

# Pairwise Kidney Exchange\*

Alvin E. Roth<sup>†</sup>  
*Harvard University*

Tayfun Sönmez<sup>‡</sup>  
*Harvard University*  
*Koç University*

M. Utku Ünver<sup>§</sup>  
*Koç University*

February 14, 2005

## Abstract

The theoretical literature on exchange of indivisible goods finds natural application in organizing the exchange of live donor kidneys for transplant. However, in kidney exchange, there are constraints on the size of feasible exchanges. Initially, kidney exchanges are likely to be *pairwise* exchanges, between just two patient-donor pairs, as these are logistically simpler than larger exchanges. Furthermore, the experience of many American surgeons suggests to them that preferences over kidneys are approximately 0-1, i.e. that patients and surgeons should be largely indifferent among healthy donors whose kidneys are compatible with the patient. This is because, in the United States, transplants of compatible live kidneys have about equal graft survival probabilities, regardless of the closeness of tissue types between patient and donor. We show that, although the pairwise constraint eliminates some potential exchanges, there is a wide class of constrained-efficient mechanisms that are strategy-proof when patient-donor pairs and surgeons have 0-1 preferences. This class of mechanisms includes deterministic mechanisms that would accommodate the kinds of priority setting that organ banks currently use to allocate cadaver organs, as well as stochastic mechanisms that allow distributive justice issues to be addressed. The results also clarify the combinatorics of pairwise exchange generally.

---

\*We thank Anna Bogomolnaia, Frank Delmonico, Hervé Moulin, Muriel Niederle, Parag Pathak, Susan Saidman, Jen Stack, and two anonymous referees for helpful comments. Sönmez gratefully acknowledges the research support of KoçBank via the KoçBank scholar program and Turkish Academy of Sciences through the Young Scientist Award Program via grant TS/TÜBA-GEBİP/2002-1-19. Roth and Ünver acknowledge support from the NSF. Any errors are our own. An earlier version with the same title appeared as NBER Working Paper 10698, August 2004.

<sup>†</sup>E-mail: [aroth@hbs.edu](mailto:aroth@hbs.edu); address: Harvard University, Department of Economics, Cambridge, MA 02138 and Harvard Business School, Boston, MA 02163.

<sup>‡</sup>E-mail: [tsonmez@hbs.edu](mailto:tsonmez@hbs.edu); address: Harvard Business School, Boston, MA 02163.

<sup>§</sup>E-mail: [uunver@ku.edu.tr](mailto:uunver@ku.edu.tr); address: Koç University, Department of Economics, College of Administrative Sciences and Economics, Rumeli Feneri Yolu, Sarıyer, İstanbul, 34450, Turkey.

# 1 Introduction

As of this writing, there are 60,752 kidney patients in the United States who are registered on a waiting list for a transplant of a cadaver kidney. The median waiting time is from 2 to almost 6 years depending on blood type (for registrations in 1999-2000), and in 2004, 3,971 patients died while on the waiting list, or were removed from it after having become too ill for a transplant. In 2004 there were 8,577 transplants of cadaver kidneys.<sup>1</sup>

Because healthy people have two kidneys (and can remain healthy on only one), it is also possible for a kidney patient to receive a live-donor transplant. There were 6,086 live-donor transplants in 2004. However, a willing, healthy donor is not always able to donate to his intended patient, because of blood type or immunological incompatibilities between them. In this case, most often, the donor is sent home, and becomes once again invisible to the health care system.

However, in a few cases, an *exchange* has been arranged between one incompatible patient-donor pair and another. In such an exchange, the donor from each pair gives a kidney to the patient from the other pair. From 2001 through 2004, there were 5 such paired exchanges in the fourteen transplant centers that make up the New England region, and, in the United States, there have even been 2 exchanges among three incompatible patient-donor pairs.<sup>2</sup> These exchanges do not violate the 1984 National Organ Transplant Act (NOTA), which prohibits the sale or purchase of human organs.<sup>3</sup>

While there is a national database of tissue types of kidney *patients*, used for allocating cadaver kidneys, there is as of this writing no national (and few regional) databases of incompatible patient-donor pairs, despite earlier proposals to set up such databases (Rapaport (1986), Ross et al (1997)).<sup>4</sup> Nor is there yet a systematic method used for arranging exchanges between incompatible pairs. However there are efforts to change this at a number of medical centers, and in September 2004 the Renal Transplant Oversight Committee of New England approved the establishment of a clearinghouse for kidney exchange, proposed by Drs. Francis Delmonico, Susan Saidman, and the three authors of this paper (cf. Roth, Sönmez, and Ünver, 2005).

Our initial work on kidney exchange, in Roth, Sönmez, and Ünver (2004), showed how to identify efficient exchanges in a way that gave patients and their surgeons dominant strategy incentives to

---

<sup>1</sup>United Network for Organ Sharing (UNOS) - The Organ Procurement and Transplantation Network (OPTN) national data, retrieved on 2/10/2005 from <http://www.optn.org/data>.

<sup>2</sup>Both of these have been arranged at the Johns Hopkins Comprehensive Transplant Center in Baltimore. Lucan et al (2003) also reports on three pair and four pair exchanges conducted in Romania.

<sup>3</sup>See the legal opinion to this effect obtained by the UNOS at [http://asts.org/ezefiles/UNOSSection\\_301\\_NOTA\\_.pdf](http://asts.org/ezefiles/UNOSSection_301_NOTA_.pdf).

<sup>4</sup>Some hospitals have started to generate their own databases of incompatible patient-donor pairs, that could be used for exchanges. We are aware of such databases in Alabama, Baltimore (Johns Hopkins) and Ohio (Medical College of Ohio).

straightforwardly reveal their preferences. And, using tissue typing statistics from the Caucasian patient population, we showed that the benefits of such an exchange could be very substantial, increasing live organ donations between unrelated donors from about 54% to as much as 91% if multiple-pair exchanges are feasible, and to as much as 75% even if only pairwise exchanges are feasible.<sup>5</sup>

However in our subsequent discussions with medical colleagues, aimed at organizing such exchanges in the New England region of the transplant system, it became clear that a likely first step will be to implement *pairwise* exchanges, between just two patient-donor pairs, as these are logistically simpler than exchanges involving more than two pairs. That is because all transplantations in an exchange need to be carried out simultaneously, for incentive reasons, since otherwise a donor may withdraw her consent after her intended recipient receives a transplanted kidney.<sup>6</sup> So even a pairwise exchange involves *four* simultaneous surgical teams, operating rooms, etc. Furthermore, the experience of American surgeons suggests to them that preferences over kidneys can be well approximated as 0-1, i.e. that patients and surgeons should be more or less indifferent among kidneys from healthy donors that are blood type and immunologically compatible with the patient.<sup>7</sup> This is because, in the United States, transplants of compatible live kidneys have about equal graft survival probabilities, regardless of the closeness of tissue types between patient and donor (Gjertson and Cecka (2000) and Delmonico (2004)).<sup>8</sup>

The present paper explores how to organize such exchanges. The theoretical groundwork for kidney exchange without a constraint on the size of exchanges, explored in Roth, Sönmez, and Ünver (2004), was laid in the papers by Shapley and Scarf (1974), Roth and Postlewaite (1977), Roth (1982a), and Abdulkadiroğlu and Sönmez (1999). The constrained exchange problem studied in the present paper is closely related to elegant results from graph theory, which will prove very useful, in ways recently pioneered by Bogomolnaia and Moulin (2004). They looked at a problem of pairwise matching with 0-1 preferences on *two-sided* graphs, i.e. on graphs in which the parties to the exchange

---

<sup>5</sup>Subsequent investigation of a database constructed by Dr. Susan Saidman of Massachusetts General Hospital, of patients who had an incompatible donor (and who were consequently on the waiting list for a cadaver kidney), showed that 18% of them could participate in live donor exchanges involving only paired exchanges among patient-donor pairs in the database, and 27% could receive transplants if larger exchanges among them were feasible.

<sup>6</sup>For legal reasons related to the NOTA, it seems unlikely that legally binding contracts can be written about the future provision of a live donor kidney.

<sup>7</sup>Bogomolnaia and Moulin (2004), on whose work the present paper builds, refer to such preferences (in settings quite different than kidney exchange) as *dichotomous*. That the surgeons with whom we are working on implementing kidney exchange have approximately dichotomous preferences was brought to our attention when we began to work with Dr. Saidman's database of current patients with incompatible donors, and as we began developing software to identify potential matches for the New England program.

<sup>8</sup>This is contrary to the "European" view which maintains that the graft survival rate increases as the tissue type mismatch decreases. See Opelz (1997, 1998).

can be a-priori divided into two sets, e.g. into buyers and sellers, each of which can only trade with the other. Bipartite graphs cannot be used to model kidney exchange, since any patient-donor pair might potentially exchange with any other. The present paper therefore generalizes such a model to the case of arbitrary graphs.

While the constraint that only pairwise exchanges be conducted means that the number of live donor transplants that can be arranged by exchange is smaller than if larger exchanges are feasible, it is still substantial. And, in the constrained problem, efficient and strategy-proof mechanisms will be shown to exist. This class of mechanisms includes deterministic mechanisms that would accommodate the kinds of priority setting that organ banks currently use for the allocation of cadaver organs, and which therefore may be especially appealing to transplant organizations. Also included are random matching mechanisms, such as the egalitarian mechanism, a stochastic mechanism that arises in connection with elementary notions of distributive justice.

## 1.1 Related Literature

This paper intersects with the literatures on transplantation, graph theory, and mechanism design. The idea of paired kidney exchange between one incompatible patient-donor pair and another was first proposed by Rapaport (1986) and then again by Ross et al (1997). UNOS initiated pilot testing of a paired kidney exchange program in 2000, and the same year the transplantation community issued a consensus statement indicating that the paired kidney exchange program is considered to be “ethically acceptable” (Abecassis et al (2000)).<sup>9</sup>

While the transplantation community approved the use of kidney exchanges to increase live kidney donation, it has provided little guidance about how to organize such exchanges. Roth, Sönmez, and Ünver (2004) proposed an efficient and strategy-proof mechanism that uses both pairwise and larger exchanges. The present paper differs from Roth, Sönmez, and Ünver (2004) in two major ways: We only consider exchanges involving two patients and their donors, and we adopt the assumption of many American transplant surgeons (including those we are working with in New England) that each patient is indifferent between all compatible kidneys (cf. Gjertson and Cecka (2000) and Delmonico (2004)). These two assumptions considerably change the mathematical structure of the kidney exchange problem, and efficient exchange becomes an application of what is known in the graph theory

---

<sup>9</sup>Another kind of exchange proposed in the transplantation literature is an *indirect exchange* between an incompatible patient-donor pair and a patient on the cadaveric waitlist (in return for a high priority on the waitlist for the donor’s patient; Ross and Woodle (2000)). Zenios (2002) focuses on the mix of direct and indirect exchanges that maximizes the welfare of the candidates in the participating pairs, whereas Zenios, Woodle, and Ross (2001) propose preferential selection of O blood-type donors of patients with multiple donors to reduce the adverse affect of indirect exchange programs on patients with no donors. Roth, Sönmez, and Ünver (2004) explore how more elaborate forms of direct and indirect exchange can increase the potential benefits of kidney exchange.

literature as the **cardinality matching problem** (see for example Korte and Vygen (2002)).<sup>10</sup> For this purpose consider an undirected graph whose vertices each represent a particular patient and her incompatible donor(s), and whose edges connect those pairs of patients between whom an exchange is possible, i.e. pairs of patients such that each patient in the pair is compatible with a donor of the other patient. Finding an efficient matching then reduces to finding a maximum cardinality matching in this undirected graph (see Lemma 1), a problem well analyzed in this literature. More specifically the Gallai (1963, 1964)-Edmonds (1965) Decomposition Lemma (henceforth GED Lemma) characterizes the set of maximum cardinality matchings. Technical aspects of our contribution heavily build on the GED Lemma.

We first concentrate on deterministic outcomes, and show that there exists a wide class of efficient and strategy-proof mechanisms that accommodate the kinds of priority setting that organ banks currently use for the allocation of cadaver organs. We then allow stochastic outcomes as well, and show that there exists an efficient and strategy-proof mechanism, the **egalitarian mechanism**, which equalizes as much as possible the individual probabilities of receiving a transplant. If stochastic mechanisms can be accepted by the transplantation community, this mechanism can serve as a basis for discussion of how to address equity issues while achieving efficiency and strategy-proofness.

Our paper builds on the closely related recent paper by Bogomolnaia and Moulin (2004). They considered *two-sided* matching i.e. matching between two sets of agents that can be specified exogenously (e.g. firms and workers), such that an agent on one side of the market can only be matched with an agent on the other side (cf. Roth and Sotomayor (1990)), modeled as a bipartite graph, with 0-1 preferences. It was their paper that made us aware of some of the graph-theoretic results that we also use here. Our results on the egalitarian mechanism generalize their corresponding results to general, not necessarily bipartite graphs. Kidney exchange cannot be modelled as a two-sided market, since any patient with incompatible donors can potentially be matched with any other.<sup>11</sup> The extensions to the general case are of interest not only because of the importance of the application to kidney exchange, but also because of the insights they give into pairwise exchange in general graphs, and the technical challenges that the generalization to arbitrary graphs presents.

---

<sup>10</sup>If we instead only consider exchanges involving two pairs (as in this paper) but assume strict preferences over compatible kidneys (as in Roth, Sönmez, and Ünver (2004)), the problem becomes an application of what is known as the **roommates problem** (Gale and Shapley (1962)). See Abeledo and Rothblum (1994), Chung (2000), Diamantoudi, Miyagawa, and Xue (2004,) and Teo and Sethuraman (2000).

<sup>11</sup>Note that, since each donor comes to the exchange in the company of his incompatible patient, there isn't an option of modeling a two-sided market in which the sides are donors and patients. There is a very small population of *undirected* donors, who wish to donate but not to a specific patient, and such a formulation might be applicable to them, but we do not consider such unattached donors here.

## 2 Pairwise Kidney Exchange

Let  $N = \{1, 2, \dots, n\}$  be a set of patients each of whom has one or more incompatible donors. Each patient is indifferent between all compatible donors and between all incompatible donors, except she strictly prefers her donor(s) to any other incompatible donor, and any compatible donor to her own donor(s). Since we are considering only pairwise exchanges in this paper, the above assumptions induce the following preference relation  $\succsim_i$  for patient  $i$  over the set of patients  $N$ :

1. For any patient  $j$  with a compatible donor for patient  $i$  we have  $j \succ_i i$ ,
2. for any patient  $j$  without any compatible donor for patient  $i$  we have  $i \succ_i j$ ,
3. for any patients  $j, h$  each of whom has a compatible donor for patient  $i$  we have  $j \sim_i h$ ,
4. for any patients  $j, h$  neither of whom has a compatible donor for patient  $i$  we have  $j \sim_i h$ .

Here  $\succ_i$  denotes the strict preference relation and  $\sim_i$  denotes the indifference relation induced by  $\succsim_i$ . A **(pairwise kidney exchange) problem** is a pair  $(N, \succsim)$  where  $\succsim = (\succsim_i)_{i \in N}$  denotes the list of patient preferences. Throughout the paper with the exception of the proofs of the results on incentives, we fix a problem  $(N, \succsim)$ .

We consider the case in which an exchange can involve only two pairs. Patients  $i, j \in N$  are **mutually compatible** if  $i \succ_j j$  and  $j \succ_i i$ . That is, two patients are mutually compatible if each one has a donor whose kidney is compatible for the other patient.

A **matching**  $\mu : N \rightarrow N$  is a function such that:  $\mu(i) = j$  if and only if  $\mu(j) = i$  for any pair of patients  $i, j \in N$ . A matching  $\mu$  is **individually rational** if for any patient  $i \in N$ ,  $\mu(i) \neq i$  implies  $\mu(i) \succ_i i$ . Let  $\mathcal{M}$  be the set of individually rational matchings for the problem  $(N, \succsim)$ . Throughout the paper we consider only individually rational matchings. That is, exchange is possible only between mutually compatible patients. A *matching* denotes an individually rational matching throughout the rest of the paper.

For each matching  $\mu \in \mathcal{M}$  and patient  $i \in N$ ,  $\mu(i) = i$  means that the patient  $i$  remains unmatched.<sup>12</sup> For any matching  $\mu \in \mathcal{M}$  and pair of patients  $i, j \in N$ ,  $\mu(i) = j$  means that patient  $i$  receives a compatible kidney from a donor of patient  $j$  and patient  $j$  receives a compatible kidney from a donor of patient  $i$ . Since exchange is possible only among mutually compatible pairs, it is

---

<sup>12</sup>A patient who is unmatched does not receive a live-donor transplant, nor does her donor donate a kidney. Such a patient may wait for a cadaver kidney, or the patient and incompatible donor may participate in an exchange arranged at a later date when other incompatible patient-donor pairs have become available. Note that when a patient is matched, only one of her donors donates a kidney (no matter how many incompatible donors the patient has).

sufficient for our purposes to keep track of the symmetric  $|N|$  by  $|N|$  **mutual compatibility matrix**  $R = [r_{i,j}]_{i \in N, j \in N}$  defined by

$$r_{i,j} = \begin{cases} 1 & \text{if } j \succ_i i \text{ and } i \succ_j j \\ 0 & \text{otherwise} \end{cases}$$

for any pair of (not necessarily distinct) patients  $i, j \in N$ .

We will refer to the pair  $(N, R)$  as the **reduced problem** of  $(N, \succ)$ . Occasionally it will be helpful to think of the reduced problem as a graph  $G = (N, R)$  whose vertices  $N$  are the patients (and their incompatible donors), and whose edges  $R$  are the connections between mutually compatible pairs of patients; i.e. there is an edge  $(i, j) \in R$  if and only if  $r_{i,j} = 1$ . (It will be clear from the context whether  $R$  is the mutual compatibility matrix or the set of edges indicating mutual compatibility.) A matching then can be thought of as a subset of the set of edges such that each patient can appear in at most one of the edges. With this alternative representation if  $(i, j)$  is an edge in the matching  $\mu$ , patients  $i$  and  $j$  are matched by  $\mu$  and, if patient  $i$  does not appear in any edge in the matching  $\mu$ , she remains unmatched.

A **mechanism** is a systematic procedure that selects a matching for each problem.

### 3 Efficient Exchange

A matching  $\mu \in \mathcal{M}$  is **Pareto-efficient** if there exists no other matching  $\eta \in \mathcal{M}$  such that  $\eta(i) \succ_i \mu(i)$  for all  $i \in N$  and  $\eta(i) \succ_i \mu(i)$  for some  $i \in N$ . In the present setting,  $\mu$  is Pareto-efficient if and only if the set  $M_\mu = \{i \in N : \mu(i) \neq i\}$  of patients matched by  $\mu$  is maximal, i.e. if there does not exist any other matching  $\eta \in \mathcal{M}$  such that  $M_\eta \supset M_\mu$ . Let  $\mathcal{E}$  be the set of Pareto-efficient matchings for the problem  $(N, \succ)$ . A well known result from abstract algebra will help clarify the structure of the set of Pareto-efficient matchings (see e.g. Lovász and Plummer, 1986 on matchings, and Korte and Vygen, 2002 on matroids).

A **matroid** is a pair  $(X, \mathcal{I})$  such that  $X$  is a set and  $\mathcal{I}$  is a collection of subsets of  $X$  (called the **independent sets**) such that

**M1** if  $I$  is in  $\mathcal{I}$  and  $J \subset I$  then  $J$  is in  $\mathcal{I}$ ; and

**M2** if  $I$  and  $J$  are in  $\mathcal{I}$  and  $|I| > |J|$  then there exists an  $i \in I \setminus J$  such that  $J \cup \{i\}$  is in  $\mathcal{I}$ .

**Proposition 1** *Let  $\mathcal{I}$  be the sets of simultaneously matchable patients, i.e.  $\mathcal{I} = \{I \subseteq N : \exists \mu \in \mathcal{M} \text{ such that } I \subseteq M_\mu\}$ . Then  $(N, \mathcal{I})$  is a matroid.*

For any matching  $\mu \in \mathcal{M}$ , let  $|\mu| = |M_\mu| = |\{i \in N : \mu(i) \neq i\}|$  denote the number of patients who are matched with another patient. The following well known property of matchings, which

follows immediately from the second property of matroids, states that the same number of patients will receive a transplant at every Pareto-efficient matching.

**Lemma 1** *For any pair of Pareto-efficient matchings  $\mu, \eta \in \mathcal{E}$ ,  $|\mu| = |\eta|$ .*

If exchange is possible among more than two pairs, the conclusion of Lemma 1 no longer holds.

**Example 1** Let  $N = \{1, 2, 3, 4\}$  and suppose preferences are such that

$$\begin{array}{ll} 2 \succ_1 4 \succ_1 1 \succ_1 3 & 1 \succ_3 3 \succ_3 2 \succ_3 4 \\ 3 \succ_2 2 \succ_2 1 \succ_2 4 & 1 \succ_4 4 \succ_4 2 \succ_4 3 \end{array}$$

Consider the following two Pareto efficient trades:

- Patient 1 receives a kidney from a donor of Patient 4 and Patient 4 receives a kidney from a donor of Patient 1. (This is the only possible pairwise trade, since only Patients 1 and 4 are mutually compatible.)
- Patient 1 receives a kidney from a donor of Patient 2, Patient 2 receives a kidney from a donor of Patient 3, and Patient 3 receives a kidney from a donor of Patient 1.

Two patients receive transplants if the first trade is carried out whereas three patients receive transplants if the second trade is carried out. ■

### 3.1 Priority Mechanisms

The experience of transplant centers is mostly with the priority allocation systems used to allocate cadaver organs. It is therefore natural to consider how priority mechanisms would function in the context of live kidney exchange.

A **priority ordering** is a permutation of patients such that the  $k^{\text{th}}$  patient in the permutation is the patient with the  $k^{\text{th}}$  priority. Without loss of generality let the priority ordering of patients be the natural ordering  $(1, 2, \dots, n)$ , i.e. patient  $k$  is the  $k^{\text{th}}$  priority patient for each  $k$ .

While we will concentrate on ordinal priorities here, priorities may depend on quantifiable patient characteristics such as the patient’s “percent reactive antibody” (PRA), which is correlated with how difficult it will be to find a compatible kidney for that patient. (So it might be desirable, for example, for a high PRA patient to have a high priority for a compatible kidney in the relatively rare event that one becomes available.) In general, we will say that a non-negative function  $\pi : N \rightarrow \mathbb{R}_+$  is a **priority function** if it is increasing in priority, i.e. if  $\pi(i) \geq \pi(i + 1)$ .

Consider a transplant center  $T$  whose decision makers wish to find the set of exchanges that maximizes a preference  $\succ_T$  defined over matchings (more specifically, over sets of matched patients).



We will say that  $\succ_T$  is a **priority preference** if it is *responsive* to the priority ordering (Roth, 1985), i.e. if  $\mu \succ_T \nu$  whenever  $M_\mu \supset M_\nu$ , or when  $M_\mu$  and  $M_\nu$  differ in only one patient, i.e.  $M_\mu \setminus M_\nu = \{i\}$ ,  $M_\nu \setminus M_\mu = \{j\}$ , for some  $i, j \in N$ , and  $i < j$ . That is, whenever  $M_\mu$  and  $M_\nu$  differ in only one patient, the matching with the higher priority patient is preferred, and adding additional matched patients to an existing matching always results in a preferred matching. (For a given priority ordering of *patients*, there remain many possible priority preferences  $\succ_T$  over *sets* of matched patients.)

A priority mechanism produces a matching as follows, for any problem  $(N, R)$  and priority ordering  $(1, 2, \dots, n)$  among the patients:

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

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

For a given problem  $(N, R)$  and priority ordering  $(1, 2, \dots, n)$ , we refer to each matching in  $\mathcal{E}^n$  as a **priority matching**, and a **priority mechanism** is a function which selects a priority matching for each problem. A priority matching matches as many patients as possible starting with the patient with the highest priority and following the priority ordering, never “sacrificing” a higher priority patient because of a lower priority patient.

By construction, a priority matching is maximal, and hence Pareto-efficient, i.e.  $\mathcal{E}^n \subset \mathcal{E}$ . Proposition 1 implies, through the second property of matroids, that the “opportunity cost” of matching a higher priority patient will never be more than one lower priority patient who could otherwise have been matched. (For example, there might be two patients each of whom is mutually compatible only with the same third patient, and so matching the higher priority of the two patients will preclude matching the lower priority patient. But it cannot happen that, by matching a higher priority patient, *two* lower priority patients are excluded who otherwise could both have been matched.) And of course, by Lemma 1, the same total number of patients will be matched at each Pareto-efficient matching, so there is no trade-off between priority allocation and the number of transplants that can be arranged. In the matroid literature, a priority mechanism is called a *greedy algorithm* (since it “greedily” takes the highest priority remaining patient at each stage). The following proposition by Rado (1957) and Edmonds (1971) from the matroid literature will be helpful in understanding the resulting priority matchings.

**Proposition 2** *For a matroid  $(N, \mathcal{I})$  and any priority function  $\pi$  on  $N$ , a priority matching  $\mu$  (obtained by a greedy algorithm with respect to  $\pi$ ) identifies an element  $M_\mu$  of  $\mathcal{I}$  that maximizes  $\sum_{i \in I} \pi(i)$  over all  $I \in \mathcal{I}$ .*

Example 1 shows that if larger exchanges were permitted, the conclusions of Proposition 2 would not carry over (e.g. suppose that patient 4 has the highest priority, so the priority mechanism chooses the pairwise trade even if the 3-way trade has a higher sum of priorities).

For ordinal priorities, Proposition 2 allows us to quickly prove the following corollary, which helps explain the appeal that priority algorithms may have to transplant centers accustomed to prioritizing their patients.

**Corollary 1** *For any priority preference  $\succ_T$  a priority matching  $\mu$  maximizes  $\succ_T$  on the set of all matchings, i.e.  $\mu \succeq_T \nu$  for all  $\nu \in \mathcal{M}$ .*

### 3.1.1 Incentives in priority mechanisms

We turn next to consider the incentives facing patients (and their surgeon advocates) in a priority matching mechanism. Two apparently different issues arise that turn out to be closely related. The first has to do with patients who have multiple incompatible donors willing to donate on their behalf. We show that a patient maximizes her chance of being included in an exchange by revealing all of her willing donors. The second issue involves revealing which compatible kidneys the patient is willing to accept. Again, we show that a patient maximizes her chance of being able to take part in an exchange by accepting her full set of compatible kidneys. That is, we show that with respect to both donors and kidneys, priority mechanisms do not give participants perverse incentives, but rather make it a dominant strategy to fully reveal which willing donors are available, and which kidneys are acceptable.

These two conclusions have the same cause. A patient enlarges the set of other patients with whom she is mutually compatible by coming to the exchange with more donors, and by being able to accept a kidney from more of those other patients' donors. And a patient's probability of being included in an exchange is monotonic in the set of other patients with whom she is compatible.

For a given set of patients and their available donors, the basic data for the problem  $(N, R)$  is determined by the tissue typing laboratory. So, once each patient has revealed a set of donors, the tissue typing lab establishes for each patient  $i$  a set of compatible kidneys  $K_i = \{j \in N : r_{ij} = 1\}$ . But a kidney exchange is a complicated event involving a patient and a donor on each side of the exchange, and so there will be no way to prevent a patient from declining a medically compatible kidney (e.g. for logistical reasons such as location of the other patient-donor pair, or, for that matter, for unspecified reasons related e.g. to the preferences of the patient's donor). So the strategy set of each agent (i.e. each patient and donor, or each surgeon acting on behalf of a patient) is the set of all possible subsets of *acceptable* kidneys  $A_i \subseteq K_i$  that she might declare. (A kidney that is declared not acceptable can be thought of as being incompatible for reasons not initially revealed by the medical data.) A mechanism can only arrange exchanges between patient-donor pairs who

are willing to accept each other’s donor kidneys. A mechanism is strategy-proof in this dimension if a patient’s probability of being included in an exchange with a compatible donor is maximized by declaring truthfully that  $A_i = K_i$ .<sup>13</sup> That is, a patient would have an incentive to declare a smaller acceptable set  $A_i \subset K_i$  only if this could cause the mechanism to include the patient in an exchange when truthful revelation of preferences ( $A_i = K_i$ ) would not. The first part of Theorem 1 states that a patient can never benefit by declaring a compatible kidney to be unacceptable under a priority mechanism.<sup>14</sup> The second part states the similar result for revelation of available donors.

**Theorem 1** *A priority mechanism makes it a dominant strategy for a patient to reveal both a) her full set of acceptable kidneys; and b) her full set of available donors.*

The proof of Theorem 1 a) is contained in the Appendix. The proof of part b) follows quickly from the proof of part a). In particular, a mechanism is **donor-monotonic** if a patient never suffers from the addition of one additional (incompatible) donor for her. The addition of one extra donor for a patient has the effect of enlarging her set of mutually compatible patients. But the proof of part a) of Theorem 1 shows that a patient never suffers from enlarging her set of mutually compatible patients in a priority mechanism, which therefore implies that priority mechanisms are donor monotonic in pairwise kidney exchange.

**Corollary 2** *A priority mechanism is donor-monotonic.*

It is worth emphasizing some limits on what these results establish. When multiple potential donors prove to be incompatible, there may be complex family preferences involved in going forward with an exchange, e.g. a family might prefer to have a sibling donate a kidney to its patient only after the possibility of a parent’s donation had been exhausted, and such a family might prefer to initially look for an exchange by only revealing the availability of the parent as a donor. Theorem 1 b) does not rule out this possibility. Similarly, it might be that, even if a patient is willing to accept all compatible kidneys, her donor is for some reason unwilling to donate a kidney to some of the

---

<sup>13</sup>For simplicity here we assume that all compatible kidneys are in fact acceptable, i.e. that patients do not have logistical or other concerns not already reflected in the mutual compatibility matrix  $R$ . More generally, we are considering a system in which patients may freely indicate in advance that there are some kidneys that they will not accept, but cannot back out of an exchange for a kidney that they have already indicated is acceptable without paying an unacceptably high cost (e.g. because this would mean that they could not participate in the current exchange and might receive very low priority for future exchanges). Consequently we do not consider any strategies at which a patient declares unacceptable kidneys to be acceptable, but only strategies in which patients may choose which set of acceptable kidneys to reveal.

<sup>14</sup>See Roth (1982a, 1982b), Svensson (1994, 1999), Abdulkadiroğlu and Sönmez (1999, 2003b), Bogomolnaia, Ehlers and Deb (2000), Papai (2000), Ehlers (2002), Ehlers, Klaus, and Papai (2002), and Ehlers and Klaus (2003), for strategy-proofness in various related models with indivisibilities.

patients who are compatible with him (i.e. with his kidney), which would reduce the set of acceptable kidneys in a way not precluded by Theorem 1 a).<sup>15</sup> What Theorem 1 *does* establish is that a priority mechanism introduces no new reasons for available donors to be concealed, or compatible kidneys to be rejected.

### 3.2 Gallai-Edmonds Decomposition

The structure of Pareto-efficient pairwise matchings has been an active area of research in combinatorial optimization. Understanding this structure will yield additional insight into priority mechanisms and also allow us to apply a central concept in distributive justice - egalitarianism - to pairwise kidney exchange. The following partition of the set of patients is key to the structure of the set of Pareto-efficient matchings. Partition  $N$  as  $\{N^U, N^O, N^P\}$  such that

$$\begin{aligned} N^U &= \{i \in N : \exists \mu \in \mathcal{E} \text{ s.t. } \mu(i) = i\}, \\ N^O &= \{i \in N \setminus N^U : \exists j \in N^U \text{ s.t. } r_{i,j} = 1\}, \text{ and} \\ N^P &= N \setminus (N^U \cup N^O). \end{aligned}$$

$N^U$  is the set of patients for each of whom there is at least one Pareto-efficient matching which leaves her unmatched.  $N^O$  is the set of patients each of whom is not in  $N^U$  (i.e., each of whom is matched with another patient at each Pareto-efficient matching) but is mutually compatible with at least one patient in  $N^U$ .  $N^P$  is the set of remaining patients (i.e., the set of patients who are matched with another patient at each Pareto-efficient matching and who are not mutually compatible with any patient in  $N^U$ ).

Consider the reduced problem  $(N, R)$ . For  $I \subset N$ , let  $R_I = [r_{i,j}]_{i \in I, j \in I}$ . We refer to the pair  $(I, R_I)$  as the **reduced subproblem** restricted to  $I$ . A reduced subproblem  $(I, R_I)$  is **connected** if there exists a sequence of patients  $i_1, i_2, \dots, i_m$  (possibly with repetition of patients) such that  $r_{i_k, i_{k+1}} = 1$  for all  $k \in \{1, 2, \dots, m-1\}$  and  $I = \{i_1, i_2, \dots, i_m\}$ . A connected reduced subproblem  $(I, R_I)$  is a **component** of  $(N, R)$  if  $r_{i,j} = 0$  for any  $i \in I$  and  $j \in N \setminus I$ . We refer to a component  $(I, R_I)$  as an **odd component** if  $|I|$  is odd and as an **even component** if  $|I|$  is even. The following result due to Gallai (1963, 1964) and Edmonds (1965) is central to our paper:

**Lemma 2 (Gallai-Edmonds Decomposition Lemma)** *Let  $(I, R_I)$  be the reduced subproblem with  $I = N \setminus N^O$  and let  $\mu$  be a Pareto-efficient matching for the original problem  $(N, R)$ .*

1. For any patient  $i \in N^O$ ,  $\mu(i) \in N^U$ .

---

<sup>15</sup>Donations of cadaver organs are often carried out under a screen of anonymity. This will be more difficult in the case of live donor exchange, since the full medical history of each donor will be an essential part of such exchange.

2. For any even component  $(J, R_J)$  of  $(I, R_I)$ ,  $J \subseteq N^P$  and for any patient  $i \in J$ ,  $\mu(i) \in J \setminus \{i\}$ .
3. For any odd component  $(J, R_J)$  of  $(I, R_I)$ ,  $J \subseteq N^U$  and for any patient  $i \in J$  it is possible to match all remaining patients in  $J$  with each other (so that any patient  $j \in J \setminus \{i\}$  can be matched with a patient in  $J \setminus \{i, j\}$ ). Moreover for any odd component  $(J, R_J)$ , either
  - (a) one and only one patient  $i \in J$  is matched with a patient in  $N^O$  under the Pareto-efficient matching  $\mu$  whereas all remaining patients in  $J$  are matched with each other so that  $\mu(j) \in J \setminus \{i, j\}$  for any patient  $j \in J \setminus \{i\}$ , or
  - (b) one patient  $i \in J$  remains unmatched under the Pareto-efficient matching  $\mu$  whereas all remaining patients in  $J$  are matched with each other so that  $\mu(j) \in J \setminus \{i, j\}$  for any patient  $j \in J \setminus \{i\}$ .

Based on the Gallai-Edmonds Decomposition Lemma (GED Lemma), we refer to  $N^U$  as the set of **underdemanded patients**,  $N^O$  as the set of **overdemanded patients** and  $N^P$  as the set of **perfectly matched patients**.<sup>16</sup>

A Pareto-efficient matching matches each perfectly matched patient with another perfectly matched patient in the same even component; each overdemanded patient is matched with an underdemanded patient; and one patient in each odd component is either matched with an overdemanded patient or remains unmatched, whereas the remaining underdemanded patients in the same odd component are matched with one another. So each even component is self sufficient whereas the odd components compete for the overdemanded patients.

Let  $\mathcal{D} = \{D_1, D_2, \dots, D_p\}$  be the partition of the set of underdemanded patients  $N^U$  such that  $(D_k, R_{D_k})$  is an odd component of  $(N \setminus N^O, R_{N \setminus N^O})$  for all  $k \in \{1, 2, \dots, p\}$ . We slightly abuse the notation and refer each  $D \in \mathcal{D}$  also as an **odd component**.

The following is an immediate implication of the GED Lemma:

**Corollary 3**  $|\mathcal{D}| > |N^O|$  whenever  $N^U$  is nonempty, and  $|\mu| = |N| - (|\mathcal{D}| - |N^O|)$  for each Pareto-efficient matching  $\mu$ .

That is, the efficient matchings each leave unmatched  $|\mathcal{D}| - |N^O|$  patients, each one in a distinct odd component. Note that Lemma 1 is an immediate corollary to the GED Lemma as well.

---

<sup>16</sup>Edmonds (1965) introduced the first polynomial-time algorithm for construction of a Pareto-efficient matching and construction of partition  $\{N^U, N^O, N^P\}$  of  $N$ . Faster algorithms were introduced in later dates. An excellent survey of combinatorial matching theory including matching algorithms are given in Lovász and Plummer (1986).

### 3.3 The Induced Two-Sided Matching Market

Loosely speaking there is competition at two levels: At the first level, odd components (of underdemanded patients) compete for overdemanded patients. With the addition of an overdemanded patient to an odd component, all the patients in the augmented odd component can be matched to one another. The second level of competition is between the members of odd components that do not secure an overdemanded patient. If the odd component is of size one, its member will remain unmatched, but members of each larger odd component compete against each other not to be the one patient in the component who remains unmatched.

In this subsection we focus on the first level of competition, between odd components for overdemanded patients. Since this competition does not involve perfectly matched patients (who are all matched to one another at every efficient outcome), we will not need to consider them for this discussion. Instead, we focus on an “induced” two-sided market, one side consisting of the overdemanded patients, and the other side consisting of the (odd) *components* of underdemanded patients.<sup>17</sup>

For each odd component  $J \in \mathcal{D}$  and overdemanded patient  $i \in N^O$ , let

$$\tilde{r}_{i,J} = \begin{cases} 1 & \text{if } \exists j \in J \text{ s.t. } r_{i,j} = 1 \\ 0 & \text{otherwise} \end{cases}$$

and let  $\tilde{R} = [\tilde{r}_{i,J}]_{i \in N^O, J \in \mathcal{D}}$ . Whenever  $\tilde{r}_{i,J} = 1$  for  $i \in N^O$  and  $J \in \mathcal{D}$ , we say there is a **link** between patient  $i$  and set  $J$ .

Given the problem  $(N, \succ)$ , we refer to the triple  $(N^O, \mathcal{D}, \tilde{R})$  as the **induced two-sided matching market**.

A **pre-matching** is a function  $\tilde{\mu} : N^O \cup \mathcal{D} \rightarrow N^O \cup \mathcal{D} \cup \{\emptyset\}$  such that

1.  $\tilde{\mu}(i) \in \mathcal{D} \cup \{\emptyset\}$  for any  $i \in N^O$ ,
2.  $\tilde{\mu}(J) \in N^O \cup \{\emptyset\}$  for any  $J \in \mathcal{D}$ ,
3.  $\tilde{\mu}(i) = J \Leftrightarrow \tilde{\mu}(J) = i$  for any pair  $i \in N^O, J \in \mathcal{D}$ , and
4.  $\tilde{\mu}(i) = J \Rightarrow \tilde{r}_{i,J} = 1$  for any pair  $i \in N^O, J \in \mathcal{D}$ .

Let  $\tilde{\mathcal{M}}$  denote the set of pre-matchings. A pre-matching  $\tilde{\mu} \in \tilde{\mathcal{M}}$  is **efficient** if it assigns each overdemanded patient  $i \in N^O$  to an odd component  $J \in \mathcal{D}$ . Let  $\tilde{\mathcal{E}}$  be the set of efficient pre-matchings. Note that  $\tilde{\mathcal{E}}$  is non-empty by the GED Lemma.

---

<sup>17</sup>So the induced two-sided matching market differs from natural two-sided markets (such as the medical labor market studied in Roth, 1984) in two ways. First, one of the sides is made up not of individual patients (and their donors), but of groups (odd components) of patients (and their donors). Second, these sides aren't specified exogenously, but are determined by the preferences (compatibilities) of the patients.

### 3.4 Priority Mechanisms Revisited

The GED Lemma allows us to see in detail how competition for compatible kidneys plays out in priority mechanisms. The outcome of a priority mechanism is Pareto-efficient and by the GED Lemma, each overdemanded as well as each perfectly matched patient is matched at each Pareto-efficient matching. So the competition in a priority mechanism is among the underdemanded patients. Moreover, for any odd-component  $J$  and any patient  $j \in J$ , it is possible to match all patients in  $J \setminus \{j\}$  among themselves. So a priority mechanism determines which odd components will be fully matched (one member with an overdemanded patient and remaining patients with each other) and which odd components will have all but one of its patients matched (all with each other). This will depend on the relative priority ordering among the set of patients each of whom is the lowest priority patient at an odd component.

For each odd component  $J \in \mathcal{D}$ , let  $\ell_J \in J$  be the lowest priority patient in  $J$ . Since if any patient is unmatched the remaining patients in each odd component can be matched among themselves, all patients in  $N^U \setminus \{\ell_J : J \in \mathcal{D}\}$  will be matched at each priority matching (i.e. if a patient remains unmatched at an odd component, she will be the lowest priority patient in the odd component). Relabel odd components in  $\mathcal{D}$  such that  $D_1 \in \mathcal{D}$  hosts the highest priority patient in  $\{\ell_J : J \in \mathcal{D}\}$ ,  $D_2 \in \mathcal{D}$  hosts the second highest priority patient in  $\{\ell_J : J \in \mathcal{D}\}$ , and so on (i.e. the components are ordered in priority order of their lowest priority patient). For each  $\mathcal{J} \subseteq \mathcal{D}$  and  $I \subseteq N^O$ , define the **neighbors** of the set of odd components  $\mathcal{J}$  among overdemanded patients in  $I$  as

$$C(\mathcal{J}, I) = \{i \in I : \exists J \in \mathcal{J} \text{ with } \tilde{r}_{i,J} = 1\}.$$

That is, each overdemanded patient in  $C(\mathcal{J}, I)$  is mutually compatible with at least one patient in  $\bigcup_{J \in \mathcal{J}} J$ . Which odd components will be fully matched and which ones will have all but its lowest priority member matched will be determined by the hierarchy among the odd components (where the priority of an odd component is determined by the priority of its lowest priority member) together with the following version of Hall's Theorem (Hall (1935)):

**Hall's Theorem:** *Let  $\mathcal{J} \subseteq \mathcal{D}$ . There exists a pre-matching which matches all odd components in  $\mathcal{J}$  with a distinct overdemanded patient in  $N^O$  if and only if*

$$\forall \mathcal{J}' \subseteq \mathcal{J}, \quad |C(\mathcal{J}', N^O)| \geq |\mathcal{J}'|.$$

Under the priority mechanism, odd components  $D_1, D_2, \dots, D_p$  are considered one at a time and the following iterative procedure will determine which odd components will have all its members matched and which ones will have all but its lowest priority member matched.

**Step 1:** If  $|C(\{D_1\}, N^O)| \geq |\{D_1\}| = 1$ , then let  $\mathcal{J}_1 = \{D_1\}$  and in this case all members of  $D_1$  will be matched. If  $|C(\{D_1\}, N^O)| < |\{D_1\}| = 1$ , then let  $\mathcal{J}_1 = \emptyset$  and in this case all members of  $D_1$  except its lowest priority patient  $\ell_1$  will be matched.

In general, at

**Step k:** If  $|C(\mathcal{J} \cup \{D_k\}, N^O)| \geq |\mathcal{J} \cup \{D_k\}|$  for every  $\mathcal{J} \subseteq \mathcal{J}_{k-1}$ , then let  $\mathcal{J}_k = \mathcal{J}_{k-1} \cup \{D_k\}$  and in this case all members of  $D_k$  will be matched. If  $|C(\mathcal{J} \cup \{D_k\}, N^O)| < |\mathcal{J} \cup \{D_k\}|$  for some  $\mathcal{J} \subseteq \mathcal{J}_{k-1}$ , then let  $\mathcal{J}_k = \mathcal{J}_{k-1}$  and in this case all members of  $D_k$  but its lowest priority patient  $\ell_k$  will be matched.

## 4 Stochastic Exchange

So far our emphasis has been on deterministic exchange. One important tool to achieve equity in resource allocation problems with indivisibilities is using lotteries and allowing for stochastic outcomes.<sup>18</sup> A **lottery**  $\lambda = (\lambda_\mu)_{\mu \in \mathcal{M}}$  is a probability distribution over the set of matchings  $\mathcal{M}$ . For each matching  $\mu \in \mathcal{M}$ ,  $\lambda_\mu \in [0, 1]$  is the probability of matching  $\mu$  in lottery  $\lambda$ , and  $\sum_{\mu \in \mathcal{M}} \lambda_\mu = 1$ . Let  $\mathcal{L}$  be the set of lotteries for the problem  $(N, \succ)$ . A **stochastic mechanism** is a systematic procedure that selects a lottery for each problem. Given a lottery  $\lambda \in \mathcal{L}$ , the **allocation matrix**  $A(\lambda) = [a_{i,j}(\lambda)]_{i \in N, j \in N}$  summarizes the total probability that patient  $i$  will be matched with patient  $j$  for any pair of patients  $i, j \in N$ . Note that two distinct lotteries can induce the same allocation matrix. Let  $\mathcal{A}$  be the set of all allocation matrices for the problem  $(N, \succ)$ . That is,  $\mathcal{A} = \{A(\lambda)\}_{\lambda \in \mathcal{L}}$ . Each lottery (and hence each allocation matrix) specifies the probability that each patient will receive a transplant. Given a lottery  $\lambda \in \mathcal{L}$ , define the utility of patient  $i$  to be the aggregate probability that she receives a transplant. Given  $\lambda \in \mathcal{L}$ , the induced **utility profile** is a non-negative real vector  $u(\lambda) = (u_i(\lambda))_{i \in N}$  such that  $u_i(\lambda) = \sum_{j \in N \setminus \{i\}} a_{i,j}(\lambda)$  for any patient  $i \in N$ . Let  $\mathcal{U}$  be the set of all feasible utility profiles for the problem  $(N, \succ)$ . That is,  $\mathcal{U} = \{u(\lambda)\}_{\lambda \in \mathcal{L}}$ .

A lottery is **ex-post efficient** if its support is a subset of the set of Pareto-efficient matchings. That is, lottery  $\lambda \in \mathcal{L}$  is ex-post efficient if  $\lambda_\mu > 0$  implies  $\mu \in \mathcal{E}$ . An allocation matrix  $A \in \mathcal{A}$  is **ex-ante efficient** if there exists no allocation matrix  $B \in \mathcal{A}$  such that  $\sum_{j \in N \setminus \{i\}} b_{i,j} \geq \sum_{j \in N \setminus \{i\}} a_{i,j}$  for all  $i \in N$  and  $\sum_{j \in N \setminus \{i\}} b_{i,j} > \sum_{j \in N \setminus \{i\}} a_{i,j}$  for some  $i \in N$ . A utility profile  $u \in \mathcal{U}$  is **efficient** if there exists no other utility profile  $v \in \mathcal{U}$  such that  $v_i \geq u_i$  for all  $i \in N$  and  $v_i > u_i$  for some  $i \in N$ .

An immediate implication of Lemma 1 (as well as the GED Lemma) is the “equivalence” of ex-ante and ex-post efficiency in the present context. This result is also stated by Bogomolnaia and Moulin (2004).

**Lemma 3** *An allocation matrix  $A \in \mathcal{A}$  is ex-ante efficient if and only if there is an ex-post efficient lottery  $\lambda \in \mathcal{L}$  such that  $A(\lambda) = A$ .*

<sup>18</sup>For other discussions of stochastic matching, see Roth and Vande Vate (1990), Roth, Rothblum and Vande Vate (1993), Roth and Xing (1997), Abdulkadiroğlu and Sönmez (1998, 2003a), Bogomolnaia and Moulin (2001), Cres and Moulin (2001), and Sönmez and Ünver (2001).



The notions for the induced two-sided matching market  $(N^O, \mathcal{D}, \tilde{R})$  can be similarly extended to allow for stochastic outcomes. A **pre-lottery**  $\tilde{\lambda} = (\tilde{\lambda}_{\tilde{\mu}})_{\tilde{\mu} \in \tilde{\mathcal{M}}}$  is a probability distribution over the set of pre-matchings  $\tilde{\mathcal{M}}$ . Let  $\tilde{\mathcal{L}}$  be the set of all pre-lotteries. A **pre-allocation (matrix)**  $\tilde{A} = [\tilde{a}_{i,J}]_{i \in N^O, J \in \mathcal{D}}$  is a non-negative matrix such that  $\sum_{J \in \mathcal{D}} \tilde{a}_{i,J} \leq 1$  for any  $i \in N^O$ ,  $\sum_{i \in N^O} \tilde{a}_{i,J} \leq 1$  for any  $J \in \mathcal{D}$ , and  $\tilde{a}_{i,J} > 0$  implies  $\tilde{r}_{i,J} = 1$  for any pair  $i \in N^O$ ,  $J \in \mathcal{D}$ . Let  $\tilde{\mathcal{A}}$  be the set of all pre-allocations. A pre-lottery  $\tilde{\lambda} \in \tilde{\mathcal{L}}$  **induces** the pre-allocation  $\tilde{A} \in \tilde{\mathcal{A}}$  if for each pair  $i \in N^O$  and  $J \in \mathcal{D}$ ,  $\tilde{a}_{i,J}$  is the cumulative probability that patient  $i \in N^O$  is matched with set  $J \in \mathcal{D}$  under the pre-lottery  $\tilde{\lambda}$ . By Lemma 2.1 in Bogomolnaia and Moulin (2002), for each pre-allocation there is a pre-lottery that induces it. A pre-lottery  $\tilde{\lambda} \in \tilde{\mathcal{L}}$  is **ex-post efficient** if its support is a subset of the set of efficient pre-matchings. That is, pre-lottery  $\tilde{\lambda}$  is ex-post efficient if  $\tilde{\lambda}_{\tilde{\mu}} > 0$  implies  $\tilde{\mu} \in \tilde{\mathcal{E}}$ . Pre-matchings, pre-lotteries and pre-allocations will be very useful in our analysis.

## 4.1 The Egalitarian Mechanism

It is rare that a cardinal representation of preferences, i.e. a utility function, has a direct interpretation in a resource allocation problem. In the present context, however, a patient's utility corresponds to the probability that she receives a transplant, and thus equalizing utilities as much as possible becomes very plausible from an equity perspective. This approach is widely known as **egalitarianism** in distributive justice.<sup>19</sup> In this section we analyze the egalitarian mechanism. The GED Lemma will be key to the construction of egalitarian utilities.

Recall that  $C(\mathcal{J}, I)$  denotes the neighbors of the set of odd components  $\mathcal{J} \subseteq \mathcal{D}$  among overdemanded patients  $I \subseteq N^O$ . For each  $\mathcal{J} \subseteq \mathcal{D}$  and  $I \subseteq N^O$ , define a real-valued function  $f$  through

$$f(\mathcal{J}, I) = \frac{|\cup_{J \in \mathcal{J}} J| - (|\mathcal{J}| - |C(\mathcal{J}, I)|)}{|\cup_{J \in \mathcal{J}} J|}.$$

Recall that at most one patient in each odd component remains unmatched at every Pareto-efficient matching and therefore no more than  $|\mathcal{J}|$  patients among patients in  $\cup_{J \in \mathcal{J}} J$  can remain unmatched at any Pareto-efficient matching. Consider a situation where only overdemanded patients in  $I \subseteq N^O$  are available to be matched with underdemanded patients in  $\cup_{J \in \mathcal{J}} J$ . By definition of a neighbor, underdemanded patients in  $\cup_{J \in \mathcal{J}} J$  can only be matched with overdemanded patients in  $C(\mathcal{J}, I) \subseteq I$  and therefore at least  $(|\mathcal{J}| - |C(\mathcal{J}, I)|)$  of these patients remain unmatched at a Pareto-efficient matching (provided that  $|\mathcal{J}| \geq |C(\mathcal{J}, I)|$ ).<sup>20</sup> Therefore if only overdemanded patients in  $I \subseteq N^O$  are available to be matched with patients in  $\cup_{J \in \mathcal{J}} J$ , the quantity  $f(\mathcal{J}, I)$  is an upper-bound of the utility (i.e. the probability of receiving a transplant) of the *least fortunate patient* in  $\cup_{J \in \mathcal{J}} J$ ,

<sup>19</sup>The two most related economic applications of egalitarianism to our setup are Bogomolnaia and Moulin (2004) and Dutta and Ray (1989).

<sup>20</sup>For the sets of odd components  $\mathcal{J}$  that we consider below,  $|\mathcal{J}| > |C(\mathcal{J}, I)|$ .

and this upper-bound can only be reached if not only all underdemanded patients in  $\bigcup_{J \in \mathcal{J}} J$  receive the same utility but also all overdemanded patients in  $C(\mathcal{J}, I)$  are matched to patients in  $\bigcup_{J \in \mathcal{J}} J$ . This motivates the following recursive construction that partitions  $\mathcal{D}$  as  $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_q\}$  and  $N^O$  as  $\{N_1^O, N_2^O, \dots, N_q^O\}$ :

**Step 1:** Let<sup>21</sup>

$$\mathcal{D}_1 = \arg \min_{\mathcal{J} \subseteq \mathcal{D}} f(\mathcal{J}, N^O) \quad \text{and} \quad N_1^O = C(\mathcal{D}_1, N^O).$$

In general, at

**Step k:** Let

$$\mathcal{D}_k = \arg \min_{\mathcal{J} \subseteq \mathcal{D} \setminus \bigcup_{\ell=1}^{k-1} \mathcal{D}_\ell} f\left(\mathcal{J}, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right) \quad \text{and} \quad N_k^O = C\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right).$$

For each  $j \in N^U$ , let  $k(j)$  be the step at which the odd component that contains patient  $j$  leaves the above procedure. That is,  $k(j) \in \{1, 2, \dots, q\}$  is such that  $j \in J \in \mathcal{D}_{k(j)}$ .

We construct the **egalitarian utility profile**  $u^E = (u_i^E)_{i \in N}$  as follows:  $u_i^E = 1$  for each  $i \in N \setminus N^U$ , and  $u_i^E = f(\mathcal{D}_{k(i)}, N_{k(i)}^O)$  for each  $i \in N^U$ .

**Theorem 2** *The egalitarian utility profile  $u^E$  is a feasible utility profile.*

Theorem 2 states that for each of the collections of under and overdemanded patients  $(\mathcal{D}_k, N_k^O)$  in the above construction, the overdemanded patients can be probabilistically matched to all of the indicated odd components of underdemanded patients, in lotteries that divide the resulting probabilities equally among the underdemanded patients, and thus achieve the upper bound on the utility of the least fortunate patients.

We next present an example that illustrates the construction of the egalitarian utilities.

**Example 2** Let  $N = \{1, \dots, 16\}$  be the set of patients and consider the reduced problem given by the graph in Figure 1. Each patient except 1 and 2 can be left unmatched at some Pareto efficient matching and hence  $N^U = \{3, \dots, 16\}$  is the set of underdemanded patients. Since both patients 1 and 2 have links with patients in  $N^U$ ,  $N^O = \{1, 2\}$  is the set of overdemanded patients.<sup>22</sup> Therefore the set of odd components is  $\mathcal{D} = \{D_1, \dots, D_6\}$  where  $D_1 = \{3\}$ ,  $D_2 = \{4\}$ ,  $D_3 = \{5\}$ ,  $D_4 = \{6, 7, 8\}$ ,  $D_5 = \{9, 10, 11\}$ ,  $D_6 = \{12, 13, 14, 15, 16\}$ . Note that the egalitarian utility of members of the 3-patient odd components  $D_4, D_5$  is at least  $2/3$ . That is because any two of the three can be matched among themselves, by the GED Lemma. Similarly the egalitarian utility of each member of the

<sup>21</sup>If there are multiple sets that minimize  $f$ , their union minimizes  $f$  as well and we pick the largest such set as the arg min. See Lemma 4 in the Appendix for a proof of this result.

<sup>22</sup>Since perfectly matched patients do not play an active role in determining the egalitarian utilities, we consider an example with no such patients.

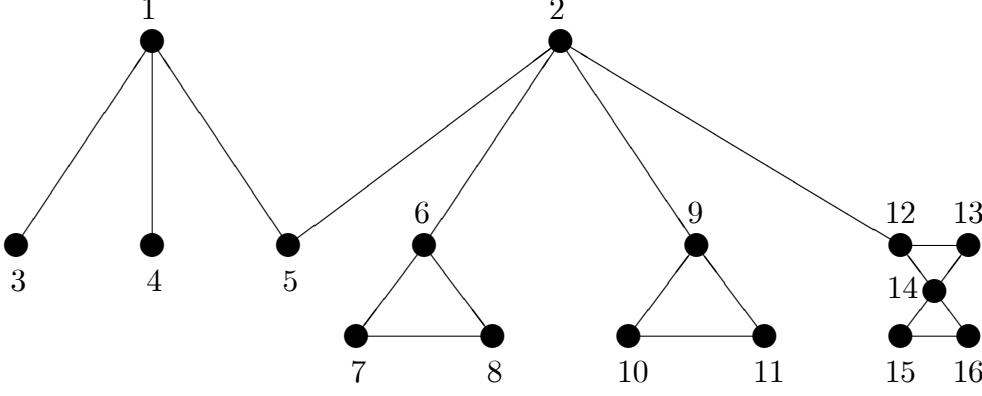


Figure 1: Graphical Representation for Example 2.

5-patient odd component  $D_6$  is at least  $4/5$ . These observations are helpful for constructing the egalitarian utility profile.

Consider  $\mathcal{J}_1 = \{D_1, D_2\} = \{\{3\}, \{4\}\}$ . Note that  $f(\mathcal{J}_1, N^O) = \frac{2-(2-1)}{2} = \frac{1}{2}$ . Since  $f(\mathcal{J}_1, N^O) = \frac{1}{2} < \frac{2}{3} < \frac{4}{5}$ , none of the multi-patient odd components is an element of  $\mathcal{D}_1$ . Moreover patient 5 has two overdemanded neighbors and  $f(\mathcal{J}, N^O) > f(\mathcal{J}_1, N^O)$  for any  $\mathcal{J} \subseteq \{\{3\}, \{4\}, \{5\}\}$  with  $\{5\} \in \mathcal{J}$ . Therefore

$$\mathcal{D}_1 = \mathcal{J}_1 = \{\{3\}, \{4\}\}, \quad N_1^O = \{1\}, \quad \text{and} \quad u_3^E = u_4^E = \frac{1}{2}.$$

Next consider  $\mathcal{J}_2 = \{D_3, D_4, D_5\} = \{\{5\}, \{6, 7, 8\}, \{9, 10, 11\}\}$ . Note that  $f(\mathcal{J}_2, N^O \setminus N_1^O) = \frac{7-(3-1)}{7} = \frac{5}{7}$ . Since  $f(\mathcal{J}_2, N^O \setminus N_1^O) = \frac{5}{7} < \frac{4}{5}$ , the 5-patient odd component  $D_6$  is not an element of  $\mathcal{D}_2$ . Moreover

$$\begin{aligned} f(\{D_3\}, N^O \setminus N_1^O) &= f(\{D_4\}, N^O \setminus N_1^O) = f(\{D_5\}, N^O \setminus N_1^O) = 1, \\ f(\{D_3, D_4\}, N^O \setminus N_1^O) &= f(\{D_3, D_5\}, N^O \setminus N_1^O) = \frac{3}{4}, \quad \text{and} \quad f(\{D_4, D_5\}, N^O \setminus N_1^O) = \frac{5}{6}. \end{aligned}$$

Therefore

$$\mathcal{D}_2 = \mathcal{J}_2 = \{\{5\}, \{6, 7, 8\}, \{9, 10, 11\}\}, \quad N_2^O = \{2\}, \quad \text{and} \quad u_5^E = \dots = u_{11}^E = \frac{5}{7}.$$

Finally since  $N^O \setminus (N_1^O \cup N_2^O) = \emptyset$ ,

$$\mathcal{D}_3 = \{\{12, 13, 14, 15, 16\}\}, \quad N_3^O = \emptyset, \quad \text{and} \quad u_{12}^E = \dots = u_{16}^E = \frac{4}{5}.$$

Hence the egalitarian utility profile is  $u^E = (1, 1, \frac{1}{2}, \frac{1}{2}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5})$ . ■

For any utility profile  $u \in \mathcal{U}$ , re-order individual utilities in an increasing order as  $(u^{(t)})_{t \in \{1, 2, \dots, n\}}$  such that  $u^{(1)} \leq u^{(2)} \leq \dots \leq u^{(n)}$ . A utility profile  $u \in \mathcal{U}$  **Lorenz-dominates** a utility profile  $v \in \mathcal{U}$  if

1. for each  $t \in \{1, 2, \dots, n\}$  we have  $\sum_{s=1}^t u^{(s)} \geq \sum_{s=1}^t v^{(s)}$ , and
2. there is some  $t \in \{1, 2, \dots, n\}$  such that  $\sum_{s=1}^t u^{(s)} > \sum_{s=1}^t v^{(s)}$ .

A utility profile is **Lorenz-dominant** if and only if it Lorenz-dominates every other utility profile. If it exists, a Lorenz-dominant utility profile is efficient and as “evenly” distributes the probability of receiving a transplant among patients as possible constrained by the mutual compatibility constraints.

**Theorem 3** *The egalitarian utility profile  $u^E$  is Lorenz-dominant.*

We illustrate Theorem 3 with a simple example.

**Example 3** Let  $N = \{1, 2, 3, 4\}$  and suppose patient 1 is mutually compatible with patient 2 as well as with patient 3 but patients 2 and 3 are not mutually compatible. The two Pareto efficient matchings are

$$\mu = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \text{ and } \nu = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

and the egalitarian utility profile is  $u^E = (1, 1/2, 1/2)$ . So re-ordering the individual utilities from smallest to largest,  $u^{E(1)} = 1/2$ ,  $u^{E(2)} = 1/2$ , and  $u^{E(3)} = 1$ . Next consider any ex-post efficient lottery  $\alpha\mu + (1 - \alpha)\nu$ . The induced utility profile is  $u^\alpha = (1, \alpha, 1 - \alpha)$  and  $u^{\alpha(1)} = \min\{\alpha, 1 - \alpha\}$ ,  $u^{\alpha(2)} = \max\{\alpha, 1 - \alpha\}$ , and  $u^{\alpha(3)} = 1$ . Hence  $u^{E(1)} > u^{\alpha(1)}$ ,  $u^{E(1)} + u^{E(2)} = u^{\alpha(1)} + u^{\alpha(2)}$ , and  $u^{E(1)} + u^{E(2)} + u^{E(3)} = u^{\alpha(1)} + u^{\alpha(2)} + u^{\alpha(3)}$  for  $\alpha \neq 1/2$  showing that  $u^E$  Lorenz dominates any other efficient utility profile. Since each inefficient utility profile is Lorenz dominated by any efficient utility profile that Pareto dominates it (and since Lorenz domination is transitive),  $u^E$  Lorenz dominates any utility profile. ■

We refer to any lottery that induces the egalitarian utility profile as an **egalitarian lottery**. Similarly we refer to any allocation matrix that induces the egalitarian utility profile as an **egalitarian allocation matrix**. We refer to a stochastic mechanism that selects an egalitarian lottery for each problem as an **egalitarian mechanism**.

Our next result states that, as for priority mechanisms, truthful revelation of private information is a dominant strategy under an egalitarian mechanism.

**Theorem 4** *An egalitarian mechanism makes it a dominant strategy for a patient to reveal both a) her full set of acceptable kidneys; and b) her full set of available donors.*

As discussed when we proved the similar result for the priority mechanisms, the second part of the Theorem follows from the first. Revelation of the full set of acceptable kidneys increases the set of patients with whom a given patient is mutually compatible, and, for the same reason, a patient never suffers because of an addition of one more incompatible donor. That is:

**Corollary 4** *An egalitarian mechanism is donor-monotonic.*

## 5 Concluding Remarks

One of the most challenging steps in implementing new market designs is addressing the constraints that arise in the field. Because all surgeries in a kidney exchange must be done simultaneously, even the simplest exchange, between two patient-donor pairs, requires four simultaneous surgeries among the two donors and two recipients. It therefore seems likely that some transplant centers are correct in anticipating that they will, at least initially, only be able to proceed with exchange among two pairs. Roth, Sönmez, and Ünver (2004, 2005) showed that, even under this constraint, implementing kidney exchange could substantially increase the number of live organ kidney transplants. The present paper shows that when exchange is constrained in this way, and when the 0-1 nature of American surgeons' preferences regarding compatible/incompatible kidneys are taken into account, it is still possible to arrange exchange in an efficient and incentive compatible manner.

Another challenge in implementing new market designs arises in meeting the perceived needs and desires of the institutions, organizations, and individuals who must adopt and use the new design. Here too, the results are encouraging, and we show that the kinds of priority allocation that already govern the allocation of cadaver kidneys can be adapted to work effectively in organizing live donor kidney exchange. Since there are virtually no constraints on the kinds of priorities that can be used in an efficient and incentive compatible mechanism, we anticipate that priority mechanisms may appeal to different transplant centers that do not necessarily agree on how patients should be prioritized.

Indeed, there is lively discussion and disagreement about, and frequent revision of the priorities that different kinds of patients should have for cadaver organs, and we don't doubt that similar discussions about live organ exchange will take place as exchange becomes more common. The debate about cadaver organs frequently refers to considerations of distributive justice, and in this connection we have discussed stochastic mechanisms of exchange, and the egalitarian mechanism in particular. We do not anticipate or propose that this be taken as the last word on distributive justice in kidney exchange, but rather intend merely to show how the tools we assemble here can be used to address the kinds of distributional questions that arise. An interesting area of future research might be the extent to which egalitarian (and weighted-egalitarian) exchange mechanisms can be approximated by priority mechanisms with appropriately chosen priorities, taking into account different patients' likelihoods of finding compatible donors with whom to exchange, the size and frequency of exchanges, etc., in the dynamic environment in which new patients and donors arrive, and exchanges are carried out periodically.

More generally, as economists start to take a more active role in practical market design, we often find we have to deal with constraints, demands, and situations different than those that arise

in the simplest theoretical models of mechanism design.<sup>23</sup> In the present paper we address some of the issues that have arisen as we try to help surgeons implement an organized exchange of live-donor kidneys among incompatible donor-patient pairs. Not only do these issues appear to allow satisfactory practical solutions, they suggest new directions in which to pursue the underlying theory.

## 6 Appendix: Proofs

The proof of Proposition 1 is a standard exercise in combinatorial optimization theory, for example see Goemans (2004). Proposition 2 is due to Rado (1957) and Edmonds (1971). Lemma 1 and Lemma 3 are corollaries of the GED Lemma (Lemma 2) and see Lovász and Plummer (1986) for a proof of the GED Lemma.

*Proof of Theorem 1a):* W.l.o.g. we will prove the theorem for the priority mechanism  $\phi$  induced by the natural ordering. Let  $R$  be a reduced problem and  $k$  be a patient. If patient  $k$  is matched with another patient under  $\phi(R)$ , then she has nothing to gain by revealing only a subset of her full set of compatible kidneys. Suppose patient  $k$  remains unmatched under  $\mu = \phi(R)$  and let  $Q$  be a reduced problem obtained from  $R$  by patient  $k$  declaring some of her compatible kidneys to be incompatible. Observe that this implies  $\mathcal{E}^{k-1}(Q) \subseteq \mathcal{E}^{k-1}(R)$ . Let  $\phi(Q) = \nu$ . Since  $\mu(k) = k$ ,  $\mu'(k) = k$  for all  $\mu' \in \mathcal{E}^{k-1}(R)$ . But then  $\mu'(k) = k$  for all  $\mu' \in \mathcal{E}^{k-1}(Q)$  as well and hence  $\nu(k) = k$  completing the proof. ■

The following Lemma is useful to construct the egalitarian utility profile:

**Lemma 4** Fix  $\mathcal{G} \subseteq \mathcal{D}$  and  $I \subseteq N^O$ . Suppose  $\mathcal{G}_1, \mathcal{G}_2 \in \arg \min_{\mathcal{J} \subseteq \mathcal{G}} f(\mathcal{J}, I)$ . Then  $\mathcal{G}_1 \cup \mathcal{G}_2 \in \arg \min_{\mathcal{J} \subseteq \mathcal{G}} f(\mathcal{J}, I)$  as well.

*Proof of Lemma 4:* Fix  $\mathcal{G} \subseteq \mathcal{D}$  and  $I \subseteq N^O$ . Suppose  $\mathcal{G}_1, \mathcal{G}_2 \in \arg \min_{\mathcal{J} \subseteq \mathcal{G}} f(\mathcal{J}, I)$ . Let  $\mathcal{G}_3 = \mathcal{G}_1 \cap \mathcal{G}_2$  and  $\mathcal{G}_4 = \mathcal{G}_1 \cup \mathcal{G}_2$ . For all  $i \in \{1, 2, 3, 4\}$  define

$$n_i = \left| \bigcup_{J \in \mathcal{G}_i} J \right|, \quad C_i = C(\mathcal{G}_i, I), \quad \text{and} \quad f_i = f(\mathcal{G}_i, I).$$

By definition we have

$$|\mathcal{G}_1| + |\mathcal{G}_2| = |\mathcal{G}_3| + |\mathcal{G}_4| \quad \text{and} \quad n_1 + n_2 = n_3 + n_4.$$

---

<sup>23</sup>See for instance Roth and Peranson (1999), Roth (2002), Wilson (2002), Abdulkadiroğlu and Sönmez (2003b), Milgrom (2004), Niederle and Roth (2004, 2005), Crawford (2005), Abdulkadiroğlu, Pathak, and Roth (2005), and Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005) for some examples. Indeed, one of the principal motivations of Roth, Sönmez, and Ünver (2004) was to organize efficient kidney exchange under the social/ethical/legal prohibitions on monetized markets.

Moreover

$$|C_1| + |C_2| \geq |C_3| + |C_4|.$$

That is because, in the LHS of the inequality not only the neighbors of  $\mathcal{G}_3 = \mathcal{G}_1 \cap \mathcal{G}_2$  (i.e. members of  $C_3$ ) are counted twice but also there may be additional common neighbors of  $\mathcal{G}_1$  and  $\mathcal{G}_2$ ; RHS of the inequality accounts for the double counting of members of  $C_3$  but not the remaining common neighbors of  $\mathcal{G}_1$  and  $\mathcal{G}_2$ .

Since  $\mathcal{G}_1, \mathcal{G}_2$  each minimize the function  $f$ ,

$$\underbrace{\frac{(n_1 - (|\mathcal{G}_1| - |C_1|))}{n_1}}_{=f_1} = \underbrace{\frac{(n_2 - (|\mathcal{G}_2| - |C_2|))}{n_2}}_{=f_2} \leq \underbrace{\frac{(n_3 - (|\mathcal{G}_3| - |C_3|))}{n_3}}_{=f_3}$$

and hence

$$\begin{aligned} f_1 n_1 &= n_1 - (|\mathcal{G}_1| - |C_1|), \\ f_1 n_2 &= n_2 - (|\mathcal{G}_2| - |C_2|), \\ f_1 n_3 &\leq n_3 - (|\mathcal{G}_3| - |C_3|). \end{aligned}$$

Adding the first two lines and subtracting the third line

$$f_1 \underbrace{(n_1 + n_2 - n_3)}_{=n_4} \geq \underbrace{(n_1 + n_2 - n_3)}_{=n_4} - \underbrace{(|\mathcal{G}_1| + |\mathcal{G}_2| - |\mathcal{G}_3|)}_{=|\mathcal{G}_4|} + \underbrace{(|C_1| + |C_2| - |C_3|)}_{\geq |C_4|}$$

and therefore  $f_1 n_4 \geq n_4 - |\mathcal{G}_4| + |C_4|$ , or equivalently

$$f_1 \geq \frac{(n_4 - (|\mathcal{G}_4| - |C_4|))}{n_4} = f_4.$$

But since  $\mathcal{G}_1$  minimizes  $f$ , we shall have  $f_4 = f_1$  and hence  $\mathcal{G}_4 = \mathcal{G}_1 \cup \mathcal{G}_2$  minimizes  $f$  as well.  $\blacksquare$

We next present two lemmata that will be useful in our proof for Theorem 2. Lemma 5 is a part of Lemma 3.2.2 in Lovász and Plummer (1986) pp 95:

**Lemma 5** *For each  $i \in N^O$ , the Gallai-Edmonds decomposition of the reduced subproblem  $(N \setminus \{i\}, R_{N \setminus \{i\}})$  is given by  $\{N^U, N^O \setminus \{i\}, N^P\}$  where  $N^U$  is the set of underdemanded patients,  $N^O \setminus \{i\}$  is the set of overdemanded patients, and  $N^P$  is the set of perfectly matched patients for the reduced subproblem  $(N \setminus \{i\}, R_{N \setminus \{i\}})$ .*

**Lemma 6** *For each  $k \in \{1, 2, \dots, q\}$ , we have*

$$i. C(\mathcal{D}_k, N_k^O) = C\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right) \text{ and } f(\mathcal{D}_k, N_k^O) = f\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right),$$

ii.  $f(\mathcal{D}_k, N_k^O) < 1$ .

*Proof of Lemma 6:* For any  $k \in \{1, 2, \dots, q\}$ , let  $J_k = \bigcup_{J \in \mathcal{D}_k} J$ . Note that  $J_k$  is the set of patients in sets of  $\mathcal{D}_k$ . Pick  $k \in \{1, 2, \dots, q-1\}$ .

i. First observe that  $C(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O) = C(\mathcal{D}_k, C(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O))$ . This together with  $N_k^O = C(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O)$  implies that  $C(\mathcal{D}_k, N_k^O) = C(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O)$ . Furthermore

$$\begin{aligned} f\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right) &= \frac{|J_k| - \left(|\mathcal{D}_k| - \left|C\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right)\right|\right)}{|J_k|} \\ &= \frac{|J_k| - \left(|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|\right)}{|J_k|} = f(\mathcal{D}_k, N_k^O). \end{aligned}$$

ii. Consider the construction of the partition  $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_q\}$  of  $\mathcal{D}$  and  $\{N_1^O, N_2^O, \dots, N_q^O\}$  of  $N^O$ . We prove Lemma 6 (ii) by iteration on steps  $k \in \{1, 2, \dots, q\}$  of the construction.

First consider  $k = 1$ . We have

$$\begin{aligned} f(\mathcal{D}_1, N_1^O) &= f(\mathcal{D}_1, N^O) && \text{by Lemma 6 (i)} \\ &\leq f(\mathcal{D}, N^O) && \text{by construction of } \mathcal{D}_1 \\ &= \frac{|N^U| - (|\mathcal{D}| - |C(\mathcal{D}, N^O)|)}{|N^U|} && \text{by definition of } f \\ &= \frac{|N^U| - (|\mathcal{D}| - |N^O|)}{|N^U|} && \text{by } C(\mathcal{D}, N^O) = N^O \\ &< 1 && \text{since } |N^O| < |\mathcal{D}| \text{ by Corollary 3} \end{aligned}$$

If  $\mathcal{D}_1 = \mathcal{D}$ , then  $q = 1$  and we are done. Otherwise we proceed with  $k = 2$ .

Next consider  $k = 2$ . Consider the reduced subproblem  $(N \setminus N_1^O, R_{N \setminus N_1^O})$ . Since  $N_1^O \subseteq N^O$ , Lemma 5 implies that the Gallai-Edmonds decomposition for the reduced subproblem  $(N \setminus N_1^O, R_{N \setminus N_1^O})$  is given by  $\{N^U, N^O \setminus N_1^O, N^P\}$ . Since  $N_1^O = C(\mathcal{D}_1, N^O)$ , there is no patient  $j \in J_1 = \bigcup_{J \in \mathcal{D}_1} J$  who is mutually compatible with a patient in  $N \setminus N_1^O$ . Therefore,  $(J, R_J)$  is a component of reduced subproblem  $(N \setminus N_1^O, R_{N \setminus N_1^O})$  for each  $J \in \mathcal{D}_1$ , which in turn implies  $\{N^U \setminus J_1, N^O \setminus N_1^O, N^P\}$  is the Gallai-Edmonds decomposition for the reduced subproblem  $(N \setminus (N_1^O \cup J_1), R_{N \setminus (N_1^O \cup J_1)})$ . Therefore,  $\mathcal{D} \setminus \mathcal{D}_1$  is the set of odd components of the reduced subproblem which is obtained by removing overdemanded patients from the reduced subproblem  $(N \setminus (N_1^O \cup J_1), R_{N \setminus (N_1^O \cup J_1)})$ .



By Corollary 3,  $|\mathcal{D} \setminus \mathcal{D}_1| > |N \setminus N_1^O|$  and at this point the proof of  $f(\mathcal{D}_2, N_2^O) < 1$  follows from the same sequence of arguments as in  $k = 1$ :

$$\begin{aligned}
f(\mathcal{D}_2, N_2^O) &= f(\mathcal{D}_2, N^O \setminus N_1^O) \\
&\leq f(\mathcal{D} \setminus \mathcal{D}_1, N^O \setminus N_1^O) \\
&= \frac{|N^U \setminus J_1| - (|\mathcal{D} \setminus \mathcal{D}_1| - |C(\mathcal{D} \setminus \mathcal{D}_1, N^O \setminus N_1^O)|)}{|N^U \setminus J_1|} \\
&= \frac{|N^U \setminus J_1| - (|\mathcal{D} \setminus \mathcal{D}_1| - |N^O \setminus N_1^O|)}{|N^U \setminus J_1|} < 1
\end{aligned}$$

If  $\mathcal{D}_1 \cup \mathcal{D}_2 = \mathcal{D}$  then  $q = 2$  and we are done. Otherwise, we iteratively proceed in a similar way with  $k = 3$ , and so on.  $\blacksquare$

*Proof of Theorem 2:* Let  $k \in \{1, 2, \dots, q\}$  and  $J \in \mathcal{D}_k$ . Recall that  $J$  is a set of underdemanded patients and  $(J, R_J)$  is an odd component of  $(N \setminus N^O, R_{N \setminus N^O})$ . Under a Pareto-efficient matching, at most one patient in  $J$  is matched with an overdemanded patient, and for each  $j \in J$ , it is possible to match the remaining  $|J| - 1$  patients in  $J \setminus \{j\}$  with each other by the GED Lemma. Therefore the set  $J$ , by itself, generates an aggregate utility of  $|J| - 1$  for its members under each efficient utility profile without any help of the overdemanded patients. Moreover,  $|J| \cdot f(\mathcal{D}_k, N_k^O)$  is the aggregate utility of patients in set  $J$  under  $u^E$  and  $|J| \cdot f(\mathcal{D}_k, N_k^O) \geq |J| - 1$  by construction of  $f$ . Therefore, if  $u^E$  is a feasible utility profile, then patients in  $J$  should be matched with overdemanded patients with a cumulative probability of  $|J| \cdot f(\mathcal{D}_k, N_k^O) - (|J| - 1)$  under any lottery  $\lambda$  that induces  $u^E$ . Let

$$\alpha_J = |J| \cdot f(\mathcal{D}_k, N_k^O) - (|J| - 1).$$

Note that  $f(\mathcal{D}_k, N_k^O) < 1$  by Lemma 6 (ii), and therefore  $f(\mathcal{D}_k, N_k^O) \geq \alpha_J$ . Also note that for each  $k \in \{1, 2, \dots, q\}$  and  $J \in \mathcal{D}_k$  we have

$$\begin{aligned}
\alpha_J = |J| \cdot f(\mathcal{D}_k, N_k^O) - (|J| - 1) &= |J| \cdot \frac{|\bigcup_{J' \in \mathcal{D}_k} J'| - (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|)}{|\bigcup_{J' \in \mathcal{D}_k} J'|} - (|J| - 1) \\
&= \frac{|\bigcup_{J' \in \mathcal{D}_k} J'| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|)}{|\bigcup_{J' \in \mathcal{D}_k} J'|} \tag{1}
\end{aligned}$$

We will show that  $u^E$  is a feasible utility profile in two major steps: In the first step (Claim 1), we will show that it is possible to assign overdemanded patients  $N^O$  to odd components  $\mathcal{D}$  such that each odd component  $J \in \mathcal{D}$  is assigned with an overdemanded patient with an aggregate probability of  $\alpha_J$ . In the second step (Claim 2) we will show that for each odd component  $J \in \mathcal{D}_k$ , it is possible to evenly divide the aggregate utility  $\alpha_J + (|J| - 1)$  among  $|J|$  members as  $f(\mathcal{D}_k, N_k^O)$ .

*Claim 1:* There exists a pre-allocation matrix  $\tilde{A} \in \tilde{\mathcal{A}}$  such that

1. For each  $i \in N^O$ ,  $\sum_{J \in \mathcal{D}} \tilde{a}_{i,J} = 1$ , and
2. For each  $k \in \{1, 2, \dots, q\}$  and  $J \in \mathcal{D}_k$ 
  - (a)  $\tilde{a}_{i,J} = 0$  for all  $i \in N^O \setminus N_k^O$ , and
  - (b)  $\sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$ .

*Proof of Claim 1:* Let  $k \in \{1, 2, \dots, q\}$ . We will show that it is possible to share the aggregate  $|N_k^O|$  units of “weight” of members of  $N_k^O$  (1 unit weight from each member) among (only) members of  $\mathcal{D}_k$  such that the share of odd component  $J \in \mathcal{D}_k$  is  $\alpha_J$ . Formally, we will show that there exists a non-negative valued matrix  $\tilde{A}^{k,k} = [\tilde{a}_{i,J}]_{i \in N_k^O, J \in \mathcal{D}_k}$  such that

1.  $\sum_{J \in \mathcal{D}_k} \tilde{a}_{i,J} = 1$  for all  $i \in N_k^O$ ,
2.  $\sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$  for all  $J \in \mathcal{D}_k$ , and
3.  $\tilde{a}_{i,J} > 0 \Rightarrow \tilde{r}_{i,J} = 1$  for any pair  $i \in N_k^O, J \in \mathcal{D}_k$ .

We will show this by defining an **auxiliary task assignment problem** and applying Hall’s Theorem to the auxiliary task assignment problem.<sup>24</sup> Given  $N_k^O$  and  $\mathcal{D}_k$ , construct the task assignment problem  $(X, \mathcal{T}, \Gamma)$  as follows:

- For each overdemanded patient  $i \in N_k^O$ , introduce  $|\bigcup_{J' \in \mathcal{D}_k} J'|$  identical agents. Let  $X_i$  be the set of the identical agents associated with patient  $i$ , and  $X = \bigcup_{i \in N_k^O} X_i$ .
- For each odd component  $J \in \mathcal{D}_k$ , introduce  $|\bigcup_{J' \in \mathcal{D}_k} J'| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|)$  identical tasks. Let  $\mathcal{T}_J$  be the set of identical tasks associated with set  $J$ , and  $\mathcal{T} = \bigcup_{J \in \mathcal{D}_k} \mathcal{T}_J$ .
- Finally, introduce a matrix  $\Gamma = (\gamma_{x,T})_{x \in X, T \in \mathcal{T}}$  such that  $\gamma_{x,T} = 1$  if  $\tilde{r}_{i,J} = 1$  for  $x \in X_i$  and  $T \in \mathcal{T}_J$ , and  $\gamma_{x,T} = 0$  otherwise. Intuitively agent  $x$  is **fit** to perform task  $T$  if and only if patient  $i$  associated with agent  $x$  has a link with the odd component  $J$  that is associated with task  $T$  in the induced two-sided matching market  $(N^O, \mathcal{D}, \tilde{R})$ .

---

<sup>24</sup>This can be interpreted as the proof of a continuous version of Hall’s theorem that deals with probabilistic assignments.

Given  $N_k^O$  and  $\mathcal{D}_k$  we refer to  $(X, \mathcal{T}, \Gamma)$  as the auxiliary task assignment problem. Note that

$$\begin{aligned}
|\mathcal{T}| &= \sum_{J \in \mathcal{D}_k} \left( \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|) \right) \\
&= |\mathcal{D}_k| \cdot \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|) \\
&= \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot |C(\mathcal{D}_k, N_k^O)| = \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot |N_k^O| = \sum_{i \in N_k^O} \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| = |X|.
\end{aligned}$$

An **auxiliary task assignment** is a bijection  $\nu : X \rightarrow \mathcal{T}$ . An auxiliary task assignment  $\nu$  is **feasible** if and only if  $\nu(x) = T$  implies that  $\gamma_{x,T} = 1$ .

Here is the point of introducing the above auxiliary task assignment problem: Each agent or task in  $X \cup \mathcal{T}$  corresponds to a fraction  $\frac{1}{|\bigcup_{J' \in \mathcal{D}_k} J'|}$  of its ‘‘owner’’ in  $N_k^O \cup \mathcal{D}_k$ . Therefore if we show that there exists a feasible auxiliary task assignment  $\nu$ , this would mean that it is possible to assign each agent in  $X$  to a distinct compatible task in  $\mathcal{T}$ , and therefore the aggregate  $|N_k^O| \cdot \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot \frac{1}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = |N_k^O|$  weight of the patients in  $N_k^O$  can be allocated among odd components of  $\mathcal{D}_k$  such that the share of the odd component  $J$  is  $(\left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|)) \cdot \frac{1}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \alpha_J$ .

We next prove that there exists a feasible auxiliary task assignment  $\nu$  for the above task assignment problem. Given  $\tau \subseteq \mathcal{T}$  define

$$C(\tau, X) = \{x \in X : \exists T \in \tau \text{ with } \gamma_{x,T} = 1\}$$

That is,  $C(\tau, X)$  is the set of agents each of whom is fit to perform at least one of the tasks in  $\tau$ . By Hall’s Theorem there exists a feasible auxiliary task assignment if and only if

$$|\tau| \leq |C(\tau, X)| \quad \text{for every } \tau \subseteq \mathcal{T}.$$

That is, no matter what subset of tasks is considered, the number of agents who are fit for at least one of these tasks should be no less than the number of the tasks in this subset. We will prove this by contradiction.

Suppose there exists a subset  $\tau \subseteq \mathcal{T}$  of tasks such that  $|\tau| > |C(\tau, X)|$ . Next construct the following set of tasks  $\tau^* \supseteq \tau$ . For any task  $T \in \tau$ , include all tasks which are identical to task  $T$  in set  $\tau^*$ . Note that since  $C(\tau^*, X) = C(\tau, X)$ , we have

$$|\tau^*| \geq |\tau| > |C(\tau, X)| = |C(\tau^*, X)| \tag{2}$$

Let  $\mathcal{J}^* \subseteq \mathcal{D}_k$  be the set of odd components each of which is associated with a task in  $\tau^*$ . Note

that  $\bigcup_{J \in \mathcal{J}^*} \mathcal{T}_J = \tau^*$  and therefore

$$\begin{aligned} |\tau^*| &= \sum_{J \in \mathcal{J}^*} \left( \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|) \right) \\ &= |\mathcal{J}^*| \cdot \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - \left| \bigcup_{J' \in \mathcal{J}^*} J' \right| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|) \end{aligned} \quad (3)$$

Moreover

$$|C(\tau^*, X)| = \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot |C(\mathcal{J}^*, N_k^O)| \quad (4)$$

By Eq (2), Eq (3), and Eq (4)

$$|\mathcal{J}^*| \cdot \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| - \left| \bigcup_{J' \in \mathcal{J}^*} J' \right| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|) > \left| \bigcup_{J' \in \mathcal{D}_k} J' \right| \cdot |C(\mathcal{J}^*, N_k^O)|;$$

rearranging the terms, we have

$$\underbrace{-\frac{|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|}{\left| \bigcup_{J' \in \mathcal{D}_k} J' \right|}}_{=f(\mathcal{D}_k, N_k^O)-1} > \underbrace{-\frac{|\mathcal{J}^*| - |C(\mathcal{J}^*, N_k^O)|}{\left| \bigcup_{J' \in \mathcal{J}^*} J' \right|}}_{=f(\mathcal{J}^*, N_k^O)-1}$$

and therefore,

$$f(\mathcal{D}_k, N_k^O) > f(\mathcal{J}^*, N_k^O). \quad (5)$$

However  $f(\mathcal{D}_k, N_k^O) = f(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O)$  and  $f(\mathcal{J}^*, N_k^O) = f(\mathcal{J}^*, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O)$  by Lemma 6, and this together with Eq (5) imply that

$$f\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right) > f\left(\mathcal{J}^*, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right)$$

contradicting the definition of  $\mathcal{D}_k$  and showing that for each  $\tau \subseteq \mathcal{T}$  we have  $|\tau| \leq |C(\tau, X)|$ . Therefore, there exists a feasible auxiliary task assignment  $\nu$  by Hall's Theorem.

We next construct matrix  $\tilde{A}^{k,k} = [\tilde{a}_{i,J}]_{i \in N_k^O, J \in \mathcal{D}_k}$  using the auxiliary task assignment  $\nu$ . For each  $J \in \mathcal{D}_k$  and  $i \in N_k^O$ , define

$$\nu_{i,J} = \{x \in X_i : \nu(x) = T \text{ for some } T \in \mathcal{T}_J\}.$$

By definition  $|\nu_{i,J}|$  is the total number of tasks associated with odd component  $J$  each of which is assigned to an agent associated with the overdemanded patient  $i$ . For each  $J \in \mathcal{D}_k$  and  $i \in N_k^O$ , let

$$\tilde{a}_{i,J} = \frac{|\nu_{i,J}|}{\left| \bigcup_{J' \in \mathcal{D}_k} J' \right|}$$

and let  $\tilde{A}^{k,k} = [\tilde{a}_{i,J}]_{i \in N_k^O, J \in \mathcal{D}_k}$ . For each odd component  $J \in \mathcal{D}_k$ , we have

$$\sum_{i \in N_k^O} \tilde{a}_{i,J} = \frac{\sum_{i \in N_k^O} |\nu_{i,J}|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \frac{|\mathcal{T}_J|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \frac{|\bigcup_{J' \in \mathcal{D}_k} J'| - |J| \cdot (|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|)}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \alpha_J$$

where the last equality holds by Eq (1). Moreover for each overdemanded patient  $i \in N_k^O$ , we have

$$\sum_{J \in \mathcal{D}_k} \tilde{a}_{i,J} = \frac{\sum_{J \in \mathcal{D}_k} |\nu_{i,J}|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \frac{|X_i|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = \frac{|\bigcup_{J' \in \mathcal{D}_k} J'|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = 1. \quad (6)$$

We conclude the proof of Claim 1 by constructing a pre-allocation matrix  $\tilde{A} \in \tilde{\mathcal{A}}$  using the matrices  $\{\tilde{A}^{k,k}\}_{k \in \{1,2,\dots,q\}}$  constructed above.

For each  $k, k'$  with  $k \neq k'$ , for each  $i \in N_k^O$  and each  $J \in \mathcal{D}_{k'}$  let  $\tilde{a}_{i,J} = 0$ . Let  $\tilde{A}^{k,k'} = [\tilde{a}_{i,J}]_{i \in N_k^O, J \in \mathcal{D}_{k'}}$ .

$$\text{Let } \tilde{A} = \left[ \tilde{A}^{k,k'} \right]_{k \in \{1,2,\dots,q\}, k' \in \{1,2,\dots,q\}} = [\tilde{a}_{i,J}]_{i \in N^O, J \in \mathcal{D}}.$$

For each  $k$  and each odd component  $J \in \mathcal{D}_k$ , we have  $\sum_{i \in N^O} \tilde{a}_{i,J} = \sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$  by Eq (??) and for each overdemanded patient  $i \in N_k^O$ , we have  $\sum_{J \in \mathcal{D}} \tilde{a}_{i,J} = \sum_{J \in \mathcal{D}_k} \tilde{a}_{i,J} = 1$  by Eq (6) concluding the proof of Claim 1.  $\square$

The next claim completes the proof of Theorem 2.

*Claim 2:* There exists an ex-post efficient lottery  $\lambda^E \in \mathcal{L}$  such that  $u(\lambda^E) = u^E$ .

*Proof of Claim 2:* By Claim 1 there exists a pre-allocation  $\tilde{A} \in \tilde{\mathcal{A}}$  such that

1. For each  $i \in N^O$ ,  $\sum_{J \in \mathcal{D}} \tilde{a}_{i,J} = 1$ , and
2. For each  $k \in \{1, 2, \dots, q\}$  and  $J \in \mathcal{D}_k$ 
  - (a)  $\tilde{a}_{i,J} = 0$  for all  $i \in N^O \setminus N_k^O$ , and
  - (b)  $\sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$ .

For each  $k \in \{1, 2, \dots, q\}$  and  $J \in \mathcal{D}_k$ , we have  $\sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$  and  $\tilde{a}_{i,J} = 0$  for all  $i \in N \setminus N_k^O$ . By Lemma 2.1 in Bogomolnaia and Moulin (2002) there exists an ex-post efficient pre-lottery  $\tilde{\lambda} \in \tilde{\mathcal{L}}$  that implements  $\tilde{A}$ . We will “build on” the pre-lottery  $\tilde{\lambda}$  to construct the lottery  $\lambda^E$ . For each pre-matching  $\tilde{\mu} \in \tilde{\mathcal{M}}$  in the support of pre-lottery  $\tilde{\lambda}$ , partition set  $\mathcal{D}$  as  $\{\mathcal{D}^m(\tilde{\mu}), \mathcal{D}^u(\tilde{\mu})\}$  where

- $\mathcal{D}^m(\tilde{\mu}) = \{J \in \mathcal{D} : \tilde{\mu}(J) \neq \emptyset\}$  is the set of matched odd components, and
- $\mathcal{D}^u(\tilde{\mu}) = \mathcal{D} \setminus \mathcal{D}^m(\tilde{\mu})$  is the set of unmatched odd components.

For each pre-matching  $\tilde{\mu} \in \tilde{\mathcal{M}}$  in the support of pre-lottery  $\tilde{\lambda}$  construct  $\prod_{J \in \mathcal{D}^u(\tilde{\mu})} |J|$  distinct matchings as follows:

Pick one patient from each  $J \in \mathcal{D}^u(\tilde{\mu})$ . Note that there are  $\prod_{J \in \mathcal{D}^u(\tilde{\mu})} |J|$  possible combinations. For each combination construct a Pareto-efficient matching  $\mu$  such that:

- each of the chosen patients is matched to herself,
- each remaining patient in each odd component  $J \in \mathcal{D}^u(\tilde{\mu})$  is matched with another patient in the same odd component  $J$ , and
- one patient in each odd component  $J \in \mathcal{D}^m(\tilde{\mu})$  is matched with an overdemanded patient  $i \in N^O$  whereas all other patients in each such odd component  $J$  is matched with another patient in  $J$ .

By the GED Lemma, there exists at least one such matching. Pick one and only one such matching for each of the  $\prod_{J \in \mathcal{D}^u(\tilde{\mu})} |J|$  possible combinations. Let  $\mathcal{M}(\tilde{\mu})$  be the resulting set of matchings. Clearly  $|\mathcal{M}(\tilde{\mu})| = \prod_{J \in \mathcal{D}^u(\tilde{\mu})} |J|$ .

We are finally ready to construct a lottery  $\lambda^E$  which induces the utility profile  $u^E$ . The lottery  $\lambda^E$  is constructed from the pre-lottery  $\tilde{\lambda}$  by simply replacing each pre-matching  $\tilde{\mu}$  in the support of  $\tilde{\lambda}$  with the uniform lottery over  $\mathcal{M}(\tilde{\mu})$ . That is:

$$\lambda_\mu^E = \begin{cases} \frac{\tilde{\lambda}_{\tilde{\mu}}}{|\mathcal{M}(\tilde{\mu})|} & \text{if } \mu \in \mathcal{M}(\tilde{\mu}) \text{ and } \tilde{\lambda}_{\tilde{\mu}} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Clearly,  $\lambda^E$  is a lottery:

$$\sum_{\mu \in \mathcal{M}} \lambda_\mu^E = \sum_{\tilde{\mu} \in \tilde{\mathcal{M}}} \left( \sum_{\mu \in \mathcal{M}(\tilde{\mu})} \lambda_\mu^E \right) = \sum_{\tilde{\mu} \in \tilde{\mathcal{M}}} \left( \sum_{\mu \in \mathcal{M}(\tilde{\mu})} \frac{\tilde{\lambda}_{\tilde{\mu}}}{|\mathcal{M}(\tilde{\mu})|} \right) = \sum_{\tilde{\mu} \in \tilde{\mathcal{M}}} \tilde{\lambda}_{\tilde{\mu}} = 1$$

Moreover, by construction  $\lambda^E$  is an ex-post efficient lottery.

We conclude the proof of Claim 2 and Theorem 2 by showing that  $u(\lambda^E) = u^E$ :

Each patient in  $N \setminus N^U$  is matched with another patient in every efficient matching by the GED Lemma. Since  $\lambda^E$  is ex-post efficient, for each patient  $i \in N \setminus N^U$  we have  $u_i(\lambda^E) = u_i^E = 1$ .

Consider a patient  $i \in N^U$ . Let  $i \in J \in \mathcal{D}_k$  for some  $k$ .

Let  $\tilde{\mu} \in \tilde{\mathcal{M}}$  be a pre-matching with  $\tilde{\lambda}_{\tilde{\mu}} > 0$ .

1. If  $J \in \mathcal{D}^m(\tilde{\mu})$  then all patients in  $J$  are matched under every matching  $\mu \in \mathcal{M}(\tilde{\mu})$ .
2. If  $J \in \mathcal{D}^u(\tilde{\mu})$  then  $|J| - 1$  patients in  $J$  are matched under every matching  $\mu \in \mathcal{M}(\tilde{\mu})$  and patient  $i$  (just as any other patient in  $J$ ) is matched with another patient in  $\frac{|J|-1}{|J|} |\mathcal{M}(\tilde{\mu})|$  of these matchings.

Since  $\sum_{i \in N_k^O} \tilde{a}_{i,J} = \alpha_J$  is the probability that the odd component  $J$  is assigned a patient in  $N_k^O$  under the pre-lottery  $\tilde{\lambda}$ , we have

$$\sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^m(\tilde{\mu})} \tilde{\lambda}_{\tilde{\mu}} = \alpha_J \quad \text{and} \quad \sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^u(\tilde{\mu})} \tilde{\lambda}_{\tilde{\mu}} = 1 - \alpha_J \quad (7)$$

Therefore

$$\begin{aligned} u_i(\lambda^E) &= \sum_{\mu \in \mathcal{M} \text{ s.t. } \mu(i) \neq i} \lambda_{\mu}^E \\ &= \sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^m(\tilde{\mu})} \left( \sum_{\mu \in \mathcal{M}(\tilde{\mu})} \frac{\tilde{\lambda}_{\tilde{\mu}}}{|\mathcal{M}(\tilde{\mu})|} \right) + \sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^u(\tilde{\mu})} \left( \sum_{\mu \in \mathcal{M}(\tilde{\mu}) \text{ s.t. } \mu(i) \neq i} \frac{\tilde{\lambda}_{\tilde{\mu}}}{|\mathcal{M}(\tilde{\mu})|} \right) \\ &= \sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^m(\tilde{\mu})} \tilde{\lambda}_{\tilde{\mu}} + \sum_{\tilde{\mu} \in \tilde{\mathcal{M}} \text{ s.t. } J \in \mathcal{D}^u(\tilde{\mu})} \frac{|J| - 1}{|J|} \tilde{\lambda}_{\tilde{\mu}} \\ &= \alpha_J + (1 - \alpha_J) \cdot \frac{|J| - 1}{|J|} = \frac{|J| - 1 + \alpha_J}{|J|} = 1 - \frac{|\mathcal{D}_k| - |C(\mathcal{D}_k, N_k^O)|}{|\bigcup_{J' \in \mathcal{D}_k} J'|} = f(\mathcal{D}_k, N_k^O) = u_i^E \end{aligned}$$

Here the fourth equality follows from Eq (7) and the sixth equality follows from Eq (1). This completes the proof of Claim 2 as well as the proof of Theorem 2.  $\square \blacksquare$

*Proof of Theorem 3:* Let  $J_k = \bigcup_{J \in \mathcal{D}_k} J$  for any  $k \in \{1, 2, \dots, q\}$ .

*Claim 1:*  $f(\mathcal{D}_k, N_k^O) < f(\mathcal{D}_{k+1}, N_{k+1}^O)$  for each  $k \in \{1, 2, \dots, q-1\}$ .

*Proof of Claim 1:* Pick  $k \in \{1, 2, \dots, q-1\}$ . Let  $I = N^O \setminus \bigcup_{\ell=1}^{k-1} N_{\ell}^O$ . Consider the construction of  $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_q\}$ . Note that  $\mathcal{D}_k \cup \mathcal{D}_{k+1} \subseteq \mathcal{D} \setminus \bigcup_{\ell=1}^{k-1} \mathcal{D}_{\ell}$ . Since

$$f(\mathcal{D}_k, I) = \min_{\mathcal{J} \subseteq \mathcal{D} \setminus \bigcup_{\ell=1}^{k-1} \mathcal{D}_{\ell}} f(\mathcal{J}, I)$$

and  $\mathcal{D}_k$  is the largest subset  $\mathcal{J} \subseteq \mathcal{D} \setminus \bigcup_{\ell=1}^{k-1} \mathcal{D}_\ell$  satisfying this equality, we have

$$\begin{aligned}
f(\mathcal{D}_k, I) &< f(\mathcal{D}_k \cup \mathcal{D}_{k+1}, I) \\
&= \frac{|J_k \cup J_{k+1}| - (|\mathcal{D}_k \cup \mathcal{D}_{k+1}| - |C(\mathcal{D}_k \cup \mathcal{D}_{k+1}, I)|)}{|J_k \cup J_{k+1}|} \\
&= \frac{|J_k| + |J_{k+1}| - (|\mathcal{D}_k| + |\mathcal{D}_{k+1}| - |C(\mathcal{D}_k \cup \mathcal{D}_{k+1}, I)|)}{|J_k| + |J_{k+1}|} && \text{by } \mathcal{D}_k \cap \mathcal{D}_{k+1} = \emptyset \\
&= \frac{|J_k| + |J_{k+1}| - (|\mathcal{D}_k| + |\mathcal{D}_{k+1}| - |C(\mathcal{D}_k, I)| - |C(\mathcal{D}_{k+1}, I \setminus N_k^O)|)}{|J_k| + |J_{k+1}|} \\
&&& \text{by } C(\mathcal{D}_k \cup \mathcal{D}_{k+1}, I) = C(\mathcal{D}_k, I) \cup C(\mathcal{D}_{k+1}, I \setminus N_k^O) \\
&&& \text{and } C(\mathcal{D}_k, I) \cap C(\mathcal{D}_{k+1}, I \setminus N_k^O) = \emptyset \\
&= \frac{\frac{|J_k| - (|\mathcal{D}_k| - |C(\mathcal{D}_k, I)|)}{|J_k|} |J_k| + \frac{|J_{k+1}| - (|\mathcal{D}_{k+1}| - |C(\mathcal{D}_{k+1}, I \setminus N_k^O)|)}{|J_{k+1}|} |J_{k+1}|}{|J_k| + |J_{k+1}|} \\
&= \frac{f(\mathcal{D}_k, I) |J_k| + f(\mathcal{D}_{k+1}, I \setminus N_k^O) |J_{k+1}|}{|J_k| + |J_{k+1}|} && \text{by definition of } f
\end{aligned}$$

Rearranging the terms in this inequality, we find  $f(\mathcal{D}_k, I) < f(\mathcal{D}_{k+1}, I \setminus N_k^O)$ . We conclude the proof of Claim 1 observing that  $f(\mathcal{D}_k, N_k^O) = f(\mathcal{D}_k, I)$  and  $f(\mathcal{D}_{k+1}, N_{k+1}^O) = f(\mathcal{D}_{k+1}, I \setminus N_k^O)$  by Lemma 6 (i).  $\square$

By Claim 1, each patient in  $J_1$  has the lowest utility under  $u^E$ , each patient in  $J_2$  has the lowest utility among the remaining patients under  $u^E$ , and so on. Since patients can only be matched with their neighbors, patients in  $J_1$  can only be matched with patients in  $N_1^O = C(\mathcal{D}_1, N^O)$  and since  $C(\mathcal{D}_1, N^O) = C(\mathcal{D}_1, N_1^O)$  by Lemma 6 (i), at least  $(|\mathcal{D}_1| - |C(\mathcal{D}_1, N_1^O)|)$  of patients in  $J_1$  remain unmatched at any matching. Equivalently the aggregate utility of patients in  $J_1$  cannot be more than  $|J_1| - (|\mathcal{D}_1| - |C(\mathcal{D}_1, N_1^O)|) = |J_1| \times f(\mathcal{D}_1, N_1^O)$ . So for any feasible utility profile,  $f(\mathcal{D}_1, N_1^O)$  (which is the lowest utility under  $u^E$ ) is the highest utility that can be received by the lowest utility patient in  $J_1$ ;  $2f(\mathcal{D}_1, N_1^O)$  is the highest sum of utilities that can be received by the lowest utility 2 patient in  $J_1$ ;  $\dots$ , and  $|J_1| \times f(\mathcal{D}_1, N_1^O)$  is the highest sum of utilities that can be received by patients in  $J_1$ . Moreover these upperbounds can only be reached, if all the overdemanded patients in  $C(\mathcal{D}_1, N_1^O)$  are committed to patients in  $J_1$ . Repeating these arguments for patients in  $J_2$ ,  $|J_1| \times f(\mathcal{D}_1, N_1^O) + f(\mathcal{D}_2, N_2^O)$  is the highest sum of utilities that can be received by the lowest utility  $|J_1|+1$  patient in  $J_1 \cup J_2$ ;  $|J_1| \times f(\mathcal{D}_1, N_1^O) + 2f(\mathcal{D}_2, N_2^O)$  is the highest sum of utilities that can be received by the lowest utility  $|J_1|+2$  patient in  $J_1 \cup J_2$ ;  $\dots$ , and proceeding in a similar way with patients in  $J_3, \dots, J_q$ , we show that the utility profile  $u^E$  Lorenz dominates any other feasible utility profile.  $\blacksquare$

The next Lemma will be useful in proving Theorem 4.



**Lemma 7** *Let  $u, v \in \mathcal{U}$  be such that  $u$  Lorenz-dominates  $v$ . Then for any  $\alpha \in (0, 1)$ , vector  $\alpha u + (1 - \alpha)v$  Lorenz-dominates  $v$ .*

*Proof of Lemma 7:* Let  $u, v \in \mathcal{U}$  be such that  $u$  Lorenz-dominates  $v$ , let  $\alpha \in (0, 1)$  and  $w = \alpha u + (1 - \alpha)v$ . Since  $u \in \mathcal{U}$  there is a lottery  $\lambda \in \mathcal{L}$  that induces  $u$ , and since  $v \in \mathcal{U}$  there is a lottery  $\gamma \in \mathcal{L}$  that induces  $v$ . Let  $\sigma = \alpha\lambda + (1 - \alpha)\gamma$ . For each  $i \in N$ , we have

$$u_i(\sigma) = \alpha u_i(\lambda) + (1 - \alpha)u_i(\gamma) = \alpha u_i + (1 - \alpha)v_i = w_i.$$

This implies that  $\sigma$  induces  $w = (w_i)_{i \in N} \in \mathcal{U}$ .

Next, we show that  $w$  Lorenz-dominates  $v$ . By the definition of order statistics, for any  $t \in \{1, 2, \dots, n\}$  and any  $t$  member subset  $N' \subseteq N$  of patients we have

$$\sum_{s=1}^t v^{(s)} \leq \sum_{i \in N'} v_i. \quad (8)$$

Moreover, since  $u$  Lorenz-dominates  $v$ ,  $\sum_{s=1}^t v^{(s)} \leq \sum_{s=1}^t u^{(s)}$  for any  $t \in \{1, 2, \dots, n\}$  and this inequality holds strictly for some  $t \in \{1, 2, \dots, n\}$ . This together with the definition of order statistics imply that for any  $t \in \{1, 2, \dots, n\}$  and any  $t$  member subset  $N'' \subseteq N$  of patients we have

$$\sum_{s=1}^t v^{(s)} \leq \sum_{s=1}^t u^{(s)} \leq \sum_{i \in N''} u_i \quad (9)$$

where the second inequality holds strictly for some  $t \in \{1, 2, \dots, n\}$ . We have  $\sum_{s=1}^t w^{(s)} = \alpha \left( \sum_{i \in N'} v_i \right) + (1 - \alpha) \left( \sum_{i \in N''} u_i \right)$  for some  $N', N'' \subseteq N$  with  $|N'| = |N''| = t$ . Therefore, Eq (8) and Eq (9) imply that

$$\sum_{s=1}^t w^{(s)} = \alpha \left( \sum_{s=1}^t v^{(s)} \right) + (1 - \alpha) \left( \sum_{s=1}^t v^{(s)} \right) \leq \alpha \left( \sum_{i \in N'} v_i \right) + (1 - \alpha) \left( \sum_{i \in N''} u_i \right) = \sum_{s=1}^t w^{(s)}.$$

where the inequality holds strictly for some  $t \in \{1, 2, \dots, n\}$ , completing the proof of Lemma 7. ■

*Proof of Theorem 4:* First we introduce some additional notation. Fix the set of agents  $N$  and hence each mutual compatibility matrix  $R$  defines a distinct reduced problem. For any reduced problem  $R$ , let

$$J_k(R) = \bigcup_{J \in \mathcal{D}_k(R)} J \quad \text{and} \quad e(R) = \max_{\mu \in \mathcal{M}(R)} |\mu|.$$

Recall that for any  $\mu \in \mathcal{M}(R)$ , we have  $\mu \in \mathcal{E}(R) \iff |\mu| = e(R)$ . For any reduced problem  $R$  and any two sets  $I, J \subseteq N$ , define **neighbors** of  $J$  among  $I$  as

$$C(J, I, R) = \{i \in I \setminus J : r_{i,j} = 1 \text{ for some } j \in J\}.$$

For a singleton set  $J = \{j\}$ , we slightly abuse the notation and use  $C(j, I, R)$  instead of  $C(\{j\}, I, R)$ .

Let  $\phi^E$  denote an egalitarian mechanism and  $R = [r_{i,h}]_{i \in N, h \in N}$  be a reduced problem. Construct the Gallai-Edmonds Decomposition  $\{N^U(R), N^O(R), N^P(R)\}$  of the set of patients  $N$ , the partition  $\mathcal{D}(R)$  of the set of underdemanded patients  $N^U(R)$ , the partition  $\{\mathcal{D}_1(R), \dots, \mathcal{D}_{q(R)}(R)\}$  of  $\mathcal{D}(R)$  and the partition  $\{N_1^O(R), \dots, N_{q(R)}^O(R)\}$  of  $N^O(R)$ , and the egalitarian utility profile  $u^E(R)$ . Note that  $u(\phi^E(R)) = u^E(R)$ .

For any patient  $j \in (N^O(R) \cup N^P(R))$  we have  $u_j^E(R) = 1$  and therefore no such patient can benefit by underreporting her set of compatible patients. Let  $j \in N^U(R)$  be such that  $j \in J \in \mathcal{D}(R)$ . Note that  $u_j^E(R) < 1$ . We will prove that patient  $j$  cannot increase her utility by declaring a mutually compatible patient to be incompatible, and repeated application of this argument will conclude the proof.

Let  $j' \in C(j, N, R)$ . Either  $j' \in J$  or  $j' \in N^O(R)$ . Let  $Q$  be the reduced problem obtained from  $R$  by patient  $j$  declaring patient  $j'$  to be incompatible. Note that  $C(j, N, Q) = C(j, N, R) \setminus \{j'\}$ ,  $C(j', N, Q) = C(j', N, R) \setminus \{j\}$ ,  $C(i, N, Q) = C(i, N, R)$  for all  $i \in N \setminus \{j, j'\}$ , and  $\mathcal{M}(Q) = \{\mu \in \mathcal{M}(R) : \mu(j) \neq j'\}$ . Construct the Gallai-Edmonds Decomposition  $\{N^U(Q), N^O(Q), N^P(Q)\}$  of the set of patients  $N$ , the partition  $\mathcal{D}(Q)$  of the set of underdemanded patients  $N^U(Q)$ , the partition  $\{\mathcal{D}_1(Q), \dots, \mathcal{D}_{q(Q)}(Q)\}$  of  $\mathcal{D}(Q)$  and the partition  $\{N_1^O(Q), \dots, N_{q(Q)}^O(Q)\}$  of  $N^O(Q)$ , and the egalitarian utility profile  $u^E(Q)$ . Note that  $u(\phi^E(Q)) = u^E(Q)$ . We will prove three claims that will be useful in our proof.

*Claim 1:*

- (i)  $e(Q) = e(R)$ ,
- (ii)  $\mathcal{E}(Q) \subseteq \mathcal{E}(R)$  and  $\mu \in \mathcal{E}(R) \cap \mathcal{M}(Q) \Rightarrow \mu \in \mathcal{E}(Q)$ .

*Proof of Claim 1:*

- (i) Since  $\mathcal{M}(Q) \subseteq \mathcal{M}(R)$  we have  $e(Q) \leq e(R)$ . Since  $j \in N^U(R)$ , there exists a Pareto-efficient matching  $\mu \in \mathcal{E}(R)$  such that  $\mu(j) = j$ . We have  $|\mu| = e(R)$  and  $\mu \in \mathcal{M}(Q)$  which implies  $e(Q) \geq |\mu| = e(R)$ . Therefore  $e(Q) = e(R)$ .
- (ii) First let  $\mu \in \mathcal{E}(Q)$ . We have  $\mu \in \mathcal{M}(Q) \subseteq \mathcal{M}(R)$ . Moreover  $|\mu| = e(Q)$  and  $e(Q) = e(R)$  by Claim 1 (i) implying that  $|\mu| = e(R)$ . Therefore  $\mu \in \mathcal{E}(R)$ .

Next let  $\mu \in \mathcal{E}(R) \cap \mathcal{M}(Q)$ . Since  $\mu \in \mathcal{E}(R)$ , we have  $|\mu| = e(R) = e(Q)$  and this together with  $\mu \in \mathcal{M}(Q)$  imply  $\mu \in \mathcal{E}(Q)$ .

*Claim 2:*  $N^O(R) \subseteq N^O(Q) \cup N^P(Q)$  and  $N^U(R) \supseteq N^U(Q)$ .

*Proof of Claim 2:* First suppose there exists a patient  $i \in N^O(R) \cap N^U(Q)$ . Then there exists a Pareto-efficient matching  $\mu \in \mathcal{E}(Q)$  such that  $\mu(i) = i$ . By Claim 1 (ii)  $\mathcal{E}(Q) \subseteq \mathcal{E}(R)$ , and therefore  $\mu \in \mathcal{E}(R)$ . This together with  $\mu(i) = i$  imply that  $i \in N^U(R)$  contradicting  $i \in N^O(R)$ . Therefore  $N^O(R) \cap N^U(Q) = \emptyset$  and hence  $N^O(R) \subseteq N^O(Q) \cup N^P(Q)$ .

Next pick a patient  $i \in N^U(Q)$ . There exists a Pareto-efficient matching  $\mu \in \mathcal{E}(Q)$  such that  $\mu(i) = i$ . By Claim 1 (ii),  $\mathcal{E}(Q) \subseteq \mathcal{E}(R)$  and therefore  $\mu \in \mathcal{E}(R)$ . This together with  $\mu(i) = i$  imply that  $i \in N^U(R)$ .  $\square$

*Claim 3:*  $u_{j'}^E(Q) \geq u_j^E(Q)$ .

*Proof of Claim 3:* Recall that patient  $j'$  is the patient who is declared to be incompatible by patient  $j$  under  $Q$ , although they are mutually compatible under  $R$ . Also recall that either  $j' \in N^O(R)$  or  $j' \in J$ . First suppose  $j' \in N^O(R)$ . By Claim 2,  $j' \in N^O(Q) \cup N^P(Q)$  and therefore in this case  $u_{j'}^E(Q) = 1 \geq u_j^E(Q)$ .

Next suppose  $j' \in J$ . Consider the support of lottery  $\phi^E(Q)$ . Since  $\phi^E(Q)$  is ex-post efficient, the support of  $\phi^E(Q)$  is a subset of  $\mathcal{E}(Q)$ . Contrary to Claim 3, suppose  $u_{j'}^E(Q) < u_j^E(Q)$ .

Since  $u_{j'}^E(Q) < 1$ , there exists a Pareto-efficient matching  $\mu \in \mathcal{E}(Q)$  with  $\phi_\mu^E(Q) > 0$  such that  $\mu(j') = j'$ . By Claim 1 (ii)  $\mathcal{E}(Q) \subseteq \mathcal{E}(R)$ , and therefore  $\mu \in \mathcal{E}(R)$ . Since  $J \in \mathcal{D}(R)$ , by the GED Lemma there can be at most one patient in  $J$  who remains unmatched. Therefore since  $\mu(j') = j'$ , we have  $\mu(j) \in J \setminus \{j\}$ . Again by the GED Lemma there exists a Pareto-efficient matching  $\nu \in \mathcal{E}(R)$  which leaves patient  $j$  unmatched instead of patient  $j'$ , but otherwise matches the same patients as in matching  $\mu$ . Since  $\nu(j) = j$  we have  $\nu \in \mathcal{M}(Q)$ . Let  $\varepsilon$  be such that  $0 < \varepsilon \leq \min \left\{ \phi_\mu^E(Q), \frac{u_j^E(Q) - u_{j'}^E(Q)}{2} \right\}$ . Next construct lottery  $\lambda$  from lottery  $\phi^E(Q)$  by subtracting  $\varepsilon\mu$  and adding  $\varepsilon\nu$ . We have

$$u_h(\lambda) = \begin{cases} u_h^E(Q) - \varepsilon & \text{if } h = j \\ u_h^E(Q) + \varepsilon & \text{if } h = j' \\ u_h^E(Q) & \text{otherwise} \end{cases}$$

Since there is ‘‘utility transfer’’ from the higher utility patient  $j$  to the lower utility patient  $j'$ , utility profile  $u(\lambda)$  Lorenz-dominates the egalitarian profile  $u^E(Q)$  under  $Q$  contradicting Theorem 3. Therefore  $u_{j'}^E(Q) \geq u_j^E(Q)$ .  $\square$

We are ready to proceed with the proof of Theorem 4. Suppose  $u_j^E(Q) > u_j^E(R)$ . We will show that this will lead to a contradiction. Let  $J \in \mathcal{D}_k(R)$ . Since  $\phi^E(Q)$  is an ex-post efficient lottery under  $Q$ ,  $\phi^E(R)$  is an ex-post efficient lottery under  $R$ , and  $e(Q) = e(R)$  by Claim 1 (i), we have

$$\sum_{i \in N} u_i^E(Q) = e(Q) = e(R) = \sum_{i \in N} u_i^E(R).$$

Therefore, since  $u_j^E(Q) > u_j^E(R)$ , there exists a patient  $h \in N$  such that  $u_h^E(Q) < u_h^E(R)$ . This implies  $u_h^E(Q) < 1$  which in turn implies  $h \in N^U(Q)$ . Since  $N^U(Q) \subseteq N^U(R)$  by Claim 2,

$h \in N^U(R)$  as well. In a way of speaking, the utility of patient  $j$  increases under  $\phi^E(Q)$  with respect to  $\phi^E(R)$  at the expense of the utility of some other patients each of whom is underdemanded under  $R$ . That is, some utility is transferred from some underdemanded patients under  $R$  to patient  $j$ . We partition  $N^U(R)$  as  $\left\{ \bigcup_{\ell=1}^k J_\ell(R), \bigcup_{\ell=k+1}^{q(R)} J_\ell(R) \right\}$ .

Consider patients in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$ . By the construction of  $u^E(R)$ , at any matching  $\mu \in \mathcal{E}(R)$  in the support of  $\phi^E(R)$ , any patient in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  either remains unmatched or she is matched with another underdemanded patient in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  or is matched with an overdemanding patient in  $\bigcup_{\ell=k+1}^{q(R)} N_\ell^O(R)$ . Since patients in  $\bigcup_{\ell=1}^k J_\ell(R)$  are handled before patients in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  during the construction of  $u^E(R)$ , there is no patient in  $\bigcup_{\ell=1}^k J_\ell(R)$  that is mutually compatible with any patient in  $\bigcup_{\ell=k+1}^{q(R)} N_\ell^O(R)$ . Since any patient in  $\bigcup_{\ell=1}^k J_\ell(R)$  and any patient in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  are in different odd components, there is no patient in  $\bigcup_{\ell=1}^k J_\ell(R)$  that is mutually compatible with a patient in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$ , either. Therefore for any  $i \in \bigcup_{\ell=1}^k J_\ell(R)$ , we have

$$C(i, N, R) \cap \bigcup_{\ell=k+1}^{q(R)} (N_\ell^O(R) \cup J_\ell(R)) = \emptyset.$$

Therefore, patients in  $\bigcup_{\ell=k+1}^{q(R)} N_\ell^O(R)$  shall be committed for patients in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  under the profile  $Q$  as well, and therefore the aggregate utility of patients in  $\bigcup_{\ell=k+1}^{q(R)} J_\ell(R)$  cannot decrease under  $Q$ . Since aggregate utility remains constant at  $e(Q) = e(R)$ , and since only patients in  $N^U(Q) \subseteq N^U(R)$  can have a utility reduction,

$$\exists h \in \bigcup_{\ell=1}^k J_\ell(R) \quad \text{s.t.} \quad u_h^E(Q) < u_h^E(R).$$

Since  $j \in J_k(R)$ , we have  $u_h^E(R) \leq u_j^E(R)$  by Claim 1 in the proof of Theorem 3. This together with  $u_j^E(Q) > u_j^E(R)$  and Claim 3 imply that

$$u_h^E(Q) < u_h^E(R) \leq u_j^E(R) < u_j^E(Q) \leq u_{j'}^E(Q). \quad (10)$$

Let  $\phi = \phi^E(R)$  and  $\varphi = \phi^E(Q)$ . Given  $\phi$ , construct lottery  $\lambda$  as follows: For any matching  $\mu$  in the support of  $\phi$ ,

1. If  $\mu(j) \neq j'$ , then do not alter this ‘‘portion’’ of the lottery (i.e. let  $\lambda_\mu = \phi_\mu$  for any  $\mu$  in the support of  $\phi$  with  $\mu(j) \neq j'$ ).
2. If  $\mu(j) = j'$ , then
  - (a) construct the matching  $\mu_{-j,j'}$  from  $\mu$  by ‘‘breaking’’ the match between  $j$  and  $j'$  (leaving each one unmatched) and preserving the rest of the matching  $\mu$ ; and

- (b) replace matching  $\mu$  with  $\mu_{-j,j'}$  for each such matching in lottery  $\phi$  (i.e. let  $\lambda_{\mu_{-j,j'}} = \lambda_\mu$  for any  $\mu$  in the support of  $\phi$  with  $\mu(j) = j'$ ).

Note that  $\lambda$  is feasible under  $Q$  and  $u_i(\lambda) = u_i(\phi) = u_i^E(R)$  for all  $i \in N \setminus \{j, j'\}$ . Given  $\varepsilon \in (0, 1)$ , let

$$\gamma^\varepsilon = \varepsilon\phi + (1 - \varepsilon)\varphi \quad \text{and} \quad \lambda^\varepsilon = \varepsilon\lambda + (1 - \varepsilon)\varphi.$$

Note that  $u_i(\lambda^\varepsilon) = u_i(\gamma^\varepsilon)$  for all  $i \in N \setminus \{j, j'\}$  by construction of  $\lambda$  from  $\phi$ .

Since  $\varphi$  is feasible under  $R$ ,  $\varphi \in \mathcal{L}(R)$ . Therefore since  $\phi$  is a Lorenz-dominant lottery for the reduced problem  $R$  by Theorem 3,  $\phi$  Lorenz-dominates  $\varphi$  which in turn implies  $\gamma^\varepsilon$  Lorenz-dominates  $\varphi$  by Lemma 7. Pick  $\varepsilon \in (0, 1)$  small enough such that  $u_j(\lambda^\varepsilon) > u_h(\lambda^\varepsilon) = u_h(\gamma^\varepsilon)$  and  $u_{j'}(\lambda^\varepsilon) > u_h(\lambda^\varepsilon) = u_h(\gamma^\varepsilon)$ . This can be done by Eq (10). Let  $h$  be the patient with  $s_1^{\text{th}}$  lowest utility under  $\varphi$  and  $s_2^{\text{th}}$  lowest utility under  $\gamma^\varepsilon$ . Let  $s = \min\{s_1, s_2\}$ . Since  $\gamma^\varepsilon$  Lorenz-dominates  $\varphi$  by Lemma 7,

$$\sum_{\ell=1}^t (u(\gamma^\varepsilon))^{(\ell)} \geq \sum_{\ell=1}^t (u(\varphi))^{(\ell)} \quad \text{for all } t \leq n.$$

Since  $\varepsilon \in (0, 1)$  is such that  $u_j(\gamma^\varepsilon) > u_j(\lambda^\varepsilon) > u_h(\gamma^\varepsilon)$  and  $u_{j'}(\gamma^\varepsilon) > u_{j'}(\lambda^\varepsilon) > u_h(\gamma^\varepsilon)$ , neither  $j$  nor  $j'$  is one of the  $s$  lowest utility agents under  $\gamma^\varepsilon$ . Therefore, since only patients  $j, j'$  are affected between lotteries  $\gamma^\varepsilon$  and  $\lambda^\varepsilon$ ,

$$\sum_{\ell=1}^t (u(\lambda^\varepsilon))^{(\ell)} \geq \sum_{\ell=1}^t (u(\varphi))^{(\ell)} \quad \text{for all } t \leq s. \quad (11)$$

We will show that Inequality (11) holds strictly for some  $t \leq s$  which in turn will contradict  $\varphi$  is Lorenz-dominant under  $Q$  completing the proof. Suppose not. Then Inequality (11) holds with equality which in turn implies

$$(u(\lambda^\varepsilon))^{(t)} = (u(\varphi))^{(t)} \quad \text{for all } t \leq s. \quad (12)$$

Observe that there is  $t \leq s$  such that the  $t^{\text{th}}$  smallest utility patient is different under these two lotteries. In particular this is the case for  $t = s$ , because (1)  $u_h(\lambda^\varepsilon) = u_h(\gamma^\varepsilon) > u_h(\varphi)$  and (2)  $h$  is the  $s^{\text{th}}$  smallest utility agent under one of the two lotteries  $\lambda^\varepsilon, \varphi$  although not in both since that would contradict Eq (12). Pick the smallest such  $t$ . That is under  $\lambda^\varepsilon$  and  $\varphi$ , not only the  $\hat{t}^{\text{th}}$  smallest utility patient is the same patient for any  $\hat{t} < t$ , but she also has the same utility. On the other hand the  $t^{\text{th}}$  lowest utility patient differs under the two lotteries although they have the same utility by Eq (12).

Now consider the lottery  $\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi$ . Since  $\lambda^\varepsilon$  and  $\varphi$  are both feasible under  $Q$ , the lottery  $\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi$  is feasible under  $Q$  as well. Under this lottery

- the smallest utility patient is the same patient as in both  $\lambda^\varepsilon, \varphi$  and she has the same utility  $(u(\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi))^{(1)} = (u(\lambda^\varepsilon))^{(1)} = (u(\varphi))^{(1)}$ ;
- $\vdots$   $\vdots$
- the  $(t-1)^{\text{th}}$  smallest utility patient is the same patient as in both  $\lambda^\varepsilon, \varphi$  and she has the same utility  $(u(\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi))^{(t-1)} = (u(\lambda^\varepsilon))^{(t-1)} = (u(\varphi))^{(t-1)}$ .

But the utility of the  $t^{\text{th}}$  smallest utility patient under  $\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi$  is strictly larger than the utility of the  $t^{\text{th}}$  smallest utility patient under  $\varphi$ . That is because, whoever she is, her utility is no less than  $(u(\varphi))^{(t)}$  under both  $\lambda^\varepsilon, \varphi$  by Eq (12) and strictly larger in at least one, since the  $t^{\text{th}}$  smallest utility patient differs under  $\lambda^\varepsilon$  and  $\varphi$ . Hence  $\varphi$  does not Lorenz-dominate  $\frac{1}{2}\lambda^\varepsilon + \frac{1}{2}\varphi$ , a feasible lottery under  $Q$ , a contradiction. Therefore Inequality (11) holds strictly for some  $t \leq s$ . But then  $\varphi$  does not Lorenz-dominate  $\lambda^\varepsilon$ , a feasible lottery under  $Q$ , leading to another contradiction and completing the proof. ■

## References

- [1] A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth (2005), “The New York City High School Match,” *American Economic Review*, Papers and Proceedings, May, forthcoming.
- [2] A. Abdulkadiroğlu, P. A. Pathak, A. E. Roth, and T. Sönmez (2005) “The Boston Public School Match,” *American Economic Review*, Papers and Proceedings, May, forthcoming.
- [3] A. Abdulkadiroğlu and T. Sönmez (1998) “Random serial dictatorship and the core from random endowments in house allocation problems.” *Econometrica* 66: 689-701.
- [4] A. Abdulkadiroğlu and T. Sönmez (1999) “House allocation with existing tenants.” *Journal of Economic Theory* 88: 233-260.
- [5] A. Abdulkadiroğlu and T. Sönmez (2003a) “Ordinal efficiency and dominated sets of assignments.” *Journal of Economic Theory* 112: 157-172.
- [6] A. Abdulkadiroğlu and T. Sönmez (2003b) “School choice: A mechanism design approach.” *American Economic Review* 93: 729-747.
- [7] M. Abecassis et al (2002) “Consensus statement on the live organ donor.” *The Journal of the American Medical Association* 284: 2919-2926.

- [8] H. Abeledo and U.G. Rothblum (1994) "Stable matchings and linear inequalities." *Discrete Applied Mathematics* 54: 1-27.
- [9] A. Bogomolnaia, L. Ehlers, and R. Deb (2002) "Incentive-compatible assignment on the full preference domain." *Journal of Economic Theory* forthcoming.
- [10] A. Bogomolnaia and H. Moulin (2001) "A new solution to the random assignment problem." *Journal of Economic Theory* 100: 295-328.
- [11] A. Bogomolnaia and H. Moulin (2002) "A simple random assignment problem with a unique solution." *Economic Theory* 19: 623-635.
- [12] A. Bogomolnaia and H. Moulin (2004) "Random matching under dichotomous preferences." *Econometrica* 72: 257-279.
- [13] K.-S. Chung (2000) "On the existence of stable roommate matchings." *Games and Economic Behavior* 33: 206-230.
- [14] H. Cres and H. Moulin (2001) "Scheduling with opting out: improving upon random priority." *Operations Research* 49: 565-577.
- [15] V.P. Crawford (2005) "The flexible-salary match: a proposal to increase the salary flexibility of the National Resident Matching Program." Working paper, UCSD.
- [16] F.L. Delmonico (2004) "Exchanging kidneys – Advances in living-donor transplantation." *The New England Journal of Medicine* 350: 1812-1814.
- [17] E. Diamantoudi, E. Miyagawa, and L. Xue (2004) "Random paths to stability in the roommate problem." *Games and Economic Behavior* 48: 18-28.
- [18] B. Dutta and D. Ray (1989) "A concept of egalitarianism under participation constraints." *Econometrica* 57: 615-635.
- [19] J. Edmonds (1965) "Paths, trees, and flowers." *Canadian Journal of Mathematics* 17: 449-467.
- [20] J. Edmonds (1971) "Matroids and the greedy algorithm." *Mathematical Programming* 1: 127-136.
- [21] L. Ehlers (2002) "Coalitional strategy-proof house allocation." *Journal of Economic Theory* 105: 298-317.
- [22] L. Ehlers and B. Klaus (2003). "Coalitional strategy-proof and resource monotonic solutions for multiple assignment problems." *Social Choice and Welfare* 21: 265-280.

- [23] L. Ehlers, B. Klaus, and S. Papai (2002) “Strategy-proofness and population-monotonicity in house allocation problems.” *Journal of Mathematical Economics* 38: 329-339.
- [24] D. Gale and L.S. Shapley (1962) “College admissions and the stability of marriage.” *American Mathematical Monthly* 69: 9-15.
- [25] T. Gallai (1963) “Kritische Graphen II.” *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 8: 373-395.
- [26] T. Gallai (1964) “Maximale Systeme unabhängiger kanten.” *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 9: 401-413.
- [27] D.W. Gjertson and J.M. Cecka (2000) “Living unrelated donor kidney transplantation.” *Kidney International* 58: 491-499.
- [28] M.X. Goemans (2004) *Lecture Notes on Topics in Combinatorial Optimization*. Massachusetts Institute of Technology Lecture Notes.
- [29] P. Hall (1935) “On representatives of subsets.” *Journal of London Mathematical Society* 10: 26-30.
- [30] B. Korte and J. Vygen (2002) *Combinatorial Optimization: Theory and Algorithms*. Second Edition. Springer: Berlin, Heidelberg, New York.
- [31] L. Lovász and M.D. Plummer (1986) *Matching Theory*. North-Holland: Amsterdam, New York, Oxford, Tokyo.
- [32] M. Lucan, P. Rotariu, D. Neculoiu, and G. Iacob (2003) “Kidney exchange program: a viable alternative in countries with low rate of cadaver harvesting.” *Transplantation Proceedings* 35: 933-934.
- [33] P. Milgrom (2004), *Putting Auction Theory to Work*, Cambridge University Press.
- [34] M. Niederle, and A. E. Roth (2004), “The gastroenterology fellowship match: How it failed, and why it could succeed once again.” *Gastroenterology*, 127, 2 August, 658-666.
- [35] M. Niederle, and A. E. Roth (2005), “The Gastroenterology Fellowship Market: Should there be a Match?,” *American Economic Review*, Papers and Proceedings, May, forthcoming.
- [36] G. Opelz (1997) “Impact of HLA compatibility on survival of kidney transplants from unrelated live donors.” *Transplantation* 64, 1473-1475.
- [37] G. Opelz, for the Collaborative Transplant Study (1998) “HLA compatibility and kidney grafts from unrelated live donors.” *Transplantation Proceedings* 30: 704-705.



- [38] S. Papai (2000) “Strategyproof assignment by hierarchical exchange.” *Econometrica* 68: 1403-1433.
- [39] R. Rado (1957) “Note on independence functions.” *Proceedings of the London Mathematical Society* 7: 300-320.
- [40] F.T. Rapaport (1986) “The case for a living emotionally related international kidney donor exchange registry.” *Transplantation Proceedings* 18: 5-9.
- [41] L. F. Ross and E. S. Woodle (2000) “Ethical issues in increasing living kidney donations by expanding kidney paired exchange programs.” *Transplantation* 69: 1539-1543.
- [42] L.F. Ross, D.T. Rubin, M. Siegler, M.A. Josephson, J.R. Thistlethwaite, Jr., and E.S. Woodle (1997) “Ethics of a paired-kidney-exchange program.” *The New England Journal of Medicine* 336: 1752-1755.
- [43] A.E. Roth (1982a) “Incentive compatibility in a market with indivisible goods.” *Economics Letters* 9: 127-132.
- [44] A.E. Roth (1982b) “The economics of matching: stability and incentives.” *Mathematics of Operations Research* 7: 617-628.
- [45] A.E. Roth (1984) “The evolution of the labor market for medical interns and residents: a case study in game theory.” *Journal of Political Economy* 92: 991-1016.
- [46] A.E. Roth (1985) “The college admissions problem is not equivalent to the marriage problem.” *Journal of Economic Theory* 36: 277-288.
- [47] A.E. Roth (2002) “The economist as engineer: game theory, experimental economics and computation as tools of design economics.” *Econometrica* 70: 1341-1378.
- [48] A.E. Roth and E. Peranson (1999) “The redesign of the matching market for American physicians: some engineering aspects of economic design.” *American Economic Review* 89: 748-780.
- [49] A.E. Roth and A. Postlewaite (1977) “Weak versus strong domination in a market with indivisible goods.” *Journal of Mathematical Economics* 4, 131-137.
- [50] A.E. Roth, U.G. Rothblum, and J.H. Vande Vate (1993) “Stable matchings, optimal assignment and linear programming.” *Mathematics of Operations Research* 18: 803-828.
- [51] A.E. Roth, T. Sönmez, and M.U. Ünver (2004) “Kidney exchange.” *Quarterly Journal of Economics* 119: 457-488.

- [52] A.E. Roth, T. Sönmez, and M. U. Ünver (2005) "A Kidney Exchange Clearinghouse in New England," *American Economic Review*, Papers and Proceedings, May, forthcoming.
- [53] A.E. Roth and M. Sotomayor (1990) *Two-sided matching: a study in game-theoretic modelling and analysis*. Cambridge University Press, Cambridge.
- [54] A.E. Roth and J.H. Vande Vate (1990) "Random paths to stability in two-sided matching." *Econometrica* 58: 1475-1480.
- [55] A. E. Roth and X. Xing (1997) "Turnaround time and bottlenecks in market clearing: decentralized matching in the market for clinical psychologists." *Journal of Political Economy* 105: 284-329.
- [56] L. Shapley and H. Scarf (1974) "On cores and indivisibility." *Journal of Mathematical Economics* 1: 23-28.
- [57] T. Sönmez and M.U. Ünver (2001) "House allocation with existing tenants: an equivalence." *Games and Economic Behavior* forthcoming.
- [58] L-G. Svensson (1994) "Queue allocation of indivisible goods." *Social Choice and Welfare* 11: 323-330.
- [59] L.-G. Svensson (1999) "Strategy-proof allocation of indivisible goods." *Social Choice and Welfare* 16: 557-567.
- [60] C.-P. Teo and J. Sethuraman (2000) "On a cutting plane heuristic for the stable roommates problem and its applications." *European Journal of Operational Research* 123: 195-205.
- [61] R.B. Wilson (2002) "Architecture of Power Markets." *Econometrica* 70: 1299-1340.
- [62] S. A. Zenios (2002) "Optimal control of a paired-kidney exchange program." *Management Science* 48: 328-342.
- [63] S. A. Zenios, E. S. Woodle, and L. F. Ross (2001) "Primum non nocere: avoiding harm to vulnerable waitlist candidates in an indirect kidney exchange." *Transplantation* 72: 648-654.