The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Balancing Lexicographic Fairness and a Utilitarian Objective with Application to Kidney Exchange

Duncan C. McElfresh[†,‡]
[†]Department of Mathematics
University of Maryland
dmcelfre@math.umd.edu

John P. Dickerson[†,‡]
[‡]Department of Computer Science
University of Maryland
john@cs.umd.edu

Abstract

Balancing fairness and efficiency in resource allocation is a classical economic and computational problem. The price of fairness measures the worst-case loss of economic efficiency when using an inefficient but fair allocation rule; for indivisible goods in many settings, this price is unacceptably high. One such setting is kidney exchange, where needy patients swap willing but incompatible kidney donors. In this work, we close an open problem regarding the theoretical price of fairness in modern kidney exchanges. We then propose a general hybrid fairness rule that balances a strict lexicographic preference ordering over classes of agents, and a utilitarian objective that maximizes economic efficiency. We develop a utility function for this rule that favors disadvantaged groups lexicographically; but if cost to overall efficiency becomes too high, it switches to a utilitarian objective. This rule has only one parameter which is proportional to a bound on the price of fairness, and can be adjusted by policymakers. We apply this rule to real data from a large kidney exchange and show that our hybrid rule produces more reliable outcomes than other fairness rules.

1 Introduction

Chronic kidney disease is a worldwide problem whose societal burden is likened to that of diabetes (Neuen et al. 2013). Left untreated, it leads to end-stage renal failure and the need for a donor kidney—for which demand far outstrips supply. In the United States alone, the kidney transplant waiting list grew from $58{,}000$ people in 2004 to over $100{,}000$ needy patients (Hart et al. 2016).[1]

To alleviate some of this supply-demand mismatch, *kidney exchanges* (Rapaport 1986; Roth, Sönmez, and Ünver 2004) allow patients with willing *living* donors to trade donors for access to compatible or higher-quality organs. In addition to these patient-donor pairs, modern exchanges include *non-directed donors*, who enter the exchange without a patient in need of a kidney. Exchanges occur in cycle- or chain-like structures, and now account for 10% of living transplants in the United States. Yet, access to a kidney exchange does not guarantee equal access to kidneys themselves; for example, certain classes of patients may be particularly disadvantaged based on health characteristics or other logistical factors. Thus, *fairness* considerations are an active topic of theoretical and practical research in kidney exchange and the matching market community in general.

Intuitively, any enforcement of a fairness constraint or consideration may have a negative effect on overall economic efficiency. A quantification of this tradeoff is known as the *price of fairness* (Bertsimas, Farias, and Trichakis 2011). Recent work by Dickerson, Procaccia, and Sandholm (2014) adapted this concept to the kidney exchange case, and presented two fair allocation rules that strike a balance between fairness and efficiency. Yet, as we show in this paper, those rules can "fail" unpredictably, yielding an arbitrarily high price of fairness.

With this as motivation, we adapt to the kidney exchange case a recent technique for trading off a form of fairness and utilitarianism in a principled manner. This technique is parameterized by a bound on the price of fairness, as opposed to a set of parameters that may result in hard-to-predict final matching behavior, as in past work. We implement our rule in a realistic mathematical programming framework and–on real data from a large, multi-center, fielded kidney exchange–show that our rule effectively balances fairness and efficiency without unwanted outlier behavior.

1.1 Related Work

We briefly overview related work in balancing efficiency and fairness in resource allocation problem. Bertsimas, Farias, and Trichakis (2011) define the price of fairness; that is, the relative loss in system efficiency under a fair allocation rule. Hooker and Williams (2012) give a formal method for combining utilitarianism and equity. We direct the reader to those two papers for a greater overview of research in fairness in general resource allocation problems.

Fairness in the context of kidney exchange was first studied by Roth, Sönmez, and Ünver (2005b); they explore concepts like Lorenz dominance in a stylized model, and show that preferring fair allocations can come at great cost. Li et al. (2014) extend this model and present an algorithm to solve for a Lorenz dominant matching. Stability in kidney exchange, a concept intimately related to fairness, was explored by Liu, Tang, and Fang (2014). The use of randomized allocation machanisms to promote fairness in stylized models is theoretically promising (Fang et al. 2015;

[1] `https://optn.transplant.hrsa.gov/data/`

Aziz et al. 2016; Mattei, Saffidine, and Walsh 2017). Recent work discusses fairness in stylized random graph models of dynamic kidney exchange (Ashlagi, Jaillet, and Manshadi 2013; Anderson et al. 2015). None of these papers provide practical models that could be implemented in a fully-realistic and fielded kidney exchange.

Practically speaking, Yılmaz (2011) explores in simulation equity issues from combining living and deceased donor allocation; that paper is limited to only short length-two kidney swaps, while real exchanges all use longer cycles and chains. Dickerson, Procaccia, and Sandholm (2014) introduced two fairness rules explicitly in the context of kidney exchange, and proved bounds on the price of fairness under those rules in a random graph model; we build on that work in this paper, and describe it in greater detail later. That work has been incorporated into a framework for learning to balance efficiency, fairness, and dynamism in matching markets (Dickerson and Sandholm 2015); we note that the fairness rule we present in this paper could be used in that framework as well.

1.2 Our Contributions

Dickerson, Procaccia, and Sandholm (2014) finds that the theoretical price of fairness in kidney exchange is small when *only* patient-donor pairs participate in the exchange. They did not include non-directed donors (NDDs). However, in modern kidney exchanges, non-directed donors (NDDs) provide many more matches than patient-donor pairs; furthermore, NDDs create more opportunities to expand the fair matching, potentially increasing the price of fairness. Here, we prove that adding NDDs to the theoretical model actually *decreases* the price of fairness, and that—with enough NDDs—the price of fairness is zero.

Real kidney exchanges are less dense and more uncertain than the (standard) theoretical model in which we prove our results. Previous approaches to incorporating fairness into kidney exchange have neglected this fact: they have been either ad-hoc—e.g., "priority points" decided on by committee (Kidney Paired Donation Work Group 2013)—or brittle (Roth, Sönmez, and Ünver 2005b; Dickerson, Procaccia, and Sandholm 2014), resulting in an unacceptably high price of fairness. This paper provides the first approach to incorporating fairness into kidney exchange in a way that both prioritizes disadvantaged participants, but also comes with acceptable worst-case guarantees on the price of fairness. Our method is easily applied as an objective in the mathematical-programming-based clearing methods used in today's fielded exchanges; indeed, using real data we show that this method guarantees a limit on efficiency loss.

Section 1.3 introduces the kidney exchange problem. Section 2 extends work by Ashlagi and Roth (2014) and Dickerson, Procaccia, and Sandholm (2014), showing that the price of fairness is small on the canonical random graph model even with NDDs. Section 3 shows that two recent fair allocation rules from the kidney exchange literature (Dickerson, Procaccia, and Sandholm 2014) can perform unacceptably poorly in the worst case. Then, Section 4 presents a new allocation rule that allows policymakers to set a limit on efficiency loss, while also favoring disadvantaged patients. Section 5 shows on real data from a large fielded kidney exchange that our method limits efficiency loss while still favoring disadvantaged patients when possible.

1.3 Preliminaries

A kidney exchange can be represented as a directed *compatibility graph* $G = (V, E)$, with vertices $V = P \cup N$ including both patient donor pairs $p \in P$ and non-directed-donors $n \in N$ (Roth, Sönmez, and Ünver 2004; 2005a; 2005b; Abraham, Blum, and Sandholm 2007). A directed edge e is drawn from vertex v_i to v_j if the donor at v_i can give to the patient at v_j. Fielded kidney exchanges consist mainly of directed cycles in G, where each patient vertex in the cycle receives the donor kidney of the previous vertex. Modern exchanges also include non-cyclic structures called chains (Montgomery et al. 2006; Rees et al. 2009). Here, an NDD donates her kidney to a patient, whose paired donor donates her kidney to another patient, and so on.

In practice, cycles are limited in size, or "capped," to some small constant L, while chains are limited in size to a much larger constant R—or not limited at all. This is because all transplants in a cycle must execute *simultaneously*; if a donor whose paired patient had already received a kidney backed out of the donation, then some participant in the market would be strictly worse off than before. However, chains need not be executed simultaneously; if a donor backs out after her paired patient receives a kidney, then the chain breaks but no participant is strictly worse off. We will discuss how these caps affect fairness and efficiency in the coming sections.

The goal of kidney exchange programs is to find a *matching* M—a collection of disjoint cycles and chains in the graph G. The cycles and chains must be disjoint because no donor can give more than one of her kidneys (although ongoing work explores multi-donor kidney exchange (Ergin, Sönmez, and Utku Ünver 2017; Farina, Dickerson, and Sandholm 2017)). The *clearing problem* in kidney exchange is to find a matching M^* that maximizes some utility function $u : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the set of all legal matchings. Real kidney exchanges typically optimize for the maximum weighted cycle cover (i.e., $u(M) = \sum_{c \in M} \sum_{e \in c} w_e$). This *utilitarian* objective can favor certain classes of patient-donor pairs while disadvantaging others. This is formalized in the following section.

1.4 The Price of Fairness

As an example for this paper, we focus on *highly-sensitized* patients, who have a very low probability of their blood passing a feasibility test with a random donor organ; thus, finding a kidney is often quite hard, and their median waiting time for an organ jumps by a factor of three over less sensitized patients.[2] Utilitarian objectives will, in general, marginalize these patients. Sensitization is determined using the Calculated Panel Reactive Antibody (CPRA) level of each patient, which reflects the likelihood that a patient will find a matching donor.

[2] https://optn.transplant.hrsa.gov/data/

Formally the sensitization of each patient-donor vertex v be $v_s \in [0, 100]$, the CPRA level of v's patient; NDD vertices are not associated with patients, so they do not have sensitization levels. Each patient-donor vertex $v \in P$ is considered highly sensitized if v_s exceeds threshold $\tau \in [0, 100]$, and lowly-sensitized otherwise. These vertex sets V_H and V_L are defined as:

- Lowly sensitized: $V_L = \{v \mid v \in P : v_s < \tau\}$
- Highly sensitized: $V_H = \{v \mid v \in P : v_s \geq \tau\}$.

By definition, highly-sensitized patients are harder to match than lowly-sensitized patients. Naturally, efficient matching algorithms prioritize easy-to-match vertices in V_L, marginalizing V_H. Let $u_f : \mathcal{M} \to \mathbb{R}$ be a *fair* utility function. Formally, a utility function is fair when its corresponding optimal match M_f^* is viewed as fair, where M_f^* is defined as:

$$M_f^* = \underset{M \in \mathcal{M}}{\arg\max}\, u_f(M)$$

Bertsimas, Farias, and Trichakis (2011) defined the *price of fairness* to be the "relative system efficiency loss under a fair allocation assuming that a fully efficient allocation is one that maximizes the sum of [participant] utilities." Caragiannis et al. (2009) defined an essentially identical concept in parallel. Formally, given a fair utility function u_f and the utilitarian utility function u, the price of fairness is:

$$\mathsf{POF}(\mathcal{M}, u_f) = \frac{u\left(M^*\right) - u\left(M_f^*\right)}{u\left(M^*\right)} \tag{1}$$

The price of fairness $\mathsf{POF}(\mathcal{M}, u_f)$ is the relative loss in (utilitarian) efficiency caused by choosing the fair outcome M_f^* rather than the most efficient outcome.

In the next section we show that the theoretical price of fairness in kidney exchange is small, even when both cycles *and chains* are used—thus generalizing an earlier result due to Dickerson, Procaccia, and Sandholm (2014) to modern kidney exchanges.

2 The Theoretical Price of Fairness with Chains is Low (or Zero)

In this section we use the random graph model for kidney exchange introduced by Ashlagi and Roth (2014) to show that the theoretical price of fairness is always small, especially when NDDs are included. A complete description of this model can be found in Appendix A[3]. Dickerson, Procaccia, and Sandholm (2014) finds that without NDDs, the maximum price of fairness is $2/33$. Adding NDDs to this model creates more opportunities to match highly sensitized patients, which could potentially lead to a higher price of fairness. However we find that including chains in this model only *decreases* the price of fairness; furthermore, when the ratio of NDDs to patient-donor pairs is high enough, the price of fairness is zero.

2.1 Price of Fairness

Ashlagi and Roth (2014) characterize efficient matchings in a random graph model without chains, and Dickerson, Procaccia, and Sandholm (2014) build on this to show that the price of fairness without chains is bounded above by $2/33$. Dickerson, Procaccia, and Sandholm (2012) extend the efficient matching of Ashlagi and Roth (2014) to include chains, but do not calculate the price of fairness. We close the gap in theory regarding the price of fairness with chains.

Given $|P|$ patient-donor pairs, we parameterize the number of NDDs $|N|$ with $\beta \geq 0$ such that $|N| = \beta|P|$. Theorems 1 and 2 state our two main results: adding chains to the random graph model does not increase the price of fairness, and when the fraction of NDDs is high enough ($\beta > 1/8$), the price of fairness is zero. The proofs of the following theorems are given in Appendix A.

Theorem 1. *Adding NDDs to the random graph model ($\beta > 0$) does not increase the upper bound on the price of fairness found by Dickerson, Procaccia, and Sandholm (2014).*

Proof Sketch: We explore every possible efficient matching on the random graph model with chains; only four of these matchings have nonzero price of fairness. For each case, we compare the price of fairness to that of the efficient matching without chains found in Dickerson, Procaccia, and Sandholm (2014), and find that the upper bound does not increase.

Theorem 2. *The price of fairness is zero when $\beta > 1/8$.*

Proof sketch: For each matching with nonzero price of fairness, $\beta \leq 1/8$. When $\beta > 1/8$, a different matching occurs, and the price of fairness is zero.

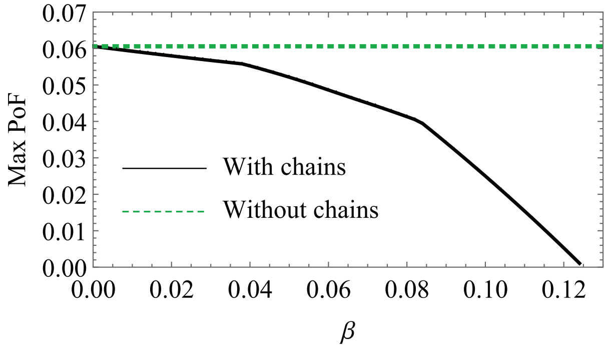


Figure 1: Price of fairness with chains. (The horizontal dotted line at $2/33$ is the price of fairness without chains.)

To illustrate these results, we compute the price of fairness when $\beta \in [0, 1/8]$. These calculations confirm our theoretical results, as shown in Figure 2.1: the price of fairness decreases as β increases, and is zero when $\beta > 1/8$.

The worst-case price of fairness is small in the random graph model, with or without NDDs. However, real exchange graphs are typically much sparser and less uniform—in reality the price of fairness can be high. In the next section, we discuss two notions of fairness in kidney exchange and determine their worst-case price of fairness.

[3] Full paper and Appendices can be found at https://arxiv.org/abs/1702.08286.

3 The Price of Fairness in State-of-the-Art Fair Rules can be Arbitrarily Bad

The price of fairness depends on how fairness is defined. This is especially true in real exchanges where the price of fairness can be unacceptably high. In this section, we discuss two kidney-exchange-specific fairness rules introduced by Dickerson, Procaccia, and Sandholm (2014): lexicographic fairness and weighted fairness. These rules favor the disadvantaged class without considering overall loss in efficiency; we show that in the worst case these rules allow the the price of fairness to approach 1 (i.e., total efficiency loss). Proofs of these theorems are in Appendix B.

3.1 Lexicographic Fairness

As proposed by Dickerson, Procaccia, and Sandholm (2014), α-lexicographic fairness assigns nonzero utility only to matchings that award at least a fraction α of the maximum possible fair utility. Letting $u_H(M)$ and $u_L(M)$ be the utility assigned to only vertices in V_H and V_L, respectively, the utility function for α-lexicographic fairness is given in Equation (2).

$$u_\alpha(M) = \begin{cases} u_L(M) + u_H(M) \\ \qquad \text{if } u_H(M) \geq \alpha \max\limits_{M' \in \mathcal{M}} u_H(M') \\ 0 \qquad \text{otherwise.} \end{cases} \tag{2}$$

Theorems 3 and 4 state that strict lexicographic fairness ($\alpha = 1$) allows the price of fairness to approach 1.

Theorem 3. *For any cycle cap L there exists a graph G such that the price of fairness of G under α-lexicographic fairness with $0 < \alpha \leq 1$ is bounded by $\mathit{POF}(\mathcal{M}, u_\alpha) \geq \frac{L-2}{L}$.*

Theorem 4. *For any chain cap R there exists a graph G such that the price of fairness of G under the α-lexicographic fairness rule with $0 < \alpha \leq 1$ is bounded by $\mathit{POF}(\mathcal{M}, u_\alpha) \geq \frac{R-1}{R}$.*

Thus, α-lexicographic fairness allows for a price of fairness that approaches 1 as the cycle and chain cap increase.

3.2 Weighted Fairness

The weighted fairness rule (Dickerson, Procaccia, and Sandholm 2014) defines a utility function by first modifying the original edge weights w_e by a multiplicative factor $\gamma \in \mathbb{R}$ such that

$$w'_e = \begin{cases} (1+\gamma)w_e & \text{if } e \text{ ends in } V_H \\ w_e & \text{otherwise.} \end{cases}$$

Then the weighted fairness rule u_{WF} is

$$u_{WF}(M) = \sum_{c \in M} u'(c),$$

where $u'(c)$ is the utility of a chain or cycle c with modified edge weights. The modified edge weights prompt the matching algorithm to include more highly-sensitized patients; as in the lexicographic case, we now show that the price of fairness approaches 1 under weighted fairness.

Theorem 5. *For any cycle cap L and $\gamma \geq L-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathit{POF}(\mathcal{M}, u_{WF}) \geq \frac{L-2}{L}$.*

Theorem 6. *For any chain cap R and $\gamma \geq R-1$, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathit{POF}(\mathcal{M}, u_{WF}) \geq \frac{R-1}{R}$.*

In the worst case, weighted fairness allows a price of fairness that approaches 1 as the cycle and chain caps increase. The price of fairness also approaches 1 as γ increases.

Theorem 7. *With no chain cap, there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathit{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

A similar result exists with cycles rather than chains.

Theorem 8. *With no cycle cap there exists a graph G such that the price of fairness of G under the weighted fairness rule is bounded by $\mathit{POF}(\mathcal{M}, u_{WF}) \geq \frac{\gamma}{\gamma+1}$.*

These bounds show that weighted fairness allows for a price of fairness that approaches 1, i.e., arbitrarily bad, as the cycle cap, chain cap, or γ increase.

We have shown that the worst-case prices of fairness approach 1 under both the lexicographic and weighted fairness rules of Dickerson, Procaccia, and Sandholm (2014). Next, we propose a rule that favors disadvantaged groups, but also strictly *limits* the price of fairness using a parameter set by policymakers.

4 Hybrid Fairness Rule

In this section, we present a hybrid fair utility function that balances lexicographic fairness and a utilitarian objective. We generalize the hybrid utility function proposed by Hooker and Williams (2012), which chooses between a Rawlsian (or maximin) objective and a utilitarian objective for multiple classes of agents.

4.1 Utilitarian and Rawlsian Fairness

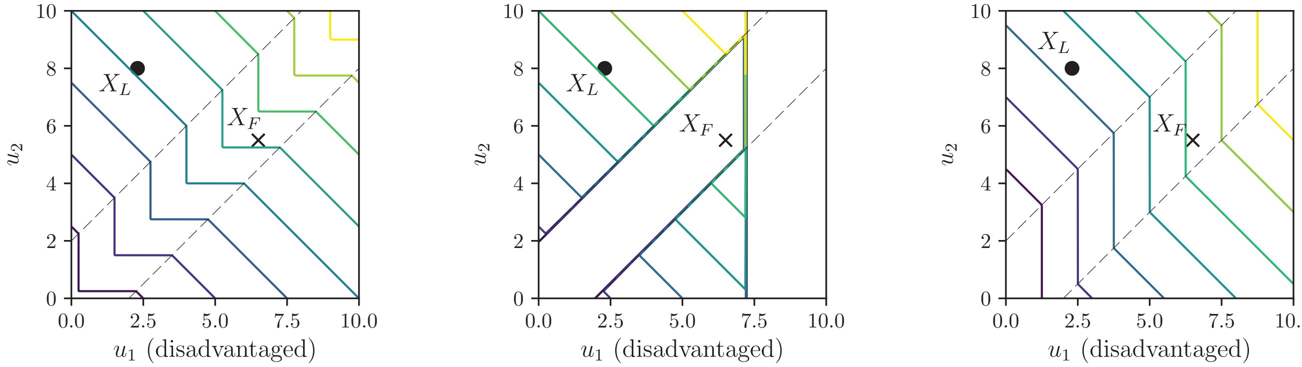
Consider two classes of agents that receive utilities $u_1(X)$ and $u_2(X)$, respectively, for outcome X. The fairness rule introduced by Hooker and Williams (2012) maximizes the utility of the worst-off class, unless this requires taking too many resources from other classes. When the inequality exceeds a threshold Δ (i.e., $|u_1(X) - u_2(X)| > \Delta$) they switch to a utilitarian objective that maximizes $u_1(X) + u_2(X)$. The utility function for this rule is

$$u_\Delta(X) = \begin{cases} 2\min(u_1(X), u_2(X)) + \Delta \\ \qquad \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ \\ u_1(X) + u_2(X) \\ \qquad \text{otherwise.} \end{cases}$$

The parameter Δ is problem-specific, and should be chosen by policymakers. Figure 2(a) shows the level sets of this utility function, with $\Delta = 2$. This utility function can be generalized by switching to a different fairness rule in the *fair region* (i.e. when $|u_1(X) - u_2(X)| \leq \Delta$). The next section generalizes this rule using lexicographic fairness.

4.2 Hybrid-Lexicographic Rule

When it is desirable to favor one class of agents g_1 over class g_2, lexicographic fairness favors g_1. We propose a rule that



(a) Hybrid Rawlsian-Utilitarian (b) Δ-Lexicographic ($u_{\Delta 1}$) (c) Relaxed Δ-Lexicographic (u_Δ)

Figure 2: Level sets for hybrid fair utility functions with $\Delta = 2$, with example outcomes X_L and X_F.

implements lexicographic fairness only when inequality between groups does not exceed Δ. This rule uses two steps: 1) determine whether inequality is small enough to use lexicographic fairness 2) choose the optimal outcome. These steps are outlined below, and formalized in Algorithm 1.

Step 1: Find all outcomes that maximize a hybrid utility function, and determine whether lexicographic fairness is appropriate.

We use a utility function to identify outcomes that satisfy either a lexicographic or utilitarian objective. Equation (3) shows one option for such a utility function, which assigns strict lexicographic utility ($\alpha = 1$) according to Equation (2) in the fair region, and utilitarian utility otherwise.

$$u_{\Delta 1}(X) = \begin{cases} u_1(X) + u_2(X) & \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ & \text{ and } u_1(X) = \max\limits_{X' \in \mathcal{X}} (u_1(X')) \\ & \\ u_1(X) + u_2(X) & \text{if } |u_1(X) - u_2(X)| > \Delta \\ & \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\mathcal{X}$ is the set of all possible outcomes. Figure 2(b) shows the contours $u_{\Delta 1}$. This utility function is clearly too harsh—it assigns zero utility to outcomes in the fair region that do not maximize u_1, and its optimal outcomes are not always Pareto efficient. Consider outcomes X_F and X_L in Figure 2(b). X_F is in the fair region but does not maximize u_1, so $u_{\Delta 1}(X_F) = 0$; X_L is in the utilitarian region but is less efficient, so $u_{\Delta 1}(X_L) = u(X_L)$. Under utility function $u_{\Delta 1}$, the less-efficient outcome X_L is chosen over X_F.

To address this problem we introduce u_Δ in Equation (4), which relaxes $u_{\Delta 1}$. For outcomes in the fair region (that is, with $|u_1 - u_2| \leq \Delta$), utility is assigned proportional to u_1. As shown in Figure 2(c), the contours of u_Δ are continuous.

$$u_\Delta(X) = \begin{cases} u_1(X) + u_2(X) - \Delta & \text{if } u_2(X) - u_1(X) > \Delta \\ 2u_1(X) & \text{if } |u_1(X) - u_2(X)| \leq \Delta \\ u_1(X) + u_2(X) + \Delta & \text{if } u_1(X) - u_2(X) > \Delta \end{cases} \tag{4}$$

Let X_{OPT} be the set of outcomes that maximize u_Δ. If any outcomes in X_{OPT} are in the utilitarian region , then any utilitarian-optimal outcome is selected. However, if any outcomes in X_{OPT} are in the fair region, then Step 2 must be used. This process is described below, and formalized in Algorithm 1.

Step 2: If any solution in X_{OPT} is in the fair region, select the lexicographic-optimal solution in the fair region.

The utility function u_Δ assigns the same utility to all solutions in the fair region with the same $u_1(X)$, no matter the value of $u_2(X)$. However, if there exist two outcomes X_A and X_B such that $u_1(X_A) = u_1(X_B)$ and $u_2(X_A) > u_2(X_B)$, then X_A is lexicographically preferred to X_B.

Algorithm 1 FairMatching

```
Input: Threshold $\Delta$, matchings $\mathcal{M}$
Output: Fair matching $M^*$
  $\mathcal{M}_{OPT} \leftarrow \arg\max_{M \in \mathcal{M}} u_\Delta(M)$
  if $|\mathcal{M}_{OPT}| > 1$ then
      Select a matching $M \in \mathcal{M}_{OPT}$
      if $M$ is in the utilitarian region then
          $M^* \leftarrow M$
      else
          $\mathcal{M}_1 \leftarrow \{M' \in \mathcal{M}_{OPT} \mid u_1(M') = u_1(M)\}$
          $M^* \leftarrow \arg\max_{M' \in \mathcal{M}_1} u_2(M')$
  else
      $M^* \leftarrow \mathcal{M}_{OPT}$
```

4.3 Hybrid Rule for Several Classes

We now generalize the hybrid-lexicographic fairness rule to more than two classes. Consider a set $\mathcal{P}$ of classes g_i, $i = 1, \ldots, |\mathcal{P}|$. Let there be an ordering $\succ$ over g_i, where $g_a \succ g_b$ indicates that g_a should receive higher priority over g_b. WLOG, let the preference ordering over g_i be $g_1 \succ g_2 \succ \cdots \succ g_P$. Let $u_i(X)$ be the utility received by group i under outcome X. As in the previous section, we 1) use a utility function to determine whether lexicographic fairness is appropriate, then 2) select either a lexicographic- or utilitarian-optimal outcome.

Step 1: To define a utility function, we observe that in Equation (4), in the utilitarian region a positive offset Δ is added if $u_1(X) > u_2(X)$, and a negative offset is added otherwise. With $|\mathcal{P}|$ classes, each solution in the utilitarian region receives a utility offset of $+\Delta$ if $u_1(X) > u_i(X)$, and $-\Delta$ otherwise, for each class $i = 2, 3, \dots, |\mathcal{P}|$. As in the previous section, these offsets ensure continuity in the utility function, and ensure that at least one of the maximizing outcomes will be Pareto optimal.

$$u_\Delta(X) = \begin{cases} |\mathcal{P}| \cdot u_1(X) \\ \qquad \text{if } \max_i(u_i(X)) - \min_i(u_i(X)) \leq \Delta, \\ \\ u_1(X) + \sum_{i=2}^{|\mathcal{P}|} (u_i(X) + \text{sgn}(u_1(X) - u_i(X))\Delta) \\ \qquad \text{otherwise} \end{cases} \tag{5}$$

Step 2: Let X_{OPT} be the set of solutions that maximize u_Δ. If all optimal solutions are in the utilitarian region, any utilitarian-optimal solution is selected. If any optimal solution is in the fair region, then the lexicographic-optimal solution in the fair region must be selected, subject to the preference ordering $g_1 \succ g_2 \succ \dots \succ g_{|\mathcal{P}|}$.

Algorithm 2 FairMatching for $|\mathcal{P}| \geq 2$ classes

```
Input: Threshold $\Delta$, matchings $\mathcal{M}$
Output: Fair matching $M^*$
  $\mathcal{M}_{OPT} \leftarrow \arg\max_{M \in \mathcal{M}} u_\Delta(M)$
  if $|\mathcal{M}_{OPT}| > 1$ then
      Select a matching $M \in M_{OPT}$
      if $M$ in utilitarian region then
          $M^* \leftarrow M$
      else
          $\mathcal{M}_1 \leftarrow \{M' \in \mathcal{M}_{OPT} \mid u_1(M') = u_1(M)\}$
          for $i = 2, \dots, |\mathcal{P}|$ do
              $\mathcal{M}_i \leftarrow \arg\max_{M' \in \mathcal{M}_{i-1}} u_i(M')$
          $M^* \leftarrow$ any matching in $\mathcal{M}_{|\mathcal{P}|}$
  else
      $M^* \leftarrow \mathcal{M}_{OPT}$
```

4.4 Price of Fairness for the Hybrid Rule

Theorem 9 gives a bound on the price of fairness for the hybrid-lexicographic rule; its proof is given in Appendix B.

Theorem 9. *Assume the optimal utilitarian outcome X_E receives utility $u(X_E) = u_E$, with most prioritized class $g_1 \in \mathcal{P}$ receiving utility u_1, and Z other classes $g_i \in \mathcal{P}$ such that $u_1(X_E) > u_i(X_E)$. Then, POF$(\mathcal{M}, u_\Delta) \leq \frac{2((|\mathcal{P}|-1)-Z)\Delta}{u_E}$.*

4.5 Hybrid Fairness in Kidney Exchange

The hybrid-lexicographic fairness rule in Equation (4) is easily applied to kidney exchange, with u_H and u_L the total utility received by highly-sensitized and lowly-sensitized patients, respectively,

$$u_\Delta(M) = \begin{cases} u_L(M) + u_H(M) - \Delta \\ \qquad \text{if } u_L(M) - u_H(M) > \Delta \\ 2u_H(M) \qquad \text{if } |u_L(M) - u_H(M)| \leq \Delta \\ u_L(M) + u_H(M) + \Delta \\ \qquad \text{if } u_H(M) - u_L(M) > \Delta \end{cases} \tag{6}$$

In the following section, we demonstrate the practical effectiveness of the hybrid-lexicographic rule by testing it on real kidney exchange data.

5 Experiments

In this section, we compare the behavior of α-lexicographic, weighted, and hybrid-lexicographic fairness. Code for these experiemnts is available on GitHub.[4] We use each rule to find the optimal fair outcomes for 314 real kidney exchanges from the United Network for Organ Sharing (UNOS), collected between 2010 and 2016. To solve the kidney exchange clearing problem (KEP) we use the PICEF formulation of Dickerson et al. (2016), with cycle cap 3 and various chain caps. In real exchanges, not all recommended edges in a matching result in successful transplants. To reflect this uncertainty, we use the concept of failure-aware kidney exchange introduced in (Dickerson, Procaccia, and Sandholm 2013): all edges in the exchange can fail with probability $(1-p)$; the matching algorithm maximizes *expected* matching weight, considering edge success probability p.

5.1 Procedure

For each UNOS exchange graph G, we use the following procedure to implement each fairness rule. We repeat the following procedure for chain caps 0, 3, 10, and 20, and for edge success probabilities $p = 0.1n$, with $n = 1, 2, \dots, 10$.

1. Find the efficient matching M_E by solving the to optimality the NP-hard kidney exchange problem (KEP) on G.
2. Find the fair matching M_F by solving the KEP on $G' = (V, E')$, where each edge $e \in E'$ has weight 1 if e ends in V_H and 0 otherwise.
3. **Weighted Fairness:** Find the γ-fair matching M_γ by solving the KEP on $G^\gamma = (V, E^\gamma)$, where each edge $e \in E^\gamma$ has weight $1 + \gamma$ if e ends in V_H and 1 otherwise. After finding M_γ, the reported utilities are calculated using edge weights of E and not E'. We use weight parameters $\gamma = 2n$, with $n = 0, 1, 2, \dots, 10$.
4. **α-Lexicographic Fairness:** Find the α-fair matching M_α by solving the KEP on G, with the additional constraint $u_H(M_\alpha) \geq \alpha u_H(M_E)$. We use parameters $\alpha = 0.1n$, with $n = 0, 1, 2, \dots, 10$.
5. **Hybrid-Lexicographic Fairness:** Find the Δ-fair matching M_Δ using the α-fair matchings M_α, and Algorithm 1. That is, $M_\Delta = \text{FairMatching}(\Delta, M_\alpha)$. We use parameters $\Delta = 0.1n \cdot u(M_E)$, with $n = 0, 1, 2, \dots, 10$.

[4] https://github.com/duncanmcelfresh/FairKidneyExchange

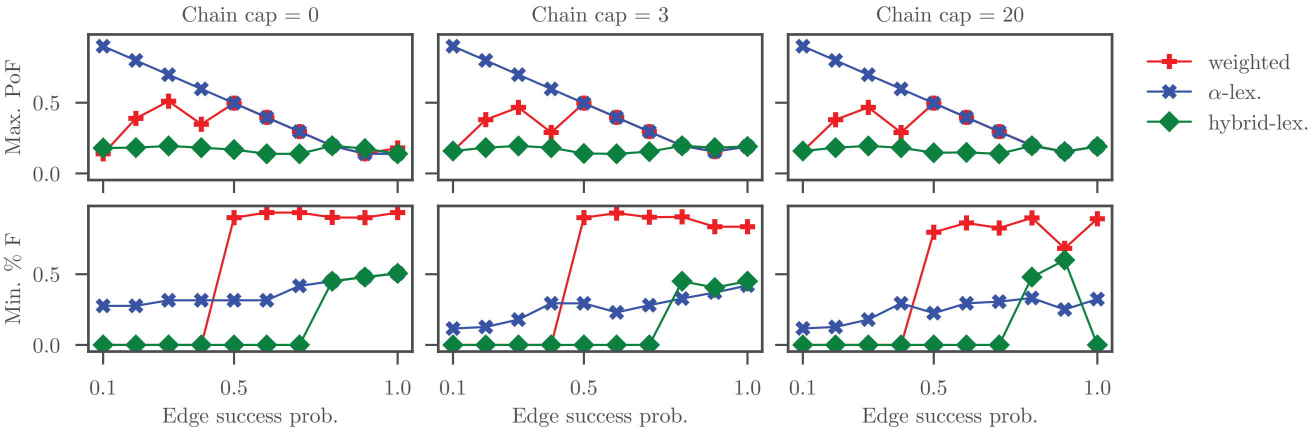


Figure 3: Worst-case price of fairness and $\%F$ for various edge success probabilities, and fairness parameters $\alpha = 0.1$, $\gamma = 0.1$, $\Delta = 0.1u(M_E)$.

Throughout this procedure, we calculate the utility of the efficient matching (u_E) and the fair matching (u_F) for each UNOS graph, and for each fairness rule—with parameters $\alpha \in [0, 1]$, $\gamma \in [0, 20]$, and $\Delta \in [0, u(M_E)]$.

There are two important outcomes of each fairness rule: Price of Fairness (PoF), and fraction of the fair score ($\%F$). To calculate PoF we use the definition in Equation (1), using u_E and u_F. We define $\%F$ as the fraction of the maximum highly sensitized utility, achieved by $M_{\{\alpha,\gamma,\Delta\}}$, defined as

$$\%F(M_{\{\alpha,\gamma,\Delta\}}, M_F) = u_H(M_{\{\alpha,\gamma,\Delta\}})/u_H(M_F).$$

PoF and $\%F$ indicate the efficiency loss and the fairness of each rule, respectively.

5.2 Results and Discussion

Each fairness rule offers a parameter that balances efficiency and fairness. Two of these rules guarantee a certain outcome: α-lexicographic guarantees fairness, but allows high efficiency loss, while hybrid-lexicographic bounds overall efficiency loss. Weighted fairness makes no guarantees.

The price of fairness can be high in real exchanges, especially when edge success probability p is small. In failure-aware kidney exchange, cycles and chains of length k receive utility proportional to p^k. Fair matchings often use longer cycles and chains than the efficient matching, in order to reach highly sensitized patients; this leads to a high price of fairness when p is small.

Even when α and γ are small, there are cases when both α-lexicographic and weighted fairness allow for a high PoF. This becomes worse with lower edge probability. Figure 3 shows the worst-case PoF and $\%F$ for each rule, for the smallest parameters tested, for a range of edge success probabilities; results for all parameter values are in Appendix C.

Hybrid-lexicographic fairness limits PoF within the guaranteed bound of 0.2; this comes at the cost of a low $\%F$—when edge success probability is small, hybrid-lexicographic fairness awards zero fair utility in the worst case. α-lexicographic fairness produces the opposite behavior: $\%F$ is always larger than the guaranteed bound of 0.1, but the worst-case price of fairness grows steadily as edge probability decreases.

Theory suggests that the price of fairness is small on denser random graphs (see Section 2). We empirically confirm this theoretical finding by calculating the worst-case price of fairness and $\%F$ for random graphs of various sizes generated from real data; these results are given in Appendix C. In this case—when the price of fairness is small—α-lexicographic fairness may be appropriate, as overall efficiency loss is not severe.

Both α-lexicographic and hybrid-lexicographic fairness are useful, depending on the desired outcome. Policymakers may choose between these rules, and set the parameters α and Δ to guarantee either a minimum $\%F$ or a maximum price of fairness.

6 Conclusion

We addressed the classical problem of balancing fairness and efficiency in resource allocation, with a specific focus on kidney exchange. Extending work by Ashlagi and Roth (2014) and Dickerson, Procaccia, and Sandholm (2014), we show that the theoretical price of fairness is small on a random graph model of kidney exchange, when both cycles and chains are used. However this model is too optimistic—real kidney exchanges are less certain and more sparse, and in reality the price of fairness can be unacceptably high.

Drawing on work by Hooker and Williams (2012), which is not applicable to kidney exchange, we provided the first approach to incorporating fairness into kidney exchange in a way that prioritizes marginalized participants, but also comes with acceptable worst-case guarantees on overall efficiency loss. Furthermore, our method is easily applied as an objective in the mathematical-programming-based clearing methods used in today's fielded exchanges. Using data from a large fielded kidney exchange, we showed that our method bounds efficiency loss while also prioritizing marginalized participants when possible.

Moving forward, it would be of theoretical and practical interest to address fairness in a realistic *dynamic* model of a matching market like kidney exchange (Anshelevich et al. 2013; Akbarpour, Li, and Gharan 2014; Anderson et al. 2015; Dickerson and Sandholm 2015). For example, how does prioritizing a class of patients in the present affect their, or other groups', long-term welfare? Similarly, exploring the effect on long-term efficiency of the single-shot Δ we use in this paper would be of practical importance; to start, Δ can be viewed as a hyperparameter to be tuned (Thornton et al. 2013).

References

Abraham, D.; Blum, A.; and Sandholm, T. 2007. Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 295–304.

Akbarpour, M.; Li, S.; and Gharan, S. O. 2014. Dynamic matching market design. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 355.

Anderson, R.; Ashlagi, I.; Gamarnik, D.; and Kanoria, Y. 2015. A dynamic model of barter exchange. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1925–1933.

Anshelevich, E.; Chhabra, M.; Das, S.; and Gerrior, M. 2013. On the social welfare of mechanisms for repeated batch matching. In *AAAI Conference on Artificial Intelligence (AAAI)*, 60–66.

Ashlagi, I., and Roth, A. E. 2014. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics* 9(3):817–863.

Ashlagi, I.; Jaillet, P.; and Manshadi, V. H. 2013. Kidney exchange in dynamic sparse heterogenous pools. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 25–26.

Aziz, H.; Filos-Ratsikas, A.; Chen, J.; Mackenzie, S.; and Mattei, N. 2016. Egalitarianism of random assignment mechanisms. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Bertsimas, D.; Farias, V. F.; and Trichakis, N. 2011. The price of fairness. *Operations Research* 59(1):17–31.

Caragiannis, I.; Kaklamanis, C.; Kanellopoulos, P.; and Kyropoulou, M. 2009. The efficiency of fair division. International Workshop on Internet and Network Economics (WINE).

Dickerson, J. P., and Sandholm, T. 2015. FutureMatch: Combining human value judgments and machine learning to match in dynamic environments. In *AAAI Conference on Artificial Intelligence (AAAI)*, 622–628.

Dickerson, J. P.; Manlove, D.; Plaut, B.; Sandholm, T.; and Trimble, J. 2016. Position-indexed formulations for kidney exchange. In *Proceedings of the ACM Conference on Economics and Computation (EC)*.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2012. Optimizing kidney exchange with transplant chains: Theory and reality. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 711–718.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2013. Failure-aware kidney exchange. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 323–340.

Dickerson, J. P.; Procaccia, A. D.; and Sandholm, T. 2014. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 1013–1020.

Ergin, H.; Sönmez, T.; and Utku Ünver, M. 2017. Multi-donor organ exchange. Working paper.

Fang, W.; Filos-Ratsikas, A.; Frederiksen, S. K. S.; Tang, P.; and Zuo, S. 2015. Randomized assignments for barter exchanges: Fairness vs. efficiency. In *International Conference on Algorithmic Decision Theory (ADT)*.

Farina, G.; Dickerson, J. P.; and Sandholm, T. 2017. Operation frames and clubs in kidney exchange. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Hart, A.; Smith, J. M.; Skeans, M. A.; Gustafson, S. K.; Stewart, D. E.; Cherikh, W. S.; Wainright, J. L.; Boyle, G.; Snyder, J. J.; Kasiske, B. L.; and Israni, A. K. 2016. Kidney. *American Journal of Transplantation (Special Issue: OPTN/SRTR Annual Data Report 2014)* 16, Issue Supplement S2:11–46.

Hooker, J. N., and Williams, H. P. 2012. Combining equity and utilitarianism in a mathematical programming model. *Management Science* 58(9):1682–1693.

Kidney Paired Donation Work Group. 2013. OPTN KPD pilot program cumulative match report (CMR) for KPD match runs: Oct 27, 2010 – Apr 15, 2013.

Li, J.; Liu, Y.; Huang, L.; and Tang, P. 2014. Egalitarian pairwise kidney exchange: Fast algorithms via linear programming and parametric flow. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 445–452.

Liu, Y.; Tang, P.; and Fang, W. 2014. Internally stable matchings and exchanges. In *AAAI Conference on Artificial Intelligence (AAAI)*, 1433–1439.

Mattei, N.; Saffidine, A.; and Walsh, T. 2017. Mechanisms for online organ matching. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Montgomery, R.; Gentry, S.; Marks, W. H.; Warren, D. S.; Hiller, J.; Houp, J.; Zachary, A. A.; Melancon, J. K.; Maley, W. R.; Rabb, H.; Simpkins, C.; and Segev, D. L. 2006. Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet* 368(9533):419–421.

Neuen, B. L.; Taylor, G. E.; Demaio, A. R.; and Perkovic, V. 2013. Global kidney disease. *The Lancet* 382(9900):1243.

Rapaport, F. T. 1986. The case for a living emotionally related international kidney donor exchange registry. *Transplantation Proceedings* 18:5–9.

Rees, M.; Kopke, J.; Pelletier, R.; Segev, D.; Rutter, M.; Fabrega, A.; Rogers, J.; Pankewycz, O.; Hiller, J.; Roth, A.; Sandholm, T.; Ünver, U.; and Montgomery, R. 2009. A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine* 360(11):1096–1101.

Roth, A.; Sönmez, T.; and Ünver, U. 2004. Kidney exchange. *Quarterly Journal of Economics* 119(2):457–488.

Roth, A.; Sönmez, T.; and Ünver, U. 2005a. A kidney exchange clearinghouse in New England. *American Economic Review* 95(2):376–380.

Roth, A.; Sönmez, T.; and Ünver, U. 2005b. Pairwise kidney exchange. *Journal of Economic Theory* 125(2):151–188.

Thornton, C.; Hutter, F.; Hoos, H. H.; and Leyton-Brown, K. 2013. Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. In *International Conference on Knowledge Discovery and Data Mining (KDD)*, 847–855. ACM.

Yılmaz, Ö. 2011. Kidney exchange: An egalitarian mechanism. *Journal of Economic Theory* 146(2):592–618.