Altruistically Unbalanced Kidney Exchange*

Tayfun Sönmez† M. Utku Ünver‡

Boston College *Boston College*

First Draft: January 2010. Revised: June 2013.

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 because of institutional constraints. We explore the structure of Pareto-efficient matchings in this framework. 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. We also thank an associate editor and a referee for their useful comments.

†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, which led to an interdisciplinary collaboration between economists and medical professionals to establish several kidney exchange programs. To explain and motivate the contribution of the current paper, it is essential to describe how this collaboration has evolved over the years, and led to new innovations in kidney exchange. In the early 2000s, economists observed that the two main types of kidney exchanges conducted in the US correspond 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* mechanism (TTCC) [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 live donors who fail to receive a transplant in the absence of kidney exchanges, fewer than 10 percent would remain without a transplant under TTCC.

When economists shared their findings with the medical community, two reservations were expressed regarding RSÜ [2004]. 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 US where doctors assume that two compatible living-donor kidneys have essentially the same survival rates [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 the discrete-optimization literature,[1] some of the techniques of which were recently imported to economic theory by Bogomolnaia 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 the New England Kidney Exchange Program (NEPKE) – the first exchange program using an optimization-based mechanism – and the Alliance for Paired Donation (APD), both of which were formed as a result of the collaboration between economists and medical professionals. Most recently, the National Kidney Paired Donation Pilot Program in the US and National Program in the UK were established based on similar principles.[3]

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

[2] See Yılmaz [2011a] for an application of this two-sided matching approach in kidney exchange.

[3] These four programs also allow for three-way exchanges based on findings of RSÜ [2007] and Saidman et al. [2006], see below.

An earlier, abstract version of the RSÜ [2005a] model was analyzed extensively 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 to the GED Theorem has a very plausible implication for pairwise kidney exchange: the same number of patients are matched at every Pareto-efficient matching. Hence, a program never matches a high-priority patient at the expense of multiple patients under the Pareto-efficient *pairwise priority mechanisms* offered by RSÜ [2005a]. This result does not hold for TTCC or more generally for mechanisms that allow larger exchanges than pairwise. Hence, from a medical ethics perspective it gives pairwise priority mechanisms an edge. However, this advantage comes at a high cost to aggregate patient welfare: compared to TTCC, 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 because of either blood-type incompatibility or tissue-type incompatibility. There are four blood types, A, B, AB, O, where 44 percent of the US 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; and
- a B blood-type donor is blood-type compatible with only B and AB blood-type patients;
- 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 US 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 across various kidney exchange programs are blood-type incompatible.

A key observation illustrating the key role of compatible pairs in kidney exchange 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 engage in exchange with any blood-type-incompatible pair. Hence each blood-type incompatible pair needs a distinct blood-type-compatible pair to engage in exchange. 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, virtually all pairs are available for exchange. This is by far the most important reason for the large aggregate patient welfare gap between the RSÜ [2004] TTCC mechanism and the 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 aggregate patient welfare. Based on these results, all major kidney exchange programs, including the pilot national kidney exchange program in the US, adopted mechanisms that allow for three-way exchanges. One negative implication of this flexibility is the loss of the feature that an equal number of patients receive transplants in all Pareto-efficient matchings. In particular, the priority mechanism used by NEPKE does not necessarily maximize the number of patients receiving transplants. That is perhaps a small price to pay in comparison to welfare gains from three-way exchanges, but there is an alternative that not only dramatically increases the welfare gains from kidney exchange but also overcomes the potential welfare loss associated with priority mechanisms. Under this new paradigm, the elegant GED structure of Pareto-efficient matchings – which no longer exists in the presence of three-way exchanges – is restored.

This alternative was advocated by RSÜ [2005b], who proposed the inclusion of compatible pairs in the kidney exchange pool. They emphasized that the inclusion of compatible pairs in the kidney exchange pool would produce the largest patient welfare gains in comparison to a number of other design modifications tailored to improve patient welfare. 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 patient welfare is due to the elimination of the above-discussed asymmetry, getting to the root of the problem.

Although RSÜ [2005b] were the first to advocate the inclusion of compatible pairs in kidney exchange, they were not the first to introduce the idea. This type of kidney exchange was originally 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 did not fully close the door on its implementation.

As a result to this strong objection, altruistically unbalanced kidney exchange received no attention until RSÜ [2005b] strongly advocated for the inclusion of compatible pairs in kidney exchange. This message has reached the transplantation community, and a number of recent papers in the transplantation literature also make a case for altruistically unbalanced kidney exchange.[4] Gentry et al. [2007] verified the large efficiency gains from the inclusion of compatible pairs in the exchange pool. Ratner et al. [2010] reported 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. The Texas Transplant Institute in San Antonio, TX is a transplant center that has successfully utilized compatible pairs in its exchange program.[5] 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 shift, in this paper, we consider a pairwise kidney exchange model in which both compatible and incompatible pairs are available for exchange. Our

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

[5]Ratner et al. [2010] also reported 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 another altruistically unbalanced kidney exchange program.

main focus is understanding the structure of Pareto-efficient matchings 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). Motivated by this observation, we propose priority allocation rules as mechanisms that match the maximum number of incompatible pairs no matter how we choose the priority order. Although incentive issues have proved to be of secondary importance in kidney exchange mechanisms, we also show that these mechanisms are incentive compatible in the Appendix (Theorem 2).

We further 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 can be minimized, 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, 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 difficult distributional choices to increase the number of patients who benefit, 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.[6]

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.[7] 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**.[8]

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

[6]The term "incompatible pairs" may lead to some ambiguity. We will use this term to refer to the plural of the term "incompatible pair." Pairs that cannot feasibly participate in an exchange will be referred to as "pairs that are not mutually compatible."

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

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

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

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

Here $(x, y) \in \mu$ means that the patient of each pair receives 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).

For any $\mu \in \mathcal{M}(N, R)$ and $(x, y) \in \mu$, define $\mu(x) \equiv y$ and $\mu(y) \equiv x$. Here pairs x and y are **matched** with each other under 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** under μ. For any matching μ, let M^μ denote the set of pairs that are matched under μ. Formally,

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

Observe that an incompatible pair receives a transplant under a matching μ only if it is matched with another pair 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** under 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 under 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 under μ. That is,

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

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 patients who receive a transplant is the same under any Pareto-efficient matching and that number is equal to the maximum number of pairs that can receive a transplant under any 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 of Proposition 1. It is straightforward to see that if $|T^{\mu}| = \max_{\eta \in \mathcal{M}} |T^{\eta}|$ then μ is Pareto efficient.

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 under μ or under 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 under μ but not under 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 under μ but not under 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 under ν but not under 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 under ν but not under η, whereas a_0 is matched under η but not under 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 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).

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 under μ or under 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 $\ell \in \{0, 1, \ldots, k-1\}$.

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.

There are three cases to consider:

Case 1. k is odd:

In this case both a_0 and a_k are matched under μ, but not under 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 under ν but not under μ. In contrast, a_0, an incompatible pair, is matched under μ but not under 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 under 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. The following corollary states that such a policy does not affect the choice of incompatible pairs.

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

3.1 The Priority Mechanisms

The experience of transplant centers is mostly with the priority allocation systems used to allocate cadaver organs. NEPKE adopted a variant of a priority allocation system for ordinary kidney exchanges. Priority mechanisms can be easily adapted to kidney exchanges that are or are not altruistically unbalanced.

Let $|N_I| = n$. A **priority order** 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

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

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 conflict between priority allocation and aggregate patient welfare maximization.

We also inspect the incentive properties of a priority mechanism in the Appendix. A priority mechanism is immune to any patient's manipulations by declaring some of the pairs whose donors are compatible with this patient to be incompatible. Through our interactions with the medical community, we have observed that manipulations of this sort do not play a significant role, since compatibility information is usually obtained from observable and verifiable medical data. Because of this, we relegate this result and its proof to the Appendix.

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[9]

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

That is, $U(N,R)$ is the set of incompatible pairs each of which remains unmatched under at least one 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 $R_K=\left[r_{x,y}\right]_{x,y\in K}$ 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 connected subproblem (L,R_L) such that $K\subsetneqq L$.

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

We refer to a component (K,R_K) of $\left(N\setminus O\left(N,R\right),R_{N\setminus O(N,R)}\right)$ as a **dependent component** if $K\subseteq N_I$ and $|K|$ is odd. We refer to a component (K,R_K) of $\left(N\setminus O\left(N,R\right),R_{N\setminus O(N,R)}\right)$ 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. We explain these definitions through the following example:

Example 1. Consider the problem in Figure 1. It consists of 17 incompatible pairs $i_0,i_1,...,i_{16}$, and two compatible pairs c_0 and c_1. Since we will analyze the effect of adding an additional compatible pair c to this problem in Example 2, denote it as (N^{-c},R^{-c}). In

[9] In this section and in the Appendix, we no longer fix a problem. Therefore, we denote the problem that we are referring to together with the notation for the relevant concepts in this section. For example, we denote the set of Pareto-efficient matchings of problem (N,R) by $\mathcal{E}(N,R)$ instead of just $\mathcal{E}$.

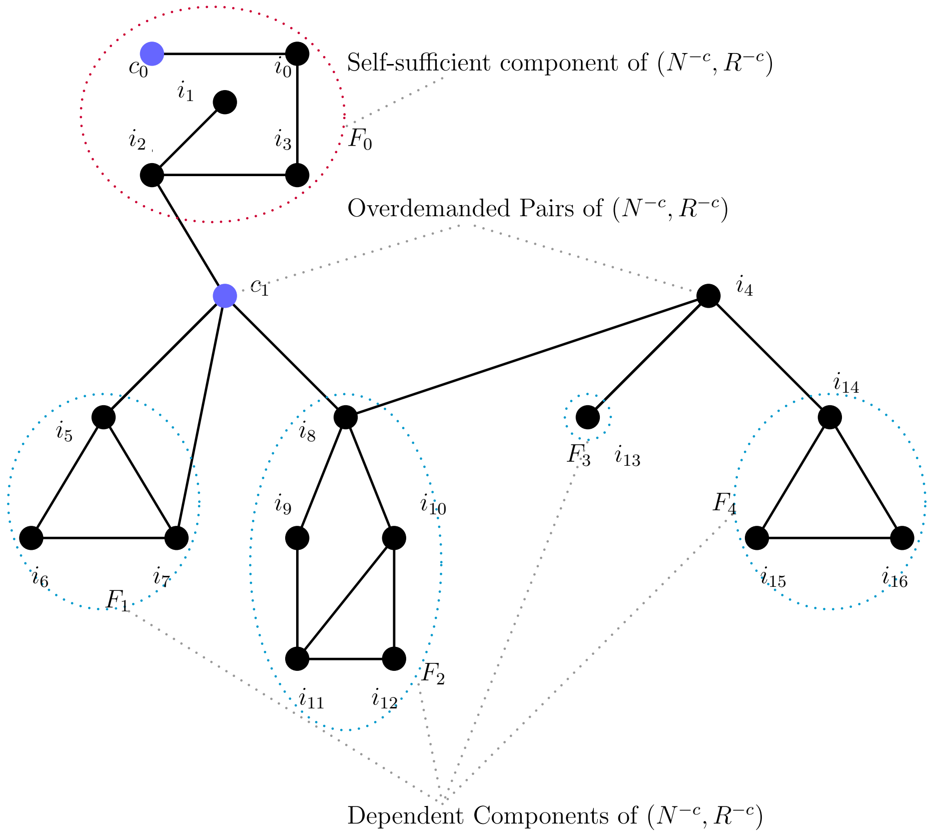

Figure 1: Problem (N^{-c}, R^{-c}) in Example 1

Figure 1, we show the demand decomposition of (N^{-c}, R^{-c}). There are two overdemanded pairs c_1 and i_4. In general, an algorithm needs to be executed to find the overdemanded pairs; however, in this example it is relatively straightforward to verify that $U(N^{-c}, R^{-c}) = \{i_5, i_6, ..., i_{16}\}$, $O(N^{-c}, R^{-c}) = \{c_1, i_4\}$, and $P(N^{-c}, R^{-c}) = \{c_0, i_0, i_1, i_2, i_3\}$. There is only one self-sufficient component $(F_0, R^{-c}_{F_0})$ where $F_0 = \{c_0, i_0, i_1, i_2, i_3\}$, and four dependent components $(F_1, R^{-c}_{F_1})$, $(F_2, R^{-c}_{F_2})$, $(F_3, R^{-c}_{F_3})$, and $(F_4, R^{-c}_{F_4})$ where $F_1 = \{i_5, i_6, i_7\}$, $F_2 = \{i_8, i_9, i_{10}, i_{11}, i_{12}\}$, $F_3 = \{i_{13}\}$, and $F_4 = \{i_{14}, i_{15}, i_{16}\}$. Thus, $\mathcal{S} = \{(F_0, R^{-c}_{F_0})\}$ and $\mathcal{D} = \{(F_1, R^{-c}_{F_1}), (F_2, R^{-c}_{F_2}), (F_3, R^{-c}_{F_3}), (F_4, R^{-c}_{F_4})\}$.

To verify these observations, note that all incompatible pairs in F_0 can be matched with each other and c_0 is not needed to be matched. On the other hand, in each of F_1,...,F_4, all but one pair can be matched with other pairs within the same set. Pairs c_1 and i_4 can be used to match the remaining pair in two of the $F_1, ..., F_4$. That is the only way to match the maximum number of incompatible pairs and hence obtain a Pareto-efficient matching (c.f. Proposition 1). Therefore, for any incompatible pair $i \in F_k$ for $k = 1, ..., 4$, there exists a Pareto-efficient matching that leaves i unmatched. These observations imply that $U(N^{-c}, R^{-c}) = F_1 \cup F_2 \cup F_3 \cup F_4$, $O(N^{-c}, R^{-c}) = \{c_1, i_4\}$, and $P(N^{-c}, R^{-c}) = F_0$.

Hence, under all Pareto-efficient matchings, all pairs but one in each of F_1, F_2, F_3, F_4 are matched with one another, and both overdemanded pairs, c_1 and i_4, are matched with pairs in F_1, F_2, F_3, F_4, and hence all incompatible pairs in F_0 are matched with other pairs in F_0. Moreover, the set of underdemanded pairs consists of the pairs in dependent components and the set of perfectly matched pairs consists of the pairs in self-sufficient components. These observations will be important to understand the implications of Theorem 1 given below. $\diamondsuit$

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 is 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 are proven within the proof of the theorem as Claims 1 and 6.

Before we prove our theorem, we illustrate our proof technique and claims through an example:

Example 2 (Continuation of Example 1)**.** We add a new compatible pair, c, to the problem in Figure 1. Let the new problem be denoted as (N, R). Two cases are possible:

Case 1. Pair c is not mutually compatible with any underdemanded pair of $(N,^{-c}, R^{-c})$:[10] In this case, either c by itself becomes a self-sufficient component of (N, R) or it joins one or more self-sufficient components of (N^{-c}, R^{-c}) to form a new self-sufficient component of (N, R). We have $O(N, R) = O(N^{-c}, R^{-c})$ and $\mathcal{D}(N, R) = \mathcal{D}(N^{-c}, R^{-c})$. Moreover, the remaining self-sufficient components of (N^{-c}, R^{-c}) become the other self-sufficient components of (N, R).

Case 2. Pair c is mutually compatible with some underdemanded pair of $(N,^{-c}, R^{-c})$:[11] Two subcases are possible:

(a) Compatible pair c, potentially together with some overdemanded and perfectly matched pairs of (N^{-c}, R^{-c}), joins some underdemanded pairs of (N^{-c}, R^{-c}) to form a new self-sufficient component: This subcase is illustrated with an example in Figure 2. Pair c is mutually compatible with the underdemanded pair i_{10} of dependent component $(F_2, R^{-c}_{F_2})$ of (N^{-c}, R^{-c}). In this case c, overdemanded pair c_0, dependent components $(F_1, R^{-c}_{F_1})$ and $(F_2, R^{-c}_{F_2})$, and self-sufficient component $(F_0, R^{-c}_{F_0})$ of (N^{-c}, R^{-c}) form the only self-sufficient component of (N, R). Overdemanded pair i_4 of (N^{-c}, R^{-c}) becomes the only overdemanded pair of (N, R). Dependent components $(F_3, R^{-c}_{F_3})$ and $(F_4, R^{-c}_{F_4})$ of (N^{-c}, R^{-c}) become the dependent components of (N, R).

(b) Compatible pair c becomes a new overdemanded pair: This subcase is illustrated with an example in Figure 3. Pair c is mutually compatible with underdemanded pairs i_{10} and i_{13} of two distinct dependent components of (N^{-c}, R^{-c}). In this case c joins the overdemanded pairs of (N^{-c}, R^{-c}) to form the set of overdemanded pairs of (N, R). The dependent and self-sufficient components of (N, R) are identical to those of (N^{-c}, R^{-c}). $\diamondsuit$

[10]This case is covered by Claim 1 in the proof of Theorem 1.
[11]This case is covered by Claim 6 in the proof of Theorem 1.

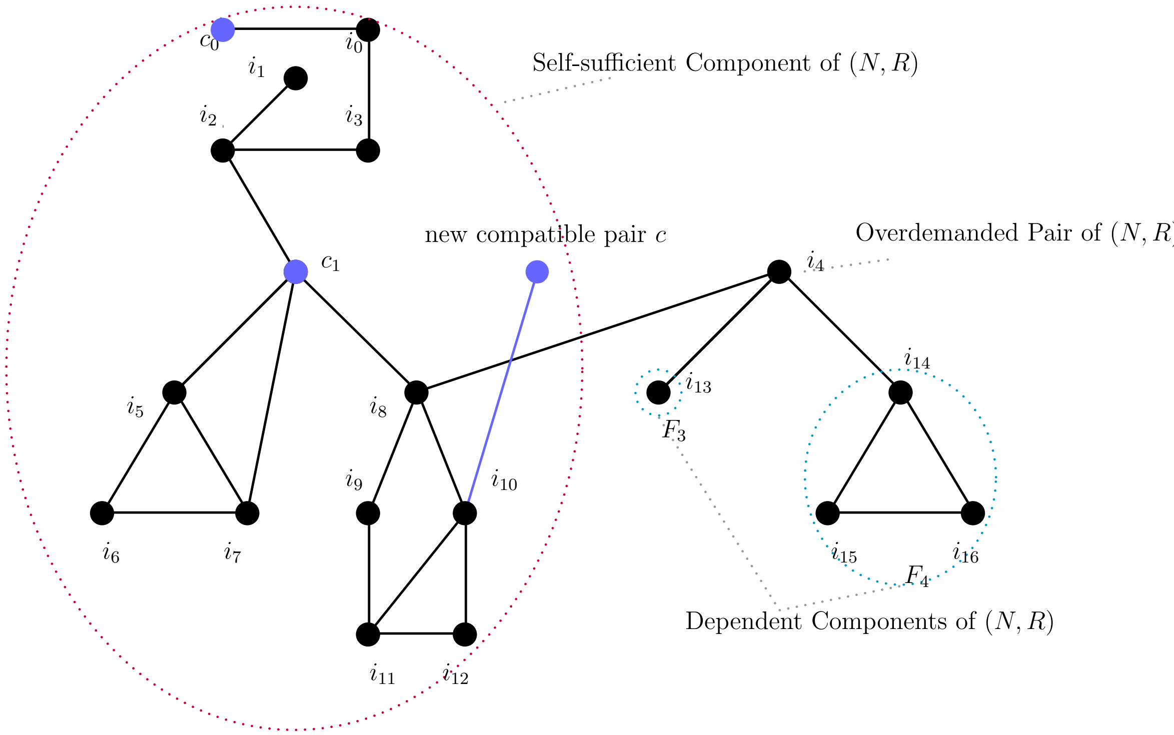

Figure 2: Problem (N, R) in Example 2 - Case 2(a)

We will rely on the following well-known result by Hall [1935] in our proof of Theorem 1 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 is 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 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:

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

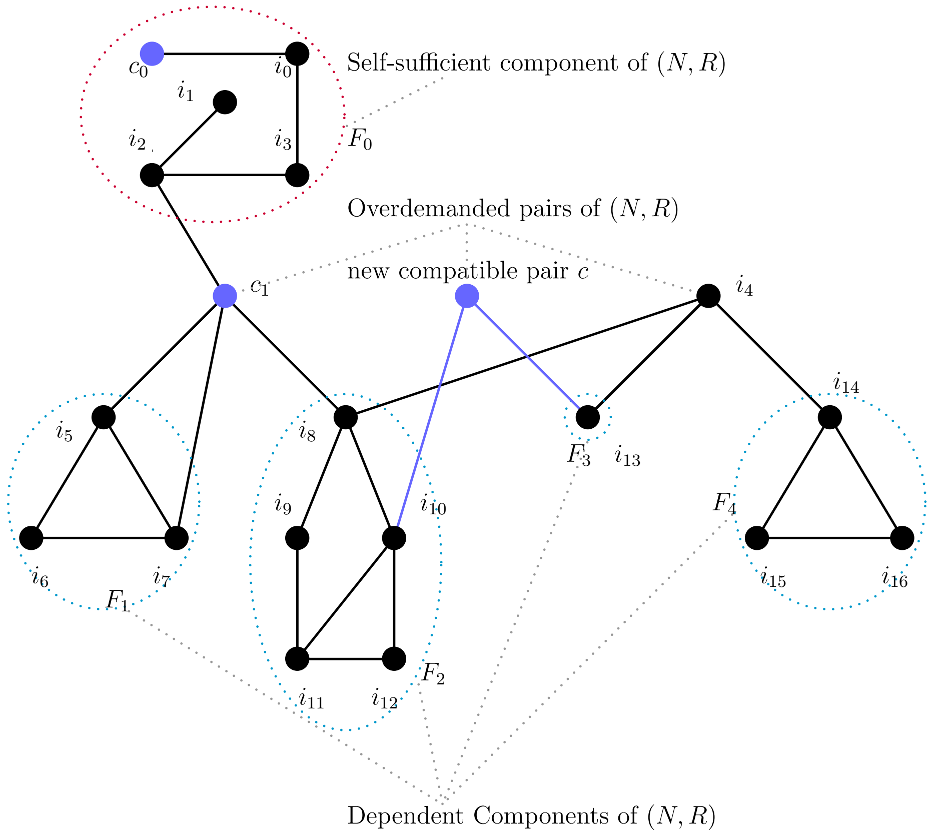

Figure 3: Problem (N, R) in Example 2 - Case 2(b)

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 $\mu' = \mu \backslash \{(\mu(c), c)\}$. 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 implies 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 implies $\eta \in \mathcal{E}(N^{-c}, R^{-c})$. Observe that $\eta(i) = i$. Thus $i \in U(N^{-c}, R^{-c})$, and hence $U(N, R) \subseteq U(N^{-c}, R^{-c})$.

Given the above three parts, we will show that Theorem 1 holds for (N, R). By induction assumption 2(b), all incompatible pairs in $P(N, R) = P(N^{-c}, R^{-c}) \cup \{c\}$ can be matched within the set without using c, which is compatible and not mutually compatible with any pair in $U(N, R) = U(N^{-c}, R^{-c})$. Moreover, by induction assumptions 1 and 3(b), matching some pair in $P(N, R)$ with a pair in $O(N, R) = O(N^{-c}, R^{-c})$ would lead to a strictly lower number of matched incompatible pairs in $U(N, R) = U(N^{-c}, R^{-c})$. Hence, Theorem 1 part 2(b) holds for (N, R). Moreover, by induction assumptions 1 and 3(b), Theorem 1 parts 1 and 3(b) also hold. Finally, Theorem 1 parts 2(a) and 3(a) follow from induction assumptions 2(a) and 3(a). $\diamondsuit$

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 $\mathcal{N}(Q, \mathcal{F})$.[12]

First, we present the following corollary to the induction 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 induction 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 each 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 induction 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 induction 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 efficient matchings of (N^{-c}, R^{-c}), contradicting the claim that such pairs are underdemanded. We showed that for all $Q \subseteq O(N^{-c}, R^{-c})$, $\quad |\mathcal{N}\big(Q, \mathcal{D}(N^{-c}, R^{-c})\big)| > |Q|$. $\diamondsuit$

Our next claim follows easily from Claim 2:

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

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

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

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

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

[12] 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 $(F, R_F^{-c}) \in \mathcal{F}$.

Claim 4. $|\mathcal{N}(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c}))| = |\hat{Q}|$.

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 $|\mathcal{N}\big(Q'' \cup Q', \mathcal{D}(N^{-c}, R^{-c})\big)| = |Q'' \cup Q'|$. Suppose not. This and Claim 3 together imply

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

Let $\mathcal{F}'' = \mathcal{N}\big(Q'', \mathcal{D}(N^{-c}, R^{-c})\big)$. 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 and $\big|\mathcal{N}\big(Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |Q'|$ 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, $\big|\mathcal{N}\big(Q''\cup Q', \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |Q''\cup Q'|$. $\diamondsuit$

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 all $F \in \mathcal{G}$, and all $Q \subseteq \Big(O(N^{-c}, R^{-c}) \cup \{c\}\Big) \setminus \hat{Q}$,*

$$|\mathcal{N}\big(Q, \mathcal{G}\setminus\{F\}\big)| \geq |Q|.$$

Proof of Claim 5 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

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

$\diamondsuit$

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 prove that $U(N, R) = U(N^{-c}, R^{-c}) \setminus \hat{F}$ (part 1 of the claim), in two steps:

1. 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).

   (a) 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 under 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 also match 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 only with 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,

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

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 under μ.

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

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

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 $|\hat{\mathcal{F}}| = \big|\mathcal{N}\big(\hat{Q}, \mathcal{D}(N^{-c}, R^{-c})\big)\big| = |\hat{Q}|$, one pair in each $D \in \hat{\mathcal{F}}$ is matched under 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,

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

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

(c) 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 $P(N^{-c}, R^{-c})$,

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

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

2. 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 in step 1 does just 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 in step 1 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 under 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 induction 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 induction assumption 2(b), all incompatible pairs in $P(N^{-c},R^{-c})$ can be matched within the set, and by the above construction in step 1(b) all incompatible pairs in $\hat{Q} \cup \hat{F}$ can be matched within this set. Hence, all incompatible pairs in $P(N,R) = P(N^{-c},R^{-c}) \cup \hat{Q} \cup \hat{F}$ can be matched within itself. Moreover, if some pair in $P(N,R)$ is matched with a pair in $O(N,R) = O(N^{-c},R^{-c}) \setminus \hat{Q}$, then the facts that (1) all pairs in $O(N,R)$ can be matched with pairs in $U(N,R) = U(N^{-c},R^{-c}) \setminus \hat{F}$ such that each pair comes from a distinct component of $\mathcal{D}(N,R) = \mathcal{D}(N^{-c},R^{-c}) \setminus \hat{\mathcal{F}}$ (established in step 1(a) above), and (2) each component in $\mathcal{D}(N,R)$ has an odd number of pairs, all of which but (any) one can be matched within the component (from induction assumption 3(a)), imply that less than the maximum possible incompatible pairs are matched in (N,R). Hence, no pair in $P(N,R)$ should be matched with a pair in $O(N,R)$ at an efficient matching of (N,R). Thus, Theorem part 2(b) should hold for (N,R). Then, the previous argument also implies that at most one pair from each component from $\mathcal{D}(N,R)$ should be matched with a pair from $O(N,R)$ and the rest should be matched within the component to maximize the pairs matched, implying Theorem 1 parts 1 and 3(b) hold for (N,R). $\diamondsuit$ $\blacksquare$

We state an immediate corollary to the theorem regarding the relationships between underdemanded pairs and dependent components, and similarly between perfectly matched pairs and self-sufficient components.

Corollary 2. *For any problem (N, R), $U(N, R) = \cup_{K \in \mathcal{D}(N,R)} K$ and $P(N, R) = \cup_{K \in \mathcal{S}(N,R)} K$.*

An immediate corollary to Claims 1 and 6 in the proof of Theorem 1 regards how the set of underdemanded pairs changes when a compatible pair is added to the exchange pool.

Corollary 3. *When a new compatible pair c is added to the exchange pool, the set of underdemanded pairs weakly shrinks. That is, if there is a Pareto-efficient matching that leaves an incompatible pair unmatched when c is added, then there exists a Pareto-efficient matching that leaves the same pair unmatched without c added.*

We conclude the section with four 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}). Moreover, c becomes either itself a self-sufficient component or a member of a self-sufficient component.

- If c is mutually compatible with a pair in $U(N^{-c}, R^{-c})$, the overdemanded set is determined through the 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 set of perfectly matched pairs, 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 be fully 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 $|\mathcal{D}(N^{-c}, R^{-c})| - |O(N^{-c}, R^{-c})|$ 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 $|\mathcal{D}(N^{-c}, R^{-c})| - |O(N^{-c}, R^{-c})| - 1$ incompatible pairs unmatched, one fewer 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 $|\mathcal{D}(N^{-c}, R^{-c})| - |O(N^{-c}, R^{-c})|$ 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 $|\mathcal{D}(N^{-c}, R^{-c})| - |O(N^{-c}, R^{-c})|$ 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 $\nu' = \nu \setminus \{(c, \mu(c))\} \in \mathcal{M}(N^{-c}, R^{-c})$ will match one more incompatible pair than μ'. 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 matching. 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.

Remark 4 In practice, demand decomposition and Theorem 1 can be used to determine which incompatible pairs will not be matched in a priority matching. Also, a priority matching can be constructed using the demand decomposition. By the theorem, at most one incompatible pair from each dependent component remains unmatched in any efficient matching including a priority matching. Thus, all incompatible pairs except the lowest priority pair in each dependent component will be matched. A procedure similar to the one reported in RSÜ [2005a], pages 163-164, can be used to determine which of the lowest priority pairs of dependent components will be matched in a priority 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. We showed 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; thus, unlike other design considerations that provide efficiency gains at the expense of harming various subsets of patients, the inclusion of compatible pairs provides much larger gains without any adverse distributional effects.

Motivated by our analysis, Yılmaz [2011b] considers the impact of inclusion of two-way list exchanges on the system rather than altruistically unbalanced kidney exchanges (c.f. RSÜ [2004]).[13] 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-type decomposition no longer exists for Pareto-efficient matchings in his framework. Moreover, the number of patients who receive live-donor transplants no longer remains the same across Pareto-efficient matchings.[14] 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.

[13]Unfortunately, from a practical point of view, list exchanges are rarely conducted and were never adopted outside of New England. The reservations are due to strong ethical concerns regarding their adverse distributional effects on deceased-donor waiting lists, and the exchange of a live-donor with a deceased donor is viewed to be unethical by many doctors.

[14]As a result, he only characterizes the structure of "p-maximum" matchings, i.e., Pareto-efficient matchings that maximize the number of ordinary kidney exchanges among the matchings that maximize the number of live-donor transplants.

Anna Bogomolnaia and Hervé 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. Kutató Int. Közl.*, 8:373–395, 1963.

Tibor Gallai. Maximale Systeme Unabhangiger kanten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 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.

László Lovász 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. Increasing the opportunity of live kidney donation by matching for two and three way exchanges. *Transplantation*, 81:773–782, 2006.

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

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

Özgür Yılmaz. 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.

A Appendix: Incentives in Priority Mechanism

Let $\mathcal{P}(\succsim_i)$ be the set of allowable manipulations that pair i can make when its true preferences are $\succsim_i$. These can be summarized as declaring some of the feasible exchanges infeasible for the pair.

Formally, for each problem with underlying preference profile $\succsim = (\succsim_x)_{x \in N}$, let $R[\succsim]$ denote the associated feasible exchange matrix. A mechanism is formally defined as a mapping ϕ from the set problems to the set of feasible matchings such that $\phi(N, R) \in \mathcal{M}(N, R)$. A mechanism ϕ is **dominant-strategy incentive compatible** if for all $i \in N$, for all $\succsim$, there is no $\succsim_i' \in \mathcal{P}(\succsim_i)$ such that $\phi(N, R[\succsim_i', \succsim_{-i}])(i) \succ_i \phi(N, R[\succsim_i, \succsim_{-i}])(i)$. It turns out that a priority mechanism is dominant-strategy incentive compatible:[15]

Theorem 2. *Any priority mechanism is dominant-strategy incentive compatible.*

Proof of Theorem 2. For each problem (N, R), consider the construction of a priority matching and sets of matchings $\mathcal{E}^0(N,R), \mathcal{E}^1(N,R)$,...,$\mathcal{E}^n(N,R)$ under the natural ordering $\pi = (1, 2, ..., n)$ where $N_I = \{1, 2, ..., n\}$ is the set of incompatible pairs (we drop π from the subscript whenever possible). We define sets of incompatible pairs $M^0(N,R), M^1(N,R), ..., M^n(N,R)$ as

$$M^0(N,R) = \emptyset \text{ and}$$
$$M^k(R) = \left\{ i \in \{1, 2, ..., k\} : \mu(i) \neq i \text{ for any } \mu \in \mathcal{E}^k(N,R) \right\} \qquad \text{for each } k \in \{1, 2, ..., n\}.$$

Note that $M^{k-1}(N,R) \subseteq M^k(R)$ for any $k \in \{1, 2, ..., n\}$.

Without loss of generality, we will prove the strategy-proofness of a priority mechanism that selects a priority matching under the natural ordering for each problem. Let ϕ be a priority mechanism for the natural ordering π and $R = [r_{y,z}]_{y,z \in N}$ be a problem. Construct sets of matchings $\mathcal{E}^0(N,R), \mathcal{E}^1(N,R)$,...,$\mathcal{E}^n(N,R)$ and sets of incompatible pairs $M^0(N,R), M^1(N,R), ..., M^n(N,R)$.

No compatible pair can manipulate profitably, as all compatible pairs receive transplants under ϕ at R.

Any incompatible pair $j \in M^n(N,R)$ is matched at $\phi(N,R) \in \mathcal{E}^n(N,R)$; hence, it cannot possibly benefit by underreporting the set of pairs that it is compatible with under ϕ. Let $j \in N \backslash M^n(R)$. Incompatible pair j is unmatched at $\phi(N,R)$. If incompatible pair j manipulates its preferences and cannot change the underlying feasible exchange matrix R, then it cannot benefit. So we focus on a manipulation of j that declares a mutually compatible pair to be incompatible. We will prove that incompatible pair j cannot receive a transplant by this manipulation, and repeated application of this argument will conclude the proof.

Let $x \in N \backslash \{j\}$ be such that $r_{j,x} = 1$. Let $R' = \left[r'_{y,z}\right]_{y,z \in N}$ be the feasible exchange matrix obtained from R by incompatible pair j declaring pair x to be incompatible. Note that $\mathcal{M}(R') = \{\mu \in \mathcal{M}(R) : \mu(j) \neq x\}$. Construct sets of matchings $\mathcal{E}^0(R'), \mathcal{E}^1(R')$,...,$\mathcal{E}^n(R')$ and sets of pairs $M^0(R'), M^1(R'), ..., M^n(R')$.

We conclude the proof with a claim that implies $M^n(R') = M^n(R)$. This together with $j \notin M^n(R)$ implies that $j \notin M^n(R')$.

[15] The proof of Theorem 2 is almost the same as that of Theorem 1 of Roth, Sönmez, and Ünver [2005a]. Nevertheless, we include it in here for the sake of completeness.

Claim 7. *For each* $k \in \{0, 1, ..., n\}$,

(i) $M^k(R') = M^k(R)$ *and*

(ii) $\mathcal{E}^k(R') = \left\{\mu \in \mathcal{E}^k(R) : \mu(j) \neq x\right\}$.

Proof of Claim 7 We prove it by induction.

- Let $k = 0$. By construction, $M^0(R') = \emptyset = M^0(R)$. Since $\mathcal{E}^0(R) = \mathcal{M}(R)$, $\mathcal{E}^0(R') = \mathcal{M}(R')$ and $\mathcal{M}(R') = \{\mu \in \mathcal{M}(R) : \mu(j) \neq x\}$, we have

$$\mathcal{E}^0(R') = \mathcal{M}(R') = \{\mu \in \mathcal{M}(R) : \mu(j) \neq x\} = \left\{\mu \in \mathcal{E}^0(R) : \mu(j) \neq x\right\}.$$

- Let $k > 0$. For all ℓ with $0 \leq \ell < k$ assume that $M^\ell(R') = M^\ell(R)$ and $\mathcal{E}^\ell(R') = \left\{\mu \in \mathcal{E}^\ell(R) : \mu(j) \neq x\right\}$.

  We will prove that $M^k(R') = M^k(R)$ and $\mathcal{E}^k(R') = \left\{\mu \in \mathcal{E}^k(R) : \mu(j) \neq x\right\}$. Consider incompatible pair k. We have either $k \in M^k(R)$ or $k \notin M^k(R)$. We consider these two cases separately:

  1. $k \notin M^k(R)$: We have $M^k(R) = M^{k-1}(R)$ and $\mathcal{E}^k(R) = \mathcal{E}^{k-1}(R)$. For all $\eta \in \mathcal{E}^{k-1}(R)$, $\eta(k) = k$. By the induction assumption $\mathcal{E}^{k-1}(R') \subseteq \mathcal{E}^{k-1}(R)$, therefore for all $\eta \in \mathcal{E}^{k-1}(R')$, we have $\eta(k) = k$; and hence, $k \notin M^k(R')$ and $\mathcal{E}^k(R') = \mathcal{E}^{k-1}(R')$. This together with the induction assumption implies

  $$\begin{aligned} M^k(R') &= M^{k-1}(R') = M^{k-1}(R) = M^k(R) \text{ and} \\ \mathcal{E}^k(R') &= \mathcal{E}^{k-1}(R') = \left\{\mu \in \mathcal{E}^{k-1}(R) : \mu(j) \neq j'\right\} = \left\{\mu \in \mathcal{E}^k(R) : \mu(j) \neq x\right\}. \end{aligned}$$

  2. $k \in M^k(R)$: We have

  $$M^k(R) = M^{k-1}(R) \cup \{k\}$$

  and

  $$\mathcal{E}^k(R) = \left\{\mu \in \mathcal{E}^{k-1}(R) : \mu(k) \neq k\right\}.$$

  We prove the two statements separately:

  (i) Let $\eta \in \mathcal{E}^n(R)$. Since $k \in M^k(R) \subseteq M^n(R)$, $\eta(k) \neq k$. Since $j \notin M^n(R)$, $\eta(j) = j$. These together with the induction assumption and $\mathcal{E}^n(R) \subseteq \mathcal{E}^{k-1}(R)$ imply that $\eta \in \left\{\mu \in \mathcal{E}^{k-1}(R) : \mu(j) \neq x\right\} = \mathcal{E}^{k-1}(R')$. This together with $\eta(k) \neq k$ implies that

  $$\mathcal{E}^k(R') = \left\{\mu \in \mathcal{E}^{k-1}(R') : \mu(k) \neq k\right\} \tag{16}$$

and

$$M^k(R') = M^{k-1}(R') \cup \{k\} = M^{k-1}(R) \cup \{k\} = M^k(R).$$

(ii) First let $\eta \in \left\{\mu \in \mathcal{E}^k(R) : \mu(j) \neq x\right\}$. Since $\eta \in \mathcal{E}^k(R) \subseteq \mathcal{E}^{k-1}(R)$ and $\eta(j) \neq x$, we have $\eta \in \left\{\mu \in \mathcal{E}^{k-1}(R) : \mu(j) \neq x\right\} = \mathcal{E}^{k-1}(R')$ where the last equality follows from the induction assumption. Since $k \in M^k(R)$ and $\eta \in \mathcal{E}^k(R)$, we have $\eta(k) \neq k$. These imply that $\eta \in \left\{\mu \in \mathcal{E}^{k-1}(R') : \eta(k) \neq k\right\} = \mathcal{E}^k(R')$ by Equation 16.
Next let $\eta \in \mathcal{E}^k(R') = \left\{\mu \in \mathcal{E}^{k-1}(R') : \mu(k) \neq k\right\}$. Since $\mathcal{E}^k(R') \subseteq \mathcal{E}^{k-1}(R') \subseteq \mathcal{E}^{k-1}(R)$ where the last set inclusion follows from the induction assumption, we have $\eta \in \mathcal{E}^{k-1}(R)$. This together with $\eta(k) \neq k$ implies that $\eta \in \left\{\mu \in \mathcal{E}^{k-1}(R) : \mu(k) \neq k\right\} = \mathcal{E}^k(R)$. Therefore $\mathcal{E}^k(R') = \left\{\mu \in \mathcal{E}^k(R) : \mu(j) \neq x\right\}$, completing the proof of the Claim as well as Theorem 2.

$\diamondsuit$ $\blacksquare$