareto-efficient and non-manipulable mechanisms in this domain.

and in particular the role of compatible pairs in this structure. In our main result (Theorem 1) we show that the GED Theorem extends to this natural structure, and in particular the number of patients who receive transplants is the same across all Pareto-efficient matchings (Proposition 1). As we have argued before, this is very plausible from a medical ethics perspective. We also show that the choice of incompatible pairs can be separated from the choice of compatible pairs under any Pareto-efficient mechanism (Proposition 2). This result implies that the number of compatible pairs needed to participate in a Pareto-efficient matching is the same, regardless of the choice of incompatible pairs who benefit from the exchange (Corollary 1). This corollary is particularly important, since policy makers may wish to minimize the number of compatible pairs participating in exchanges, and Corollary 1 implies that this potential policy puts no restriction on the choice of incompatible pairs. In contrast to RSÜ [2005a], which builds on the discrete-optimization literature, here we have no results that we can directly utilize from the earlier literature, although the original GED Theorem provides us with a convenient starting point for the inductive proof of our main result. Our proof technique is also of independent interest as it allows us to carry out a useful comparative static exercise: We fully characterize the impact of the addition of one compatible pair to a problem, and among other things, we show that the entire patient population (weakly) benefits from the inclusion of a compatible pair. In contrast to the use of three-way exchanges that require kidney exchange programs to make hard distributional choices to increase the number of patients who benefit from kidney exchange, inclusion of compatible pairs in the pool benefits the whole population and in particular hard-to-match O blood-type patients.

2 The Model

A **pair** consists of a patient and a donor. A pair is **compatible** if the donor of the pair can medically donate her kidney to the patient of the pair and **incompatible** otherwise. Let N_I be the set of incompatible pairs and N_C be the set of compatible pairs. Let $N = N_I \cup N_C$ be the set of all pairs. The donor of pair x is compatible with the patient of pair y if the donor of pair x can medically donate a kidney to the patient of pair y. Two distinct pairs $x, y \in N$ are **mutually compatible** if the donor of pair x is compatible with the patient of pair y and the donor of pair y is compatible with the patient of pair x.

For any pair $x \in N$, let $\succsim_x$ denote its preferences over N. Let $\succ_x$ denote the strict preference relation and $\sim_x$ denote the indifference relation associated with $\succsim_x$. The preferences of a pair are dictated by the patient of the pair who is indifferent between all compatible kidneys and who strictly prefers any compatible kidney to any incompatible kidney. In addition, the patient of an incompatible pair strictly prefers remaining unmatched (i.e. keeping his donor's incompatible kidney) to any other incompatible kidney. Therefore, for any incompatible pair $i \in N_I$,

- $x \sim_i y$ for distinct $x, y \in N$ with a compatible donor for the patient of pair i,
- $x \succ_i i$ for any $x \in N$ with a compatible donor for the patient of pair i,
- $i \succ_i x$ for any $x \in N$ without a compatible donor for the patient of pair i,

and for any compatible pair $c \in N_C$,

- $x \sim_c y$ for distinct $x, y \in N$ with a compatible donor for the patient of pair c,
- $c \succ_c x$ for any $x \in N$ without a compatible donor for the patient of pair c.

Throughout the paper we assume that two-way exchanges are feasible only when at least one of the pairs is incompatible.[4] A two-way exchange is **ordinary** if it is an exchange between two incompatible pairs that are mutually compatible. A two-way exchange is **altruistically unbalanced** if it is an exchange between an incompatible and a compatible pair that are mutually compatible.

The **feasible exchange matrix** $R = [r_{x,y}]_{x,y \in N}$ identifies all feasible exchanges where

$$r_{x,y} = \begin{cases} 1 & \text{if } y \in N \setminus \{x\},\ x, y \text{ are mutually compatible, and } x \text{ or } y \in N_I \\ 0 & \text{otherwise.} \end{cases}$$

For any $x, y \in N$ with $r_{x,y} = 1$, we refer to the pair (x, y) as a **feasible exchange**.

[4] Clearly there is no benefit from an exchange between two compatible pairs in our model.

An **altruistically unbalanced kidney exchange problem** (or simply a **problem**) (N, R) consists of a set of pairs and its feasible exchange matrix.

A **matching** is a set of mutually exclusive feasible exchanges. Formally, given a set N of pairs, a matching is a set $\mu \subseteq 2^{N^2}$ such that

1. $(x, y) \in \mu$ and $(x, y') \in \mu$ implies $y = y'$,

2. $(x, y) \in \mu$ and $(x', y) \in \mu$ implies $x = x'$, and

3. $(x, y) \in \mu$ implies $r_{x,y} = 1$.

Here $(x, y) \in \mu$ means that the patient of both pairs receive a kidney from the donor of the other pair. Let $\mathcal{M}(N, R)$ denote the set of all matchings for a given problem (N, R).[5]

For any $\mu \in \mathcal{M}(N, R)$ and $(x, y) \in \mu$, define $\mu(x) \equiv y$ and $\mu(y) \equiv x$. Here x and y are **matched** with each other in μ. For any $\mu \in \mathcal{M}(N, R)$ and $x \in N$ with no $y \in N \setminus \{x\}$ such that $(x, y) \in \mu$, define $\mu(x) \equiv x$. Here x is **unmatched** in μ. For any matching μ, let M^μ denote the set of pairs that are matched in μ. Formally,

$$M^\mu = \{x \in N : \mu(x) \neq x\}.$$

Observe that an incompatible pair receives a transplant in a matching μ only if it is matched in μ whereas a compatible pair receives a transplant whether it is matched or not. For any matching μ, let T^μ denote the set of all pairs who receive a **transplant** in μ. Formally,

$$T^\mu = \{x \in N_I : \mu(x) \neq x\} \cup N_C.$$

Let I^μ refer to the set of incompatible pairs that are matched in μ. That is,

$$I^\mu = M^\mu \cap N_I = T^\mu \cap N_I.$$

Similarly let C^μ refer to the set of compatible pairs that are matched in μ. That is,

$$C^\mu = M^\mu \cap N_C.$$

[5]The ordering of pairs in a feasible exchange is not important, thus $(x, y) = (y, x)$ in our notation.

3 Pareto-Efficient Matchings

Throughout this section, fix a problem (N, R). For any $\mu, \nu \in \mathcal{M}$, μ **Pareto-dominates** ν if $\mu(x) \succsim_x \nu(x)$ for all $x \in N$ and $\mu(x) \succ_x \nu(x)$ for some $x \in N$. A matching $\mu \in \mathcal{M}$ is **Pareto efficient** if there exists no matching that Pareto-dominates μ. Let $\mathcal{E} \subseteq \mathcal{M}$ be the set of Pareto-efficient matchings.

When there are no compatible pairs, it is well-known that the same number of incompatible pairs is matched at each Pareto-efficient matching. In our model, what is critical is who receives a transplant (rather than who is matched). In our first result, we show that the number of the pairs who receive a transplant is the same in any two Pareto-efficient matchings and that number is the maximum number of pairs that can receive a transplant in a matching:

Proposition 1 *A matching $\mu \in \mathcal{M}$ is Pareto efficient if and only if $|T^\mu| = \max_{\eta \in \mathcal{M}} |T^\eta|$. Hence, for any two Pareto-efficient matchings $\mu, \nu \in \mathcal{E}$, $|T^\mu| = |T^\nu|$.*

Proof. [Proof of Proposition 1] First, we show that if $|T^\mu| = \max_{\eta \in \mathcal{M}} |T^\eta|$ then μ is Pareto efficient. Suppose that $\mu \in \mathcal{M}$ is such that $|T^\mu| = \max_{\eta \in \mathcal{M}} |T^\eta|$ and suppose that there exists a matching $\nu \in \mathcal{M}$ that Pareto-dominates μ. Then all pairs receiving a transplant in μ also receive a transplant in ν and at least one other incompatible pair that does not receive a transplant in μ receives it in ν. Thus, $|T^\nu| > |T^\mu|$, contradicting $|T^\mu| = \max_{\nu \in \mathcal{M}} |T^\nu|$.

Next, we show that for two matchings $\mu, \nu \in \mathcal{M}$ that are such that $|T^\mu| > |T^\nu|$, there exists a matching that Pareto-dominates ν. This will prove that if a matching μ is Pareto efficient then $|T^\mu| = \max_{\eta \in \mathcal{M}} |T^\eta|$. Let $\mu, \nu \in \mathcal{M}$ be such that $|T^\mu| > |T^\nu|$. Let $a_0 \in T^\mu \setminus T^\nu$. Since patients of compatible pairs always receive a transplant, $a_0 \in N_I$ and therefore $a_0 \in M^\mu$. Construct the sequence $\{a_0, a_1, \ldots, a_k\} \subseteq M^\mu \cup M^\nu$ as follows:

$$a_1 = \mu(a_0), \quad a_2 = \nu(a_1), \quad \ldots \quad a_k = \begin{cases} \mu(a_{k-1}) & \text{if } k \text{ is odd} \\ \nu(a_{k-1}) & \text{if } k \text{ is even} \end{cases}$$

and where the last element of the sequence, a_k, is unmatched either in μ or in ν (i.e. $a_k \in (M^\mu \setminus M^\nu) \cup (M^\nu \setminus M^\mu)$). Observe that by construction, a_0 is matched in μ but not in ν, whereas $a_1, \ldots, a_{k-1}$ are all matched in both μ and ν. Also observe that $(a_\ell, a_{\ell+1})$ is a feasible exchange for any $\ell \in \{0, 1, \ldots, k-1\}$.

There are three cases to consider:

Case 1. $a_k \in T^\nu \setminus T^\mu$:

This case, indeed, does not help us to construct a matching that Pareto-dominates ν. However, since

(i) $|T^\mu| > |T^\nu|$, and

(ii) any pair that is not at the two ends of the sequence receives a transplant in both μ and ν,

there exists $a_0 \in T^\mu \setminus T^\nu$ such that the last element of the above constructed sequence a_k is such that $a_k \notin T^\nu \setminus T^\mu$. Hence Case 1 cannot cover all situations.

Case 2. $a_k \in M^\mu \setminus M^\nu$:

Since a_k is matched in μ but not in ν, k is odd. Consider the following matching $\eta \in \mathcal{M}$:

$$\eta = (\nu \setminus \{(a_1, a_2), (a_3, a_4), \ldots, (a_{k-2}, a_{k-1})\}) \cup \{(a_0, a_1), (a_2, a_3), \ldots, (a_{k-1}, a_k)\}.$$

We have $T^\eta = T^\nu \cup \{a_0, a_k\}$. Since $a_0 \notin T^\nu$, matching η Pareto-dominates matching ν.

Case 3. $a_k \in N_C$ and $a_k \in M^\nu \setminus M^\mu$:

Since a_k is matched in ν but not in μ, k is even. Consider the following matching $\eta \in \mathcal{M}$:

$$\eta = (\nu \setminus \{(a_1, a_2), (a_3, a_4), \ldots, (a_{k-1}, a_k)\}) \cup \{(a_0, a_1), (a_2, a_3), \ldots, (a_{k-2}, a_{k-1})\}.$$

Observe that a_k is matched in ν but not in η whereas a_0 is matched in η but not in ν. But since $a_k \notin N_I$, $T^\eta = T^\mu \cup \{a_0\}$ and therefore matching η Pareto-dominates matching ν.

Since there exists $a_0 \in T^\mu \setminus T^\nu$ where either Case 2 or Case 3 applies, matching ν is Pareto inefficient. ∎

Our next result shows that the choice of compatible pairs to be matched at a Pareto-efficient matching can be separated from the choice of incompatible pairs.

Proposition 2 *Let $\mu, \nu \in \mathcal{E}$ be two Pareto-efficient matchings. Then there exists a Pareto-efficient matching $\eta \in \mathcal{E}$ such that $M^{\eta} = C^{\mu} \cup I^{\nu}$.*

Proof. [Proof of Proposition 2] Let μ, ν be as in the statement of the proposition. By Proposition 1, $|T^{\mu} \setminus T^{\nu}| = |T^{\nu} \setminus T^{\mu}|$. If $T^{\mu} = T^{\nu}$ then $\eta = \mu$ and we are done. Otherwise let $a_0 \in T^{\mu} \setminus T^{\nu}$. Note that $a_0 \in N_I$ (since only incompatible pairs can receive a transplant in one matching but not in another). We will construct a matching that matches a_k together with all elements of M^{μ} except $a_0 \in N_I$. Repeated application of this construction yields the desired matching η.

Construct the sequence $\{a_0, a_1, \ldots, a_k\} \subseteq M^{\mu} \cup M^{\nu}$ as follows:

$$a_1 = \mu(a_0), \quad a_2 = \nu(a_1), \quad \ldots \quad a_k = \begin{cases} \mu(a_{k-1}) & \text{if } k \text{ is odd} \\ \nu(a_{k-1}) & \text{if } k \text{ is even} \end{cases}$$

and where the last element of the sequence, a_k, is unmatched either in μ or in ν (i.e. $a_k \in (M^{\mu} \setminus M^{\nu}) \cup (M^{\nu} \setminus M^{\mu})$). Observe that $(a_{\ell}, a_{\ell+1})$ is a feasible exchange for any $\ell \in \{0, 1, \ldots, k-1\}$.

There are three cases to consider:

Case 1. k is odd:

In this case both a_0 and a_k are matched in μ, but not in ν. Consider the matching

$$\nu' = (\nu \setminus \{(a_1, a_2), (a_3, a_4), \ldots, (a_{k-2}, a_{k-1})\}) \cup \{(a_0, a_1), (a_2, a_3), \ldots, (a_{k-1}, a_k)\}.$$

By construction, $M^{\nu'} = M^{\nu} \cup \{a_0, a_k\}$. Moreover, while a_k may not be an incompatible pair, a_0 is, and hence $T^{\nu} \subset T^{\nu'}$. Therefore ν' Pareto-dominates ν, contradicting the Pareto efficiency of ν.

Case 2. k is even with $a_k \in N_C$:

In this case a_k, a compatible pair, is matched in ν but not in μ. In contrast, a_0, an incompatible pair, is matched in μ but not in ν. Consider the matching

$$\nu' = (\nu \setminus \{(a_1, a_2), (a_3, a_4), \ldots, (a_{k-1}, a_k)\}) \cup \{(a_0, a_1), (a_2, a_3), \ldots, (a_{k-2}, a_{k-1})\}.$$

By construction, $M^{\nu'} \setminus M^{\nu} = \{a_0\}$, whereas $M^{\nu} \setminus M^{\nu'} = \{a_k\}$. Since a_0 is an incompatible pair while a_k is not, $T^\nu \subset T^{\nu'}$. Therefore ν' Pareto-dominates ν, contradicting the Pareto efficiency of ν.

Since Cases 1 and 2 each yield a contradiction, for each $a_0 \in T^\mu \setminus T^\nu$, the last element a_k of the above constructed sequence $\{a_0, a_1, \ldots, a_k\}$ should be an incompatible pair and k should be even. We next consider this final case.

Case 3. k is even with $a_k \in N_I$:

In this case a_k is matched in ν, and therefore, by construction, $a_k \in T^\nu \setminus T^\mu$. Consider the matching

$$\mu' = (\mu \setminus \{(a_0, a_1), (a_2, a_3), \ldots, (a_{k-2}, a_{k-1})\}) \cup \{(a_1, a_2), (a_3, a_4), \ldots, (a_{k-1}, a_k)\}.$$

By construction, $M^{\mu'} = (M^\mu \setminus \{a_0\}) \cup \{a_k\}$. So in comparison with matching μ, matching μ' matches incompatible pair a_k instead of incompatible pair a_0. Observe that $|T^{\mu'} \cap T^\nu| = |T^\mu \cap T^\nu| + 1$ while $C^\mu = C^{\mu'}$. If $|T^\nu \setminus T^\mu| = 1$, then $\eta = \mu'$ is the desired matching and we are done. Otherwise, since Case 3 is the only viable case we can repeat the same construction for any $a_0 \in T^\mu \setminus T^\nu$ to obtain the desired matching η.

■

In the present context, the involvement of compatible pairs in exchange is purely altruistic and it may therefore be plausible to minimize the number of compatible pairs matched at Pareto-efficient matchings. An immediate corollary of Proposition 2 is that the number of compatible pairs who exchange kidneys can be minimized without affecting the choice of incompatible pairs.

Corollary 1 *Let $\mu \in \mathcal{E}$. Then there exists $\eta \in \mathcal{E}$ be such that $I^\eta = I^\mu$ and $|C^\eta| \le |C^\nu|$ for any $\nu \in \mathcal{E}$.*

3.1 The Priority Mechanisms

The experience of transplant centers is mostly with the priority allocation systems used to allocate cadaver organs. NEPKE has recently adopted a variant of a priority allocation system for ordinary kidney exchanges. Priority mechanisms can be easily adapted to the present context.

Let $|N_I| = n$. A priority ordering is a one-to-one and onto function $\pi : \{1, \ldots, n\} \to N_I$. Here incompatible pair $\pi(k)$ is the k^{th} highest priority pair for any $k \in \{1, \ldots, n\}$.

For any problem, the **priority mechanism** induced by π picks any matching from a set of matchings $\mathcal{E}_\pi^n$ which is obtained by refining the set of matchings in n steps as follows:

- Let $\mathcal{E}_\pi^0 = \mathcal{M}$ (i.e. the set of all matchings).
- In general for $k \leq n$, let $\mathcal{E}_\pi^k \subseteq \mathcal{E}_\pi^{k-1}$ be such that

$$\mathcal{E}_\pi^k = \begin{cases} \left\{\mu \in \mathcal{E}_\pi^{k-1} : \mu\left(k\right) \neq k\right\} & \text{if } \exists \mu \in \mathcal{E}_\pi^{k-1} \text{s.t. } \mu\left(k\right) \neq k \\ \mathcal{E}_\pi^{k-1} & \text{otherwise} \end{cases}.$$

Each matching in $\mathcal{E}_\pi^n$ is referred to as a **priority matching** and they all match the same set of incompatible pairs. By construction, each matching in $\mathcal{E}_\pi^n$ is Pareto efficient. Observe that by Proposition 1 there is no trade-off between priority allocation and the number of transplants that can be arranged. In our model, all patients are indifferent between any two matchings in $\mathcal{E}_\pi^n$, and hence, the priority mechanism can pick any one of them. Nevertheless, there can be other considerations affecting this selection, such as minimizing the number of compatible pairs that are matched.

4 The Structure of Pareto-Efficient Matchings and Comparative Statics

For any problem (N, R), partition the set of pairs $N = N_I \cup N_C$ as $\{U(N, R), O(N, R), P(N, R)\}$ where

$$\begin{aligned} U(N,R) &= \left\{x \in N_I : \exists \mu \in \mathcal{E}\left(N,R\right) \text{ s.t. } \mu\left(x\right) = x\right\}, \\ O(N,R) &= \left\{x \in N \setminus U(N,R) : \exists y \in U(N,R) \text{ s.t. } r_{y,x} = 1\right\}, \\ P(N,R) &= N \setminus \big(U(N,R) \cup O(N,R)\big). \end{aligned}$$

That is, $U(N, R)$ is the set of incompatible pairs each of which remains unmatched at a Pareto-efficient matching. We refer to $U(N, R)$ as **the set of underdemanded pairs**. Set $O(N, R)$ is the set of pairs that are not underdemanded and have a mutually compatible underdemanded pair. We will refer to $O\left(N, R\right)$ as **the set of**

overdemanded pairs. Set $P(N,R)$ is the remaining set of pairs, and we will refer to it as **the set of perfectly matched pairs**. Theorem 1, which we will shortly state, will justify this terminology. We refer to this decomposition of pairs as the **demand decomposition** of problem (N,R).

For any $K \subset N$, let $R_K = [r_{x,y}]_{x,y \in K}$ be the **feasible exchange submatrix** for the pairs in K. We refer to (K, R_K) as a **subproblem** of (N,R). A subproblem (K, R_K) is **connected** if for any $x, y \in K$ there exist $x^1, x^2, ..x^m \in K$ with $x^1 = x$ and $x^m = y$ such that for all $\ell \in \{1, ..., m-1\}$, $r_{x^\ell, x^{\ell+1}} = 1$. A connected subproblem (K, R_K) is a **component** of (N,R) if there is no other connected subproblem (L, R_L) such that $K \subsetneq L$.

Consider the subproblem $\left(N \setminus O(N,R), R_{N \setminus O(N,R)}\right)$ obtained by removal of all pairs in $O(N,R)$.

We refer to a component (K, R_K) of $\left(N \setminus O(N,R), R_{N \setminus O(N,R)}\right)$ as a **dependent component** if $K \subseteq N_I$ and $|K|$ is odd. We refer to a component (K, R_K) of $\left(N \setminus O(N,R), R_{N \setminus O(N,R)}\right)$ as a **self-sufficient component** if $K \cap N_C \neq \varnothing$ or $|K|$ is even. We will justify this choice of terminology in the theorem presented below. Let $\mathcal{D}$ denote the set of dependent components. Let $\mathcal{S}$ denote the set of self-sufficient components.

The following result characterizes the structure of the set of Pareto-efficient matchings for problem (N,R).

Theorem 1 *Given a problem (N,R), let (K, R_K) be the subproblem with $K = N \setminus O(N,R)$ (i.e. the subproblem where all overdemanded pairs are removed) and let μ be a Pareto-efficient matching for the original problem (N,R). Then,*

1. *For any pair $x \in O(N,R)$, $\mu(x) \in U(N,R)$.*

2.
   (a) *For any self-sufficient component (L, R_L) of (K, R_K), $L \subseteq P(N,R)$, and*

   (b) *for any incompatible pair $i \in L \cap N_I$, $\mu(i) \in L \setminus \{i\}$.*

3.
   (a) *For any dependent component (J, R_J) of (K, R_K), $J \subseteq U(N,R)$, and for any pair $i \in J$, it is possible to match all remaining pairs in J with each other.*

(b) *Moreover, for any dependent component* (J, R_J) *of* (K, R_K), *either*

   i. *one and only one pair* $i \in J$ *is matched with a pair in* $O(N, R)$ *in the Pareto-efficient matching* μ, *whereas all remaining pairs in* J *are matched with each other (so that all pairs in* J *are matched), or*

   ii. *one pair* $i \in J$ *remains unmatched in the Pareto-efficient matching* μ, *whereas all remaining pairs in* J *are matched with each other (so that only* i *remains unmatched among pairs in* J*).*

Our proof strategy will be based on an induction on the number of compatible pairs, as this approach helps us to execute a very useful comparative static exercise on how the structure of Pareto-efficient matchings evolves with the addition of a single compatible pair to the pool of pairs. These comparative static results will be proven inside the proof of the theorem as Claims 1 and 6.

We will invoke a well-known result from graph theory due to Hall [1935] in our proof. We first state this theorem here and then prove Theorem 1:

Hall's Theorem *Consider a graph with two finite sets* X, $\mathcal{Y}$ *such that each member of* X *is connected with some members of* $\mathcal{Y}$. *For any* $X' \subseteq X$, *let* $\mathcal{N}(X', \mathcal{Y}) \subseteq \mathcal{Y}$ *denote the set of members of* $\mathcal{Y}$ *each of which are connected with at least one member of* X'. *Then, we can match each* $x \in X$ *with a distinct connected member of* $\mathcal{Y}$ *if and only if*

$$\forall X' \subseteq X, \quad |\mathcal{N}(X', \mathcal{Y})| \geq |X'|.$$

Proof. [Proof of Theorem 1] We use an induction on the number of compatible pairs. Fix $s \geq 0$. Let N have $s + 1$ compatible pairs including pair c. Let $N^{-c} = N \setminus \{c\}$. Clearly N^{-c} has s compatible pairs. Let (N^{-c}, R^{-c}) be the problem such that $R^{-c} = R_{N^{-c}}$. The initial step, i.e., the case with no compatible pairs, was proven by Gallai [1963, 1964] and Edmonds [1965], and we refer to this result as the Gallai-Edmonds Decomposition (or GED for short) Theorem. Now, for induction, we make the following assumption:

Inductive Assumption Theorem 1 holds for problem (N^{-c}, R^{-c}).

Since $U(N, R) \subseteq N_I$, $c \notin U(N, R)$. Depending on whether it is mutually compatible with an underdemanded pair of (N^{-c}, R^{-c}) or not, our proof strategy will differ.

Below we show that when the latter is the case, nothing changes for the demand decomposition except c becoming a perfectly matched pair of (N, R):

Claim 1 *If c is not mutually compatible with any pair in $U(N^{-c}, R^{-c})$ then*

1. $U(N, R) = U(N^{-c}, R^{-c})$,
2. $O(N, R) = O(N^{-c}, R^{-c})$, *and*
3. $P(N, R) = P(N^{-c}, R^{-c}) \cup \{c\}$.

Moreover, Theorem 1 holds for problem (N, R).

Proof of Claim 1 We will prove $U(N, R) = U(N^{-c}, R^{-c})$, Part 1, which will immediately prove Parts 2 and 3 of the claim.

- First, we will show that $U(N, R) \supseteq U(N^{-c}, R^{-c})$: Let $\eta' \in \mathcal{E}(N^{-c}, R^{-c})$. We must show that $\eta' \in \mathcal{E}(N, R)$. Suppose not. Then there exists a matching $\mu \in \mathcal{M}(N, R)$ that Pareto-dominates η' under (N, R). Observe that $\mu(c) \neq c$, for otherwise $\mu \in \mathcal{M}(N^{-c}, R^{-c})$ and it would Pareto-dominate η' under (N^{-c}, R^{-c}) as well. Therefore, since c is not mutually compatible with any pair in $U(N^{-c}, R^{-c})$, $\mu(c) \in O(N^{-c}, R^{-c}) \cup P(N^{-c}, R^{-c})$. Let $\mu' = \mu \setminus \{(\mu(c), c)\}$. Since μ Pareto-dominates η' under (N, R), $|I^{\mu}| > |I^{\eta'}|$. Hence, $|I^{\mu'}| \geq |I^{\eta'}|$. As $\mu' \in \mathcal{M}(N^{-c}, R^{-c})$, by Proposition 1, this inequality should hold with equality and $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$. Recall that compatible pairs can only be matched with incompatible pairs. Thus, $\mu(c)$ is an incompatible pair. However, $\mu(c)$ is unmatched under μ', contradicting $\mu(c) \in O(N^{-c}, R^{-c}) \cup P(N^{-c}, R^{-c})$. Thus, $\eta' \in \mathcal{E}(N, R)$. This implies $\mathcal{E}(N, R) \supseteq \mathcal{E}(N^{-c}, R^{c})$, which in turn implies $U(N, R) \supseteq U(N^{-c}, R^{-c})$.

- Next, we will show that $U(N, R) \subseteq U(N^{-c}, R^{-c})$: We have already shown that $\mathcal{E}(N, R) \supseteq \mathcal{E}(N^{-c}, R^{-c})$. This together with Proposition 1 imply for any $\mu \in \mathcal{E}(N, R)$ and $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$, $|I^{\mu}| = |I^{\mu'}|$. Let $i \in U(N, R)$ and $\nu \in \mathcal{E}(N, R)$ such that $\nu(i) = i$. Let $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$. Observe that c is not matched under μ' and $\mu' \in \mathcal{E}(N, R)$. By Proposition 2, there exists $\eta \in \mathcal{E}(N, R)$ such that $M^{\eta} = C^{\mu'} \cup I^{\nu}$. Since $c \notin M^{\eta}$, $\eta \in \mathcal{M}(N^{-c}, R^{-c})$. Moreover since $I^{\eta} = I^{\nu}$, $\mathcal{E}(N, R) \supseteq \mathcal{E}(N^{-c}, R^{-c})$ along with Proposition 1 imply $\eta \in \mathcal{E}(N^{-c}, R^{-c})$. Observe that $\eta(i) = i$. Thus $i \in U(N^{-c}, R^{-c})$, and hence $U(N, R) \subseteq U(N^{-c}, R^{-c})$.

Finally, we will show that Theorem 1 holds for (N, R). Parts 1,2,3 of the claim along with the inductive assumption – Theorem 1 Parts 2(a) and 3(a) for (N^{-c}, R^{-c}) – together imply that Parts 2(a) and 3(a) of Theorem 1 hold for problem (N, R). By the inductive assumption – Theorem 1 Parts 1 and 3(b) for (N^{-c}, R^{-c}) – the maximum number of pairs in $U(N^{-c}, R^{-c}) = U(N, R)$ will be matched in a matching of (N^{-c}, R^{-c}) or (N, R), if we match each pair in $O(N^{-c}, R^{-c}) = O(N, R)$ with a pair in $U(N^{-c}, R^{-c}) = U(N, R)$, and for each $D \in \mathcal{D}(N^{-c}, R^{-c})$, we match (1) at most one pair in D with a pair in $O(N^{-c}, R^{-c}) = O(N, R)$, and (2) $|D| - 1$ pairs of D with each other. Thus, Theorem 1 Parts 1 and 3(b) also hold for problem (N, R). These parts in turn imply that Theorem 1 Part 2(b) holds for problem (N, R). $\diamond$

Claim 1 covers the easier of the two cases. We will next build the machinery needed for the harder case through a series of claims.

For any $Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$ and $\mathcal{F} \subseteq \mathcal{D}(N^{-c}, R^{-c})$, let

$$\mathcal{N}(Q, \mathcal{F}) \equiv \{F \in \mathcal{F} \ : \ \exists\, a \in Q \text{ and } i \in F \text{ such that } r_{i,a} = 1\}.$$

That is, the "neighbors" of pairs in Q among dependent components of $\mathcal{F}$ are represented by the set $\mathcal{N}(Q, \mathcal{F})$. [6]

First, we present the following corollary to the inductive assumption:

Claim 2 *For all* $Q \subseteq O(N^{-c}, R^{-c}), \quad |\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)| > |Q|$.

Proof of Claim 2 Suppose that for some $Q \subseteq O(N^{-c}, R^{-c}), \quad |\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)| \leq |Q|$. Then, by the inductive assumption, as all overdemanded pairs are matched in all efficient matchings of (N^{-c}, R^{-c}) to underdemanded pairs (by Part 1), with at most one from a dependent component (by Part 3), it should be the case that $|Q| = |\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)|$. But then, as all overdemanded pairs are always matched in an efficient matching, each pair in Q will be matched with a pair in a distinct component of $\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)$ (by Part 3(a) of the inductive assumption) and all remaining pairs in a component of $\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)$ will be matched with another pair in the component (by Part 3(b) of the inductive assumption), and in particular, all pairs in all components of $\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)$ will always be matched at all effi-

[6] For simplicity, when it is not ambiguous, we will simply refer to a component by its set of pairs, i.e., we will refer to $F \in \mathcal{F}$ instead of $(F, R_F^{-c}) \in \mathcal{F}$.

cient matchings of (N^{-c}, R^{-c}), contradicting that such pairs are underdemanded. We showed that for all $Q \subseteq O(N^{-c}, R^{-c})$, $\quad |\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| > |Q|$. $\diamond$

Our next claim easily follows from Claim 2:

Claim 3 *Let c be mutually compatible with a pair in $U(N^{-c}, R^{-c})$. Then for all $Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$,* $\quad |\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| \geq |Q|$.

Proof of Claim 3 If $Q = \{c\}$, then by the hypothesis of the claim and the inductive assumption that implies $U(N^{-c}, R^{-c}) = \bigcup_{D \in \mathcal{D}(N^{-c}, R^{-c})} D$, we have $|\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| \geq 1 = |Q|$. If $Q \neq \{c\}$, then let $Q' = Q \setminus \{c\}$. We have $|\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| \geq |\mathcal{N}(Q', \mathcal{D}(N^{-c}, R^{-c}))| \geq |Q'| + 1 \geq |Q|$, where the second inequality follows from Claim 2. $\diamond$

We are ready to identify pairs whose roles in the structure of Pareto efficient matchings will differ between problems (N, R) and (N^{-c}, R^{-c}). Let c be mutually compatible with an underdemanded pair of (N^{-c}, R^{-c}). Define

- $\hat{Q} \equiv \bigcup \{Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\} : \quad |\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| = |Q|\}$;
- $\hat{\mathcal{F}} = \mathcal{N}(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c}))$; and
- $\hat{F} = \bigcup_{F \in \hat{\mathcal{F}}} F$.

Observe that by Claim 2, either $c \in \hat{Q}$ or $\hat{Q} = \emptyset$.

Claim 4 $|\mathcal{N}(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c}))| = |\hat{Q}|$. *Thus, $Q \subseteq \hat{Q}$ for any $Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$ such that* $|\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| = |Q|$.

Proof of Claim 4 Suppose $Q', Q'' \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$ are such that $|\mathcal{N}(Q, \mathcal{D}(N^{-c}, R^{-c}))| = |Q|$ for each $Q \in \{Q', Q''\}$. It suffices to show that $|\mathcal{N}(Q'' \cup Q', \mathcal{D}(N^{-c}, R^{-c}))| = |Q'' \cup Q'|$. Suppose not. This and Claim 3 together imply

$$|Q'' \cup Q'| < |\mathcal{N}(Q'' \cup Q', \mathcal{D}(N^{-c}, R^{-c}))|. \tag{1}$$

Let $\mathcal{F}'' = \mathcal{N}(Q'', \mathcal{D}(N^{-c}, R^{-c}))$. Observe that

$$|\mathcal{F}''| = |Q''|, \tag{2}$$

and

$$\mathcal{N}\big(Q'' \cup Q', \mathcal{D}(N^{-c}, R^{-c})\big) = \mathcal{F}'' \cup \mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}''\big). \tag{3}$$

By Relations 1 and 3,

$$|Q''|+|Q'\setminus Q''| = |Q''\cup Q'| < \big|\mathcal{N}\big(Q''\cup Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |\mathcal{F}''|+\big|\mathcal{N}\big(Q'\setminus Q'', \mathcal{D}(N^{-c}, R^{-c})\setminus\mathcal{F}''\big)\big| \tag{4}$$

Relation 4 and Relation 2 together imply

$$|Q' \setminus Q''| < \big|\mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}''\big)\big|. \tag{5}$$

Let $\mathcal{F}^{\cap} = \mathcal{N}\big(Q'' \cap Q', \mathcal{D}(N^{-c}, R^{-c})\big)$. Then

$$\underbrace{\mathcal{N}\big(Q'' \cap Q', \mathcal{D}(N^{-c}, R^{-c})\big)}_{=\mathcal{F}^{\cap}} \subseteq \underbrace{\mathcal{N}\big(Q'', \mathcal{D}(N^{-c}, R^{-c})\big)}_{=\mathcal{F}''}. \tag{6}$$

and

$$\mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}''\big) \subseteq \mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}^{\cap}\big). \tag{7}$$

Relations 5 and 7 imply

$$|Q' \setminus Q''| < \big|\mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}^{\cap}\big)\big|. \tag{8}$$

Also observe that

$$\mathcal{N}\big(Q', \mathcal{D}(N^{-c}, R^{-c})\big) = \underbrace{\mathcal{N}\big(Q'' \cap Q', \mathcal{D}(N^{-c}, R^{-c})\big)}_{=\mathcal{F}^{\cap}} \cup \mathcal{N}\big(Q' \setminus Q'', \mathcal{D}(N^{-c}, R^{-c}) \setminus \mathcal{F}^{\cap}\big). \tag{9}$$

Relation 9 along with $\big|\mathcal{N}\big(Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |Q'|$ imply

$$|Q''\cap Q'|+|Q'\setminus Q''| = |Q'| = \big|\mathcal{N}\big(Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |\mathcal{F}^{\cap}|+\big|\mathcal{N}\big(Q'\setminus Q'', \mathcal{D}(N^{-c}, R^{-c})\setminus\mathcal{F}^{\cap}\big)\big|. \tag{10}$$

Finally, we obtain the contradiction we have sought: Relations 8 and 10 imply

$$|Q'' \cap Q'| > |\mathcal{F}^{\cap}| = \big|\mathcal{N}\big(Q'' \cap Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big|,$$

contradicting Claim 3. Thus, $\big|\mathcal{N}\big(Q'' \cup Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |Q'' \cup Q'|$. $\diamond$

Next define

- $\mathcal{G} = \mathcal{D}(N^{-c}, R^{-c}) \setminus \hat{\mathcal{F}}$.

We will use the following claim to invoke Hall's Theorem to prove Theorem 1 for the harder of our two cases.

Claim 5 *Let c be mutually compatible with a pair in $U(N^{-c}, R^{-c})$. Then for any $F \in \mathcal{G}$,*

$$\forall Q \subseteq \Big(O(N^{-c}, R^{-c}) \cup \{c\}\Big) \setminus \hat{Q}, \quad |\mathcal{N}\big(Q, \mathcal{G} \setminus \{F\}\big)| \geq |Q|.$$

<u>*Proof of Claim 5*</u> Fix $Q \subseteq \Big(O(N^{-c}, R^{-c}) \cup \{c\}\Big) \setminus \hat{Q}$ and $F \in \mathcal{G}$.

If $|\mathcal{N}(Q, \mathcal{G})| < |Q|$, then

$$|\mathcal{N}\big(\hat{Q} \cup Q, \mathcal{D}(N^{-c}, R^{-c})\big)| = \underbrace{|\mathcal{N}\big(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c})\big)|}_{=|\hat{\mathcal{F}}|=|\hat{Q}|} + |\mathcal{N}(Q, \mathcal{G})| < |\hat{Q}| + |Q| = |\hat{Q} \cup Q|,$$

contradicting Claim 3 as $\hat{Q} \cup Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$.

If $|\mathcal{N}(Q, \mathcal{G})| = |Q|$, then

$$|\mathcal{N}\big(\hat{Q} \cup Q, \mathcal{D}(N^{-c}, R^{-c})\big)| = \underbrace{|\mathcal{N}\big(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c})\big)|}_{=|\hat{\mathcal{F}}|=|\hat{Q}|} + |\mathcal{N}(Q, \mathcal{G})| = |\hat{Q}| + |Q| = |\hat{Q} \cup Q|.$$

But this contradicts the maximality of $\hat{Q}$ (i.e. the second part of Claim 4) since $\hat{Q} \cup Q \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$. Hence,

$$\forall Q \subseteq O(N^{-c}, R^{-c}) \setminus \hat{Q}, \quad |\mathcal{N}(Q, \mathcal{G})| > |Q|.$$

Thus, we have

$$\forall Q \subseteq O(N^{-c}, R^{-c}) \setminus \hat{Q}, \quad |\mathcal{N}\big(Q, \mathcal{G} \setminus \{F\}\big)| \geq |Q|.$$

$\diamond$

Finally, using the above preparatory claims, we characterize the demand decomposition when c is mutually compatible with a pair in $U(N^{-c}, R^{-c})$:

Claim 6 *Let c be mutually compatible with a pair in $U(N^{-c}, R^{-c})$. Then*

1. $U(N,R) = U(N^{-c}, R^{-c}) \setminus \hat{F}$,
2. $O(N,R) = \Big(O(N^{-c}, R^{-c}) \cup \{c\}\Big) \setminus \hat{Q}$, *and*
3. $P(N,R) = P(N^{-c}, R^{-c}) \cup \hat{Q} \cup \hat{F}$.

Moreover, Theorem 1 holds for problem (N,R).

Proof of Claim 6

- First, we will show that $U(N,R) \supseteq U(N^{-c}, R^{-c}) \setminus \hat{F}$: Recall that $\mathcal{G} = \mathcal{D}(N^{-c}, R^{-c}) \setminus \hat{\mathcal{F}}$. Fix $i \in U(N^{-c}, R^{-c}) \setminus \hat{F}$. By Part 3(a) of Theorem 1 for (N^{-c}, R^{-c}), $i \in F$ for some $F \in \mathcal{G}$. In several steps, we will construct a matching $\mu \in \mathcal{M}(N,R)$, which leaves i unmatched, and show that it is efficient under (N,R).

  ⋆ By Claim 5

$$\forall Q \subseteq \big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}, \quad |\mathcal{N}(Q, \mathcal{G} \setminus \{F\})| \geq |Q|. \tag{11}$$

  By Relation 11 and Hall's Theorem, we can match each pair in $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$ with a pair in a distinct component of $\mathcal{G} \setminus \{F\}$. Let μ match such pairs with each other. At this point some components of $\mathcal{G} \setminus \{F\}$ have only one pair matched in μ, whereas the rest have all pairs unmatched. By Part 3(a) of Theorem 1 for (N^{-c}, R^{-c}), we can also match still-unmatched $|D| - 1$ pairs in any component $D \in \mathcal{G} \setminus \{F\}$ with each other and all pairs in $F \setminus \{i\}$ with each other. Let μ match also such pairs with each other. Observe that $\mu(i) = i$.

  By the definition of $\mathcal{D}(N^{-c}, R^{-c})$ and construction of $\hat{Q}$, any pair that belongs to any dependent component D in $\mathcal{G}$ is mutually compatible with only pairs in D or $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$. Also recall that each dependent component in $\mathcal{G}$ consists of an odd number of incompatible pairs. Thus, so far,

$$\mu \in \arg\max_{\nu \in \mathcal{M}(N,R)} \Big| T^{\nu} \cap \big[\big(U(N^{-c}, R^{-c}) \setminus \hat{F}\big) \cup \big[\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}\big]\big]\Big|, \tag{12}$$

  i.e., the maximum possible number of pairs in the set $\big(U(N^{-c}, R^{-c}) \setminus \hat{F}\big) \cup \big[\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}\big]$ receive a transplant in μ.

⋆ Claim 3 together with $\hat{Q} \subseteq O(N^{-c}, R^{-c}) \cup \{c\}$ imply

$$\forall Q \subseteq \hat{Q}, \qquad |\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)| \geq |Q|. \tag{13}$$

Hence, we can invoke Hall's Theorem through Relation 13 once again and match each pair in $\hat{Q}$ with an incompatible pair in a distinct dependent component in $\hat{\mathcal{F}}$. Let μ match such pairs with each other. At this point, as $|\hat{F}| = \big|\mathcal{N}\big(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |\hat{Q}|$, one pair in each $D \in \hat{\mathcal{F}}$ is matched in μ. By Part 3(a) of Theorem 1 for (N^{-c}, R^{-c}), we can also match yet-unmatched $|D| - 1$ pairs in each dependent component $D \in \hat{\mathcal{F}}$ with each other. Let μ further be constructed to match such pairs with each other. Thus, μ matches all pairs in $\hat{Q} \cup \hat{F}$ with each other, and so far μ is well defined. Moreover,

$$\mu \in \arg \max_{\nu \in \mathcal{M}(N,R)} |T^{\nu} \cap (\hat{Q} \cup \hat{F})|, \tag{14}$$

i.e., the maximum possible number of pairs in the set $\hat{Q} \cup \hat{F}$ receive a transplant in μ.

⋆ By Part 2(b) of Theorem 1 for (N^{-c}, R^{-c}), we can further construct μ such that all incompatible pairs in $P(N^{-c}, R^{-c})$ are matched with other pairs in $P(N^{-c}, R^{-c})$. Hence, μ is well defined and $\mu \in \mathcal{M}(N, R)$. Moreover, having matched all incompatible pairs in $P(N^{-c}, R^{-c})$,

$$\mu \in \arg \max_{\nu \in \mathcal{M}(N,R)} |T^{\nu} \cap P(N^{-c}, R^{-c})|, \tag{15}$$

i.e., the maximum possible number of pairs in the set $P(N^{-c}, R^{-c})$ receive a transplant in μ.

By Equations 12, 14, and 15, $|T^{\mu}| = \max_{\nu \in \mathcal{M}(N,R)} |T^{\nu}|$. This together with Proposition 1 implies $\mu \in \mathcal{E}(N, R)$. Since $\mu(i) = i$, we have $i \in U(N, R)$.

- Next, we will show that $U(N, R) \subseteq U(N^{-c}, R^{-c}) \setminus \hat{F}$: It is possible to match all incompatible pairs in $\hat{F} \cup \hat{Q} \cup P(N^{-c}, R^{-c})$ with other pairs in the same set, as the matching μ constructed above does that. By the definition of $\mathcal{D}(N^{-c}, R^{-c})$ and construction of $\hat{Q}$, any pair that belongs to any dependent component

D in $\mathcal{G} = \mathcal{D}(N^{-c}, R^{-c}) \setminus \hat{\mathcal{F}}$ is mutually compatible with only pairs in D or $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$. Also recall that such a component D consists of an odd number of incompatible pairs. Thus, to maximize the number of incompatible pairs matched under (N, R), we need to match all pairs in $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$ with pairs in $U(N^{-c}, R^{-c}) \setminus \hat{F}$, at most one pair from each $D \in \mathcal{G}$ with a pair in $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$, and $|D| - 1$ pairs of D with each other. This is possible, as matching μ constructed above does just that. Hence, as by Proposition 1 any efficient matching $\nu \in \mathcal{E}(N, R)$ maximizes the number of incompatible pairs matched, we should have all incompatible pairs in $\hat{F} \cup O(N^{-c}, R^{-c}) \cup P(N^{-c}, R^{-c})$ matched in ν, implying any unmatched incompatible pair under ν must belong $N_I \setminus \big\{\hat{F} \cup O(N^{-c}, R^{-c}) \cup P(N^{-c}, R^{-c})\big\} = U(N^{-c}, R^{-c}) \setminus \hat{F}$. Hence $U(N, R) \subseteq U(N^{-c}, R^{-c}) \setminus \hat{F}$.

Thus, $U(N, R) = U(N^{-c}, R^{-c}) \setminus \hat{F}$, i.e., Part 1 of the claim holds, which in turn implies Parts 2 and 3 of the claim. These parts and the inductive assumption – Theorem 1 Parts 2(a) and 3(a) for (N^{-c}, R^{-c}) – together imply that Parts 2(a) and 3(a) of Theorem 1 hold for problem (N, R). Recall that to maximize the number of incompatible pairs matched under (N, R), we need to match all pairs in $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$ with pairs in $U(N^{-c}, R^{-c}) \setminus \hat{F}$, at most one pair from each $D \in \mathcal{G} = \mathcal{D}(N^{-c}, R^{-c}) \setminus \hat{\mathcal{F}}$ with a pair in $\big(O(N^{-c}, R^{-c}) \cup \{c\}\big) \setminus \hat{Q}$, and $|D| - 1$ pairs of D with each other. This with Proposition 1 implies that Part 1 and Part 3(b) of Theorem 1 hold for problem (N, R), which in turn implies Part 2(b) of Theorem 1 holds for problem (N, R). $\diamond$ $\blacksquare$

We conclude the section with three remarks on the structure of Pareto-efficient matchings:

Remark 1 Observe that Claims 1 and 6 in the proof of the theorem give a complete picture of the evolution of the demand decomposition when a new compatible pair c joins the pool:

- If c is not mutually compatible with a pair in $U(N^{-c}, R^{-c})$, the overdemanded set and the set of dependent components do not change; on the other hand, each self-sufficient component of (N, R) is either a self-sufficient component of (N^{-c}, R^{-c}) or a super component that includes c and possibly some other self-sufficient components of (N^{-c}, R^{-c}).

- If c is mutually compatible with a pair in $U(N^{-c}, R^{-c})$, the overdemanded set is determined through removal of pairs in $\hat{Q}$ from the prior overdemanded set union $\{c\}$, and
  - ⋆ each dependent component of (N, R) is a dependent component of (N^{-c}, R^{-c}) that is not covered by the set $\hat{\mathcal{F}}$;
  - ⋆ each self-sufficient component of (N, R) is either a self-sufficient component of (N^{-c}, R^{-c}) or is a super component containing the pairs in $\hat{Q} \cup \hat{F}$, which have newly joined the perfectly matched set, and possibly some self-sufficient components of (N^{-c}, R^{-c}).

Remark 2 The number of incompatible pairs (un)matched in an efficient matching (when a compatible pair joins the pool) can fully be determined through the same statistic prior to the addition of compatible pair c: In the problem (N, R),

- If c is not mutually compatible with a pair in $U(N^{-c}, R^{-c})$, then any efficient matching of (N, R) leaves $|U(N^{-c}, R^{-c})| - (|O(N^{-c}, R^{-c})| - |\mathcal{D}(N^{-c}, R^{-c})|)$ incompatible pairs unmatched, the same number as an efficient matching of the problem (N^{-c}, R^{-c}).
- if c is mutually compatible with a pair in $U(N^{-c}, R^{-c})$, then any efficient matching of (N, R) leaves $|U(N^{-c}, R^{-c})| - (|O(N^{-c}, R^{-c})| - |\mathcal{D}(N^{-c}, R^{-c})|) - 1$ incompatible pairs unmatched, one less than an efficient matching of the problem (N^{-c}, R^{-c}).

This remark requires a short proof:

The first bullet point is proven as follows: By $\mathcal{E}(N^{-c}, R^{-c}) \subseteq \mathcal{E}(N, R)$ and Proposition 1, $|T^{\mu'}| = |T^{\mu}|$ for all $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$ and $\mu \in \mathcal{E}(N, R)$. By Theorem 1, any efficient matching of (N^{-c}, R^{-c}) leaves $|U(N^{-c}, R^{-c})| - (|O(N^{-c}, R^{-c})| - |\mathcal{D}(N^{-c}, R^{-c})|)$ incompatible pairs unmatched. So does any efficient matching of (N, R).

The second bullet point is shown as follows: Let $i \in U(N^{-c}, R^{-c})$ be such that $r_{i,c} = 1$. For an efficient matching $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$ with $\mu'(i) = i$, we have $\mu = \mu' \cup \{(i, c)\} \in \mathcal{M}(N, R)$. Thus, μ matches one more incompatible pair than μ', which leaves $|U(N^{-c}, R^{-c})| - (|O(N^{-c}, R^{-c})| - |\mathcal{D}(N^{-c}, R^{-c})|)$ incompatible pairs unmatched by Theorem 1. Suppose $\mu \notin \mathcal{E}(N, R)$. Then, by Proposition 1, there is a matching

$\nu \in \mathcal{M}(N, R)$ that matches more incompatible pairs than μ. Matching ν would necessarily match c. Matching $\nu' = \nu \setminus \{(c, \mu(c))\}$ will match one more incompatible pair than μ' and $\nu' \in \mathcal{M}(N^{-c}, R^{-c})$. This through Proposition 1 contradicts $\mu' \in \mathcal{E}(N^{-c}, R^{-c})$. Thus, $\mu \in \mathcal{E}(N, R)$, completing the proof of Remark 2.

Remark 3 One other observation regards Corollary 1, which states the possibility of minimizing the number of compatible pairs matched in any efficient matchings. Theorem 1 immediately places a constraint on which compatible pairs need to be matched at all efficient matchings, while we have more flexibility in deciding which ones to match or not: The compatible pairs in $N_C \cap O(N, R)$ should be matched at every efficient matching. On the other hand, the number of required compatible pairs in $N_C \cap P(N, R)$ can be optimized so that the minimum number of compatible pairs in this set are matched at an efficient matching.

5 Conclusion

Motivated by the increased willingness of the transplantation community to consider altruistically unbalanced kidney exchanges, we analyzed the impact of including compatible pairs in kidney exchange pools. We have shown that the GED structure that is available in the absence of compatible pairs is also preserved when compatible pairs are present. Not only is the elegant structure of the set of Pareto efficient matchings preserved, the role played by compatible pairs is also highly intuitive and structured. We have shown that the inclusion of each compatible pair benefits the entire patient population, and thus unlike other design considerations that provide efficiency gains at the expense of harming various subsets of patients, inclusion of compatible pairs provides much larger gains without any adverse distributional effects.

Motivated by our analysis, Yilmaz [2011b] considers the impact of inclusion of two-way list exchanges to the system rather than altruistically unbalanced kidney exchanges. The idea is the integration of incompatible pairs who are willing to exchange the donor's live kidney with a deceased donor kidney. He shows that the graph-theoretic structure of his model can be interpreted as an extension of the graph-theoretic structure of our model. However, despite the close relation between the two models, he shows that a GED-like decomposition no longer exists in his framework. In particular, the number of patients who receive live donor transplants no longer remain the same across Pareto-efficient matchings. His analysis shows that the GED structure cannot be taken for granted even in a relatively small modification to our model.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88:233–260, 1999.

Anna Bogolomania and Herve Moulin. Random matching under dichotomous preferences. *Econometrica*, 72:257–279, 2004.

Francis L. Delmonico. Exchanging kidneys: Advances in living-donor transplantation. *New England Journal of Medicine*, 350:1812–1814, 2004.

Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17: 449–467, 1965.

Tibor Gallai. Kritische graphen ii. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 8: 373–395, 1963.

Tibor Gallai. Maximale systeme unabhangiger kanten. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 9:401–413, 1964.

S. E. Gentry, D. L. Segev, M. Simmerling, and R. A. Montgomery. Expanding kidney paired donation through voluntary participation by compatible donors. *American Journal of Transportation*, 7(10):2361–2370, Oct 2007.

David W. Gjertson and J. Michael Cecka. Living unrelated donor kidney transplantation. *Kidney International*, 58:491–499, 2000.

Philip Hall. On representatives of subsets. *Journal of London Mathematical Society*, 10:26–30, 1935.

Bernhard Korte and Jens Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer, 2002.

Laszlo Lovasz and Michael D. Plummer. *Matching Theory*. North Holland, 1986.

Antonio Nicolò and Carmelo Rodriguez-Álvarez. Age-based preferences: Incorporating compatible pairs into paired kidney exchange. Working paper, 2011.

Lloyd E. Ratner, Abbas Rana, Emily R. Ratner, Victoria Ernst, Joan Kelly, Donald Kornfeld, David Cohen, and Ilona Wiener. The altruistic unbalanced paired kidney exchange: Proof of concept and survey of potential donor and recipient attitudes. *Transplantation*, 89:15–22, 2010.

Laine Friedman Ross and E. Steve Woodle. Ethical issues in increasing living kidney donations by expanding kidney paired exchange programs. *Transplantation*, 69: 1539–1543, 2000.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125:151–188, 2005a.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. A kidney exchange clearinghouse in new england. *American Economic Review Papers and Proceedings*, 95:376–380, 2005b.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review*, 97:828–851, 2007.

Susan L. Saidman, Francis L. Delmonico, Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Utilizing list exchange and non-directed donation through "chain" paired kidney donations. *American Journal of Transplantation*, 6:2694–2705, 2006.

David Steinberg. Compatible-incompatible live donor kidney exchanges. *Transplantation*, 91:257–260, 2011.

Özgür Yilmaz. Kidney exchange: An egalitarian mechanism. *Journal of Economic Theory*, 146:592–618, 2011a. Working paper.

Özgür Yilmaz. Paired kidney donation and listed exchange. Working Paper, 2011b.

Stefanos A. Zenios, E. Steve Woodle, and Lainie Friedman Ross. Primum non nocere: Avoiding increased waiting times for individual racial and blood-type subsets of kidney wait list candidates in a living donor/cadaveric donor exchange program. *Transplantation*, 72:648–654, 2001.