# Multi-Modal Donor Exchange*

Anand Siththaranjan†

March 28, 2025

*Preliminary and Incomplete*

This paper is frequently updated. Contact me for the latest version.

**Abstract**

This paper develops a general model of paired donor exchange that integrates multiple donation technologies. Through integration, we are able to enrich the set of potential patient-donor matches over benchmark models while reducing the aggregate risks placed on donors. Under reasonable assumptions of risk preferences that hold in a range of settings, a pairwise exchange mechanism that exhibits various desirable qualities, such as efficiency, stability and strategy-proofness, is constructed. The strength of this mechanism lies in allowing lower risk donations to be undertaken when possible. This mechanism can be applied to the integration of multiple organ exchanges, kidney exchange with desensitization, and exchange with multiple donors under observable risk. We complement the latter application with a more general study without observability assumptions, finding positive results for three-way exchanges under mild assumptions. To showcase the benefits and challenges of integration, a case-study of kidney-liver exchanges is provided. In particular, theoretical results on welfare improvement guarantees over non-integrated exchanges are shown, as well as impossibility results when certain model assumptions are weakened. In simulation using Korean patient population data, it is further demonstrated that there approximately is an 10-20% relative increase in transplants over the baseline that varies with the proportion of risk-tolerant donors.

*We thank Haluk Ergin for his support throughout this project. Furthermore, we are grateful for helpful comments from Tak-Huen Chau, Yuichiro Kamada, Scott Kominers, Jianmeng Lyu, Chris Shannon, and Quitzé Valenzuela-Stookey, as well as assistance from Jason Choi.

†University of California, Berkeley, Department of Electrical Engineering and Computer Science, Berkeley, 2594 Hearst Ave, Berkeley, CA 94720, USA. E-mail: anandsranjan@berkeley.edu

# 1 Introduction

For patients suffering from organ failure, such as renal or liver failure, transplantation from a willing donor is an effective means of improving the patient's life expectancy. However, a key concern is in the donors' health, which can be negatively affected by undergoing the transplant procedure. As such, it is of utmost importance for doctors to ensure that risks and benefits are appropriately balanced and agreed upon by both parties. As opposed to direct donation, which relies on willing and compatible donors, when patients have willing but incompatible donors, paired donor exchange has emerged as a means to take advantage of a *coincident of wants*: when there are two pairs in such a situation, but the patient of each pair is compatible with the donor of the other pair, then they can *exchange* donors and undergo their respective operations.

This setting has been well studied for exchanges in many specific organ markets, such as for kidneys (Roth et al., 2005), livers (Ergin et al., 2020) and lungs (Ergin et al., 2017). Yet in practice, it is possible to allow for exchanges across these markets (CMU, 2019). For example, consider allowing a pair that requires a kidney donation and pair that requires a liver donation, to exchange donors who then donate different organs. By allowing donors to donate organs different than what their patient requires, this can increase the number of compatible donors for that patient and thus allow for more transplant occur. Though this has been done in the real world, there is no centralized mechanism that allows for such an option. Rather than force search costs onto pairs, we propose developing an exchange that integrates multiple modes of donations, such as both kidney and liver donation, into a single paired exchange.

Furthermore, donors whose only options were to undergo riskier transplants in order for their patient to receive a new organ now have more options available to them. In the case of kidneys and livers, the latter is a riskier donation both in mortality and morbidity. Consider a kidney patient whose donor is willing to donate either a kidney or liver, but there is no other kidney patient they can be matched with. As well, there is a liver patient whose donor can donate either organ, and they have the option of being matched in an exchange with a kidney or liver patient. In this case, we can reduce unnecessary risks to the liver patient by matching them with kidney patient. Though the latter is taking on a greater risk than the former, it is the only available option to them. By combining exchanges, we have the option of limiting donor risks through such trades.

This paper takes a market design approach to implementing paired exchange across different modes of donation. That is, by identifying a *coincidence of wants with different* *gives* through integrating new types of donation, such as different modes of donating a certain organ as well as donating different organs, how can we design *desirable* mechanisms. We take a unifying perspective on these problems by creating an abstracted model and showing it's application to various domains under different assumptions. The literature has largely studied how to design organ exchanges for specific organ pools, yet different organs can introduce unique challenges. For example, liver exchanges can allow for two modes of transplants: left and right lobe transplant (Ergin et al., 2020). Both modes of transplants require blood type and size compatibility, just as in a kidney transplant, but since right lobes are bigger than left lobes, it increases the donation opportunities for the recipient. However, right lobe transplants are riskier than left lobe transplants for donors, and thus pairs may have different levels of risk they are willing to undertake. This can lead to non-trivial incentives issues that relate to truthful reporting of a pair's willingness to undergo different donation modes. Similar, due to improvements in medical technology, the standard model of kidney exchange that typically has a single mode of donation can be expanded. Andersson and Kratz (2020) studies the use of immunosuppressant technology to allow patients to be compatible with any patient. This can be seen as a new donation mode, and as emphasized in their work, this mode is less desirable to receive than being matching with an already compatible donor.

A key insight is that many risks are objective. For example, it is well known that the morbidity and mortality risk of the operations are ordered by kidney, left-lobe liver and right-lobe liver. Furthermore, the immunosupressant studied in Andersson and Kratz (2020) is non-toxic, and under the assumption that the cost is sufficiently moderate, it is reasonable to assume that taking an immunosuppressant and being matched with a incompatible donor as a patient is less preferred than a compatible match. Hence we can leverage these assumptions on the commonality of preferences when designing mechanisms. The main piece of private information is willingness level, that is at what point is a modality too risky.

The existence of risks across different types of donations points to another social objective, that is to reduce the risk undertaken by donors by allowing patients to still receive their needed organ while their donor donates via a less risky mode. Integrating these markets has the dual purpose of increasing the number of transplants while reducing aggregate donor risks, hence careful design is required such that both *welfare improvements* can be achieved alongside common market design objectives such as efficiency, weak-core stability, strategyproofness and individual rationality. A mechanism that satisfies these former objectives across a range of environments will be referred to as *desirable*.

In the context of organ exchanges, limitations on cycle size are also vital for practical implementation. To see why, first note that there cannot be legal enforcement of organ exchanges when one donor has already fulfilled their obligation in an exchange when the other hasn't. As such, all donors and patients in an exchange must have simultaneous surgeries. Due to the practical and technical complexity of many simultaneous exchanges, it is more feasible to prioritize pairwise exchanges. That is, exchanges involving only two patient-donor pairs.

We begin our exploration by studying *dual-mode* exchanges, and propose a condition called *weak acyclicity*, a necessary but not sufficient condition for acyclicity in Ergin et al. (2020), that allows their mechanism to be applied in novel settings. In particular, we show this condition applies to kidney donation with a) immunosuppressant technology, and b) two donors. We identify a partial converse result that shows when the existence of certain cycles determines that desirable mechanisms are impossible to construct. We show how this result applies to dual-donor liver exchange.

We then study how to develop mechanisms for more general *multi-modal* exchanges. We identify an ideal property of a compatibility graph that strengthens the notion of acyclicity in Ergin et al. (2020), which we term *separability*. We find that this creates a structural decomposition within the space of modalities, and allows us to develop simple algorithms for multi-modal exchanges. When a multi-modal environment is not completely acyclic, but satisfies such properties in a certain partitional sense, we show that if there are desirable mechanisms for each element of a partition of the modality space, then we can create a modular meta-mechanism that retains the desirable properties if preferences on modalities are common. The promise of this approach lies in settings with arbitrary organs and transplant modalities with an underlying common preference restriction, which can be used to model the integration of exchanges for different organs. As new modes of organ donation are developed in the medical field, both in existing organ exchanges and potentially new ones, having a mechanism that can seamlessly integrate desirable mechanisms specific to each exchange into a single desirable mechanism for the whole exchange allows our approach to be robust to future technological advances. Though it is clear that market integration can have positive welfare effects in this context, the idea that mechanisms specific to each market can be combined to form a new desirable mechanism is novel.

This property naturally emerges when integrating different organ exchanges. An application of key interest is the simple but practical model of integrating kidney and liver exchanges, hence we provide an in-depth qualitative analysis of our solution in this setting. We first study the setting without right-lobe donation technology, where our mechanism generalizes the work of Watanabe (2022), and our solution applied to the same setting is structurally simpler and intuitive. Our theoretical results leverage the commonality of preferences to prove that our mechanism has various desirable properties.

To emphasize the importance of the assumptions within our problem and the implicit qualities of our solution, namely the existence of a common risk preference and the necessity of prioritizing based on organ risk, we then present two impossibility results. First, we find that when preferences can be arbitrary, it is impossible to find a desirable pairwise exchange. Secondly, we consider whether it is possible to find a desirable mechanism that is *neutral* to the labelling of a group of patients as being kidney or liver patients. This alludes to a notion of fairness in treatment of different patient group that is absent from our proposed mechanism due to its prioritization structure. We find that it is impossible to construct such a pairwise exchange mechanism.

To characterize any welfare improvements over the status quo baseline, that is pairwise matchings in each organ exchange independently, we characterize two metrics. One is the number of transplants sustained in a matching, and another is the distribution of transplant types (kidney or liver) for patients matched in the baseline. We show that on both dimensions, we can find a priority order such that our mechanism weakly improves on the status quo. We complement this with simulation results that quantify, using Korean patient data, anticipated gains from integrating the exchanges across various assumptions on risk-tolerance of patients. We find, across such tolerances, anywhere from a 10 to 20 percent relative increase in the number of transplants over the baseline.

We complement our general analysis of multi-modal environments with common preferences by exploring kidney exchange with multiple donors but without any preference assumptions. We provide a negative result when restricted to pairwise exchanges, but a positive result for three-way exchanges under some mild assumptions on the population of distribution of agents.

**Outline.** This paper is organize as follows. Section 2 describes the general model contained within this paper, as well as our desiderata and domain-specific background. Section 3 provides theoretical characterization of exchanges with two donation modes, specifically kidney donation with desensitization (Section 3.1.1 and two risk-ordered donors 3.1.2. Sec-

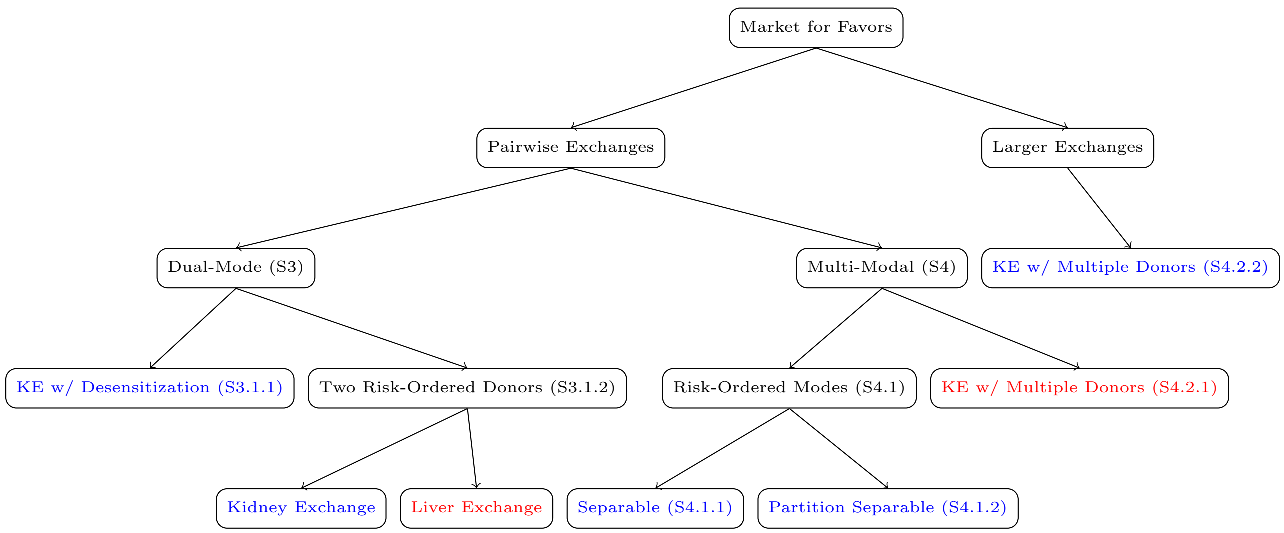



Figure 1: Diagram of Results and Applications. Result type: positive and negative.

tion 4 studies specific environments with more than two modes, providing results on risk-ordered integration of exchanges (Section 4.1) and kidney exchange with multiple donors (Section 4.2.2). Section 5 provides an in-depth analysis of the integrating kidney and liver exchanges in particular, with a focus on existence and uniqueness of desirable mechanisms as well as a welfare analysis. Section 6 provides a simulation analysis of integrating kidney and liver markets. Our theoretical results are summarized in Figure 1.

## 1.1 Related Literature

**Kidney Exchange.** The kidney exchange literature in market design begins with Roth et al. (2004), which provides a modification of the top-trading cycles algorithm (TTC) to allow chains of kidney donation instigated by a deceased donor to be performed. Practical considerations such as pairwise cycle restrictions with dichotomous preferences were explored in Roth et al. (2005), where tools from matching theory were leveraged. More general cycle restrictions we subsequently studied in Roth et al. (2007), where the goal was to identify transplant maximizing exchanges under different size restrictions. When there are multiple donors, Roth et al. (2005) also show that it is dominant strategy to bring all donors when one has dichotomous preferences. On the other hand, when there are preferences over donors we show the non-existence of desirable pairwise mechanisms. We study some weaker desiderata and three-cycle mechanisms to provide positive results.

**Other Exchanges.** Other types of organ donation and technologies have been studied, including liver exchanges (Ergin et al., 2020), dual-donor exchanges like lung and kidney-liver (Ergin et al., 2017), ABO-incompatible kidney donation via sensitization (Andersson and Kratz, 2020), and multi-donor kidney exchange (Gilon et al., 2019). We generalize the results of Ergin et al. (2020) to allow for wider application to dual-mode environments - that is environments with two means of donating or receiving - such as desensitization and two-donor kidney exchange under risk-preference assumptions. We complement the latter by studying the more general setting without the preference assumption. Though our negative results on this echo the impossibility results of Gilon et al. (2019), we use biological restrictions on the set of feasible matchings to identify such a solution whereas their work does not account for this and is in fact not possible in the kidney model. We also study multi-donor liver exchange, using our characterization of dual-mode environments with desirable mechanisms to show the non-existence of such a mechanism in this setting.

**Market Integration.** Similar to our work is that of Watanabe (2022), who take an approach similar to that of Ergin et al. (2020) and apply this to approach to the integration of kidney and liver exchanges. Our algorithm takes a similar form to theirs in that it prioritizes less risky donations first, however the algorithm is distinct in the timing of matches. In particular, we are able to finalize matches as the algorithm proceeds. In their algorithm however, they construct a reduced compatibility graph that they then compute a matching over. A benefit of our approach is that we do not need to consistently check that every agent added is matchable, rather we compute maximum matchings within or between different partitions of the compatibility graph. As well, the structure of our algorithm allows it to be clear how to generalize beyond the kidney-liver model, as we explore in later sections. We also provide a stronger theoretical characterization of our solution by showing weak-core stability and welfare improvements over the baseline. Dickerson and Sandholm (2017) also study a similar problem of integrating kidney and liver markets, but primarily focus on a computational approach to transplant maximization and ignores incentives.

# 2 Model

Let $\mathcal{I}$ be the set of agents, where $N = |\mathcal{I}|$, and $t_i \in \mathcal{T}_i$ as the type of agent $i$. Let $\mathcal{M} = \{m_1, m_2, ...\}$ denote the set of modalities. Unless said otherwise, we assume agents have strict preferences over $\mathcal{M} \cup \emptyset$ such that $m_k \succ_i m_{k+1}$ and $m_1 \succ_i \emptyset$. Let the space of preference be given by $\mathcal{R}$. For a subset of agents $A \subseteq \mathcal{I}$, $m \in \mathcal{M}$ and a preference profile $\succ \in \mathcal{R}^N$, we let $A_\emptyset(m| \succ)$ be the set of agents that do not find $m_i$ individually rational:

$$A_\emptyset(m| \succ) = \{i \in A | \emptyset \succ_i m_i\}$$

Let $\tau_m : \mathcal{T} \times \mathcal{T} \to \{0, 1\}$ be the compatibility function for modality $m \in \mathcal{M}$. That is, an agent $i$ can donate via mode $m$ to agent $j$ if and only if $\tau_m(t_i, t_j) = 1$. Otherwise $\tau_m(t_i, t_j) = 0$. We use the notation $i \to_m j$ to mean that $i$ can donate to $j$ via modality $m$, i.e. $\tau_m(t_i, t_j) = 1$. Correspondingly, if $i$ cannot donate to $j$ via modality $m$, then we use the notation $i \not\to_m j$.

Define an **exchange problem** to be the tuple $\mathcal{E} = (\mathcal{I}, \mathcal{T}, \mathcal{M}, \tau, \mathcal{R})$, and the analogous family of exchange problems to be given by $\mathbf{E} = \{\mathcal{E} = (\{1, \ldots, n\}, \mathcal{T}^n, \mathcal{M}, \tau, \mathcal{R}) | n \in \mathbb{N}, \tau_m : \mathcal{T} \times \mathcal{T} \to \{0, 1\}\}$.

The compatibility graph with respect to an exchange is an edge-labelled directed graph $G_\mathcal{E} = (V, E)$, where $V = \mathcal{I}$ is the set of vertices and $E \subseteq V \times V \times \mathcal{M}$ the set of labelled edges such that $(i, j, m) \in E$ if and only if $i \to_m j$. Let $G_C^m$ denote the directed graph induced by only considering edges corresponding to modality $m$, and $\bar{G}_C^m$ as the undirected graph where an edge between $(i, j)$ exists if $(i, j, m), (j, i, m) \in E$. We say a cycle is an $n$-cycle if it is of length $n$. A matching is a set of disjoint 2-cycles in the compatibility graph. Denote the set of matchings by $\mathbf{M}$, which is implicitly determined by the exchange problem $\mathcal{E}$ being considered. An exchange is a set of disjoint cycles with arbitrary length in the compatibility graph. Denote the set of exchanges by $\mathbb{E}$. Given an $F$ an exchange or matching, denote $\mathcal{I}(F)$ as the set of agents involved in a matching or exchange respectively.

Given a graph $G = (V, E)$, the induced subgraph with respect $A \subseteq V$ is $G(A) = (A, E_A)$ where $E_A = E \cap (A \times A)$. Operations on this graph, such as graph addition and deletion, are defined in Ergin et al. (2020).

We let $\texttt{MaxMatch}(G|\Pi)$ define a maximum cardinality matching of the graph $G$ with priority determined by $\Pi$. For a bipartite graph $G$ with partitions $A$ and $B$, define similarly $\texttt{BipartiteMatch}(G|\Pi, A, B)$. Though both functions maximize the cardinality of a matching in $G$, the computational complexity of the underlying algorithm are different because `BipartiteMatch` exploits the structure of a bipartite graph, whereas `MaxMatch` applies to general graphs.

For $(i, m) \in \mathcal{I} \times \mathcal{M}$ and $A \subseteq \mathcal{I} \times \mathcal{M}$, let $\texttt{Matchable}(i|G, A)$ output `True` if $i \in \mathcal{I}$ can be matched via $m$ in $G$ while ensuring that for all $(j, m') \in A$, $j$ can be matched via $m'$.

Otherwise it will return `False`.

## 2.1 Mechanisms

Given a family of exchange problems $\mathbf{E}$, a family of matching mechanisms is $\phi : \mathbf{E} \times \mathcal{R}^N \to \mathbf{M}$ and a family of exchange mechanisms is $\psi : \mathbf{E} \times \mathcal{R}^N \to \mathbb{E}$. A mechanism with respect to a specific $\mathcal{E} \in \mathbf{E}$ will be denoted $\phi_{\mathcal{E}}$, and if $\mathcal{E}$ or $\mathbf{E}$ is clear from context, we will simply use $\phi$ and refer to it as a mechanism.

## 2.2 Desiderata

In this section, we describe various desiderata. Fix $\mathcal{E} \in \mathbf{E}$ and $\succeq \in \mathcal{R}$.

**Desiderata 1.** A matching $M$ is **individually rational** if for all $i \in \mathcal{I}$, $M \succeq_i \emptyset$.

**Desiderata 2.** A matching M is **Pareto efficient** if there does not exist $M' \in \mathbf{M}$ such that $M \succeq_i M'$ for all $i \in \mathcal{I}$ and there is some $j \in \mathcal{I}$ such that $M \succ_j M'$.

**Desiderata 3.** A matching M is **weak-core stable** if there does not exist non-empty $C \subseteq \mathcal{I}$ and a matching $M' \in \mathbf{M}$ such that for all $i \in C$, $M' \succ_i M$, and for all $i \in C$ and $j = M'_{\mathcal{I}}(i)$, $i \to_{M'_{\mathcal{M}}(i)} j$.

A generally weaker condition is the following:

**Desiderata 4.** A matching M is **pairwise stable** if there does not exist $i, j \in \mathcal{I}$ and $m, m'$ such that $i \to_m j$, $j \to_{m'} i$, $m \succ_i M_{\mathcal{M}}(i)$, and $m' \succ_j M_{\mathcal{M}}(j)$.

A similar definition holds for exchanges, however it is with respect to the full set of exchanges $\mathbb{E}$. Note that when indifferences exist in a model, only the weak-core and not the strong-core[1] is guaranteed to exist.

**Desiderata 5.** A mechanism $\rho$ is **strategy-proof** if for all $i \in \mathcal{I}$, $\succ \in \mathcal{R}^N$ and $\succ'_{-i} \in \mathcal{R}^{-i}$, then $\rho(\succ) \succeq_i \rho(\succ_i, \succ'_{-i})$.

That is, it is a weakly dominant strategy for agents to report preferences truthfully. We say a mechanisms $\phi_{\mathcal{E}}$ is **satisfactory** if for any $\mathcal{E} \in \mathbf{E}$, $\phi_{\mathbf{E}}$ is strategy-proof, and all elements of its range are Pareto-efficient, individually rational, and weak-core stable. For a

[1] The strong core requires that in a blocking coalition, all agents weakly improve and at least one agent strictly improves over their original allocation.

family of mechanisms $\phi$, if for every $\mathcal{E} \in \mathbf{E}$ we have that $\phi_{\mathcal{E}}$ is satisfactory, then we say that $\phi$ is **desirable**. The following simple result highlights that, in the setting of matchings, pairwise stability is sufficient for weak-core stability:

**Proposition 1.** *An individually rational pairwise stable matching is weak-core stable (among matchings).*

## 2.3 Background

We provide some institutional background that motivate common assumptions, considerations and restrictions in the literature.

**Simultaneous Operations.** An important point is when an exchange between multiple pairs is done, it tends to be the case that they are done simultaneously. The reason for this is that if a the patient of a pair has received an organ from another donor, and their donor has not yet donated, there cannot be any punishment to the donor or the patient should the former choose not to donate. That is, we cannot force the donor to donate, nor take back the organ transplanted to the patient. Thus doing sequential or asynchronous operations can lead to a holdup problem. This problem is alleviated when considering chains of donation that are kickstarted by a deceased donation, though in our setting we only consider paired donation.

**Small Exchanges.** The literature often considers pairwise exchanges, that is the restriction that any pair $i$ whose donor donates to the patient of pair $j$ also has the donor of pair $j$ donating to the patient of pair $i$. Though it is possible to do larger size exchanges, and this has been done in reality, there are a number of considerations that make pairwise exchange a good starting point in theory and practice. Due to the requirement of having simultaneous operations, smaller size exchanges are preferred as they can impose a prohibitively large logistical and medical burden on the hospital performing the multiple operations. Though it is also common to have three-way exchanges, since we are introducing different methods of donation in a single model that may require different expertise and thus complications running simultaneous operations, we consider pairwise exchanges primarily and larger exchanges when desirable pairwise exchanges are impossible. Beginning with such an approach in novel environments is common in the literature. Furthermore, from a theoretical and algorithmic standpoint, attempting to maximize the size of a pairwise exchange - which corresponds to having as many transplants via pairwise exchange as possible - is very computationally tractable as it corresponds to finding a maximum matching in a graph. However hardness results are abound as we go beyond pairwise exchanges. Though under very specific assumptions we can make tractable progress towards identifying maximum larger-cycle exchanges, we consider this scope for future work.

**Tissue-Type Compatibility.** From a medical point of view, an important factor in the ability of a donor to be able to donate to a patient is their mutual biological compatibility. This is a product of different types of compatibility, such as blood-type, size and tissue-type compatibility. For all donors and patients, their individual blood-type and size are known, and their mutual compatibility can be inferred from this. As such, we can incorporate this information into a mechanism by treating this as a restriction on the set of feasible exchanges. However, tissue-type compatibility is more difficult to ascertain. In general, the blood of a patient and their prospective donor must be mixed together and tested to determine this. Thus to incorporate this as a restriction on the feasible set requires that every pair of patient and their potential donors be tested in this fashion. This is impractical in reality, and instead we tend to assume that all patients and donors from different pairs are incompatible. That being said, patients and donors from the same pair are often already tested for tissue-type compatibility in the first place, hence we allow pairs in the market to be tissue-type compatible or incompatible. The reason we usually have this information is that donors brought by a patient in a pair are usually those who want to donate specifically to that patient. As such, they must have already checked compatibility prior to entering the paired exchange mechanism. If they are compatible by all requirements but enter the exchange nevertheless, we assume they are *transplant maximizers*: there are willing to engage in an exchange in order to allow other pairs that are incompatible to benefit from a donor swap.

**Algorithmic Implementation.** To find an ideal pair exchange, designers of paired exchange mechanism construct a compatibility graph with nodes as pairs and directed edges representing the ability of a donor in one pair to donate to the patient of another. Furthermore, weights are added to edges in the graph that corresponds to factors like likelihood of transplantation success and priority. Once this has been done, we maximum **weighted** exchange is computed, with relevant cycle length restrictions placed on the set of feasible exchanges.

### 2.3.1 Kidney Exchange

The standard models of kidney exchange primarily consider blood-type compatibility as the main determinant of biological compatibility. A patient's blood type is given by the presence of two antigens, either $A$ or $B$. If there are missing both antigens, we say they have blood type $O$. Hence the set of possible blood types considered are $A$, $B$, $AB$, or $O$. A person $i$ can donate to another person $j$ if whenever $j$ is missing some antigen $k \in \{A, B\}$, then so is $i$. For example, $O$ donors are compatible with all patients as they are missing every antigen, however $O$ patients are only compatible with $O$ donors. On the other hand, $AB$ donors can only donate to $AB$ patients as they are missing no antigen, but as patients they can receive from any donor.

**Desensitization.** Though an ideal kidney exchange is one where the patients and donors are blood-type compatible with one another, there exists technology by which blood-type incompatible donations can be performed. As Andersson and Kratz (2020) note, though the graft survival rates for such a transplant are identical to compatible donations, the main issues with such an approach related to monetary costs of the immunosuppressant, longer waiting time prior to transplantation, and the need for additional medical treatment. As such, it is reasonable to assume that this is a less preferred mode of receiving a kidney than a compatible donation.

**Multiple Donors.** Some patients may have multiple willing but (potentially) incompatible donors that they want to be considered for a paired exchange. In practice, this information is incorporated by adding additional edges in the compatibility graph for every donor brought to the exchange. However, it is not possible to express preferences over these donors, and by computing a maximum weight matching it is not clear which donor may be picked. As a result, patients may strategize about which donor they may choose to bring to the market. Furthermore, given that the exchange mechanism is run repeatedly, they might choose to only bring additional less preferred donors after failing to match with their most preferred donor. Ultimately, this results in an induced game where patients strategize on which set of donors to bring. In general we want patients to bring all their feasible donors rather than stagger their entry throughout time, so as to find maximum exchanges and improve overall welfare.

### 2.3.2 Liver Exchange

In liver exchange, a donor donates a portion of their liver, referred to as a lobe, rather than their whole organ as in kidney exchange. Further more, liver exchange differs from kidney exchange in two key ways: biological compatibility and modes of donation. The main features of compatibility with liver donation is blood type as well as size compatibility. By the latter we mean that a potential donor is compatible with a patient if the lobe they donate is large than what the patient requires. This is important as donors are also able to choose which lobe - the left or right - they donate, given that they differ in size. However, it is known that the right lobe is more dangerous in terms of mortality and morbidity to donate than the left lobe. Given this, we assume that agents will prefer donating their left lobe over their right lobe, as in Ergin et al. (2020). Similarly, kidney donation is also known to be safer on both metrics than liver donation, whether left or right lobe. Thus we maintain the same assumption on preferences when comparing kidney to liver donation[2].

**Desensitization.** Though desensitization is possible, it is mainly done with donations from brain dead patients (Egawa et al., 2023). We discount this as a possibility due to its limited current implementation for living donation, but note that it might have scope for exploration in future work.

# 3 Dual-Mode Exchanges

In this section, we study the setting where our exchange only has two modes, i.e. $\mathcal{M} = \{m_1, m_2\}$. This section provides a generalization of Ergin et al. (2020), which studies a model that characterizes the existence of desirable mechanisms for liver exchange. In their work, the two modes of interest corresponded to donating a left lobe as opposed to a right lobe. It is assumed that objective medical risks associated with donation determines an individuals preference ordering between the two modes. We consider a similar assumption as their environment, and find weaker conditions under which their results hold in a new novel environments such as kidney exchanged with risk-ordered donors and kidney exchange with desensitization. We further find a partial converse result that shows our condition nearly guarantees the non-existence of a desirable mechanism, which is applied to liver

[2]Note this is it not necessarily obviously true. Since the liver regrows whereas the kidney does not, it is plausible to imagine that some agents might have the opposite preference based on the need to "feel whole".

exchange with multiple donors.

In characterizing the existence of desirable mechanisms, we begin with the same graph construction as Ergin et al. (2020). Given $G_C$ and a set of agents $A \subseteq \mathcal{I}$, construct the following digraph $G(A) = (A, E')$: let $(i, j) \in E'$ if and only if

1. $i$ can donate to $j$ via $m_1$, i.e. $i \rightarrow_{m_1} j$, and

2. $j$ can donate to $i$ only via $m_2$, i.e. $j \not\rightarrow_{m_1} i$ and $j \rightarrow_{m_2} i$.

We refer to $G(A)$ as the **precedence digraph**, as in Ergin et al. (2020). We say a pair of agents $i, j \in \mathcal{I}$ is **mutually compatible** via $m \in \mathcal{M}$ if $i \leftrightarrow_m j$. We say that $G(A)$ is **weakly acyclic** if every cycle contains a pair of agents mutually compatible via $m_1$. Furthermore, it is **acyclic** if there are no cycles. This definition of acyclicity follows from Ergin et al. (2020). Let `TopOrder`$(G)$ return a topological order on $G$ if it exists.

**Observation 1.** *$G(\mathcal{I})$ is weakly acyclic if it is acyclic.*

This follows from the fact that if there are no cycles, then it is vacuously true that all cycles contain a pair mutually compatible via $m_1$ donation. We now informally describe the algorithm of Ergin et al. (2020), whose properties we aim to generalize:

1. Compute a maximum match using some priority order via $m_1$ and promise matched agents a match via $m_1$

2. For $k \in$ `TopOrder`: (reverse topological order)

   (a) If $k$ is `Matchable` (while preserving promises), then promise a match via $m_1$

   (b) Transform to $m_2$ otherwise (if $m_2$ is feasible)

3. Compute a maximum match via $m_2$ while maintaining promises.

**Proposition 2.** *If $G(\mathcal{I})$ is weakly acyclic, then the Preference Adaptive algorithm of Ergin et al. (2020) is well-defined.*

*Proof.* Assume $G(\mathcal{I})$ is weakly acyclic. For contradiction, assume there is a cycle $C$ in $\tilde{J}_K$ (using the notation of Ergin et al. (2020)), and thus there does not exist a topological order. Otherwise the operation `TopOrder` is well-defined and the algorithm of Ergin et al. (2020) applies. Denote $G'$ as the graph. Since $G'$ is a subgraph of $G(\mathcal{I})$, then all cycles in the former are also cycles in the latter. Given that $G(\mathcal{I})$ is weakly acyclic, then any cycle contains an $m_1$ mutually compatible pair. Hence there exists $i, j \in C \cap (\tilde{J}_K)$, and thus $i, j \notin J_K$. This is a contradiction as, assuming $i < j$ in the $\Pi_L$ order, then $i$ would not be transformed to $m_2$ since it would be matchable with $j$. Thus $G'$ is acyclic, and there exists a topological order. □

This result considers the potential claim that, without acyclicity, the application of Ergin et al. (2020) to our setting is not valid as a topological order would not exist. The key insight here is that the construction of the order is only necessary **after** agents matched mutually via $m_1$ have been made. Thus a weakly acyclic graph admits a topological order once these initial agents have been matched. Given this, we find weak acyclicity is sufficient for this mechanism to be desirable:

**Theorem 1.** *If $G(\mathcal{I})$ is weakly acyclic then the Preference Adaptive algorithm of Ergin et al. (2020) is a desirable mechanism.*

Since the mechanism is the same as that of Ergin et al. (2020), our proof primarily builds on their analysis by extending their results to weak-core stability, and showing the other properties hold under weaker conditions. We have shown that the condition of weak acyclicity is sufficient, but how about necessity? The following results shows a partial converse to this end.

**Proposition 3.** *Consider a $\mathcal{E} \in \boldsymbol{E}$ that induces a preference digraph that contains a simple cycle with no mutually compatible $m_1$ pairs, and any mutually compatible $m_2$ pairs are adjacent*[3]*. Then there is no satisfactory mechanism.*

This result finds that under some additional constraints on the position of mutually compatible $m_2$ pairs, if there does exist a cycle after removing all mutually compatbile $m_1$ pairs then there is no satisfactory mechanism, much less a desirable mechanisms. The proof follows by contradiciton, where we consider the environment that induces this cycle and assume that there does exists an efficient matching from a strategyproof mechanism. Due to the ability of pairs to only match with agents before or ahead of them in the cycle, we can use the constraints of efficiency and strategyproofness to identify a contradiction. The following illustrates the impossibility result through an example with three agents.

**Example 1** (An environment without a desirable mechanism)**.** Consider the precedence digraph in Figure 2 with the cycle $L_1 \to L_2 \to L_3 \to L_1$, and assume there is a desirable

[3]That is, they have a (possibly directed) edge between them.

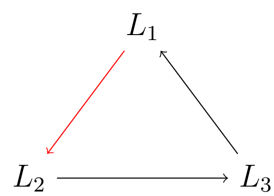


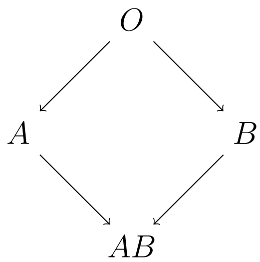

Figure 2: Precedence digraph with the matching for the preference profile $WWW$ in Example 1 highlighted.

mechanism. Since this is a cycle of length 3, and no two agents that are adjacent can be $m_1$-mutually compatible, there are no $m_1$-mutually compatible pairs in this cycle. Let $XYZ$ for $X, Y, X \in \{U, W\}$ be a preference profile, where $W$ means that $m_2$ is IR (i.e. willing) and $U$ when it is not (i.e. unwilling). Note that even if there were mutually compatible $m_2$ pairs, such a matching would never be efficient. Given this and the fact that at most two agents can be matched in a matching, we can denote a matching by $(L_i, L_j)$, where $L_i$ donates via $m_1$ and $L_j$ donates via $m_1$.

First consider the case the profile $WWW$. Without loss of generality, assume that $(L_1, L_2)$ is the matching given by the mechanism, where $L_1$ donates by $m_1$ and $L_2$ donates by $m_2$. Now consider $UUW$, and observe that $(L_2, L_3)$ is the only individually rational and efficient match. For $UWW$, the possible individually rational, efficient matches are $(L_2, L_3)$ and $(L_1, L_2)$. The former would contradict strategyproof-ness of the mechanism, as $L_1$ can misreport from $U$ to $W$ and be strictly better off. However the latter would also contradict strategyproof-ness, as a misreport by $L_2$ from $W$ to $U$ would give them a strictly better outcome. As no match is feasible, this mechanism cannot be desirable. $\triangle$

## 3.1 Applications

We provide a general characterization of settings that are sufficient, and nearly necessary, to admit desirable mechanisms. This approach builds of previous work, and an open question is how practically necessary was our generalization. For example, is it the case that other applications of interest already induce an acylic preference digraph? We motivate the strength of our results through the following applications, two of which provide applications that admit non-acyclic but weakly acyclic environments, and one of which leverages our converse result to show general non-existence of a desirable mechanism for the given setting.




(a) Compatibility Graph

| $X \leq_B Y$ | $O$ | $A$ | $B$ | $AB$ |
|---|---|---|---|---|
| $O$ | ✓ | ✓ | ✓ | ✓ |
| $A$ | | ✓ | | ✓ |
| $B$ | | | ✓ | ✓ |
| $AB$ | | | | ✓ |

(b) Table description

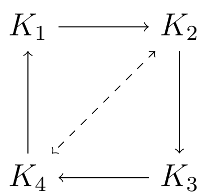
Figure 3: Blood type compatibility.

### 3.1.1 Incompatible Donation via Desensitization

The prototypical model of kidney donation considers agents that consist of patient-donor pairs who are incompatible either due to tissue-type incompatibility or blood-type incompatibility. With the assumption that all patients are tissue-type compatible with all other donors, the problem of maximizing the number of pairwise exchanges is the same as identifying the maximum match in a compatibility graph. Because certain blood types are more rare than others and thus have heterogeneous demand in the exchange market, certain agents fare better than others due to their blood type. To overcome this barrier, the development of novel desensitization has allowed patients to receive transplants from ABO-incompatible donors. In the context of efficient matching, this problem has been studied by Andersson and Kratz (2020). We extend this line of work by showing how this model satisfies weak-acyclicity strictly, and thus there exists a desirable mechanism.

The model is formally described as follows. Consider agents $i \in \mathcal{I}$ with type $X_i - Y_i \in \mathcal{B} \times \mathcal{B}$, where $X_i$ is the blood type of the patient and $Y_i$ is the blood type of the donor. Their ability to participate in an exchange with $j \in \mathcal{I}$ is by one of two modes. The first mode, $m_1$ is the standard exchange mode, where $i \to_{m_1} j$ if $j$'s donor is able to donate to $i$'s patient. That is $X_i \leq_{\mathcal{B}} Y_j$, where the blood type order $\leq_{\mathcal{B}}$ is shown in Figure 3. The second mode is $m_2$, which allows $i$'s patient to undergo desensitization in order to become compatible with $j$'s donor. Note that by doing so, a patient is compatible with all donors. However, this is less desirable than compatiblity via the non-desensitization mode, as noted in Andersson and Kratz (2020).

Andersson and Kratz (2020) note the connection between the strategic incentives in kidney donation with desensitization and those in liver exchange as studied in Ergin et al. (2020). However the following example highlights how the application of the same mechanism will not obviously preserve the same properties. The approach of Ergin et al. (2020)




(a)

| $\mathcal{I}$ | Patient | Donor 1 |
|---|---|---|
| $K_1$ | A | A |
| $K_2$ | B | A |
| $K_3$ | B | B |
| $K_4$ | A | B |

(b) Example observable characteristics

Figure 4: Isolated component, and example observable characteristics. Assume all agents in the component have preferences $K \succ L \succ \emptyset$, that is they find all donations acceptable.

utilizes the existence of a specific acyclic directed graph, which is to draw an edge from $i$ to $j$ if $i$ can interact via $m_1$ but not $m_2$[4]. Again we can consider the induced graph as in Ergin et al. (2020), where a directed edge from $i$ to $j$ means that $j$'s donor is compatible with $i$, but $j$ can only receive a donation from $i$ if they undergo desensitization. Alternatively, we can frame this as $j$ is compatible with $i$ but $i$ is not compatible with $j$, since by being desensitized they are able to receive a donation from any blood type. The following example shows that the graph is not always acyclic.

**Example 2.** *Consider the graph in Figure 4. There is clearly a cycle given by the following patient-donor pairs:*

$$\{(A, A), (A, B), (B, B), (B, A)\}$$

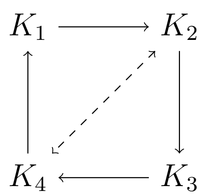
△

However, we can observe that once all donation possibilities via regular, that is without desensitization, pairwise exchanges have been exhausted and we've removed those agents, we would have removed $K_2$ and $K_4$. Thus the remaining graph is acyclic. In the following result, we show that in general the preference digraph is weakly acyclic but not acyclic by showing that only cycles similar to that in Example 2 can exist.

**Proposition 4.** *The digraph is always weakly acyclic, but not always acyclic.*

As a result, we are able to apply the mechanism of Ergin et al. (2020) to this setting, and thus achieve desirable outcomes.

[4]Note that since it is the patient that undergoes desensitization, we frame the modality as an interaction rather than what they donate.




(a)

| $\mathcal{I}$ | Patient | Donor 1 | Donor 2 |
|---|---|---|---|
| $K_1$ | A | A | B |
| $K_2$ | A | B | A |
| $K_3$ | B | B | O |
| $K_4$ | B | A | B |

(b) Example observable characteristics

Figure 5: Isolated component, and example observable characteristics. Assume all agents in the component have preferences $K \succ L \succ \emptyset$, that is they find all donations acceptable.

#### 3.1.2 Two-Donor Exchange

In this section, we interpret the modes of interaction with another agent as having multiple donors by which one can donate the corresponding organ. We consider the case where different donors face different, objective medical risks. Hence they can be ranked in order of the risk imposed on them by undergoing the surgery. We assume the agents preference over which donor donates is reflected in this risk, and without loss we assume that the first (listed) donor is preferred over the second donor in terms of who should donate. This model is studied with respect to kidney and liver exchange and we show a possibility result with the former and am impossibility result for the latter as an application of our more general results.

**Kidney Exchange.** We now consider the standard kidney exchange model, but allow for an agent to list two donors rather than one. This occurs in practice. An agent is a triple composed of a patient and two donors. We let an agent $i$'s type be given by $X_i - Y_i - Z_i$, where $X_i$ is the patient's blood type, and $Y_i$ and $Z_i$ are the blood types of the first and second donor respectively.

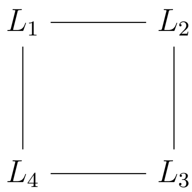
Our modes of donation are $m_1$ and $m_2$, where donation via $m_l$ from agent $i$ to $j$ means that $i$'s $l$-th donor can donate to $j$. Hence a directed edge in our digraph $i \to j$ can be interpreted as $i$'s first donor being compatible with $j$'s patient, but only $j$'s second donor is compatible with $i$.

The following example takes the same structure as the previous example:

**Example 3.** *Consider the graph. There is clearly a cycle given by the following patient-donor pairs:*

$$\{(A, A, B), (A, B, A), (B, B, O), (B, A, B)\}$$

$\triangle$




(a)

| $\mathcal{I}$ | Patient | Donor 1 | Donor 2 |
|---|---|---|---|
| $L_1$ | $A, 1$ | $AB, 2$ | $B, 1$ |
| $L_2$ | $AB, 2$ | $A, 2$ | $A, 1$ |
| $L_3$ | $A, 2$ | $O, 1$ | $O, 2$ |
| $L_4$ | $B, 1$ | $O, 1$ | $O, 2$ |

(b) Example observable characteristics

Figure 6: Counterexample for liver exchange with multiple donors.

**Proposition 5.** *The digraph is weakly acyclic but not acyclic.*

The setting of objective preferences over donors is not without contention. In particular, there may be other considerations beyond medical risk, such as different donors having different abilities to take time off work due to familial obligations. In such a case, this is not reflected in observable medical risks. We provide some insight into the problem when preferences orderings are not observable, providing a negative result for pairwise cycles and a positive result for three cycles.

**Liver Exchange.** Could we extend this to other organ exchanges, such as liver exchange? The model is as follows. Let $\mathcal{S} = \{1, ..., S\} \subseteq \mathbb{R}_+$. An agent $i$'s type is $X_i - Y_i - Z_i$ where $X_i, Y_i, Z_i \in \mathcal{B} \times \mathcal{S}$ refer to the blood-size type of the patient, the first donor, and the second donor respectively. As before, we assume the donors are ordered by risk. Let $X^{\mathcal{B}}$ and $X^{\mathcal{S}}$ refer to the blood type and size of $X$. $Y$ can donate to $X$ if they are blood and size compatible:

1. $X^{\mathcal{B}} \leq_{\mathcal{B}} Y^{\mathcal{B}}$, and
2. $X^{\mathcal{S}} \leq Y^{\mathcal{S}}$

**Proposition 6.** *The setting of multiple donors cannot be implemented in liver exchange.*

*Proof.* Use previous theorem on partial converse in conjunction with the example in Figure 6b. □

Since this is a negative result in a restricted setting, that is of a known common preference ordering, we then have an impossibility result in the general environment with potentially differing preference orderings as well.

# 4 Multi-Modal Exchanges

In this section, we study how to go beyond dual-mode exchanges. We specify a general property call *partition separability*, which applies to a range of environments, as well as study the specific application of multi-donor kidney exchange with more than two donors and heterogeneous preferences.

## 4.1 Risk-Ordered Integration of Multiple Exchanges

We first consider what we call risk-ordered exchanges, where there is an ordered set of different organ exchanges, each potentially with multiple modes. We further assume that agents share a common preference ordering consistent with the exchange ordering over these modes, which is motivated by objective medical risks being ordinally identical across agents and the determinant of their preferences. By placing restrictions on the compatibility structure through properties we call *separability* and *partition separability*, we are able to model risk-ordered exchanges where, respectively,

1. each exchange has only one donation mode
2. if a mode in one exchange is preferred to another mode in a different exchange, then this is the case for every mode in both exchanges.

Though the latter case subsumes the former, we provide a detailed description of the former for intuition. In either case, we identify desirable mechanisms. We conclude this subsection by formally describing the application of integrating multiple organ exchanges.

### 4.1.1 Warm-Up: Separable Exchanges

To provide intuition for the main results of this section, we begin by studying what we call *separable exchanges*. Consider a setting with $N$ modes (i.e. $|\mathcal{M}| = N$). We say $G_C$ is **separable** if there exists $\{A_n\}_{n=1}^N$ such that $\mathcal{I} = \oplus_{n=1}^N A_n$ and for all $n \in \{1, \ldots, N\}$, $i \to_{m_n} j$ only if $j \in A_n$. We can observe a connection between seperability and acyclicity as follows:

**Proposition 7.** *If $G_C$ is separable, then $G$ is acyclic with respect to any distinct $m_i$ and $m_j$.*

*Proof.* Fix distinct $m_i, m_j \in \mathcal{M}$. Consider the graph $G_{m_i,m_j}(\mathcal{I})$. Assume for contradiction that there is a cycle such that $C = (i_0, \ldots, i_{K-1})$. For any $k \in \mathbb{N}$, $i_{k \mod K} \to i_{(k+1) \mod K}$

implies, by separability, that $i_{k \mod K} \to_{m_i} i_{(k+1) \mod K}$ and $i_{(k+1) \mod K} \to_{m_j} i_{k \mod K}$. By separability, $i_{(k+1) \mod K} \in A_i$ and $i_{k \mod K} \in A_j$. Since this is true for arbitrary $k$, then we have that for any $k \in \{0, ..., K-1\}$, $i_k \in A_i \cap A_j$. This is a contradiction since $\{A_n\}$ is a partition of $\mathcal{I}$, and thus $A_i \cap A_j = \emptyset$. Thus there cannot exist a cycle. $\square$

In effect, separability requires that acyclicity holds for any pair of modes, regardless of the order of the mode.

Note that if $G_C$ is separable, when designing a mechanism we can simply use the disjoint sets given by $A_i = \{j \in \mathcal{I} | \exists i \in \mathcal{I},\ i \to_{m_i} j\}$ and not consider agents outside of these sets as they would not be compatible via any modality.

Consider the following mechanism, which leverages the separability of $G_C$. Assume $G_C$ is separable, and consider $\{A_i\}$ described above as the partition. Fix a preference profile $\succ \in \mathcal{R}^N$, and let $\phi$ be a mechanism such that $\phi(\succ)$ is the output of the following algorithm. Let $M = \emptyset$.

1. Process through $n \in \{1, \ldots, N\}$:

   (a) Remove unwilling $m_n$ agents: for each agent $i \in \mathcal{I}$

   $$\emptyset \succ_i m_n \implies \mathcal{I} \leftarrow \mathcal{I} - \{i\}$$

   (b) Consider $\bar{G}_n = G_C^{m_n}(A_n \cap \mathcal{I})$. Find a maximum match via $m_n$ within $\bar{G}_n$:

   $$M \leftarrow M \cup \texttt{MaxMatch}(\bar{G}_n | \Pi)$$

   (c) Remove matched agents: $\mathcal{I} \leftarrow \mathcal{I} - \mathcal{I}(M)$.

   (d) Process through $o \in \{n+1, \ldots, N\}$

      i. Remove unwilling $m_o$ agents from $A_n \cap \mathcal{I}$: for each agent $i \in A_n \cap \mathcal{I}$

      $$\emptyset \succ_i m_o \implies \mathcal{I} \leftarrow \mathcal{I} - \{i\}$$

      ii. Consider $\bar{G}_{n,o} = G_C((A_n \cup A_o) \cap \mathcal{I})$. Find a maximum bipartite match between $A_n$ and $A_0$ within $\bar{G}_n$:

      $$M \leftarrow M \cup \texttt{BipartiteMatch}(\bar{G}_{n,o} | \Pi, A_n, A_o)$$

      iii. Remove matched agents: $\mathcal{I} \leftarrow \mathcal{I} - \mathcal{I}(M)$.

Intuitively, the mechanism operates as follows. We order agents by risk, treating those that require lower risk donations to have higher priority. In the first stage, we compute a maximum match between all agents in the highest priority class. In doing so, these agents get their best option, as not only receive the organ required but also donate via the least risky mode. We proceed to prioritize these agents by considering an individually rational bipartite match between agents in the highest priority class with those in the second highest priority class. From the perspective of agents in the highest priority class, as all opportunities to match with another top priority class agent have been exhausted (since the match in the first stage was maximum), then this is their second best option. On the other hand, agents in the second highest priority class that are matched get their best choice. As we repeat this procedure of bipartite matching until we reach the lowest priority class, though we have prioritized the highest class, each class on the other side of the bipartite matching get their best option. Thus they cannot be improved upon. At this point, we have exhausted all feasible matching opportunities for the highest class, and we can repeat this procedure by replacing the highest with the second highest class in order to exhaust their opportunities. It is this ordering of matching that is associated with the common preference structure that allows us to find a mechanism with ideal properties:

**Theorem 2.** *If $G_C$ is separable, then $\phi$ is a desirable mechanism.*

We are able to decompose our mechanism into a combinations of general and bipartite matching due to the separable structure of $G_C$. Can we generalize this process beyond separability? The following section provides such a generalization.

### 4.1.2 Partition Separability

We say $G_C$ is **partition separable** if there exists a partition $\{\mathcal{M}_k\}_{k=1}^K$ of $\mathcal{M}$ such that

1. $\mathcal{M}_k$ is **contiguous**, that is if $m_a, m_b \in \mathcal{M}_k$ then for all $c$ such that $a \leq c \leq b$, $m_c \in \mathcal{M}_k$, and

2. there exists $\{A_k\}_{k=1}^K$ such that $\mathcal{I} = \oplus_{k=1}^K A_k$ and for all $k \in \{1, \ldots, K\}$ and $m \in \mathcal{M}_K$, $i \rightarrow_m j$ only if $j \in A_k$.

Contiguity allows us to say that, under a common preference assumption, that the partition is contiguous with respect to this preference. The second condition is analogous to that in separability, whereby to donate via mode in some partition element $k$, the agent receiving via that mode must belong to the associate partition of agents. We provide some insight into weakening contiguity by considering other restrictions in Section **??** of the Appendix. Again, we have a connection between acyclicity and partition separability:

**Corollary 1.** *If $G_C$ is partition separable, then for $k \neq k'$, $G$ is acyclic with respect to any $m_k \in \mathcal{M}_k$ and $m_{k'} \in \mathcal{M}_{k'}$.*

*Proof.* This follows from the same argument as in the previous proposition. □

Consider the following mechanism, which leverages the weak separability of $G_C$. Assume $G_C$ is separable, and let $\{(\mathcal{M}_k, A_k)\}_{k=1}^K$ be described as above. Fix a preference profile $\succ \in \mathcal{R}^N$, and let $\psi$ be a mechanism such that $\phi(\succ)$ is the output of the following algorithm. Let $M = \emptyset$.

1. Process through $k \in \{1, \ldots, K\}$:

   (a) Remove unwilling $\mathcal{M}_k^1$ agents: for each agent $i \in \mathcal{I}$

   $$\emptyset \succ_i \mathcal{M}_k^1 \implies \mathcal{I} \leftarrow \mathcal{I} - \{i\}$$

   (b) Consider $\bar{G}_k = G_C^{\mathcal{M}_k}(\mathcal{I} \cap \cup_{l \in \mathcal{M}_l} A_l)$. Find a matching via $\phi_k$:

   $$M \leftarrow M \cup \phi_k(\succ, \bar{G}_k)$$

   (c) Remove matched agents: $\mathcal{I} \leftarrow \mathcal{I} - \mathcal{I}(M)$.

   (d) Process through $l \in \{1 + \sum_{p=1}^{k-1} |\mathcal{M}_p|, \ldots, \sum_{p=1}^{k} |\mathcal{M}_p|\}$:

      i. Process through $o \in \{1 + \sum_{p=1}^{k} |\mathcal{M}_p|, \ldots, N\}$

         A. Remove unwilling $m_l$ agents from $A_o \cap \mathcal{I}$: for each agent $i \in A_o \cap \mathcal{I}$

         $$\emptyset \succ_i m_l \implies \mathcal{I} \leftarrow \mathcal{I} - \{i\}$$

         B. Consider $\bar{G}_{l,o} = G_C((A_l \cup A_o) \cap \mathcal{I})$. Find a maximum bipartite match between $A_l$ and $A_o$ within $\bar{G}_{l,o}$:

         $$M \leftarrow M \cup \texttt{BipartiteMatch}(\bar{G}_{l,o} | \Pi, A_l, A_o)$$

         C. Remove matched agents: $\mathcal{I} \leftarrow \mathcal{I} - \mathcal{I}(M)$.

**Theorem 3.** *If $G_C$ is partition separable and for each $k \in \{1, \dots, K\}$ there exists $\phi_k$ desirable with respect to $G_C(A_k)$, then $\psi$ is a desirable mechanism.*

The intuition behind this mechanism follows that of the separable setting. The main difference in this setting is the modular structure that takes advantage of desirable exchange-specific mechanisms instead of computing a maximum match.

The utility of this formulation is in its robustness to future medical developments. Given that new technologies in medicine are continuously being developed, allowing for novel donation modalities and thus more ways by which individuals can receive or donate an organ, this poses an issue in paired exchange when risks differ across modes. We have already identified how this occurs between exchanges, for example kidney and liver, and within exchanges, such was with left and right lobe liver donation. Market designers working in this domain must often attend to the specific structure of the problem, such as biological compatibility, to create desirable matching mechanisms. By allowing said designers to focus on individual organ exchanges and, under the assumption of a common risk-order, not on the integration of multiple organ exchanges, our mechanism can easily develop alongisde new technologies. An example of early stage research on new donation modes includes intestinal transplant. Though currently not commonly done due to the increased donor risks and lower quality of the transplanted organ, should the technology become sufficiently safe for donors and effective for patients, it is likely to be objectively riskier for donors than donating a kidney or liver. As such, this would satisfy assumptions within our model. Future work should study how to relax our risk-ordering assumption, which is less likely to hold as more modes are introduced. The following section describes major applications where our separability and partition separability structures are satisfied.

### 4.1.3 Application: Multiple Organ Exchanges

Where does this structure appear? Between organ exchanges! Let a family of organ exchanges $\{\mathbb{E}_\alpha\}_{\alpha \in \mathcal{A}}$ induce an exchange $\mathbb{E}_\mathcal{A}$ as follows:

1. An agent $i$ belonging to $\mathbb{E}_\alpha$ with type $T_i$ has new type $T_i - \alpha$ in $\mathbb{E}_\mathcal{A}$
2. The set of modes in $\mathbb{E}_\mathcal{A}$ is $\mathcal{M}_\mathcal{A} = \cup_{\alpha \in \mathcal{A}} \mathcal{M}_\alpha$
3. An agent $i$ can donate to an agent $j$ in $\mathbb{E}_\alpha$ only by some mode $m \in \mathcal{M}_\alpha$

**Corollary 2.** *Consider this family of organ exchange problems* $\{\mathbb{E}_\alpha\}_{\alpha\in\mathcal{A}}$*, where* $\mathcal{A}$ *is a finite ordered set. Let* $G$ *be the compatibility graph induced by* $\mathbb{E}_{\mathcal{A}}$*. If each exchange* $\mathbb{E}_\alpha$ *has a*

1. *single mode, then* $G$ *is separable.*
2. *(potentially) multiple modes, then* $G$ *is partition separable.*

*Proof.* This proof follows from acyclicity between exchanges, which is a result of the fact that donating via a mode belong to a certain exchange can only be donated to an agent in that exchange. □

This observation allows us to see that when families of organ exchange problems have a common ordering, then our previous results highlight the existence of a desirable mechanism. An example of this would be in integrating kidney and liver exchanges:

**Example 4** (Kidney-Liver Exchange)**.** *An agent* $i$*'s type is given by* $X_i - Y_i - O_i$ *where* $X_i, Y_i \in \mathcal{B}\times\mathcal{S}$ *is the patient and donors blood-size type (as in the liver exchange model), and* $O_i \in \{K, L\}$ *refers to the organ required by the patient. There are two modes of donation,* $m_1 = K$ *and* $m_2 = L$*. Hence* $i \to_{m_1} j$ *if* $X_i \leq_{\mathcal{B}} X_j$ *and* $O_i = K$*, and* $i \to_{m_2} j$ *if* $X_i \leq_{\mathcal{B}} X_j$*,* $Y_i \leq Y_j$ *and* $O_i = L$*.* △

We provide a more in depth analysis of kidney-liver exchange without right lobe donations in a subsequent section. We remove the possibility of right lobe donation primarily for analytic simplicity. Our mechanism, though more general, presents a similar structure as that of Watanabe (2022) in terms of the resulting matching. We detail a comparison in later sections.

## 4.2 Kidney Exchange with Multiple Donors

We give some insight into the multi-donor kidney exchange problem. Unlike before, we do not assume that the donors are ordered by risk. Instead we consider the more general setting of any preference over donors, and show some tractability in finding a desirable mechanism. Though weak acyclicity proved a key role in the dual-mode applications, it is not clear how to leverage a similar structure for this multi-donor application since this goes beyond multiple modes. Furthermore, we no longer have a common preference assumption. As such we take an alternate approach based on the structure of the underlying compatibility relation. We describe the some necessary assumptions as well as the algorithm below, and later sections provide some further intuition and discussion.

**Comment on strategyproofness.** It might be plausible to reason that our definition of desirable can be weakened. In particular, we could consider a constrained notion of strategyproofness where agents cannot say that a mode is feasible if it is their true preferences state that it is not feasible. One could motivate this as follows: blood type is verifiable, hence when a donor is proposed as a feasible option, their blood type can be readily checked. Note however that an infeasible mode is said to be feasible, but it is never used, then there may be no means of ensuring that this mode was actually feasible. This can be problematic if our mechanism relies on modes that are not used and strictly worse for the agent than the mode they are implemented with in the matching. In such a case, an agent can affect the outcome by changing the order among their comparatively worse modes. Thus, such an assumption of verifiability warrants caution, though it may be a practical consideration nevertheless.

### 4.2.1 Impossibility of Pairwise Mechanisms

Past work on the same problem of kidney exchange with multiple donors. By example, they aim to show the impossibility of a desirable mechanism under a pairwise cycle constraint. However, as we show below, their example is not possible under the standard blood type compatibility model. In particular, this compatiblity structure restricts the set of feasible trades. We say a set of pairwise matches are **rationalizable** if there exists a choice of blood types for patients and donors that make these matches as exactly the feasible set of matches.

**Proposition 8.** *The set of feasible pairwise matches given in Figure 1 of Gilon et al. (2019) is not rationalizable.*[5]

This result is illustrative about the importance of accounting for the blood-type compatibility structure. Though we nevertheless find an impossibility result for (pairwise) matching mechanisms in the subsequent proposition, we will later leverage some properties this compatibility structure for three-cycle mechanisms.

**Proposition 9.** *There does not exist a desirable matching mechanism for kidney exchange with multiple donors.*

[5] Figure 1 of Gilon et al. (2019) was used to prove the analogous negative result in their Theorem 3.

#### 4.2.2 A Two-and-Three Cycle Mechanism

We now explore the existence of desirable mechanisms when cycles of size at most three can be used. We consider various assumptions either for simplicity in the design of the mechanism, or based in realistic assumptions on the distribution of agents.

Our first assumption simplifies the design of our mechanism by removing from consideration an uncommon blood type[6].

**Assumption 1.** *There are no patients with an AB blood type.*

This next assumption is common in the literature, as stated in Roth et al. (2007), which is motivated by the intuition that there are many $O-A$ and $O-B$ pairs due to blood-type incompatibility between patient and donor[7] and the high proportion of $O$ blood types[8]

**Assumption 2** (Long-side of the Market)**.** *At least one of each type in* $O-b$ *for* $b \in \{A, B\}$ *is unmatched in any feasible matching.*

This final assumption is for simplicity of the mechanism, similar to that in Roth et al. (2007) except we preclude the possibility of no $b-b*$ agents:

**Assumption 3.** *There are at least two* $O-O*$, $B-B*$, *and* $A-A*$.

Let `Max-3-Match` compute a maximum exchange with cycles of size at most 3, analogous to `MaxMatch`. Without loss of generality, we assume $n_{A-B*} \geq n_{B-A*}$, and our algorithm is analogous in the case where $n_{A-B*} < n_{B-A*}$[9]. When we say that an agent $i$ of type $X-YZ*$ **drop** their preferred donor, we transform their type to be $X-Z*$ and eliminate their donor $Y$ from being considered. We use the following notation to represent this: $i \leftarrow X-Z*$.

Consider the following mechanism:

1. Maximum 3-match $\mathcal{I}_{B-B*}$
2. Maximum 2-match $\mathcal{I}_{A-B*}$ and $\mathcal{I}_{B-A*}$
3. Maximum $O-A* \rightarrow A-B* \rightarrow B-O*$

[6] Approximately 4% in the US.

[7] This leverages the idea that patients tend to enter the exchange after trying to use their donors but failing due to compaitiblity.

[8] Approximately 48% in the US.

[9] We detail this case in the Appendix, though it is not needed in the proof of strategyproofness of our algorithm in the edge-case where $n_{A-B*} = n_{B-A*}$ and a $A-B, A$ agent deviates to $A-A$ for example.

4. If no $A - B*$
    (a) continue
5. If no $B - O*$
    (a) $A - B*$ drop their preferred donor
6. Apply serial dictator where agents are $O - A*$ and $O - B*$ and objects are $A - O*$ and $B - O*$. Afterward, remaining $O - b*$ drops their $A$ or $B$ donors.
7. Maximum 3-match $\mathcal{I}_{A-A*}$
8. Maximum 3-match $\mathcal{I}_{O-O*}$

We now state the main properties of this mechanism:

**Theorem 4.** *This mechanism is PE, IR, and SP.*

When there are no cycle constraints, variants of TTC that account for indifferences can easily be applied to our problem by the following. Let preferences over donation modes induce preferences over agents where agent $i$ is strictly preferred to agent $j$ by agent $k$ if $k$ can donate to $i$ using a donation mode that is strictly better than any donation mode that $k$ can use to donate to $j$. Thus approaches that leverage top trading cycles can be used.

However, in our environment we require exchanges to be composed of two or three way cycles, that is cycles in an exchange are of length at most three. It is well known in the literature that imposing cycle constraints makes it such that there is generally no mechanism that is efficient, strategyproof, and individually rational for house exchange. So why is it possible in this environment to find such a mechanism when there is a cycle constraint? The following result gives some intuition by showing that if there is a top trading cycle, then there is one that has length two:

**Proposition 10.** *Consider a graph where all agents point to their favourite feasible agent. If there is a cycle, then there must be a two cycle.*

This result, and its proof, show that the structure of compatibility relation and donor preferences induces a certain structure on preferences over agents that is somewhat compatible with the concept of a top trading cycle. Though this provides some intuition for why it is possible, note however that we cannot use the TTC algorithm itself due to indifferences, and some approaches that adapt TTC to allow for indifferences cannot clearly be adapted to leverage this 2-cycle existence property. Further work should explore how to exploit such structure to show existence of satisfactory mechanisms in general.

We conclude by identifying some useful properties that our mechanism satisfies.

**Limited Donor Lists.** We say that a mechanism is **$k$-donor equivalent** for $A \subseteq \mathcal{I}$ if for any pair $i \in A$, their outcome in the mechanism is the same than if they truncated their preference to their top $k$ alternatives.[10]

**Outcome Maximum.** We say that a mechanism is **outcome maximum** if given only the donors used in the outcome of the exchange, it is a maximum exchange.

**Proposition 11.** *The mechanism is*

1. *outcome maximum,*
2. 1*-donor equivalent for* $\mathcal{I}_{B-*}$*, and*
3. 2*-donor equivalent for* $\mathcal{I}_{A-*}$*.*

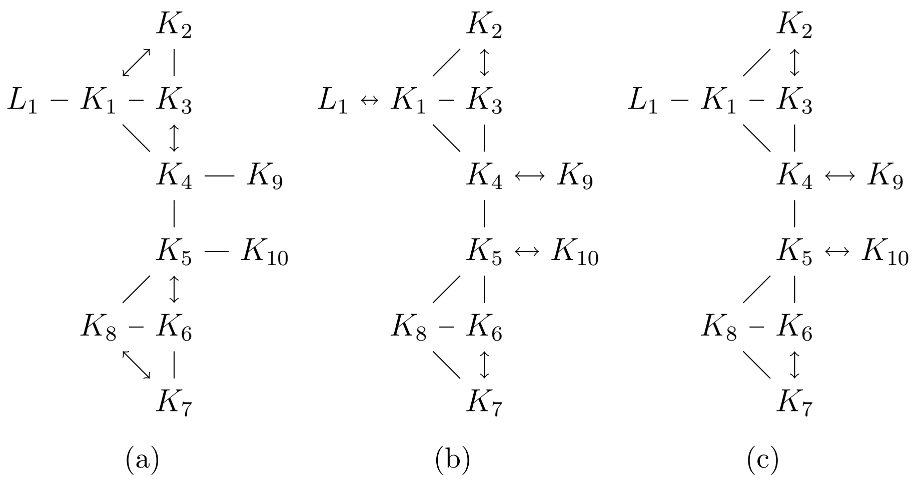
*Proof.* The first claim can be seen by observing the worst case outcomes throughout the algorithm, noting that we proceed down an agent's donor list. For the second claim, we refer to Roth et al. (2007) where our outcome is consistent with their findings for maximum two- and three-way cycles. □

# 5 Analysis of Kidney-Liver Exchanges

We provide a simple illustration of the approach, challenges and benefits by looking solely at integrating kidney and liver exchanges. We will focus on liver exchanges with only left-lobe transplantation allowed, for simplicity. Before studying these properties, we will compare our approach to previous work. We will denote our proposed mechanism by $f$.

Combining organ exchanges have been studied academically in previous work by Watanabe (2022) and Dickerson and Sandholm (2017). In comparison to the former, our mechanism is distinct in its structure, generalizable to new settings, modular in that it can utilize

[10]Formally, for a given preference $\succ_i$ over $\mathcal{M} \cup \{\emptyset\}$, denote the **$k$-truncated preference** $\succ_i^k$ such that there exists $m^l \in \mathcal{M}$ distinct where $m^1 \succ_i \cdots \succ_i m^{|\mathcal{M}|}$ and $m^1 \succ_i^k \cdots \succ_i^k m^k \succ_i^k \emptyset \succ_i^k m^{k+1}$. Then a mechanism $\phi$ is **$k$-donor equivalent** for $A$ if for all preference profiles $\succ$ and for all agents $i \in A$, then $\phi(\succ) = \phi(\succ_i^k, \succ_{-i})$.




| $\mathcal{I}$ | Patient | Donor |
|---|---|---|
| $K_1$ | AB, 1 | A, 6 |
| $K_2$ | A, 1 | B, 1 |
| $K_3$ | AB, 1 | A, 1 |
| $K_4$ | A, 1 | O, 1 |
| $K_5$ | B, 1 | O, 1 |
| $K_6$ | AB, 1 | B, 1 |
| $K_7$ | B, 1 | AB, 1 |
| $K_8$ | AB, 1 | B, 1 |
| $K_9$ | O, 1 | A, 1 |
| $K_{10}$ | O, 1 | B, 1 |
| $L_1$ | AB, 5 | A, 3 |

(d) Example types

Figure 7: Isolated component, and example types. Assume all agents in the component have preferences $K \succ L \succ \emptyset$, that is they find all donations acceptable. $-$ indicates biological feasibility, and $\leftrightarrow$ indicates edges in a matching.

other organ-specific mechanisms, and leverages properties of the compatibility graph in order to efficiently match individuals throughout the algorithm, rather than at the end.

In Dickerson and Sandholm (2017), they consider a setting without preferences or strategic behaviour, and allow for chains, cycles of length greater than 2, and altruistic donors. Whereas the goals of their analysis lie in the potential transplantation gains from trade in a general context, ours has a similar goal with added considerations due to preferences over donation that informs efficiency, stability and strategyproofness.

## 5.1 (Non-)Uniqueness

Is our mechanism unique? Consider the following mechanism $g$:

1. Identify components in $\mathcal{G}$ as in Figure 7d and match all agents according to Figure 7b if individually rational, and Figure 7c otherwise.

2. Denote matched agents as $\mathcal{I}_{K\leftrightarrow L}$ and remove them from them for $\mathcal{I}$: $\mathcal{I} \leftarrow \mathcal{I} - \mathcal{I}_{K\leftrightarrow L}$.

3. Match $\mathcal{I}$ according to $f$.

Intuitively, this mechanism matches all subsets of agents that belong to the isolated component and have the described structure, and for the remaining agents implements $f$.

Observe that the matches specified in Figure 7a is a maximum match that would arise in our proposed mechanism $f$.

**Theorem 5.** *$g$ is a distinct mechanism from $f$ that satisfies PE, IR, SP, and weak-core stable.*

*Proof.* Clearly $g$ is individually rational. It is distinct from $f$ by observing that, in the isolated component, the kidney pairs would be matched by $f$ instead of to the liver pairs as in $g$.

Consider agents in the isolated component, if any. If it is individually rational for both pairs of K-L pairs to be matched with one another, then $g$ will match them. Otherwise, they are not matched at this stage and they will be potentially matched under $f$. Observe that as the $L$ pairs get their best outcome, they have no incentive to misreport. If the $K$ pair misreports to say they are unwilling to match with a $L$ pair, then they will remain unmatched since they belong to an isolated component (hence are mutually compatible with no other pair outside of the component) and their only potential match would be the other $K$ pair, which is already matched. Thus they would remain unmatched, which is strictly worse for them according to their true preferences. Since $f$ is IC, this whole mechanism is IC.

The mechanism is PE as the pairs in the isolated component who can be improved, that is the $K$ pairs, can only be improved by unmatching with the $L$ pairs, who have no other matching possibilities. Since $f$ is PE, the whole mechanism is PE.

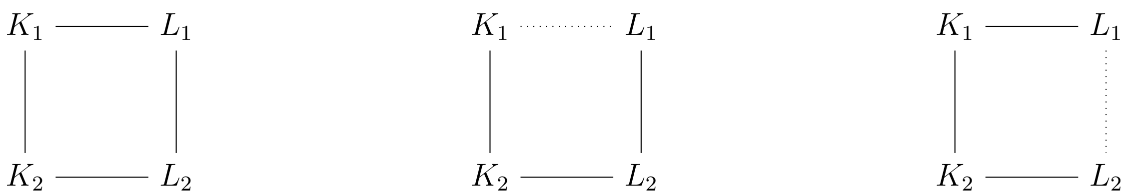
To see pairwise stability, observe that in the isolated component the only agents that $K_1$ can't form a blocking pair with any agent $K_i$ since they are all matched to some $K_j$ and thus would not be strictly improved upon. □

## 5.2 Impossibility Results

In this section we state some impossibility results in various cases. We collate the results in the following theorem, whose details are explained subsequently.

**Theorem 6.** *Consider a pairwise mechanism that is strategyproof, Pareto efficient, individually rational, and pairwise stable. Then it cannot*

1. *hold for heterogeneous preferences,*
2. *maximize the number of transplants, or*
3. *satisfy neutrality*




(a) (b) (c)

Figure 8: Example compatibility graph observable characteristics.

**Heterogeneous Preferences and Pairwise Exchanges.** Our solution thus far relies on this assumption of a common risk ordering. A natural question is whether this assumption was necessary. That is, can we find an efficient, IR and strategyproof mechanism when preferences over transplants can be arbitrary? Our following result shows that this is not possible:

**Proposition 12.** *Let preferences be arbitrary and only allow for pairwise exchanges. Then there is no efficient, IR and strategyproof mechanism.*

*Proof.* Consider Figure 8a. Let the pairs have the following preferences:

$$L \succ_{K_1} K \succ_{K_1} \emptyset$$
$$K \succ_{K_2} L \succ_{K_2} \emptyset$$
$$L \succ_{L_1} K \succ_{L_1} \emptyset$$
$$K \succ_{L_2} L \succ_{L_2} \emptyset$$

There are two possible matchings: $\{(K_1, L_1), (K_2, L_2)\}$ or $\{(K_1, K_2), (L_1, L_2)\}$.

For the first matching, consider the case if $L_1$ reports $L \succ_{L_1} \emptyset \succ_{L_1} K$. To ensure strategyproofness, we cannot allow our efficient (IR) matching under this new preference profile to match $L_1$ to $L_2$. If we match $K_2$ to $L_2$, then the former would have an incentive to report $K \succ_{K_2} \emptyset \succ_{K_2} L$, which by efficiency and IR would result in $K_2$ being matched to $K_1$. This would be a profitable deviation and thus not possible by strategyproofness. Hence the only possible match is between $K_1$ and $K_2$. But this match is not efficient as $L_1$ and $L_2$ can also be matched.

For the second matching, consider the case if $L_2$ reports $K \succ_{L_2} \emptyset \succ_{L_2} L$. As before, we cannot allow our mechanism to match $K_2$ and $L_2$ by strategyproofness. If $K_1$ and $K_2$ are matched, then $K_1$ reporting $L \succ_{K_1} \emptyset \succ_{K_1} K$ would be a profitable deviation as, by efficiency and IR, $K_1$ and $L_1$ must be matched. Thus the only possible match would be

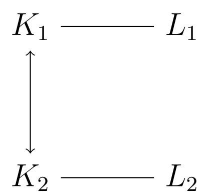



Figure 9: $\mathcal{G}$

between $K_1$ and $L_1$. However, again this match would not be efficient as $K_2$ and $L_2$ can also be matched.

Since neither matchings are possible, then no mechanism that satisfies the stated properties exists. □

Note that the restriction to pairwise exchanges is important. In the given example, the cycle $(K_1, L_1, L_2, K_2)$ would result in all agents getting their preferred choice. We explore in later sections how removing limits on cycle size can give positive results when there are heterogeneous preferences.

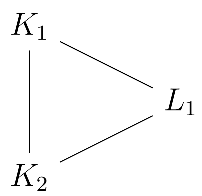
**Transplant Maximization.** One of our welfare criteria is transplant maximization. Is it possible to have a mechanism that implements a maximal matching of the compatibility graph, subject to individual rationality?

**Proposition 13.** *There is no transplant maximal, strategyproof and IR mechanism, nor a transplant maximal, pairwise stable and IR mechanism.*

*Proof.* Observe that the transplant maximal matching given by $\{(K_1, L_1), (K_2, L_2)\}$ is not strategy proof or pairwise stable. □

**Equal Treatment of Exchange Pools.** Implicitly, we are favoring kidney pairs over liver pairs as a consequence of the risk ordering. This could be viewed unfavourably, and we may want to consider a mechanism that treats either exchange pool equally. We say a mechanism $g$ satisfies neutrality if swapping organ labels (without changing preferences) does not change the outcome. Note that $f$ is not neutral:

**Example 5.** Consider the example in Figure 10, where the set of agents is $\mathcal{I} = \{K_1, K_2, L_1\}$ and the compatibility graph in Figure 10a is generated by the observable characteristics in Figure 10b. Assume all donations are individually rational. Then $f$ results in the following matching: $K_1$ and $K_2$ matched via kidney donation. However if we swapped the




(a) $\mathcal{G}$

| $\mathcal{I}$ | Patient | Donor |
|---|---|---|
| $K_1$ | $(B, 2)$ | $(A, 3)$ |
| $K_2$ | $(A, 2)$ | $(B, 3)$ |
| $L_1$ | $(AB, 2)$ | $(O, 1)$ |

(b) Example observable characteristics

Figure 10: Compatibility graph and example observable characteristics. Note that the patient of all pair is compatible with the donor of other pairs, but not with their own donor.


(a) $\mathcal{G}$ (b) $\mathcal{G}'$

| $\mathcal{I}$ | Patient | Donor |
|---|---|---|
| $K_1, L'_1$ | $(B, 2)$ | $(A, 2)$ |
| $L_1, K'_1$ | $(A, 2)$ | $(B, 2)$ |
| $L_2, K'_2$ | $(B, 2)$ | $(A, 2)$ |

(c) Example observable characteristics

Figure 11: Compatibility graphs and example observable characteristics. Note that the patient of all pairs is not compatible with their own donor.

organ labels, then for some $K \in \{K_1, K_2\}$, the following is the matching: $K_1$ and $K$ are matched. △

We can see that there is no mechanism satisfying our desiderata while also being organ-anonymous:

**Proposition 14.** *There is no IC, IR, neutral and PE mechanism, and there is no IR, neutral, PE and pairwise stable mechanism.*

*Proof.* Consider the environments in Figure 11. An IR, PE and either IC or pairwise-stable mechanism must match $(K_1, L_1)$ in $\mathcal{G}$ and $(K'_1, K'_2)$ in $\mathcal{G}'$. Otherwise, pairs can misreport their preferences to force a better match. Alternatively, there are blocking pairs. Note that swapping labels maintains the same compatibility graph, but the outcomes are different. Hence such a mechanism cannot be neutral. □

## 5.3 Welfare Comparison with Non-Integrated Exchanges

A motivation for integrating exchange pools is to improve the number of transplants and reduce the risks taken by donors. We characterize this by comparing our mechanisms with a baseline efficient matching mechanism for each exchange pool separately.

For $o \in \{K, L\}$, let $b^o$ be an efficient matching mechanism for agents in $\mathcal{I}_o$ that uses some priority order. Denote the joint mechanism by $b$, and $f$ as our kidney-liver mechanism. We consider the following metrics. For $R \in \mathcal{R}$, $A \subseteq \mathcal{I}$, $D \subseteq \{K, L, \emptyset\}$, and $\phi$ a mechanism, $\mathcal{I}_\phi : \mathcal{R} \to \mathbb{R}$ and $N_\phi : \mathcal{R} \to \mathbb{R}$ are defined as

$$\mathcal{I}_\phi^{A,D}(R) = \{i \in A | \phi_m[R](i) \in D\}$$

and $N_\phi^{A,D}(R) = |\mathcal{I}_\phi^{A,D}(R)|$. In words, $\mathcal{I}_\phi^{A,D}(R)$ is the set of agents in $A$ who are matched via some mode $m \in D$. When $A = \mathcal{I}$ or $D = \{K, L\}$, we will use suppress reference to these variables. When comparing mechanisms, we consider the following criteria:

1. If for all $R \in \mathcal{R}$, $N_\phi(R) \geq N_\psi(R)$, then we say that $\phi$ **weakly increases the number of transplants** over $\psi$.

2. If for all $R \in \mathcal{R}$,

   (a) $N_\phi^{\mathcal{I}^L,L}(R) \leq N_\psi^{\mathcal{I}^L,L}(R)$ (less donors in the liver pool donate livers),

   (b) $N_\phi^{\mathcal{I}^L,K}(R) \geq N_\psi^{\mathcal{I}^L,K}(R)$ (more donors in the liver pool donate kidneys), and

   (c) $N_\phi^{\mathcal{I}^K,K}(R) \geq N_\psi^{\mathcal{I}^K,K}(R)$ (more donors in the kidney pool donate kidneys),

   then we say that $\phi$ **weakly reduces (unnecessary) donor risks** over $\psi$.

To interpret the second comparison criteria, we say that unnecessary donor risks are reduced when there are less liver donations and more kidney donations from pairs in the liver pool. We use the phrase unnecessary as indicating that more liver pairs can receive a liver while undergoing a safe donation. Furthermore, note that kidney patients only donate livers after exhausting their kidney donation opportunities, hence any liver donation on their part is necessary.

**Proposition 15.** *Let $f$ and $b$ use the same priority order. Then $f$ weakly increases the number of transplants and reduces donor risks over $b$.*

Clearly there are environments where our claim holds strictly.

# 6 Simulations

In this section, we study in simulation the welfare impact of proposal to integrate more donation modes into paired donor exchange. Our main welfare metric of interest when evaluating a matching is in the number of pairs matched. More sophisticated mechanisms may be able to account for other welfare-relevant criteria, however that is beyond the scope of this work. We primarily focus on integrating kidney and liver markets, and for simplicity focus on left-lobe only donation.

| **Parameter** | **Value** |
|---|---|
| **Blood Type Probabilities (ABO)** | O: 0.37<br>A: 0.33<br>B: 0.21<br>AB: 0.09 |
| **Mean Sizes (cm)** | Female (F): 157.40<br>Male (M): 170.70 |
| **Standard Deviation of Sizes (cm)** | Female (F): 5.99<br>Male (M): 6.40 |
| **Sex Probabilities** | Female (F): 0.3<br>Male (M): 0.7 |
| **Kidney to Liver Ratio** | $6274/32016 \approx 0.196$ |

Table 1: Configuration Parameters

We consider aggregate population statistics that determine biological compatibility from South Korean patients, detailed in Figure 1, as in Ergin et al. (2020). For the same reason as these authors, we consider this population due to the country being a world leader in living liver transplantation. Because of this, gains from integration will be shown in a realistic setting.

To construct our simulated population, we randomly sample $n$ kidney (patient-donor) pairs and $m$ liver pairs according to the kidney-to-liver patient ratio listed. We only consider patients that are incompatible with their donors, and their sexes are drawn randomly from the listed probabilities. Given this, their blood types and relevant sizes are drawn randomly, the latter of which is drawn from a normal distribution with the mean and standard deviation as given. Finally, there is some independent probability $p$ of an agent being willing to donate a liver. For different choices of $n$ and willingness probability $p$ we give a heatmap of the relative increase (Figure 12) and absolute increase (Figure 13) of our integrated mechanism over the baseline mechanism, i.e. maximum matches in each organ market separately. We report the average values over 100 random simulation for these metrics.

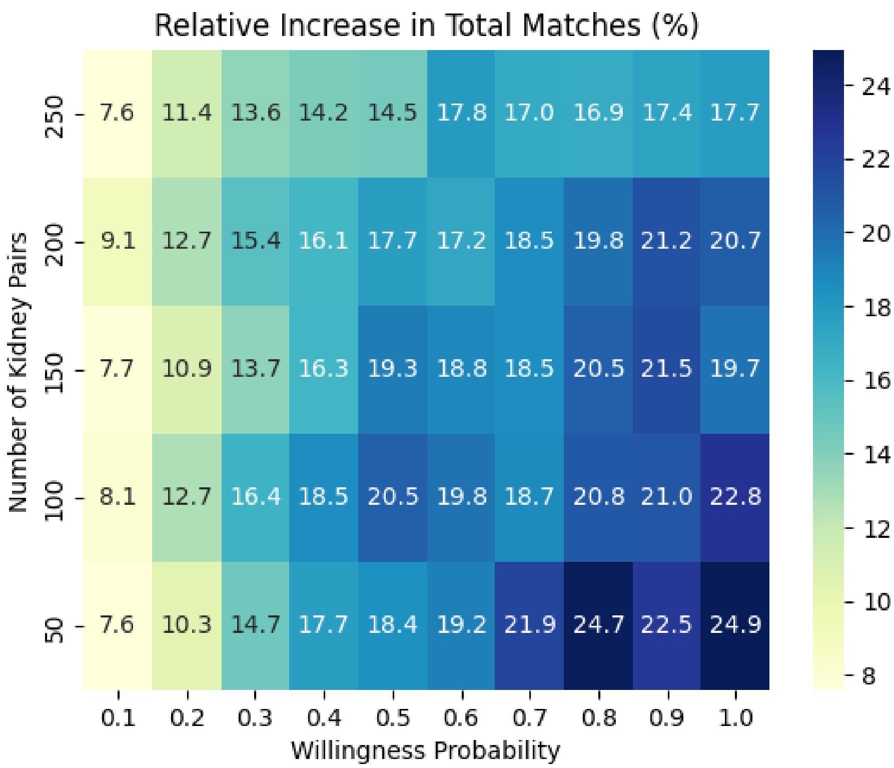



Figure 12: Relative Increase in Transplants

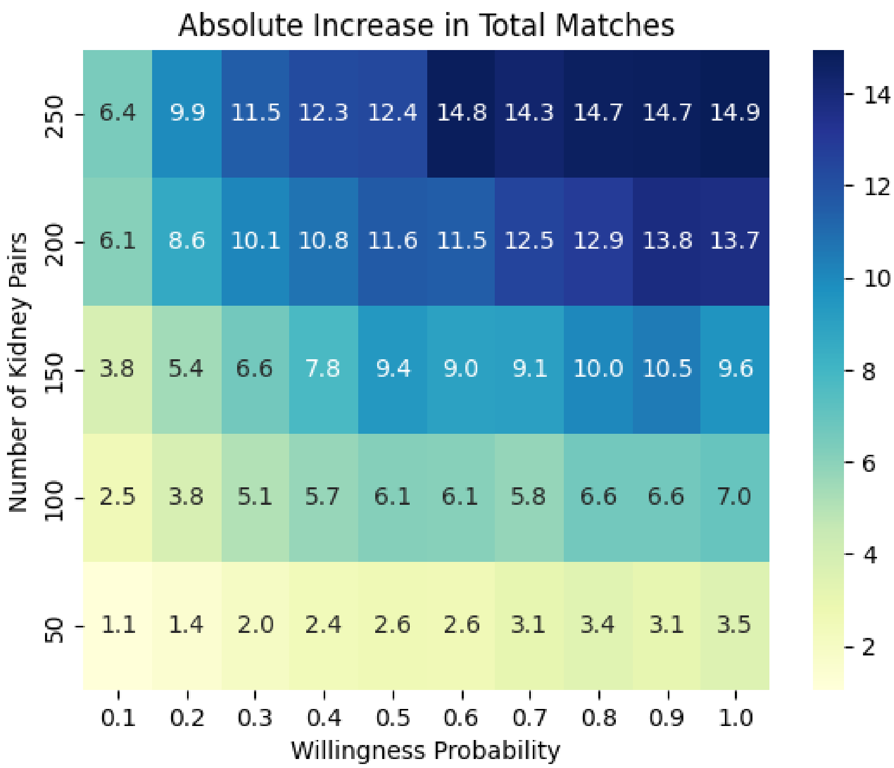



Figure 13: Absolute Increase in Transplants

# 7 Conclusion

In this work we consider a range of environments in paired organ exchange that share the feature of incorporating new donation or transplantation technologies. Future work should consider a more in-depth empirical analysis to understand the welfare gains of incorporating richer preference information into the design of paired exchange mechanisms.

# References


Andersson, Tommy and Jörgen Kratz (2020), “Pairwise kidney exchange over the blood group barrier.” *The Review of Economic Studies*, 87(3), 1091–1133. 3, 7, 12, 17

CMU (2019), “Computer science idea triggers first kidney-liver transplant swap.” `https://www.cmu.edu/news/stories/archives/2019/may/kidney-liver-swap.html`. Accessed: 2024-05-26. 2

Dickerson, John P. and Tuomas Sandholm (2017), “Multi-organ exchange.” *J. Artif. Int. Res.*, 60(1), 639–679. 7, 30, 31

Egawa, Hiroto, Hideki Ohdan, and Kazuhide Saito (2023), “Current status of abo-incompatible liver transplantation.” *Transplantation*, 107(2), 313–325. 13

Ergin, Haluk, Tayfun Sönmez, and M Utku Ünver (2017), “Dual-donor organ exchange.” *Econometrica*, 85(5), 1645–1671. 2, 7

Ergin, Haluk, Tayfun Sönmez, and M Utku Ünver (2020), “Efficient and incentive-compatible liver exchange.” *Econometrica*, 88(3), 965–1005. 2, 3, 4, 7, 8, 13, 14, 15, 17, 18, 37, 42

Gilon, Oren, Tal Gilon, and Assaf Romm (2019), “Kidney exchange with multiple donors.” *Available at SSRN 3451470*. 7, 27

Roth, Alvin E, Tayfun Sönmez, and M Utku Ünver (2004), “Kidney exchange.” *The Quarterly journal of economics*, 119(2), 457–488. 6

Roth, Alvin E, Tayfun Sönmez, and M Utku Ünver (2007), “Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences.” *American Economic Review*, 97(3), 828–851. 6, 28, 30

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005), “Pairwise kidney exchange.” *Journal of Economic Theory*, 125(2), 151–188, URL `https://www.sciencedirect.com/science/article/pii/S0022053105001055`. 2, 6, 62

Watanabe, Moyuru (2022), “Multiorgan exchange: A liver-kidney paired exchange perspective.” *SSRN*. 5, 7, 26, 30

# A Proofs

## A.1 Proof of Proposition 1

*Proof.* Consider a pairwise stable matching $M \in \mathbf{M}$. Assume for contradiction that $M$ is not weak core stable. Then there is a set of agents $U$ and a matching $M'$ that is strictly improving for all agents in $U$. Let $i$ be an agent matched in $M'$. Note that there must be some agent matched since $M$ was an IR matching and all agents in $U$ were strictly improved upon. Thus there are at least two agents $i$ and $j$ who are matched to each other in $M'$. Since $i$ and $j$ strictly preferred to be matched each other than their matching in $M$, they form a blocking pair. This contradicts the pairwise stability of $M$. Thus, $M$ is weak-core stable. $\square$

## A.2 Proof of Theorem 1

*Proof.* Let $\phi$ be the Preference Adaptive algorithm of Ergin et al. (2020). IR, strategyproofness and efficiency directly follows from the proof in Ergin et al. (2020).

Recall that individual rationality and pairwise stability are a necessary and sufficient condition for a matching to be weak-core stable. Assume for contradition that there is a preference profile $\succ$ such that $M = \phi(\succ)$ is not weak-core stable. Then there is a pair $i, j$ and $m, m' \in \mathcal{M}$ such that $i$ is compatible via $m$ with $j$, $j$ is compatible via $m'$ with $i$, $m \succ_i M_{\mathcal{M}}(i)$, and $m' \succ_j M_{\mathcal{M}}(j)$.

Note that we do not need to consider the case where either agent receives a match via $m_1$, that is $M_{\mathcal{M}}(i) = m_1$ or $M_{\mathcal{M}}(j) = m_1$, as there is no means by which either agent can strictly improve their match.

Consider the case where $M_{\mathcal{M}}(i) = M_{\mathcal{M}}(j) = \emptyset$. Given that they were not matched in $M$, and the final step output a maximum matching, it must be that either $i$ or $j$ are unwilling to be matched via $m_2$. Without loss of generality, assume it is $i$. Then $m_2 \prec_i \emptyset = M_{\mathcal{M}}(i)$, so it must be that $m = m_1$. Since a maximum match via $m_1$ bilaterally was found in the first step, it must be that $m' = m_2$ and $i \notin \mathcal{I}_{m_1 \leftrightarrow m_1}$. Thus both agent are in $\mathcal{I} - \mathcal{I}_{m_1 \leftrightarrow m_1}$. Since $i \to j$, it must be that $j$ is transformed first. When $i$ is reached when processing the topological order, it must be that in $M$, $i$ is removed from the graph given that they are unwilling and unmatched. However, $i$ is matchable with $j$, and there cannot be any promises that cannot be any promises that restrict this as both agents are unmatched at the end of the algorithm. Hence $i$ would be promised an exchange via $m_1$, which is a contradiction.

Now consider the case where $M_{\mathcal{M}}(i) = m_2$ and $M_{\mathcal{M}}(j) = \emptyset$. Since both agents strictly improve their outcome, it must be that $m = m_1$ and $m' \in \{m_1, m_2\}$. If $m' = m_1$, then $i$ and $j$ are both matchable via $m_1$ with each other while keeping all promises in the second step. This is a contradiction as they would have both been promised a match via $m_1$. Now consider $m' = m_2$. Since $i$ is not matched via $m_1$, they must have not been promised a match in the second step. Since $j$ is unmatched and $i \in \mathcal{I} - \mathcal{I}_{m_1 \leftrightarrow m_1}$, then $j$ would be transformed to an $m_2$ agent, since they must be willing for $m'$ to be individually rational, and $i$ would be matchable with them via $m_1$. This is a contradiction as $i$ is not matched via $m_1$. Note that a symmetric argument applies for $M_{\mathcal{M}}(j) = m_2$ and $M_{\mathcal{M}}(i) = \emptyset$.

Finally, consider the case where $M_{\mathcal{M}}(i) = M_{\mathcal{M}}(j) = m_2$. Then it must be that $m = m' = m_1$. By the same logic in the previous paragraph, this is not possible. As in all cases it is impossible for there to be a blocking pair, it cannot be that there is a blocking pair. Hence the mechanism is pairwise stable, and consequently weak-core stable. □

## A.3 Proof of Proposition 3

*Proof.* Fix $\mathcal{E} \in \mathbf{E}$. We proceed by considering two cases: the simple cycle has even length, or odd length.

First consider a simple cycle $C = (i_1, j_1, ..., i_K, j_K)$ of length $2K$. Note that it must be that $K > 1$, since otherwise we have that $i_1 \to i_2 \to i_1$, which is a contradiction as $i_1$ can donate to $i_2$ via $m_1$ but the latter cannot donate to the former via $m_2$. Assume for contradiction that there is a desirable mechanism $f$. Note that $C$ contains agents of distinct types. Let $\mathcal{I} = C$, and assume all agents find $m_2$ feasible. By efficiency, there is some two successive agents matched. Without loss let this be $i_1$ and $j_2$. Furthermore, $i_1$ donates via $m_1$ and $j_2$ donates via $m_2$ since no agents are mutually compatible via $m_1$, and if they were both matched via $m_2$ there would be a Pareto improvement by having one agent donate via $m_1$ since by assumption, that is possible. If $j_1$ report that $m_2$ is not IR, then $j_1$ is unmatched in $C'$ by strategy-proofness. If $i_2$ is unmatched, then there would be a Pareto improvement by matching $j_1$ and $i_2$. Thus $i_2$ must be matched with $j_2$. If $j_2$ misreports, then $i_3$ and $j_3$ must be matched by the same argument. Inductively continuing this process of progressive misreporting, we find that if $j_K$ misreports, then $i_1$ must be matched with $j_1$. However this is not individually rational for $j_1$ since they reported that $m_2$ is not individually rational. Thus there does not exists a desirable mechanism.

Now consider a simple cycle $C = (i_1, j_1, ..., i_K, j_K, i_{K+1})$ of length $2K+1$ for $K \geq 1$, since cycles of length 1 would result in a contradiction as before. Again, assume for contradiction that there is a desirable mechanism. If all agents are willing, then by efficiency there must be some matched agent. Let $(i_1, j_1)$ be matched without loss of generality. Now assume $i_1$ becomes unwilling. We now proceed with an induction argument to show that for any $K$, there is a contradiction. Consider the base case, $K = 1$. Hence we have that $(i_1, i_2)$ are matched, and $i_3$ is unmatched. If $i_2$ says they are unwilling, then we have that they must be matched with $i_3$ by Pareto efficiency. However, this would contradict strategproofness. Let the inductive hypothesis be that if we have a chain of length $2K + 1$ where $(i_1, j_1)$ are matched when $i_1$ is willing, then there is a contradiction. Consider $2K + 3$. Then either $(i_1, j_1)$ or $(j_1, i_2)$ are part of the matching. If this did not hold, then it must be that neither $i_1$ or $j_1$ are matched with any agent, due to individual rationality. However there would be a Pareto improvement to matching both agents together, which is a contradiction. If it is the case that $(j_1, i_2)$ are matched, then observe that there are an even number of agents (apart from $i_1$) and thus we can use the argument from the even cycle setting to obtain a contradiction. Hence consider the case that $(i_1, j_1)$ are matched. If $j_1$ is unwilling, then by strategyproofness, they cannot be matched with $i_2$. To maintain Pareto efficiency, we have that $(i_2, j_2)$ are matched. Observe that the chain $(i_2, \ldots, i_{K+2})$ is of length $2K + 1$, hence we obtain a contradiction through the inductive hypothesis. □

## A.4 Proof of Proposition 4

*Proof.* Note that if $i \to j$, then $j$ can donate to $i$ but $i$ can't donate to $j$. For simplicity, and to be in line with previous definition of the direction of the arrow, we flip the direction of every arrow to mean that $i$ can donate to $j$ but $j$ can't donate to $i$. This preserves all cycles. First we show that the above cycle is the only cycle in the graph. Consider some cycle $C = (i_1, i_2, ...)$. Since $i_2$ cannot donate to $i_1$ via $m_1$, then it must be that $i_2$'s donor $A$, $B$ or $AB$. Furthermore this must be true for every agent's donor since this a cycle. Since $i_1$'s donor is $A$, $B$ or $AB$, then consider the following cases: $i_1$'s donor is

- $A$: $i_2$ is $A$ or $AB$
- $B$: $i_2$ is $B$ or $AB$
- $AB$: $i_2$ is $AB$

If $i_2$ is

- $A$: $i_3$'s donor is $B$ or $AB$

- $B$: $i_3$'s donor is $A$ or $AB$
- $AB$: not possible as any type can donate to $AB$

Hence we can conclude that no agent has type $AB$. Thus agents can only have type $A$ or $B$, since there are no donors that can donate to $O$ agents. If there is an $AB$ donor, there must be an $AB$ agent. Hence there are no $AB$ donors either. Thus if $i_1$ is

- $A$: $i_2$'s donor is $B$
- $B$: $i_2$'s donor is $A$

Consider $i_1$ having patient-donor type $A - A$. Then $i_2$ must be either $A - B$. $i_3$ must then be $B - B$, $i_4$ is $B - A$, and $i_5$ is $A - A$. Since all possible types are in this cycle, and their successor is uniquely determined, this is the only cycle. Since there is a cycle, the graph is not acyclic. To see that it is weakly acyclic, note that $B - A$ and $A - B$ are mutually compatible. Hence all cycles in this graph have $m_1$-mutually compatible. □

## A.5 Proof of Proposition 5

*Proof.* Consider some cycle $C = (i_1, i_2, ...)$. Since $i_1$ is not compatible with $i_2$'s first donor, then $i_2$ must be of type $A$, $B$ or $O$ since if they were $AB$, then they would be compatible any donor. Since this is a cycle, then average agent is of type $A$, $B$ or $O$. Furthermore, given that $i_1$'s first donor can donate to $i_2$, then $i_1$'s first donor must be of type $A$, $B$ or $O$. Again, this must be true for all agents. Since $i_2$'s first donor is not compatible with $i_1$, then $i_2$'s first donor cannot be $O$ as they are compatible with all agents. This implies that $i_1$'s first donor cannot be $O$, hence $i_2$ cannot be $O$ either. Thus every patient and every first donor is either $A$ or $B$. Consider the patient-donor tuple $A - A - X$, where $X$ is the the type of the second donor. If corresponded to $i_1$, then $i_2$ must be $A - B - X$. Furthermore, $i_3$ will be $B - B - X$, $i_4$ will be $B - A - X$, and $i_5$ will be $A - A - X$. All cycles must have this form as any agent will have a type in this cycle, and their successor will have the same form. Clearly such a cycle exists (as given by the previous example), hence the digraph is acyclic. To see that this is weakly acyclic, note that $i_2$ and $i_4$ are $m_1$-mutually compatible since they can donate to each other via their first donor. Hence all cycles in this graph have $m_1$-mutually compatible agents. □

## A.6 Proof of Theorem 2

*Proof.* Clearly the mechanism is individually rational.

We now prove Pareto efficiency. Assume for contradiction that there is a matching $M'$ that Pareto dominates the $M = \phi(\succ)$. Let $i$ be such that $M'_{\mathcal{M}}(i) = m' \succ m = M_{\mathcal{M}}(i)$. We proceed by strong induction on $\mathcal{I}_{m_k \leftrightarrow m_l}$, the set of agents matched via an $m_k$-$m_l$ swap, to show that no $i$ in these set can be improved upon. $k, l$ are ordered as follows: $(1,1), \ldots, (1, |\mathcal{M}|), (2,2), \ldots, (2, |\mathcal{M}|), \ldots, (k,k), \ldots, (k, |\mathcal{M}|), \ldots, (|\mathcal{M}|, |\mathcal{M}|)$. In other words it is generated by the following process: $L = \{\}$

- For $k \in \{1, \ldots, |\mathcal{M}|\}$:
  - For $l \in \{k, \ldots, |\mathcal{M}|\}$
    - $L \leftarrow L \cup \{(k,l)\}$

Let $\rhd$ represent the order over $L$. Let $\mathcal{I}_{\emptyset \leftarrow m_k}$ be the set of agents in $A_k$ that were unmatched in $M$. Observe that all agents that are matched must belong to one of these sets.

Consider the base case $k = l = 1$. Clearly no agent in this set can be strictly improved as all agents are matched via $m_1$, the best modality to be matched via. Assume the induction hypothesis holds, that no set prior to $(k,l)$ can have an agent strictly improved upon. Now consider the successor to $(k,l)$. There are two possibilities: $l < |\mathcal{M}|$ or $l = |\mathcal{M}|$.

Consider the first case, then the successor to $(k,l)$ is $(k, l+1)$. Consider an agent $i \in \mathcal{I}_{m_k \leftrightarrow m_{l+1}}$ that donates via $m_k$ in $M$. To improve, they must be matched via $m_{k'}$. Furthermore, by separability, any agent they match with must donate via $m_{l+1}$. Assume that in $M'$ they are in $\mathcal{I}'_{m_{k'} \leftrightarrow m_{l+1}}$. Either $\mathcal{I}_{m_{k'} \leftrightarrow m_{l+1}} \subseteq \mathcal{I}'_{m_{k'} \leftrightarrow m_{l+1}}$ or not. In the former case, note that since our algorithm iterates in the same order $\rhd$, then agent $i$ is unmatched in the step of the algorithm that matches agents in $\mathcal{I}_{m_{k'} \leftrightarrow m_{l+1}}$. This is a contradiction as we compute a maximum match, however we can strictly increase the size of the match by including $i$ and $M_{\mathcal{I}}(i)$. In the latter case, we have some agent $j$ previously in $\mathcal{I}_{m_{k'} \leftrightarrow m_{l+1}}$ that is no longer in $\mathcal{I}'_{m_{k'} \leftrightarrow m_{l+1}}$. Since $M'$ Pareto dominates $M$, it must be that $j$ is strictly improved upon. This contradicts the inductive hypothesis. A similar argument can be made for the successor being $(k+1, l)$.

Now consider $i \in \mathcal{I}_{\emptyset \leftarrow m_k}$. To strictly improve on them, they must be matched via some modality $m_l$. Hence by separability they are in $\mathcal{I}'_{m_l \leftarrow m_k}$. We can apply a similar argument to the previous paragraph to show that it must be that some agent previously in $\mathcal{I}_{m_l \leftarrow m_k}$ is strictly improved upon (due to maximality of the match). Again, this is a contradiction as we showed that no agent previously matched can be strictly improved upon. As all agents belong in some $\mathcal{I}_{\emptyset \leftarrow m_k}$ or $\mathcal{I}_{m_l \leftarrow m_k}$, and no agents from these sets can be strictly improved upon, then the existence of a Pareto dominating matching $M'$ is a contradiction. Thus $M$ is Pareto efficient.

Now we show strategyproofness of the mechanism. Fix $\succ$ and $M = \phi(\succ)$. Consider some deviation by agent $i$ to $\succ'_i \neq \succ_i$, and let the resulting matching be $M' = \phi(\succ')$ where $\succ' = (\succ'_i, \succ_{-i})$. Let $m_k = M_{\mathcal{M}}(i)$, and let $m_l$ and $m_n$ be the least modality such that $m_l \succ_i \emptyset$ and $m_n \succ'_i \emptyset$. By individual rationality, we have that $k \geq l$. Consider the case where $n > k$. Then $M'_{\mathcal{M}}(i) = \emptyset$ is the outcome of the algorithm, that is $i$ is unmatched and thus a strictly worse outcome since $M$ was individually rational. Hence this is not a profitable deviation. If $k > n$, then there is no change in the outcome. That is $M_{\mathcal{M}}(i) = M'_{\mathcal{M}}(i)$. Hence there is no profitable deviation. Thus there is no profitable deviation for $i$, and thus $\phi$ is strategyproof.

We conclude with showing weak-core stability. Recall that it is sufficient for us to show the pairwise stability holds, as weak-core stability is equivalent to individual rationality and pairwise stability for matchings. Assume for contradiction that there is a pair of agents $i$ and $j$ that form a blocking pair. That is, there exists $m$ and $m'$ such that $i \rightarrow_m j$, $j \rightarrow_{m'} i$, $m \succ M_{\mathcal{M}}(i)$, and $m' \succ M_{\mathcal{M}}(j)$. By separability, $j \in A_m$ and $i \in A_{m'}$. Without loss of generality, assume $m \succeq m'$. Consider the step of the algorithm corresponding to $\mathcal{I}_{m \leftrightarrow m'}$. Observe that the size of this set can be improved by matching $i$ and $j$, and that neither agent would be matched at this stage as $(M_{\mathcal{M}}(i), M_{\mathcal{M}}(j)) \rhd (m, m')$ since $m \succ M_{\mathcal{M}}(i)$ and $m \succeq m'$. This contradicts the set being maximum, which must be case since the algorithm implements a maximum (bipartite or general) match amongst all agents that find such a match feasible and have not been previously matched, and thus this is not possible. Hence the mechanism is pairwise stable, and along with individual rationality implies weak-core stability. □

## A.7 Proof of Theorem 3

*Proof.* Clearly the mechanism is individually rational.

We now prove Pareto efficiency. Assume for contradiction that there is a matching $M'$ that Pareto dominates the $M = \psi(\succ)$. Let $i$ be such that $M'_{\mathcal{M}}(i) = m' \succ m = M_{\mathcal{M}}(i)$. We proceed by strong induction on $\mathcal{I}_{m_k \leftrightarrow m_l}$, the set of agents matched via an $m_k$-$m_l$ swap for $m_k$ and $m_l$ in different elements of the partition $\{\mathcal{M}_i\}$, and $\mathcal{I}_{\mathcal{M}_k}$, the agents in $\cup_{l \in \mathcal{M}_l} A_l$ matched under $\phi_k$, to show that no $i$ in any of these set can be strictly improved upon.

Consider the following order $\triangleright$:

$$\mathcal{M}_1, (\mathcal{M}_1^1, \mathcal{M}_{>1}^1), \ldots, (\mathcal{M}_1^1, \mathcal{M}_{>1}^{-1}),$$
$$(\mathcal{M}_1^2, \mathcal{M}_{>1}^1), \ldots, (\mathcal{M}_1^{-1}, \mathcal{M}_{>1}^1), \ldots,$$
$$(\mathcal{M}_1^{-1}, \mathcal{M}_{>1}^{-1}), \ldots, \mathcal{M}_k,$$
$$(\mathcal{M}_k^1, \mathcal{M}_{>k}^1), \ldots, (\mathcal{M}_k^{-1}, \mathcal{M}_{>k}^{-1}), \ldots, \mathcal{M}_{-1}$$

In other words, it is generated by the following process: $L = \{\}$

- For $k \in \{1, \ldots, |\mathcal{M}|\}$:
- $L \leftarrow L \cup \{\mathcal{M}_k\}$
- For $l \in \{i, \ldots, |\mathcal{M}_k|\}$
  - For $m \in \cup_{n>k} \mathcal{M}_n$:
    - $L \leftarrow L \cup \{(\mathcal{M}_k^l, m)\}$

Let $\mathcal{I}_{\emptyset \leftarrow \mathcal{M}_k}$ be the set of agents in $A_k$ that were unmatched in $M$. Observe that all agents that are matched must belong to one of the sets described.

Consider the base case $a = \mathcal{M}_1$. No agent in the set $A_1$ can be strictly improved as all agents are matched via a modality in $\mathcal{M}_1$ since the $\phi_1$ is Pareto efficient, because otherwise an improvement would mean that some agents can be matched by a better modality $m'$ without making anyone else worse off. Due to the risk ordering of $\mathcal{M}$, there are no better modalities than those in $\mathcal{M}_1$ and thus no agent can be improved upon to a modality outside of $\mathcal{M}_1$. Furthermore, no agent can be improved to a modality $m' \in \mathcal{M}_1$ better than $m$. If this were not the case, then this would contradict the Pareto efficiency of $\phi_1$.

Assume the induction hypothesis holds, that no set prior to and including $a \in L$ can have an agent strictly improved upon (while ensuring all other agents are at least weakly improved upon). Now consider the successor to $a$, denoted $b$. Either

- $a = \mathcal{M}_k$, and $b = (\mathcal{M}_k^1, \mathcal{M}_{>k}^1)$, or
- $a = (\mathcal{M}_k^l, \mathcal{M}_{>k}^m)$ such that $m < |\mathcal{M}_{>k}|$, and $b = (\mathcal{M}_k^l, \mathcal{M}_{>k}^{m+1})$, or
- $a = (\mathcal{M}_k^l, \mathcal{M}_{>k}^m)$ such that $l < |\mathcal{M}_k|$ and $m = |\mathcal{M}_{>k}|$, and $b = (\mathcal{M}_k^{l+1}, \mathcal{M}_{>k}^1)$, or
- $a = (\mathcal{M}_k^l, \mathcal{M}_{>k}^m)$ such that $l = |\mathcal{M}_k|$ and $m = |\mathcal{M}_{>k}|$, and $j = \mathcal{M}_{k+1}$.

Note that in every case, for an agent $i$ matched in $b$, to improve upon them requires that they be matched to a willing agent $j$ (that they are not matched with) where $j \rightarrow_{\hat{m}_j} i$ for $\hat{m}_j \in \mathcal{M}(i)$, $i \rightarrow_{\hat{m}_i} j$ for $\hat{m}_i \in \mathcal{M}(j)$. If $\mathcal{M}(i) = \mathcal{M}(j)$, then this would contradict efficiency of $\phi_k$ as $i$ and $j$ would be matched in $\phi_k$, unless an agent already matched in $\phi_k$ were improved upon. By the inductive, assumption, no previously matched agent can be improved upon. If $\mathcal{M}(i) \neq \mathcal{M}(j)$ then it must be that $\hat{m}_i \in \mathcal{M}_{<k}$. This would contradict the efficiency of the bipartite match between willing agents, unless (again) an agent matched via this were also strictly improved upon. By same argument as before, this is not possible. We can apply a similar argument for unmatched agents.

By the inductive argument, this match is efficient. To see strategyproofness, observe that any misreport outside of the agent's partition element to a more preferred mode does not increase any agent's chance to be matched, and any misreport to a less preferred more either does not change the outcome for the agent, or causes them to be matched via a mode less preferred to being unmatched. Thus we can conclude on strategyproofness of the mechanism. Furthermore, misreports within the agent's partition element does not affect the algorithm until the the mechanism corresponding to that partition is used. Since the mechanism is strategyproof, there is no profitable deviation.

To see pairwise stability, and thus also weak-core stability, note that if two agents preferred to be matched to one another over their partner, then by virtue of the algorithm, if they were in the same partition then this would contradict pairwise stability of the corresponding mechanism, and if they were part of different partition elements then they would have been bipartite matched earlier in the algorithm. □

## A.8 Proof of Proposition 8

*Proof.* Observe the following. First, no two $P_i$ have type $AB$. If this were not the case, for example $P_1 = P_2 = AB$, then every donor of $P_1$ and $P_2$ could donate to the other patient. This contradicts with the set of feasible pairwise matches, which does not have this property for any pair of patient-donor groups. This implies that $d_i^1 \neq AB$ for all $i$. To see this, note that for each $d_i^1$, they can donate to two other patients. If for contradiction we had that for some $i$, $d_i^1 = AB$, then we would have that there are two $AB$ patients because $AB$ donors can only donate to $AB$ patients. Furthermore, we have that for $i \in \{2, 3, 4\}$, $P_i \neq AB$. To see this, consider $P_2 = AB$ for contradiction. We have that $d_2^1 - d_4^2$ is a feasible pairwise exchange. As such, $d_2^1$ can donate to $P_4$, and since all donors can donate to $P_2 = AB$, then so can $d_4^1$. However $d_2^1 - d_4^1$ is not a feasible pairwise exchange, which is a contradiction. We can apply the same argument to the case where $P_3 = AB$ or $P_4 = AB$ (but note that it does not apply to $P_1$). Finally, observe that $P_i \neq O$ for $i \in \{2, 3, 4\}$. If this were not true, for example $P_2 = O$, then we would have that $d_1^1 = d_4^2 = O$. Given that $O$ donors can donate to any agent, we would have that $d_1^1 - d_4^2$ would be a feasible exchange. This is a contradiction with the set of feasible exchanges given. The same argument applies to $P_3 = O$ or $P_4 = O$.

Using these facts, we proceed by considering various possible cases. First consider the case that $P_1 = AB$, thus any donor can donate to $P_1$. This implies $d_1 \neq O$, which follows because if $d_1 = O$, then we would have that any exchange is feasible for $P_1$ as $P_1$ can receive from any donor as they are $AB$ and $d_1^1$ can donate to any patient as they are $O$. Recall that $d_1^1 \neq AB$, hence $d_1^1 \in \{A, B\}$. Consider the case that $d_1^1 = A$, and thus $P_2 = P_3 = P_4 = A$ (because none of these patients can be $AB$). Hence we must have that $d_2^2 \in \{A, O\}$ because $d_2^2 - d_3^1$ is feasible and $P_3 = A$, which means $d_1^1 - d_2^2$ is feasible and thus a contradiction. A similar proof applies when we instead assume that $d_1 = B$. As we have considered all the possible cases for choice of blood type of $d_1$ given $P_1 = AB$, and they all lead to contradictions, we cannot have that $P_1 = AB$.

Now consider the case where $P_1 = O$. As $O$ patients can only receive from $O$ donors, then we have that $d_2^1 = d_3^1 = d_4^1 = O$. This would imply that $d_3^1 - d_4^1$ is feasible, which is a contradiction. Hence we must have that $P_1 \in \{A, B\}$. Recall that we also have $P_2 \in \{A, B\}$.

Consider $P_1 = P_2 = A$. Thus we must have that $d_1^1, d_4^2 \in \{O, A\}$. If $d_1^1 = O$, then this would imply that $d_1^1 - d_4^2$ is feasible, which is a contradiction. If $d_1^1 = A$, then $P_4 = A$ (as it cannot be $AB$). Furthermore, if $d_4^2$ can donate to $P_2 = A$, then it can also donate to $P_1 = A$. And since $d_1^1$ can donate to $P_4$, we have that $d_1^1 - d_4^2$ is feasible, which is a contradiction. The same idea applies to $P_1 = P_2 = B$.

Now consider $P_1 = A$ and $P_2 = B$ (the same idea applies for $P_1 = B$ and $P_2 = A$). Note that $d_1^1, d_4^2 \in \{B, O\}$, because both donors can donate to $P_2 = B$. If $d_1^1 = d_4^2 = O$, then we would have $d_1^1 - d_4^2$ is feasible, which is a contradiction. If $d_1^1 = B$ and $d_4^2 = O$, then we have that $d_1^1 - d_4^2$ is feasible, which is also a contradiction. Consider $d_1^1 = O$ and $d_4^2 = B$. Because $d_1^1 - d_4^1$ is feasible and $P_1 = A$, it must be that $d_4^1 \in \{O, A\}$. If $d_4^1 = O$, then $d_4^1 - d_2^1$ would be feasible, which a contradiction. Hence it must be that $d_4^1 = A$. Given that $d_1^1 = O$ can donate to $P_2$, but $d_2^2$ cannot donate to $P_1 = A$ because $d_1^1 - d_2^2$ is not feasible, it must be that $d_2^2 \in \{B, AB\}$. First consider $d_2^2 = B$. In this case, note that $P_4 \in \{O, A\}$ because $d_4^2$ is compatible with $P_2$ but $d_2^2 = B$ is not compatible with $P_4$ as $d_2^2 - d_4^2$ is not feasible. If $P_4 = O$, then we must have that $d_2^1 = O$ as $d_2^1 - d_4^2$ is feasible. But this is a contradiction because we would have that $d_3^1 - d_2^1$ is feasible. If $P_4 = A$, then this would imply that $d_2^1 = \{A, O\}$. It must be that $d_2^1 = A$, as if $d_2^1 = O$ then we would have that $d_2^1 - d_3^1$ is feasible, which is a contradiction. Given this, we must have that $P_3 \in \{B, O\}$ so that $d_2^1 - d_3^1$ is not feasible (as $d_2^1 = A$). By our previous observation, it cannot be that $P_3 = O$. Hence if $P_3 = B$, we would get a contradiction as $d_3^2 - d_4^2$ would be feasible because $d_3^2$ is compatible with $P_4$ given the set of feasible exchanges, and $d_4^2 = B$ and $P_3 = B$ by assumption. Thus we cannot have that $d_2^2 = B$, so it must be that $d_2^2 = AB$. This would imply that $P_3 = AB$, which is a contradiction with out observation that $P_2, P_3, P_4 \neq AB$.

Now we consider the final case, that is $d_1^1 = d_4^2 = B$. This implies that $P_4 \in \{B, O\}$ as $d_1^1 - d_4^1$ is feasible and $d_1^1 = B$ by assumption. If $P_4 = B$, then $d_3^2 \in \{B, O\}$ as $d_2^3 - d_4^1$ is feasible and these are the only feasible blood types that can donate to $P_4 = B$. In either case, we would have that $d_2^2 - d_2^3$ is feasible because $P_2 = B$, which is a contradiction. Thus it must be that $P_4 = O$. However this contradicts our earlier observation that $P_4 \neq O$. As such it cannot be that $P_1 = A$ and $P_2 = B$ (or by an analogous argument that $P_1 = B$ and $P_2 = A$). As we have gone through all cases, and shown that there is no choice of blood types that are consistent with this set of feasible exchanges, then this set is not rationalizable. □

## A.9 Proof of Proposition 9

*Proof.* Consider the following agents, that is patients with donors (that may or may not be feasible):

- $\{i, j, k\}$ each with an $O$ patient and $\{O, B\}$ donors.
- $l$ with an $B$ patient and an $O$ donor.

Observe that for any of $\{i, j, k\}$ to match with each other, it must be through their $O$ donor, and to match with $l$ can be through either donor. However to use their $B$ donor, they must match with $l$. We will define a match by the agents participating in it, with their favourite donor used implicitly. For example $(i, j)$ means that the donate to each other via their $O$ donors. We say a match is valid if it is efficient and IR given the preferences considered.

Assume for contradiction that there is a desirable mechanism. Fix $l$'s preference as

$O \succ \emptyset$. Now consider the following preference profile $\succ_1$ for $\{i, j, k\}$:

$$i : B \succ \emptyset \succ O$$
$$j : B \succ \emptyset \succ O$$
$$k : B \succ \emptyset \succ O$$

Without loss of generality, let the valid allocation in the desirable mechanism match $(i, l)$ together. Consider the following profile $\succ_2$:

$$i : O \succ B \succ \emptyset$$
$$j : B \succ \emptyset \succ O$$
$$k : B \succ \emptyset \succ O$$

Since $(i, j)$ and $(i, k)$ are not IR, then to maintain strategyproofness we must have $(i, l)$ match. Now consider the following $\succ_3$:

$$i : O \succ B \succ \emptyset$$
$$j : B \succ O \succ \emptyset$$
$$k : B \succ \emptyset \succ O$$

The valid matches are $\{(i, j), (k, l)\}$ and $(j, l)$. However choosing the latter would violate strategyproofness as there would be a profitable deviation for $j$ to misreport in preference profile $\succ_3$ into $\succ_2$ where they go from being unmatched to being part of an individually rational match. Hence let the match here be $\{(i, j), (k, l)\}$. Now consider the following $\succ_4$:

$$i : O \succ \emptyset \succ B$$
$$j : B \succ O \succ \emptyset$$
$$k : B \succ \emptyset \succ O$$

Note that $(j, l)$ is valid but not strategyproof, as otherwise $i$ will misreport from $\succ_4$ to $\succ_3$ and get from being unmatched to matched with $j$. Thus the only valid match is

$\{(i,j),(k,l)\}$. Now consider the following $\succ_5$:

$$i : O \succ \emptyset \succ B$$
$$j : B \succ \emptyset \succ O$$
$$k : B \succ \emptyset \succ O$$

Note that $(j,l)$ is valid but again not strategyproof, as otherwise $j$ in $\succ_4$ would misreport in $\succ_5$ to $\succ_4$ and go from $(i,j)$ to $(j,l)$, which is a strictly better outcome for them. Hence the only valid outcome is $(k,l)$. Now consider the following $\succ_6$:

$$i : O \succ \emptyset \succ B$$
$$j : B \succ \emptyset \succ O$$
$$k : B \succ O \succ \emptyset$$

Observe if $(k,l)$ is not the valid match chosen here, then $k$ in $\succ_6$ will misreport to $\succ_5$ to get this match and thus strictly improve. Hence $(k,l)$ is the outcome here. Now consider the following $\succ_7$:

$$i : O \succ B \succ \emptyset$$
$$j : B \succ \emptyset \succ O$$
$$k : B \succ O \succ \emptyset$$

There are two valid outcomes here: $\{(i,k),(j,l)\}$ and $(k,l)$. Note that the former would not be strategyproof, as otherwise we would have $i$ misreport from $\succ_6$ to $\succ_7$ and thus go from being unmatched to being matched in a valid outcome. However the latter would also not be strategyproof, as $k$ in $\succ_2$ would misreport to $\succ_7$ and go from being unmatched to being matched in a valid outcome. Thus there is no way of choosing a valid match. Hence there is no desirable mechanism. □

## A.10 Proof of Theorem 4

*Proof.* Note that individual rationality clearly holds.

First we prove efficiency. We will construct a sequence , $i_1 \to i_2 \to \dots$, where $i_n \to i_{n+1}$ means that $i_n$ is matched with $i_{n+1}$'s partner in $M'$, and $i_{n+1}$ has a new partner in $M'$. We see this as $i_n$ *stealing* $i_{n+1}$'s partner in $M$. For three cycles, we don't allow two different agents to steal from the same agent, and we specify which agent is $i_{n+1}$. Note that $i_0$ steals first, in the sense that the agent they are pointing to can no longer use that agent. Hence this kicks off a chain of stealing from different agents where no two agents steal from the same agent. Thus there are infinitely many agents (a contradiction).

First consider the case that there are no $A - B*$ after step 3, hence $i_0$ is either $O - B*$ or $O - A*$ as all $A - *$ get their first choice and cannot be strictly improved upon.

Note that for any agent stolen from that was in a three cycle in $M$, only one item is stolen from them. Hence no agent is stolen from twice.

We proceed by cases.

Consider $i_0$ is $O - B*$ ummatched in $M$ and improved to $B$ in $M'$, or $i_0$ is $O - A*$ unmatched in $M$ and improved to $A$ in $M'$.

1. The following are what can be stolen from whom:
   (a) $B - O*$ are stolen from
      i. $O - B*$
      ii. $O - A, B*$
      iii. $O - A*$ (M: $B - O \to O - A \to A - B$)
   (b) $A - O*$[11] are stolen from
      i. $O - A*$
      ii. $O - B, A*$
2. We now specify the rule for stealing:
   (a) If $i_n = O - B*$ or $i_n = O - A, B*$ and is matched via $B$ in $M'$, then
      i. if in $M'$ they are $O - B \leftrightarrow B - O$: they steal $B - O*$
      ii. if in $M'$ they are $O - B \to B - A \to A - O$: they steal $A - O*$
   (b) If $i_n = O - A*$ or $i_n = O - B, A*$ is matched via $A$ in $M'$
      i. if in $M'$ they are $O - A \to A - B \to B - O$: they steal $B - O*$
      ii. if in $M'$ they are $O - A \leftrightarrow A - O$: they steal $A - O*$
3. All agents point to distinct agents and $i_0$ cannot be stolen from as they are unmatched, hence there are no cycles.

[11] Note that we need not consider $A - B, O*$ as all $A - *$ get their top choice in the case where there are no $A - B*$ after step 3.

Consider $i_0$ is $O - B, A*$ matched with $A - O$ in $M$ (and thus improved to $B$).

1. The following are what can be stolen from whom:
    (a) $B - O*$ are stolen from
        i. $O - B*$
        ii. Not possible is $O - A, B*$, as this would contradict PE of Serial Dictator in the construction of $M$
        iii. $A - B*$ (M: $B - O \to O - A \to A - B$)
    (b) $B - A*$ are stolen from
        i. $A - B*$
2. We now specify the rule for stealing:
    (a) If $i_n = O - B*$ and is matched via $B$ in $M'$, then
        i. if in $M'$ they are $O - B \leftrightarrow B - O$: they steal $B - O*$
        ii. if in $M'$ they are $O - B \to B - A \to A - O$: they steal $B - A*$
    (b) If $i_n = O - B, A*$ strictly improves to $B$ in $M'$
        i. if in $M'$ they are $O - B \to B - A \to A - O$: they steal $B - A*$
        ii. if in $M'$ they are $O - B \leftrightarrow B - O$: they steal $B - O*$
    (c) If $i_n = A - B*$ is matched via $B$ in $M'$
        i. if in $M'$ they are $O - A \to A - B \to B - O$: they steal $B - O*$
        ii. if in $M'$ they are $A - B \leftrightarrow B - A$: they steal $B - A*$
3. $i_0$ is not stolen from and thus there is no cycle.

Consider $i_0$ is $O - A, B*$ matched with $B - O$ in $M$ (and thus improved to $A$).

1. The following are what can be stolen from whom:
    (a) $A - O*$ are stolen from
        i. $O - A*$ (matched via $A$ in $M$)
        ii. Not possible is $O - B, A*$, as this would contradict PE of Serial Dictator in the construction of $M$

(b) $A - B*$ are stolen from

i. $B - A*$

ii. $O - A*$ $(A - B \to B - O \to O - A)$

2. We now specify the rule for stealing:

(a) If $i_n = O - A, B*$ strictly improves to $A$ in $M'$

i. if in $M'$ they are $O - A \to A - B \to B - O$: they steal $A - B*$

ii. if in $M'$ they are $O - A \leftrightarrow A - O$: they steal $A - O*$

(b) If $i_n = O - A*$ and is matched via $A$ in $M'$ (they were matched via $A$ in $M$, hence also match via $A$ in $M'$), then

i. if in $M'$ they are $O - A \leftrightarrow A - O$: they steal $A - O*$

ii. if in $M'$ they are $O - A \to A - B \to B - O$: they steal $A - B*$

(c) If $i_n = B - A*$ is matched via $A$ in $M'$

i. if in $M'$ they are $O - B \to B - A \to A - O$: they steal $A - O*$

ii. if in $M'$ they are $A - B \leftrightarrow B - A$: they steal $A - B*$

3. $i_0$ is not stolen from and thus there is no cycle.

Consider $i_0$ is $O - B, A*$ unmatched in $M$, improved to $A$ in $M'$, or $i_0$ is $O - A, B*$ unmatched in $M$, improved to $B$ in $M'$.

1. The following are what can be stolen from whom:

(a) $B - O*$ are stolen from

i. $O - B*$

ii. $O - A, B*$

iii. $O - A*$ (M: $B - O \to O - A \to A - B$)

(b) $A - O*$ are stolen from

i. $O - A*$

ii. $O - B, A*$

2. We now specify the rule for stealing:

(a) If $i_n = O - B, A*$ is matched via $A$ in $M'$

i. if in $M'$ they are $O-A \to A-B \to B-O$ or $O-B \leftrightarrow B-O$: they steal $B-O*$

ii. if in $M'$ they are $O-A \leftrightarrow A-O$: they steal $A-O*$

(b) If $i_n = O-B*$ or $i_n = O-A, B*$ and is matched via $B$ in $M'$, then

i. if in $M'$ they are $O-B \leftrightarrow B-O$: they steal $B-O*$

ii. if in $M'$ they are $O-B \to B-A \to A-O$: they steal $B-A*$

(c) If $i_n = O-A*$ or $i_n = O-B, A*$ and is matched via $A$ in $M'$, then

i. if in $M'$ they are $O-A \leftrightarrow A-O$: they steal $A-O*$

ii. if in $M'$ they are $O-A \to A-B \to B-O$: they steal $B-O*$

3. Each agent points to a distinct agent, and since $i_0$ is unmatched in $M$, there is no cycle.

Now consider the case where there are no $B-O*$ after step 3. Hence $i_0$ is $A-B*$, $O-A*$ or $O-B*$.

Consider $i_0$ is $A-B*$ matched with $A-A*$, $O-A*$, $O-B, A*$ or unmatched in $M$, and they are improved to $B$.

1. The following are what can be stolen from whom:

   (a) $B-O*$ are stolen from

      i. $A-B*$ (in $M$, $B-O$ is only matched via $A-B \to B-O \to O-A$, so they must be matched via $B$ in $M'$)

   (b) $B-A*$ are stolen from

      i. $A-B*$ (must be matched via $B$ in $M'$ as in $M$ they are)

2. We now specify the rule for stealing:

   (a) If $i_n = A-B*$ matches via $B$ in $M'$

      i. if in $M'$ they are $O-A \to A-B \to B-O$: they steal $B-O*$

      ii. if in $M'$ they are $A-B \leftrightarrow B-A$: they steal $B-A*$

3. $i_0$ is not stolen from and thus there is no cycle.

Note that we need not consider the case where $i_0$ is $A-B, A*$ or $A-B, O*$ that are unmatched in $M$ and improved to their second donor in $M'$ as they are always guaranteed their second donor in $M$. Also the case where $O-A, B*$ is matched with $B-O*$ in $M$ is not possible due to the assumption that there is no $B-O*$ after step 3.

Consider $i_0$ is $O-A*$ unmatched in $M$, $i_0 = O-A, O$ or $O-A, B, O$ is matched with $O-O$ in $M$ and is improved to their $A$ donor in $M'$.

1. The following are what can be stolen from whom:
   (a) $A-O*$ are stolen from
      i. $O-A*$ (must be matched via their top donor)
      ii. $O-B, A*$ (must be matched via their top two donors)
   (b) $A-B, O*$ (matched via their second donor in $M'$, so they must be matched via their second donor in $M$ to preserve efficiency)
      i. $O-A*$ (must be matched via their top donor)
      ii. $O-B, A*$ (must be matched via their top two donors)
   (c) $B-O*$ are stolen from
      i. $O-A*$ (In $M$, they are matched in $A-B \to B-O \to O-A$, so they get their top donor in $M'$)
2. We now specify the rule for stealing:
   (a) If $i_n = O-A*$ or $i_n = O-B, A*$ matches via $A$ in $M'$
      i. if in $M'$ they are $O-A \to A-B \to B-O$: they steal $B-O*$
      ii. if in $M'$ they are $O-A \leftrightarrow A-O$: they steal $A-O*$ or $A-B, O*$ (matched via second donor in $M'$)
   (b) If $i_n = O-B, A*$ matches via $B$ in $M'$
      i. if in $M'$ they are $O-B \leftrightarrow B-O$: they steal $B-O*$
      ii. if in $M'$ they are $O-B \to B-A \to A-O$: they steal $A-O*$ or $A-B, O*$ (via second donor)
3. $i_0$ is not stolen from and thus there is no cycle.

Consider $i_0$ is $O-A, B*$ unmatched in $M$ and is improved to their $B$ donor in $M'$.

1. The following are what can be stolen from whom:
   (a) $B - O*$ are stolen from
      i. $A - B*$ (In $M$, they are matched in $A - B \rightarrow B - O \rightarrow O - A$, so they get their top donor in $M'$)
   (b) $B - A*$ are stolen from
      i. $A - B*$ (matched via their top donor)
2. We now specify the rule for stealing:
   (a) If $i_n = O - A, B*$ matches via $B$ in $M'$
      i. if in $M'$ they are $O - B \rightarrow B - A \rightarrow A - O$: they steal $B - A*$
      ii. if in $M'$ they are $O - B \leftrightarrow B - O$: they steal $B - O*$
   (b) If $i_n = A - B*$ matches via $B$ in $M'$
      i. if in $M'$ they are $A - B \rightarrow B - O \rightarrow O - A$: steal $B - O*$
      ii. if in $M'$ they are $A - B \leftrightarrow B - A$: steal $B - A*$
3. $i_0$ is not stolen from and thus there is no cycle.

Consider $i_0$ is $O - B*$ unmatched in $M$, or $i_0 = O - B, A*$ is matched with $A - O$, and is improved to their $B$ donor in $M'$.

1. The following are what can be stolen from whom:
   (a) $B - O*$ are stolen from
      i. $A - B*$ (In $M$, they are matched in $A - B \rightarrow B - O \rightarrow O - A$, so they get their top donor in $M'$)
   (b) $B - A*$ are stolen from
      i. $A - B*$ (matched via their top donor)
2. We now specify the rule for stealing:
   (a) If $i_n = O - B*$ matches via $B$ in $M'$
      i. if in $M'$ they are $O - B \rightarrow B - A \rightarrow A - O$: they steal $B - A*$
      ii. if in $M'$ they are $O - B \leftrightarrow B - O$: they steal $B - O*$

(b) If $i_n = A - B*$ matches via $B$ in $M'$

i. if in $M'$ they are $A - B \to B - O \to O - A$: steal $B - O*$

ii. if in $M'$ they are $A - B \leftrightarrow B - A$: steal $B - A*$

3. $i_0$ is not stolen from and thus there is no cycle.

Consider $i_0$ is $O - B, A*$ unmatched in $M$ and is improved to their $A$ donor in $M'$.

1. The following are what can be stolen from whom:

(a) $B - O*$ are stolen from

i. $O - A*$ (in $M$, $O - A \to A - B \to B - O$)

(b) $A - B, O*$ (matched via second donor) or $A - O*$

i. $O - A*$

ii. $O - B, A*$ (matched second donor in $M$, so must be top two in $M'$)

2. We now specify the rule for stealing:

(a) If $i_n = O - B, A*$ or $i_n = O - A*$ matches via $A$ in $M'$

i. if in $M'$ they are $O - A \to A - B \to B - O$: they steal $B - O*$

ii. if in $M'$ they are $O - A \leftrightarrow A - O$: they steal $A - O*$ or $A - B, O*$ (can't be $A - B, O, A$ matched via $A$ in $M$ as then $i_0$ would have selected them)

(b) If $i_n = O - B, A*$ matches via $B$

i. if in $M'$ they are $O - B \to B - O$: they steal $B - O*$

ii. if in $M'$ they are $O - B \to B - A \to A - O$: they steal $A - O*$ or $A - B, O*$

3. $i_0$ is not stolen from and thus there is no cycle.

Now we show strategyproofness. Consider an agent $B - B*$, which is guaranteed to be matched in the first step via $B$. Thus they have no incentive to deviate. Consider an agent $B - A*$, who are guaranteed to be matched in the second step via $A$ and thus have no incentive to deviate. Similarly, $B - O*$ is guaranteed to be matched via $O$ either in step 3 or step 6, due to our long-side assumption.

Consider $A - B*$. Those matched in step 2 or step 3 get their best match, hence have no incentive to deviate. If there are no $B - O*$, then all $A - B*$ have been matched (due to the assumption that there is a surplus of $O-A*$) hence there are no $A-B*$ agents that have an incentive to deviate. If there are $A-B*$ that remain, then all $B-O*$ have been matched. All $B-*$ have been matched at this point, and the algorithm does not allow $A-B*$ to use their $B$ donor. Note that they have no incentive to list other donors as first, as only after their donation opportunities via $B$ have been exhausted will any match for an $A-*$ agent via some other donor will be allowed. As this is done according to their priority order, it does not improve their chances of being matched via some other donor (apart from $B$). Hence $A-B*$ can drop their preferred donor and does have a strict incentive to deviate.

Note that in step 6 onwards, all agents apart from $O-*$ get their best option of those that remain (due to the surplus of $O-A*$ and $O-B*$ if any), and thus have no incentive to deviate. Since $O-*$ are effectively undergoing a serial dictator procedure (accounting for weak preferences) they have no incentive to deviate. Since we have considered every type of agent, and none have an incentive to deviate, this mechanism is strategyproof.

□

## A.11 Proof of Proposition 10

*Proof.* First consider the case where there is an $AB$ patient. If the $AB$ patient has a donor compatible with some other patient, then they must be pointing to someone. Since anyone can donate to $AB$ patients, then everyone is pointing to $AB$. Hence there is a two-cycle. If the $AB$ patient has no donor that can feasibly donate to another patient, then they are not part of any cycle.

Now consider the case where there is no $AB$ patient. This means that any cycle cannot utilize $AB$ donors, as $AB$ donors can only donate to $AB$. Hence assume that there are no $AB$ donors. Assume there is no two cycle but there is a cycle of length $n$.

First we will show that there cannot be an $O$ donor as an patient's top choice, nor an $O$ patient, for any agent in a cycle. Assume for contradiction that there is an agent $i_1$ that points with an $O$ donor in the cycle. Hence they point to every agent. Since $i_n \to i_1$, hence $i_1 \leftrightarrow i_n$. Since there are no $O$ donors, and only $O$ donors can donate to $O$ patients, then there can be no $O$ patients.

We proceed by assuming that there are only $A$ and $B$ donors and patients, and consider different cases based on the number of agents $n$ in the cycle.

If there are only two agents, and thus all these agents are in the cycle, then this is a contradiction.

If there are at least five agents in the cycle, denoted $i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow ...$, then observe we can't have the any three consecutive agents have the same type. For example, if $i_1, i_2, i_3 \in \mathcal{I}_A$, then $i_2 \rightarrow i_1$ and thus there is a two cycle, which is a contradiction.

First consider the case where $i_1, i_2 \in \mathcal{I}_A$. This implies that $i_3 \in \mathcal{I}_B$. If the cycle is of length three, then we are done. If $i_3 \rightarrow i_4 \in \mathcal{I}_A$, then $i_3 \rightarrow i_2$, which introduces a two cycle. Hence $i_4 \in \mathcal{I}_B$. If the cycle is of length four, then we are done. If $i_4 \rightarrow i_5 \in \mathcal{I}_B$, then $i_4 \rightarrow i_3 \in \mathcal{I}_B$, another contradiction. If $i_4 \rightarrow i_5 \in \mathcal{I}_A$, then $i_4 \rightarrow i_2 \in \mathcal{I}_A$ and $i_2 \rightarrow i_4 \in \mathcal{I}_B$ since $i_2 \rightarrow i_3 \in \mathcal{I}_B$. This final contradiction shows that $i_1, i_2$ cannot both be in $\mathcal{I}_A$. A similar argument holds for $i_1, i_2 \in \mathcal{I}_B$.

Now consider the case where $i_1 \in \mathcal{I}_A$ and $i_2 \in \mathcal{I}_B$. If there are only three or four agents, then we are done (assuming in the latter case there is no three consecutive agents of the same type). Assume there are at least five agents: $i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow i_4 \rightarrow i_5$. If $i_3 \in \mathcal{I}_A$ and $i_4 \in \mathcal{I}_B$, then $i_3 \rightarrow i_2$, giving a contradiction. By the previous argument on consecutive types, it cannot be that $i_3, i_4 \in \mathcal{I}_B$. If $i_3, i_4 \in \mathcal{I}_A$, then $i_5 \in \mathcal{I}_B$, otherwise there will be three agents of consecutive types. Then $i_4 \rightarrow i_2$ and $i_2 \rightarrow i_4$ gives a two cycle and thus a contradiction. Now consider $i_3 \in \mathcal{I}_B$ and $i_4 \in \mathcal{I}_A$, then $i_1 \rightarrow i_3$ and $i_3 \rightarrow i_1$, another contradiction. A similar argument applies for $i_1 \in \mathcal{I}_B$ and $i_2 \in \mathcal{I}_A$.

If there are exactly three agents, it is clear that the cycles must be of the following form: $A \rightarrow B \rightarrow A$, or $B \rightarrow A \rightarrow B$. Note that, for example, the latter is equivalent to $A \rightarrow B \rightarrow B$. However then it must be that there is a two-cycle given by $A \leftrightarrow B$ in both cases, thus this is not possible. If there are four agents, since there can be no three consective agents, it must be that cycles are either of the form: $A \rightarrow B \rightarrow A \rightarrow B$, or $A \rightarrow B \rightarrow B \rightarrow A$. However neither are possible as there is a two cycle, which is a contradiction.

□

## A.12 Proofs of Proposition 15

*Proof.* Let $\mathcal{I}^b_{K \leftrightarrow K}$ and $\mathcal{I}^b_{L \leftrightarrow L}$ be the agents in $\mathcal{I}^K$ and $\mathcal{I}^L$, respectively, that are matched by $b^K$ and $b^L$. First note that $\mathcal{I}_{K \leftrightarrow K}$ has the same cardinality as $\mathcal{I}^b_{K \leftrightarrow K}$ as maximal matchings are maximum matchings (Roth et al., 2005). Note that we can construct $f$ (via choice of priority order) such that a pair $i$ in $\mathcal{I}^b_{L \leftrightarrow L}$ not matched in $\mathcal{I}_{L \leftrightarrow L}$ is such that they must be matched to a kidney patient, i.e. $i \in \mathcal{I}_{K \leftrightarrow L}$. Worst case, the matched pair of $i$ in $\mathcal{I}^b_{L \leftrightarrow L}$ is unmatched in $f$, but for every such case there is a kidney pair is who is matched in $f$ but not in the baseline. Hence there must be at least as many transplants in $f$ then in the baselines.

Now we show the second condition. We have already shown that more kidney patients donate kidneys in the previous paragraph. For a similar reason, less liver patients donate livers. If this did not hold, then the matching in the baseline would not have been maximum. Finally, since no liver patients donate kidneys in the baseline, the claim that more liver patients donate kidneys holds trivially. □