

Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

**Strategyproof Voluntary Pairwise Exchange with
Dichotomous Preferences**

Ryo Tsukamoto

26 October, 2026

DP2026-022

<https://ies.keio.ac.jp/en/publications/28148/>

Keio University



Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office-group@keio.jp
26 October, 2026

Strategyproof Voluntary Pairwise Exchange with
Dichotomous Preferences

Ryo Tsukamoto

Keio-IES DP2026-022

26 October, 2026

JEL Classification: D45, D47, D71

Keywords: Dichotomous preference, Kidney exchange, Matching, Matroid

Abstract

This study investigates pairwise endowment exchange without monetary transfers under dichotomous preferences, motivated by living donor kidney exchange problem. In contrast to the standard strict preference domain—where no mechanism satisfies efficiency, incentive compatibility, and individual rationality simultaneously—these properties are jointly achievable in this setting. We adopt a definition of mechanisms as welfare-based correspondences rather than single-valued functions, treating welfare-equivalent outcomes as identical. We provide the first characterization of mechanisms satisfying these properties and show that all mechanisms in this class closely resemble the priority mechanism (Roth, Sönmez, and Ünver, 2005), supporting its use in pairwise kidney exchange.

Ryo Tsukamoto

Department of Economics, Keio University

r.tsukamoto72@keio.jp

Strategyproof Voluntary Pairwise Exchange with Dichotomous Preferences*

Ryo Tsukamoto[†]

October 26, 2026

Abstract

This study investigates pairwise endowment exchange without monetary transfers under dichotomous preferences, motivated by the living donor kidney exchange problem. In contrast to the standard strict preference domain—where no mechanism satisfies efficiency, incentive compatibility, and individual rationality simultaneously—these properties are jointly achievable in this setting. We adopt a definition of mechanisms as welfare-based correspondences rather than single-valued functions, treating welfare-equivalent outcomes as identical. We provide the first characterization of mechanisms satisfying these properties and show that all mechanisms in this class closely resemble the priority mechanism (Roth, Sönmez, and Ünver, 2005), supporting its use in pairwise kidney exchange.

Keywords: Dichotomous preference, Kidney exchange, Matching, Matroid

JEL classification: D45; D47; D71

1 Introduction

This study examines voluntary endowment exchange without monetary transfers, focusing on settings where only pairwise exchanges are allowed. Such size restrictions are practically relevant when larger exchanges are logistically infeasible, as in living-donor organ exchange. However, under this restriction, no mechanism

*I am grateful to Hideo Konishi, Morimitsu Kurino, Tayfun Sönmez, Utku Ünver for their helpful comments and instructions. I also thank seminar participants at Boston College, Keio University, and The University of Osaka.

[†]Keio University. Email: r.tsukamoto72@keio.jp

can satisfy Pareto efficiency, strategyproofness, and individual rationality simultaneously on the standard strict preference domain (Nicolò and Rodríguez-Álvarez, 2012). We instead consider the dichotomous preference domain (Bogomolnaia and Moulin, 2004), under which the priority mechanism (PM) proposed by Roth, Sönmez, and Ünver (2005) satisfies all three properties.¹ Since its introduction, the PM played a central role in the US when pairwise exchanges were considered. Although programs in the US have since adopted more sophisticated rules, such as three- or four-way exchanges or the incorporation of deceased-donor chains, simple pairwise exchange remains relevant in other countries, including France, Italy, Sweden, and India—see Biró et al. (2021) for European countries and Kute et al. (2025) for India. The PM has also inspired mechanism design in other contexts (Ergin, Sönmez, and Ünver, 2020; Andersson et al., 2021; Manjunath and Westkamp, 2021; Han, Kesten, and Ünver, 2024).

Despite this theoretical and practical relevance, the theoretical understanding of the PM remains limited. This paper aims to address this gap by providing a characterization result. Following Roth, Sönmez, and Ünver (2005), we define a mechanism as a correspondence resembling a welfarist solution. In fact, the PM is a multi-valued correspondence: it selects all welfare-equivalent matchings (i.e., it is Pareto-indifferent) and never selects two matchings with different welfare distributions (i.e., it is essentially single-valued). Adopted in the literature (e.g., Roth, Sönmez, and Ünver, 2005; Andersson et al., 2021; Manjunath and Westkamp, 2021), this definition avoids unnecessary complications that arise from artificially forcing single-valuedness on the dichotomous domain. Such a definition is common in the literature on stochastic mechanism design, where welfare-equivalent lotteries are treated collectively rather than individually.

In practice, this interpretation may be natural if the mechanism designer cares only about the set of beneficiaries. For example, many kidney exchange programs give priority to patients who are hard to match with a biologically compatible donor, but do not necessarily specify where the graft comes from. Our main result is the axiomatic characterization of a class of this welfare-based mechanisms by consistency, strategyproofness, Pareto efficiency, and respect for improvements.²

The axioms are motivated by practical and institutional concerns in pairwise living donor exchange.

Consistency requires that, after any set of matched agents leaves the market,

¹Bogomolnaia and Moulin (2004) study the two-sided matching problem with the same preference domain and investigate stochastic solutions.

²These axioms imply individual rationality.

the mechanism induces the same outcome for the remaining agents as it did before their departure. This property is particularly relevant in paired living donor exchange since not all surgeries can be performed simultaneously due to logistical constraints.³ Once a matching consisting of several exchanges is formed and announced, the actual transplant surgeries are typically carried out in stages. For example, the first exchange may be performed today, and the second is scheduled for next week, depending on the availability of doctors, operating rooms, and other resources. Suppose that, after the first exchange is completed, those two patient-donor pairs leave the market. If the mechanism were inconsistent, then a patient who remains in the market could legitimately complain that she would now be matched if the mechanism were reapplied to the reduced market. Consistency rules out such ex post objections and requires that the remaining allocation continue to be justified even after some matched pairs depart.

Pareto efficiency corresponds to maximizing the number of patients who receive acceptable kidneys under the pairwise exchange restriction, which aligns with the primary objective of most kidney exchange programs worldwide.⁴

Strategyproofness requires that patient-donor pairs have no incentive to misreport their types. Here, a pair's type includes not only medical information but also other aspects relevant to organ exchange. For instance, distance between pairs may matter because patients or donors may be unwilling or unable to travel to the partner hospital. An alternative to donor travel is the shipment of organs. However, some patients or their physicians may be reluctant to accept the additional logistical or medical risks associated with organ shipment. Thus, even if two pairs are medically compatible, their exchange may not be mutually acceptable once travel constraints or shipment preferences are taken into account. In such an environment, it is desirable that pairs truthfully report all relevant information, rather than strategically concealing acceptable partners by pretending not to accept certain modalities.⁵

Respect for improvements captures another incentive consideration of pairs and corresponds to the notion of *donor monotonicity* introduced by Roth, Sönmez, and

³Sönmez and Ünver (2005) make the same point in the kidney exchange problem.

⁴In general, Pareto efficiency corresponds only to the maximality of the set of matched agents in the sense of set inclusion. Under the pairwise exchange restriction, however, all maximal matchings have the same cardinality by a classical result in graph theory. Thus, in this setting, Pareto efficiency is equivalent to maximizing the number of matched agents. Kratz (2024) studies this equivalence beyond two-way exchanges.

⁵For example, OPTN guidelines include informed consent policies under which recipients must indicate whether they accept the risks associated with shipped organs. See Klaassen et al. (2025) for a discussion of the advantages and disadvantages of the two transportation modalities.

Ünver (2005).⁶ In practice, a patient may register multiple donors in the exchange system. Adding an additional donor can only expand the set of feasible matches available to that patient. From the viewpoint of the exchange program, which aims to facilitate as many transplants as possible, there is no reason to penalize a patient for bringing additional donors into the system. Respect for improvements formalizes this idea: if a patient’s potential exchange partners increase because more donors become available, then the mechanism should not make the patient worse off.

We call the class of mechanisms satisfying these axioms the *tripolar priority mechanisms* (TPMs). TPMs generalize the use of priorities in the priority mechanism (PM). Under the PM, whenever a triage situation arises, a single priority ordering over agents is used to resolve the conflict. In contrast, TPMs allow priorities to depend on the agent whose endowment is contested: when a triage occurs over the endowment of an agent, an agent-specific priority ordering is used to determine the winner. In this sense, TPMs allow for a form of priority design.

Our main result shows, however, that this flexibility is extremely limited. The agent-specific priority orderings must be closely related: aside from the top two positions, all priority orders must coincide. Consequently, much like the PM, a TPM determines matches sequentially according to a priority order, except that the ranking among the top three agents may vary across preference profiles. In this sense, all mechanisms satisfying the axioms remain essentially equivalent to the PM. We therefore interpret our result as providing theoretical support for the use of the PM in pairwise donor exchange and related applications.

1.1 Related Literature

This work contributes to the literature on matching mechanism design with dichotomous preferences and its applications. We are primarily motivated by the pairwise kidney exchange problem initiated by Roth, Sönmez, and Ünver (2005), which builds on earlier work on matching with dichotomous preferences (Bogomolnaia and Moulin, 2004) and living-donor kidney exchange problem (Roth, Sönmez, and Ünver, 2004). See Sönmez and Ünver (2025) for a recent survey of kidney exchange problem. Subsequent studies consider richer preference structures, including age-based preferences and half-compatibility (Nicolò and Rodríguez-

⁶Biró et al. (2024) also study respect for improvements in the housing market model (Shapley and Scarf, 1974), emphasizing its practical relevance in the form of donor monotonicity in the kidney exchange problem.

Álvarez, 2017; Andersson and Kratz, 2020).⁷ These works highlight practical features of living-donor exchange and also show that individual rationality, efficiency, and strategyproofness cannot always be satisfied simultaneously with their richer preference domains. Hatfield (2005) considers strategyproof single-valued mechanisms with arbitrary exchange restrictions but still within dichotomous preferences. While his result applies to a broader class of models than ours, it does not provide a characterization. More generally, PM-like mechanisms have been studied in various contexts, including multi-unit exchanges and balanced reallocation (Han, Kesten, and Ünver, 2024; Manjunath and Westkamp, 2021; Andersson et al., 2021). Unlike these previous studies, our contribution is primarily theoretical: we provide the first axiomatic characterization of PM-like mechanisms under pairwise exchanges with dichotomous preferences, showing that the class of mechanisms satisfying our axioms essentially coincides with the PM.

More broadly, our paper lies between characterization results in the standard strict preference domain and those in general preference domains with indifference. In the strict preference domain, the celebrated *top trading cycle* as a mechanism—attributed to David Gale by Shapley and Scarf (1974)—is characterized by Pareto efficiency, strategyproofness, and individual rationality (Ma, 1994). See Afacan, Hu, and Li (2024) for a recent survey of this literature. With the pairwise exchange restriction, Root and Ahn (2025) provide the characterization of group strategyproof and efficient mechanisms.

In general preference domains, these axioms can still be satisfied simultaneously (Alcalde-Unzu and Molis, 2011; Jaramillo and Manjunath, 2012); however, no known characterization in this domain uses all three axioms. For example, Bogomolnaia, Deb, and Ehlers (2005) show that standard characterization approaches based on *non-bossiness*—a common method in the strict preference domain (e.g., Svensson, 1999; Pápai, 2000; Pycia and Ünver, 2017)—cannot be combined with rich endowment structures.⁸ Ehlers (2014) studies top trading cycles with fixed tie-breaking and provides a characterization of this class of single-valued mechanisms; however, mechanisms in this class are not necessarily Pareto efficient. These results highlight the technical challenges of extending axiomatic characterizations beyond strict preferences. In contrast, by restricting attention to the dichotomous preference domain and limiting exchanges to pairwise swaps, our analysis restores

⁷Balbusanov (2020) studies random mechanisms that promote equity under strict preferences with pairwise exchange constraints.

⁸In the strict preference domain, non-bossiness and strategyproofness are equivalent to group strategyproofness (Pápai, 2000; Takamiya, 2001). However, Ehlers (2002) shows that group strategyproofness is incompatible with Pareto efficiency in the general preference domains.

an axiomatic characterization based on the standard desiderata.

From a technical perspective, the problem we study is closely related to the maximum cardinality matching problem in undirected graphs. It is well-known that the set of covered vertices by some matching possesses a well-known combinatorial property called a *matroid* (Whitney, 1935). This structural feature facilitates the analysis of Pareto efficient matchings in our model, which is challenging in the presence of indifferences as discussed above.

Finally, our result can also be seen as a new characterization of consistent solutions for matching problems; see Thomson (1996) and Thomson (2011) for surveys. The literature provides numerous characterizations of the Core as a solution using variants of consistency and converse consistency, in the two-sided market (Sasaki and Toda, 1992; Toda, 2006; Özkal-Sanver, 2004) and the one-sided market (Klaus, 2011; Klaus and Can, 2013; Klaus, 2017). Our objective, however, is not to characterize any particular solution concept. Rather, we characterize a class of mechanisms using consistency as an axiom motivated by practical considerations. In this respect, the closest to ours are Sönmez and Ünver (2010) and Ergin (2000), who study consistent solutions in the house allocation setting with and without property rights and emphasize the practical relevance of consistency.

The rest of the paper is organized as follows. In section 2 we introduce the model and the axioms. In section 3 the new class of mechanisms is defined. The main characterization result is stated in section 4. Section 5 concludes. Omitted proofs and lemmas are found in Appendix A.

2 Model

We consider the pairwise kidney exchange model (Roth, Sönmez, and Ünver, 2005), which can also be viewed as the roommate problem (Gale and Shapley, 1962), under dichotomous preferences.

2.1 Notations and Definitions

Let $N = \{1, 2, \dots, n\}$ denote the set of potential agents. We assume $n \geq 5$. Each agent i has a dichotomous preference over the set of agents N . A dichotomous preference consists of three indifference classes. The top class contains the agents whom i finds acceptable; the second class consists solely of agent i herself, which can be interpreted as remaining unmatched. The bottom class contains all the agents whom i finds unacceptable. Formally, we use a subset of agents $R_i^N \subset$

$N \setminus \{i\}$ to denote both a preference of i and the set of agents i deems acceptable under the preference. Let $\mathcal{R}_i^N$ be the set of all dichotomous preferences of agent i over N and let $\mathcal{R}^N := \prod_{i \in N} \mathcal{R}_i^N$ be the set of preference profiles. We define the preference of agents when only a subset of agents are present in the market to account for the variation of the market participants. Given $R_i^N \in \mathcal{R}_i^N$ and $S \subset N$, let $R_i^N|_S := R_i^N \cap S$ be the restriction of R_i^N to S . Let $\mathcal{R}_i^S$ be the set of preferences of agent i restricted to S i.e.,

$$\mathcal{R}_i^S := \{R_i^N \cap S \mid R_i^N \in \mathcal{R}_i^N\}.$$

We use R_i^S to denote a generic element of $\mathcal{R}_i^S$. Define $\mathcal{R}^S := \prod_{i \in S} \mathcal{R}_i^S$. Given $R^S \in \mathcal{R}^S$ and $T \subset S \subset N$, we use R_T^S to denote the restriction of R^S to T and $R_{-T}^S := R_{S \setminus T}^S$.

We focus on voluntary matching, in which two agents are matched only if they are mutually acceptable. A graph can summarize mutual acceptability. Formally, the *mutual acceptability graph* $G(R^S) := (S, E(R^S))$ is an undirected graph in which agents are the vertices and an edge between two agents if and only if they are mutually acceptable:

$$E(R^S) := \{\{i, j\} \subset S \mid j \in R_i^S, i \in R_j^S, \text{ and } i, j \in S\}$$

We also use $A_i(R^S) := \{j \in S \mid \{i, j\} \in E(R^S)\}$ to denote the set of mutually acceptable partners for i at the profile R^S .

A *submarket* consists of a subset of agents $S \subset N$ and their preference profile $R^S \in \mathcal{R}^S$. A *matching* for S is $\alpha : S \rightarrow S$ such that $\alpha(\alpha(i)) = i$ for all $i \in S$. Under a matching, each agent is either paired with another agent (when $\alpha(i) \neq i$) or remains single (when $\alpha(i) = i$.) Let $M(S)$ be the set of matchings for S . A *solution* is a correspondence associating a set of matchings with each submarket; formally,

$$F : \cup_{S \subset N} \mathcal{R}^S \rightarrow (\cup_{S \subset N} 2^{M(S)}) \setminus \{\emptyset\}$$

such that $F(R^S) \in 2^{M(S)} \setminus \{\emptyset\}$ for each $S \subset N$ and $R^S \in \mathcal{R}^S$.

Two matchings α and β in $M(S)$ are said to be **welfare equivalent** at profile R^S if the following two conditions hold for each $i \in S$:

$$\begin{aligned} \alpha(i) \in R_i^S &\Leftrightarrow \beta(i) \in R_i^S \\ \alpha(i) = i &\Leftrightarrow \beta(i) = i. \end{aligned}$$

Under two welfare-equivalent matchings, the status of each agent—whether matched with an acceptable partner or not—is the same.

A solution F is **Pareto indifferent** if for any S , R^S , $\alpha \in F(R^S)$ and $\beta \in M(S)$, if β is welfare equivalent to α , then $\beta \in F(R^S)$. A solution F is **essentially single-valued** if for any S , R^S , and $\alpha, \beta \in F(R^S)$, α and β are welfare equivalent. Following Roth, Sönmez, and Ünver (2005), we define a mechanism as a solution satisfying the two properties; that is, a *mechanism* is a solution F that is Pareto indifferent and essentially single-valued.⁹

A mechanism specifies a welfare distribution—it determines “who obtains an acceptable match,” and then collects all welfare-equivalent matchings. This formulation resembles the welfarist solution, which evaluates outcomes solely in terms of welfare distributions; however, the axiom of consistency requires reference to the specific matchings that implement a given welfare distribution. Therefore, our analysis departs from the welfarist framework.

Let

$$N^F(R^S) := \{i \in S \mid \alpha(i) \in R_i^S \text{ for all } \alpha \in F(R^S)\}$$

be the set of agents who obtain an acceptable match under mechanism F at profile R^S .

We use the following notation for a linear order of agents:

$$>': i_1, i_2, \dots, i_k.$$

With this, we mean $>'$ is an order over a subset of agents $\{i_1, i_2, \dots, i_k\}$ in which i_ℓ is the ℓ th ranked agent for each $\ell \in \{1, 2, \dots, k\}$ and $k \in \{1, 2, \dots, n\}$. We also use $>'_{i \leftrightarrow j}$ to denote the order obtained by switching the rankings of i and j in $>'$ without changing the relative rankings of the other agents. Let

$$>: 1, 2, 3, \dots, n$$

be the natural ascending order $1, 2, 3, \dots, n$ of agents.

Abusing notation, we identify matching $\alpha \in M(S)$ with the set of all its matched pairs:

$$\alpha = \{\{i, \alpha(i)\} \mid i \in S \text{ and } \alpha(i) \neq i\}.$$

⁹Bogomolnaia, Deb, and Ehlers (2005) use the two properties as axioms. Following Roth, Sönmez, and Ünver (2005), Ergin, Sönmez, and Ünver (2020), Manjunath and Westkamp (2021), and Andersson et al. (2021), we instead impose them as requirements for a solution to be a mechanism.

2.2 Axioms of Mechanisms

We are interested in the following five axioms.

F is **Pareto efficient** if for any $S \subset N$ and $\alpha \in F(R^S)$, there is no $\beta \in M(S)$ such that

- (1) : for each $i \in S$, $\beta(i) R_i^S \alpha(i)$
- (2) : for some $j \in S$, $\beta(j) P_j^S \alpha(j)$.

In the current setting, mechanism F is Pareto efficient if and only if the set of matched agents $N^F(R^S)$ is maximal for each profile R^S . Furthermore, it is known that the number of matched agents is the same under any two maximal matchings in $G(R^S)$ for each profile R^S . Thus, Pareto efficiency is equivalent to maximizing the number of matched agents.

F is **consistent** if for any $S \subset N$, $R^S \in \mathcal{R}^S$, and $\alpha \in F(R^S)$, for each $T \subset S$ s.t. $\alpha(T) \subset T$, the restriction to $S \setminus T$ remains chosen i.e., $\alpha_{S \setminus T} \in F(R^S|_{S \setminus T})$. Consistency requires that when matched agents leave the market with their partners, the mechanism treats the remaining agents in the same manner.

F is **strategyproof** if no agent can benefit by misreporting her preference regardless of other participants and their preference profile; Formally, for any $S \subset N$, $R^S \in \mathcal{R}^S$, and $Q_i^S \in \mathcal{R}_i^S$, we have $\alpha(i) R_i^S \beta(i)$ for all $\alpha \in F(R^S)$ and $\beta \in F(R_{-i}^S, Q_i^S)$.¹⁰

F is said to **respect for improvements** if an agent getsThe following lemma completes the characterization weakly better off when she is found acceptable by weakly more (in a set-inclusion sense) agents; Formally, for any $S \subset N$, $i, j \in S$, $R^S \in \mathcal{R}^S$, and $Q_j^S \in \mathcal{R}_j^S$ such that $Q_j^S \in \{R_j^S, R_j^S \cup \{i\}\}$, we have $\beta(i) R_i^S \alpha(i)$ for all $\alpha \in F(R^S)$ and $\beta \in F(R_{-j}^S, Q_j^S)$. This axiom requires that the mechanism should not punish an agent when weakly more agents find her acceptable and the two preference profiles are otherwise the same. In the context of living donor exchange, this property is called donor monotonicity (Roth, Sönmez, and Ünver, 2005).

F is **individually rational** if for any $S \subset N$, $i \in S$ and $R^S \in \mathcal{R}^S$, we have $\alpha(i) \in R_i^S \cup \{i\}$ for all $\alpha \in F(R^S)$. That is, under individually rational mechanism, no agents are forced to match with an unacceptable partner.

Individual rationality is implied by Pareto efficiency and respect for improvements.

¹⁰Bogomolnaia, Deb, and Ehlers (2005) introduced two variants of strategyproofness for correspondences: weak and strong strategyproofness. The above definition is strong strategyproofness in their terminology. Under essential single-valuedness, the two notions are equivalent; therefore, we simply call it strategyproofness.

Lemma 1. *If F is Pareto efficient and respects for improvements, then F is individually rational.*

Proof. Assume that there are preference profile R^S , agent i , matching $\alpha \in F(R^S)$ s.t. $\alpha(i) = j$ and $j \notin R_i^S \cup \{i\}$. By Pareto efficiency, $i \in R_j^S$ because otherwise a matching that leaves i and j unmatched while treats the other agents the same as in α Pareto dominates α . Let $Q_h^S := R_h^S \setminus \{i\}$ for all $h \neq i$. By Pareto efficiency, $\beta(i) \in R_i^S \cup \{i\}$ for any matching $\beta \in F(Q_{-i}^S, R_i^S)$. R^S is obtained from (Q_{-i}^S, R_i^S) by replacing Q_h^S with R_h^S one at a time and these replacements should not make i worse off by respect for improvements. This contradicts $\alpha(i) \notin R_i^S \cup \{i\}$. $\square$

3 Tripolar Priority Mechanism

This section introduces a new class of mechanisms. Each mechanism in this class implements the PM with priorities that are allowed to vary across preference profiles. Accordingly, the class of mechanisms we study is defined by restricting the extent to which these priorities may deviate from a fixed reference priority order. Our characterization result shows that such deviations must be highly limited.

Intuitively, the PM uses a single priority order to break the tie whenever triage cases occur: the highest priority agent wins. The new mechanism is slightly more flexible in that it may apply different priority orderings to different triage cases depending on whose endowment is over-demanded. However, these endowment-specific priorities must be close to each other in sharp contrast to school choice Abdulkadiroğlu and Sönmez (2003) in which the priority orderings of schools can be completely different. In this sense, our characterization result also demonstrates the limitation of priority design in the one-sided matching model.

Given a reference priority order, for notational convenience we relabel agents when necessary and represent the reference priority order by the natural ascending order $>: 1, 2, 3, \dots, n$. This entails no loss of generality. Throughout the rest of the paper, including the appendix, we assume that the reference priority order is the natural ascending order $>$.

3.1 Priority Mechanism

We first introduce the PM and then extend its domain to all submarkets to account for the variable population. Given a mutual acceptability graph $G(R^N)$ and a reference priority order of agents $>: 1, 2, \dots, n$, we calculate

- $\mathcal{E}^0 := \{\alpha \in M(S) \mid \alpha(i) \in R_i^N \cup \{i\} \text{ for all } i \in N\}$ (i.e., the set of all individually rational matchings).
- In general, for $k \in \{1, 2, \dots, n\}$, let $\mathcal{E}^k \subset \mathcal{E}^{k-1}$ be such that

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

The PM is defined as follows:

$$\text{PM}^>(R^N) := \mathcal{E}^n.$$

For a given preference profile R^S in the market with S , define

$$\text{PM}^>(R^S) := \{\alpha \in M(S) \mid \exists \beta \in \text{PM}^>(R^S, (\emptyset)_{-S}) \text{ s.t. } \beta|_S = \alpha\},$$

where the preference profile $(R^S, (\emptyset)_{-S})$ means that each agent $i \in N \setminus S$ finds no other acceptable i.e., $R_i = \emptyset$ while each agent $i \in S$ submits $R_i = R_i^S$.

3.2 Agent-Specific Priority Profile

Given reference order $>: 1, 2, \dots, n$, an **agent-specific priority profile** $\overline{>} := \{>_i\}_{i \in N}$ is a collection of linear orders with each $>_i$ defined for $N \setminus \{i\}$ such that

- $>_1: 2, 3, 4, 5, \dots, n$ or $>_1: 3, 2, 4, 5, \dots, n$
- $>_2: 1, 3, 4, 5, \dots, n$ or $>_2: 3, 1, 4, 5, \dots, n$
- $>_i: 1, 2, 3, 4, \dots, i-1, i+1, \dots, n$ or $>_i: 2, 1, 3, 4, \dots, i-1, i+1, \dots, n$, for $i = 3, 4, \dots, n$
- $\exists i \neq j \in \{1, 2, 3\}$ s.t. $>_i: j, k, 4, 5, \dots, n$ and $>_j: i, k, 4, 5, \dots, n$, where $k \in \{1, 2, 3\} \setminus \{i, j\}$

The priority associated with agent i (or the endowment of agent i), denoted by $>_i$, is an ordering over the rest of the agents. If all orders $>_i$ have a common linear extension over N i.e., $\cup_{i \in N} >_i$ is acyclic, then we have a single priority order over all agents N that is applied to all endowments; this corresponds to the PM.

The class of agent-specific priority profiles is quite small. Given the reference priority order $>$, each agent-specific priority $>_i$ can differ only in the relative rankings among the top two agents; otherwise, it must coincide with the reference

priority order. Note that $>_i$ is defined over $N \setminus \{i\}$. Therefore, the top two agents in $>_1$ are agent 2 and 3 (similarly for $>_2$). As a result, agent 1, 2, and 3 play some special roles. Also, there cannot be an “odd length cycle” among the top three agents in the sense that two of the three who both prioritize one another over the third (this corresponds to the fourth bullet point in the above definition).

We now define the new mechanism. The **tripolar priority mechanism (TPM)** with $\succ$, denoted by $\text{TPM}^\succ$, is a mechanism defined as follows:

$$\text{TPM}^\succ(R^S) := \begin{cases} \text{PM}^{>_{1 \leftrightarrow 2}}(R^S) & \text{if } A_1(R^S) = A_2(R^S) = \{i\} \text{ and } >_i: 2, 1, \dots \\ \text{PM}^{>_{1 \leftrightarrow 3}}(R^S) & \text{if } A_1(R^S) = A_3(R^S) = \{2\} \text{ and } >_2: 3, 1, \dots \\ \text{PM}^{>_{2 \leftrightarrow 3}}(R^S) & \text{if } A_2(R^S) = A_3(R^S) = \{1\} \text{ and } >_1: 3, 2, \dots \\ \text{PM}^{>_{1 \leftrightarrow 3}}(R^S) & \text{if } \{1, 2, 3\} \text{ forms an isolated odd cycle in } G(R^S), \\ & >_2: 3, 1, \dots, \text{ and } >_3: 2, 1, \dots \\ \text{PM}^{>_{2 \leftrightarrow 3}}(R^S) & \text{if } \{1, 2, 3\} \text{ forms an isolated odd cycle in } G(R^S), \\ & >_1: 3, 2, \dots, \text{ and } >_3: 1, 2, \dots \\ \text{PM}^\succ(R^S) & \text{otherwise.} \end{cases}$$

The TPM resembles the PM. Depending on the preference profile R^S and the induced mutual compatibility graph $G(R^S)$, only the ranking among the top three agents can vary. These priority flips correspond to the difference among the agent-specific priorities. For example, suppose the only acceptable partner for both agent 1 and agent 2 is agent 4. In this case, the mechanism matches the agent with a higher priority in agent 4’s priority, $>_4$. If agent 1 ranks higher than agent 2 in $>_4$, agent 1 is matched with agent 4; otherwise, agent 4 is matched with agent 2. This situation corresponds to the first line of the above definition of the TPM with $i = 4$.

The two examples illustrate how the TPM works.

Example 1. Suppose that there are 3 agents, $N = \{1, 2, 3\}$, and the mutual acceptability graph is given in Figure 1.

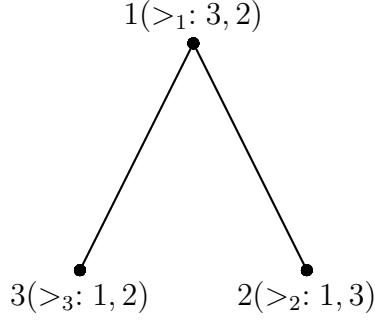


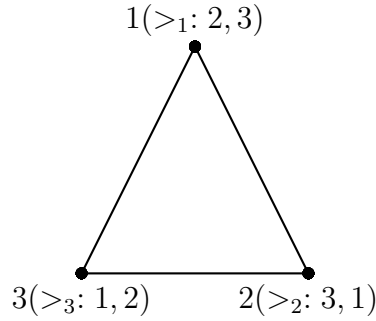
Figure 1: $G(R^N)$ for Example 1

Since the third line of the definition of the TPM applies —namely, $A_2(R^S) = A_3(R^S) = \{1\}$ and $>_1: 3, 2, \dots$ —we have $\text{TPM}^\bar{>}(R^N) = \text{PM}^{>_{2 \leftrightarrow 3}}(R^N)$. In this case, agent 1's priority, $>_1$, ranks agent 3 above agent 2. Therefore, agent 3 matches agent 1, and agent 2 remains single.

The next example explains the problem with an odd length cycle in the priorities of top three agents.

Example 2. Suppose that there are 3 agents, $N = \{1, 2, 3\}$, and the mutual acceptability graph is given in Figure 2.

Figure 2: $G(R^N)$ for Example 2



In this case, there is no way to avoid violating priorities without giving up Pareto efficiency. If we force the mechanism to match agents 1 and 2, agent 3 can benefit by declaring agent 1 unacceptable. The fourth condition on agent-specific priority profiles rules out this possibility.

4 Main Result

The main result of this paper is the following theorem.

Theorem 1. *A mechanism F satisfies consistency, Pareto efficiency, strategyproofness, and respect for improvements if and only if $F = \text{TPM}^{\succ}$ for some agent-specific priority profile $\succ$.*

This characterization implies that all mechanisms satisfying the axioms closely resemble the PM, thereby providing support for its use in pairwise donor exchange problems. Although TPM allows for some degree of priority design among the top three agents, this flexibility may be negligible relative to the typical size of living donor organ exchange markets. For instance, the UK Living Kidney Sharing Scheme routinely include around 250–300 patient–donor pairs per matching run in the pool, with about 15–25% being new registrants each quarter (Biró et al., 2021). The national program in Spain involves over 100 participating pairs per run (Andersson and Kratz, 2020). In the United States, the National Kidney Registry’s pool size, defined as the number of unmatched patient–donor pairs (a proxy for registered pairs), is on the order of 100–300 pairs, meaning that a single snapshot of the quarterly pool already contains at least one hundred incompatible pairs (National Kidney Registry, 2018).

Our result also relates to recent work highlighting the usefulness of the PM-based approaches in (balanced) exchanges of multiple goods (Andersson et al., 2021; Manjunath and Westkamp, 2021). The characterization highlights the strong structural implications of the axioms: even in the simplest pairwise exchange environment, mechanisms satisfying these properties must closely resemble PM-type rules.

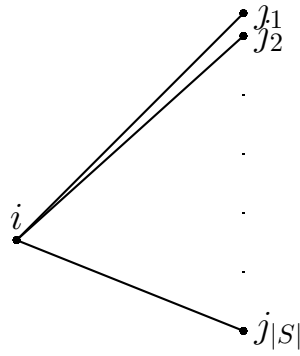
4.1 Proof Sketch

The formal proof is provided in the Appendix. Here we outline the structure of the proof. It is routine (though somewhat lengthy) to verify that TPMs satisfy all the axioms. To establish necessity, the argument proceeds in three steps.

First, we construct an agent-specific priority order $\succ_i$. For each $S \subset N \setminus \{i\}$, consider preference profiles $R^{(i,S)}$ whose associated graph $G(R^{(i,S)})$ is a complete bipartite graph with i on one side and agents in S on the other, with edges between i and every agent in S (Figure 3).¹¹ Any Pareto efficient and individually rational mechanism F must match agent i with exactly one agent in S , leaving the others

¹¹ $R^{(i,S)}$ denotes a class of preference profiles because the graph $G(R^{(i,S)})$ can arise from multiple profiles. An edge represents mutual acceptability, whereas the absence of an edge may reflect either mutual unacceptability or one-sided acceptability. Hence different preference profiles may induce the same graph $G(R^{(i,S)})$

Figure 3: $G(R^{(i,S)})$



single. This induces a choice function

$$C_i : 2^{N \setminus \{i\}} \rightarrow N \setminus \{i\}$$

such that $C_i(S) \in S$ for each $S \subset N \setminus \{i\}$. Intuitively, $C_i(S)$ identifies the agent selected when a triage situation arises among agents in S over the endowment of agent i . Consistency of mechanism F implies that C_i is rationalizable i.e., there is no cycle in its revealed preference relation (Lemma 2). Hence, there exists a linear order on $N \setminus \{i\}$ that rationalizes C_i , which we adopt as $>_i$.¹²

Second, we characterize the admissible heterogeneity in the profile of priorities $\succ = (>_i)$. The restriction arises from situations in which priorities must jointly determine which agents obtain the endowments of two agents. To see this, consider preference profiles $R^{(\{i,j\}, \{k_1, k_2, k_3\})}$ whose graph is a complete bipartite graph with two agents $\{i, j\}$ on one side and three agents $\{k_1, k_2, k_3\} \subset N \setminus \{i, j\}$ on the other (Figure 4).

Pareto efficiency implies that one of k_1, k_2, k_3 remains single; without loss of generality, suppose it is k_3 . By Pareto indifference and essential single-valuedness,

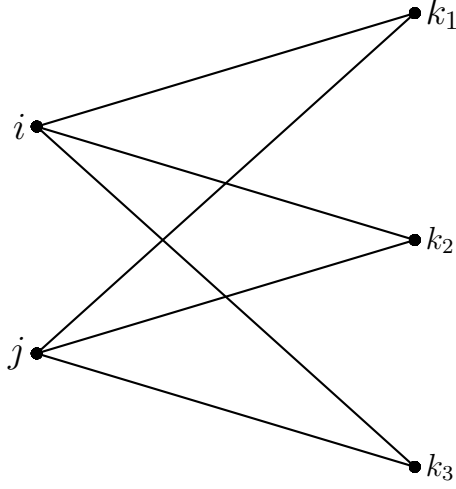
$$F(R^{(\{i,j\}, \{k_1, k_2, k_3\})}) = \{\alpha_1, \alpha_2\}$$

where $\alpha_1 = \{\{i, k_1\}, \{j, k_2\}\}$, $\alpha_2 = \{\{i, k_2\}, \{j, k_1\}\}$. If $\{i, k_1\}$ leave the market under α_1 , consistency implies

$$\alpha_1 \upharpoonright_{\{j, k_2, k_3\}} = \{\{j, k_2\}\} \in F(R^{(j, \{k_2, k_3\})}).$$

¹²See, e.g., Chambers and Echenique (2016) for fundamental results in revealed preference theory.

Figure 4: $G(R^{\{i,j\},\{k_1,k_2,k_3\}})$



Hence $k_2 >_j k_3$. By symmetry, $k_2 >_i k_3$.

Thus, the top two agents in $\{k_1, k_2, k_3\}$ must be identical under $>_i$ and $>_j$ although their relative order may differ. Extending this argument shows that any pair of priorities can differ only in their top two positions (Lemma 4 and Lemma 5).

Consider the top two agents 1 and 2. Since $>_1$ is defined on $N \setminus \{1\}$, the top two positions are filled by agent 2 and another agent: let's call her agent 3. Similarly, agents 1 and 3 occupy the top two positions of $>_2$. If a cycle such that $1 >_2 3 >_1 2 >_3 1$ arises, the mechanism becomes manipulable. Thus, we get the fourth restriction on agent-specific priorities (Lemma 6). In sum, the constructed order profile $(>_i)_{i \in N}$ is an agent-specific priority profile.

Third, we show that $F(R^S) = \text{TPM}^{\bar{>}}(R^S)$ for every preference profile R^S . If this fails at some profile R^S , mechanism F must violate the priorities $(>_i)_{i \in N}$, which in turn creates a profitable manipulation for the agent who benefits from the priority violation (Lemma 7).

4.2 Independence of Axioms

We present four mechanisms that satisfy all but one of the axioms in the main result.

First, a mechanism can satisfy consistency, Pareto efficiency, and strategyproof-

ness without respecting improvements. Consider the following variant of the PM:

$$F(R^S) = \begin{cases} \text{PM}^{>2 \leftrightarrow 3}(R^S) & \text{if } R^S \in \mathcal{P} \\ \text{PM}^{>}(R^S) & \text{otherwise,} \end{cases}$$

where the class of preference profiles $\mathcal{P}$ is defined as follows: For each $R^S \in \mathcal{P}$

- (1) : $4, 5 \in A_1(R^S)$
- (2) : $4 \in A_2(R^S)$ and $2 \notin R_1^S \cup R_5^S$
- (3) : $5 \in A_3(R^S)$ and $3 \notin R_1^S \cup R_4^S$.

Agents 2 and 3 have different potential partners (4 and 5, respectively). However, only one of them can be matched with 4 or 5 because 1 is guaranteed to be matched with one of these agents. If there is another potential partner $i \in A_1(R^S) \cup A_2(R^S) \cup A_3(R^S) \setminus \{4, 5\}$, then all of $\{1, 2, 3, 4, 5\}$ are matched anyway.

Now consider a profile $R^S \notin \mathcal{P}$ such that $\{4, 5\} = A_1(R^S)$, $4 = A_2(R^S)$, $5 = A_3(R^S)$, $2 \notin R_1^S \cup R_5^S$, but $3 \in R_1^S \cup R_4^S$. In this case, 3 will be left single by F . Thus, 3 benefits from $3 \notin R_1^S \cup R_4^S$. Note that this mechanism is individually rational.

Second, a mechanism satisfying consistency, Pareto efficiency, and respect for improvements may violate strategyproofness. Consider the following variant of the PM:

$$F(R^S) = \begin{cases} \text{PM}^{>1 \leftrightarrow 2}(R^S) & \text{if } 2 \in R_1^S \\ \text{PM}^{>}(R^S) & \text{otherwise.} \end{cases}$$

That is, whenever agent 1 finds agent 2 acceptable, the priorities of 1 and 2 are swapped. This creates a problem when agents 1 and 2 share a unique mutually acceptable partner, say agent i , while agent 2 finds agent 1 unacceptable ($1 \notin R_2^S$). In this case, agent 1 can benefit by hiding the acceptability of agent 2.

Third, the null mechanism defined by $F(R^S) = \{\emptyset\}$, for all R^S satisfies all the axioms except Pareto efficiency.

Finally, the following “bossy” variant of the PM satisfies all axioms except consistency:

$$F(R^S) = \begin{cases} \text{PM}^{>3 \leftrightarrow 4}(R^S) & \text{if } \{1, 2\} \in E(R^S) \\ \text{PM}^{>}(R^S) & \text{otherwise.} \end{cases}$$

5 Concluding Remarks

We characterize a class of mechanisms in voluntary pairwise matching problems with dichotomous preferences. This result provides theoretical support for the use of the PM in real-world living-donor exchanges, where pairwise exchanges play central roles in some countries. Our results may further help explain the distinctive features of the PM, in line with recent studies demonstrating the usefulness of PM variants in contexts where impossibility results occur under the standard strict preference domain.

Our characterization also highlights important limitations of the PM approach in the context of living-donor exchange. In particular, TPM can accommodate only a very limited degree of heterogeneity in individualized priorities. As a consequence, PM-like mechanisms are poorly suited for exchange programs with explicit distributional objectives. For example, the Spanish National Kidney Exchange Program prioritizes cycles within the same region over otherwise identical cycles spanning multiple regions. Similarly, many kidney exchange programs incorporate age differences among donors within a cycle.¹³ In smaller markets, such as France, however, the limited scope for priority design may still be sufficient to address these distributional policy goals.

Appendix A: Omitted Proofs

Throughout Appendix A, we use F to denote a mechanism satisfying consistency, Pareto efficiency, strategyproofness, and respect for improvements.

A.1 Lemmas

Since F is essentially single-valued, we know that the set of agents who are matched with mutually acceptable partners is the same across all matchings chosen by F . Let $N^F(R^S)$ be the set of all matched agents under F at the preference profile R^S .

We first construct agent-specific priority profile $\succ := \{\succ_i\}_{i \in N}$. For each i and $S \subset N \setminus \{i\}$, consider an equivalent class of preference profiles $R^{(i,S)}$ such that $G(R^{(i,S)})$ is the bipartite graph with agent i being the unique agent (node) on one side and i is connected with all agents in S who are on the other side of the graph.

¹³To incorporate such objectives, these programs typically rely on integer optimization technologies.

Since F is Pareto efficient and individually rational, one of the agents in S is matched with i while the rest of agents remain single at any preference profile of $R^{(i,S)}$.¹⁴ This induces a choice function

$$C_i : 2^{N \setminus \{i\}} \rightarrow N \setminus \{i\}$$

such that $C_i(S) \in S$ for each $S \subset N \setminus \{i\}$. Intuitively, $C_i(S)$ identifies the agent selected when a triage situation arises among agents in S over the endowment of agent i . $>_i$ is defined as the revealed preference relation of C_i . Formally, we define $>_i$ as

$$>_i := \{ (j, k) \in N^2 \mid \exists S \subset N \setminus \{i\} \text{ s.t. } C_i(S) = j \text{ and } k \in S \}.$$

Since F is consistent, $>_i$ is a linear order on $N \setminus \{i\}$.

Lemma 2. $>_i$ is acyclic.

Proof. Assume not so. Then, there are distinct agents $j_1, \dots, j_K \in N \setminus \{i\}$ and distinct (opportunity) sets $S_1, \dots, S_K \subset N \setminus \{i\}$ s.t. $j_k, j_{k+1} \in S_k$ and $C_i(S_k) = j_k$ for all $k \in \{1, 2, \dots, K\}$, where $j_{K+1} := j_1$. Let $S^* := \{j_1, \dots, j_K\}$ and suppose without loss of generality $C_i(S^*) = j_1$. This means that $N^F(R^{(i,S^*)}) = \{j_1, i\}$ and hence $\alpha(S^* \setminus \{j_1, j_K\}) = S^* \setminus \{j_1, j_K\}$ and $\alpha(i) = j_1$ for all $\alpha \in F(R^{(i,S^*)})$. By consistency, $\alpha|_{\{i, j_1, j_K\}} \in F(R^{(i,S^*)}|_{\{i, j_1, j_K\}}) = F(R^{(i, \{j_1, j_K\})})$.

The same preference profile $R^{(i, \{j_1, j_K\})}$ can be reached from $R^{(i, S_K)}$ as follows. From $C_i(S_K) = j_K$, we know that $N^F(R^{(i, S_K)}) = \{i, j_K\}$, and so $\beta(S_K \setminus \{j_1, j_K\}) = S_K \setminus \{j_1, j_K\}$ and $\beta(i) = j_K$ for all $\beta \in F(R^{(i, S_K)})$. By consistency, $\beta|_{\{i, j_1, j_K\}} \in F(R^{(i, S_K)}|_{\{i, j_1, j_K\}}) = F(R^{(i, \{j_1, j_K\})})$.

Finally, $j_1 = \alpha(i)$ and $j_K = \beta(i)$ contradict essential single-valuedness of F at $R^{(i, \{j_1, j_K\})}$ as desired. $\square$

Since $>_i$ is complete, the acyclicity of $>_i$ implies the transitivity of $>_i$. Thus, we have well-defined orders $\{>_i\}_{i \in N}$.

By construction of $>_i$, the following lemma is an immediate consequence.

Lemma 3. Suppose $>_i$ is given as $>_i: j_1, j_2, \dots, j_{n-1}$ and $\{i, j_k\} \in E(R^S)$. If $j_k \notin N^F(R^S)$, then for all $\alpha \in F(R^S)$, $\alpha(i) \in \{j_1, \dots, j_{k-1}\}$.

Proof. Assume not so. Then, there is $\alpha \in F(R^S)$ s.t. $\alpha(i) \in \{j_{k+1}, \dots, j_{n-1}\} \cup \{i\}$. By Pareto efficiency, i must be matched and so $\alpha(i) \in \{j_{k+1}, \dots, j_{n-1}\}$. As

¹⁴Consistency of F ensures that F chooses the same set of matchings at all profiles inducing the same bipartite graph $G(R^{(i,S)})$.

$\alpha(S \setminus \{i, \alpha(i), j_k\}) = S \setminus \{i, \alpha(i), j_k\}$, Consistency implies

$$\alpha|_{\{i, \alpha(i), j_k\}} \in F(R^S|_{\{i, \alpha(i), j_k\}}).$$

We divide into two cases.

Case 1 $\{j_k, \alpha(i)\} \notin E(R^S)$: In this case, $R^S|_{\{i, \alpha(i), j_k\}}$ actually belongs to the class $R^{(i, \{\alpha(i), j_k\})}$. Since $>_i$ is the revealed preference of C_i , $C_i(\{\alpha(i), j_k\}) = j_k$. This means that $\beta(i) = j_k$ for all $\beta \in F(R^S|_{\{i, \alpha(i), j_k\}})$, which contradicts essential single-valuedness of F at $(R^S|_{\{i, \alpha(i), j_k\}})$ as $\alpha(i) \neq j_k$.

Case 2 $\{j_k, \alpha(i)\} \in E(R^S)$: In this case, $\{j_k, \alpha(i)\} \in E(R^S|_{\{i, \alpha(i), j_k\}})$ and consistency imply that $N^F(R^S|_{\{i, \alpha(i), j_k\}}) = \{i, \alpha(i)\}$. Consider $Q_{j_k}^{\{i, \alpha(i), j_k\}} := \{i\}$. If

$$j_k \in N^F(Q_{j_k}^{\{i, \alpha(i), j_k\}}, (R^S|_{\{i, \alpha(i), j_k\}})_{\{i, \alpha(i)\}}),$$

then j_k can manipulate F at $R^S|_{\{i, \alpha(i), j_k\}}$ via $Q_{j_k}^{\{i, \alpha(i), j_k\}}$. Therefore, j_k should remain single at this new profile. However, this is a contradiction because

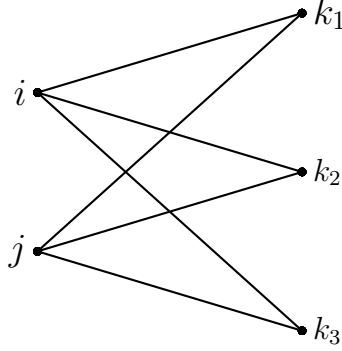
$$(Q_{j_k}^{\{i, \alpha(i), j_k\}}, (R^S|_{\{i, \alpha(i), j_k\}})_{\{i, \alpha(i)\}})$$

is in the class $R^{(i, \{\alpha(i), j_k\})}$. □

The Lemma 3 states that if agent j_k is left unmatched by F even though he has a mutually acceptable candidate i , then agent i must be matched with those with higher priority than j_k according to $>_i$. Note that Lemma 3 explains an agent's unfavorable outcome via $>_i$ only when the agent is mutually acceptable with i . However, due to indifferences, triage decisions may arise between agents who do not share any mutually acceptable partners as long as they are connected in the compatibility graph. In this sense, the role of the priority ordering $>_i$ is not limited to the above situation; rather, it must coordinate with other priority orderings to jointly determine triage outcomes. Consequently, these priority orderings must be closely aligned with one another. To see this, consider the following situation. Three agents, $\{k_1, k_2, k_3\}$, demand the endowments of i and j , and the mutual acceptability graph $G(R^{(\{i, j\}, \{k_1, k_2, k_3\})})$ is a complete bipartite graph with i and j on one side and $\{k_1, k_2, k_3\}$ on the other (Figure 5).

In this situation, which two of $\{k_1, k_2, k_3\}$ are matched with either i or j constitutes the only welfare-relevant information. The role of the priority $>_i$ is to coordinate with the priority $>_j$ to jointly triage among agents demanding both i and j . To ensure this joint role, some restrictions on the heterogeneity of priorities

Figure 5: $G(R(\{i,j\},\{k_1,k_2,k_3\}))$



must be imposed. The following interplay of Pareto indifference and consistency shows that the situation already imposes quite stringent restrictions:

Consider the above graph $G(R(\{i,j\},\{k_1,k_2,k_3\}))$. Since F is Pareto efficient, without loss of generality, we can suppose that only k_3 is left single by F . Because of Pareto indifference, $F(R(\{i,j\},\{k_1,k_2,k_3\})) = \{\alpha_1, \alpha_2\}$ where

$$\begin{aligned}\alpha_1 &= \{ \{i, k_1\}, \{j, k_2\} \} \quad , \text{ and} \\ \alpha_2 &= \{ \{i, k_2\}, \{j, k_1\} \}.\end{aligned}$$

If we pick up α_1 and let $\{i, k_1\}$ leave the market, the remaining market is $G(R^{(j,\{k_2,k_3\})})$, and, by consistency, the same match should be recommended by F i.e., $\alpha_1|_{\{j,k_2,k_3\}} = \{\{j, k_2\}\} \in F(R^{(j,\{k_2,k_3\})})$, which means that $k_2 >_j k_3$. By symmetric reasoning, we deduce that $k_2 >_i k_3$; therefore, the top two agents in one priority $>_i$ must be the top two in another priority $>_j$ as well (the relative ranking among the top two can be different in two priorities). The next lemma 4 generalizes this observation. We use the following notation: for $i \in N$, $S \subset N \setminus \{i\}$, and $\ell \in \{1, 2, \dots, |S|\}$,

$$\sigma_i(\ell | S) := \text{“}\ell\text{-th ranked agent with respect to } >_i \text{ among } S\text{.”}$$

Lemma 4. *For any distinct pair $i, j \in N$, and for any subset of other agents $S \subset N \setminus \{i, j\}$ with $|S| \geq 3$, the following statements hold:*

- (a) $\{\sigma_i(1 | S), \sigma_i(2 | S)\} = \{\sigma_j(1 | S), \sigma_j(2 | S)\}$, and
- (b) $\sigma_i(k | S) = \sigma_j(k | S)$ for all $k \in \{3, 4, \dots, |S|\}$.

Proof. (a): We prove this by contradiction. Assume that there are $i \neq j$, $S \subset N \setminus \{i, j\}$ with $|S| \geq 3$ such that $k_1 := \sigma_i(1 | S)$, $k_2 := \sigma_i(2 | S)$, and $k_3 \in \{\sigma_j(1 | S), \sigma_j(2 | S)\}$ with all distinct. Consider a preference profile R^S in the

equivalent class $R^{\{i,j\},\{k_1,k_2,k_3\}}$. By Pareto efficiency, all but one of $\{k_1, k_2, k_3\}$ are matched. We divide into three cases:

Case 1 k_1 is the unmatched: By Pareto indifference of F ,

$$\alpha = \{ \{i, k_2\}, \{j, k_3\} \} \in F(R^S).$$

By consistency,

$$\alpha|_{\{i,k_1,k_2\}} \in F(R^S|_{\{i,k_1,k_2\}}).$$

However, $R^S|_{\{i,k_1,k_2\}}$ is in $R^{(i,\{k_1,k_2\})}$ and so k_1 should be matched as $k_1 >_i k_2$, which is a contradiction.

Case 2 k_2 is the unmatched: Note that $k_1 >_i k_2 >_i k_3$. By Pareto indifference of F ,

$$\alpha = \{ \{i, k_3\}, \{j, k_1\} \} \in F(R^S).$$

By consistency,

$$\alpha|_{\{i,k_2,k_3\}} \in F(R^S|_{\{i,k_2,k_3\}}).$$

However, $(R^S|_{\{i,k_2,k_3\}})$ is in $R^{(i,\{k_2,k_3\})}$ and so k_2 should be matched as $k_2 >_i k_3$, which is a contradiction.

Case 3 k_3 is the unmatched: Note that either $k_3 >_j k_2$ or $k_3 >_j k_1$. Without loss of generality, suppose $k_3 >_j k_2$. By Pareto indifference of F ,

$$\alpha = \{ \{i, k_1\}, \{j, k_2\} \} \in F(R^S).$$

By consistency,

$$\alpha|_{\{j,k_2,k_3\}} \in F(R^S|_{\{j,k_2,k_3\}}).$$

However, $(R^S|_{\{j,k_2,k_3\}})$ is in $R^{(j,\{k_2,k_3\})}$ and so k_3 should be matched as $k_3 >_j k_2$, which is a contradiction.

(b): Assume that $\sigma_i(k|S) \neq \sigma_j(k|S)$ for some $k \in \{3, 4, \dots, |S|\}$ and $S \subset N$ with $|S| \geq 3$. Let

$$S' := (\{\sigma_i(m|S) \mid m = 1, 2, k\} \cup \{\sigma_j(m|S) \mid m = 1, 2, k\}) \setminus \{\sigma_i(2|S)\}.$$

Then,

$$\{\sigma_i(1|S'), \sigma_i(2|S'), \sigma_j(2|S')\} = \{\sigma_i(1|S), \sigma_i(k|S), \sigma_j(k|S)\}.$$

Note that all of the three in the above sets are distinct because

$$\sigma_i(1 \mid S) \neq \sigma_i(k \mid S) \neq \sigma_j(k \mid S)$$

and

$$\sigma_j(k \mid S) \notin \{\sigma_j(1 \mid S), \sigma_j(2 \mid S)\} = \{\sigma_i(1 \mid S), \sigma_i(2 \mid S)\}.$$

One can apply (a) to S' because

$$|S'| \geq |\{\sigma_i(1 \mid S'), \sigma_i(2 \mid S'), \sigma_j(2 \mid S')\}| = 3.$$

But this is a contradiction since $\sigma_j(2 \mid S') \notin \{\sigma_i(1 \mid S'), \sigma_i(2 \mid S')\}$. $\square$

We apply this implied restriction to every pair of priorities to obtain the following common structure among all priorities.

Lemma 5. $\exists D \subset N$ such that $|D| = 3$ and the following two hold:

$$\begin{aligned} (1): \quad D &= \{\sigma_i(m \mid N \setminus \{i\}) \mid m = 1, 2, 3\} \quad \text{for } i \notin D \\ (2): \quad D \setminus \{i\} &= \{\sigma_i(m \mid N \setminus \{i\}) \mid m = 1, 2\} \quad \text{for } i \in D \end{aligned}$$

Proof. Let

$$\mathcal{A} := \left\{ A \subset N \mid \begin{array}{ll} (1): & A = \{\sigma_i(m \mid N \setminus \{i\}) \mid m \in \{1, 2, \dots, |A|\}\} \quad \text{for } i \notin A \\ (2): & A \setminus \{i\} = \{\sigma_i(m \mid N \setminus \{i\}) \mid m \in \{1, 2, \dots, |A| - 1\}\} \quad \text{for } i \in A \end{array} \right\}.$$

Observe that $N \in \mathcal{A}$. Take an arbitrary agent $i_0 \in N$ and fix it. Let $i_1 := \sigma_{i_0}(n-1 \mid N \setminus \{i_0\})$. Then, for each pair $j, k \in N \setminus \{i_0, i_1\}$,

$$\sigma_j(n-1 \mid N \setminus \{j\}) = \sigma_k(n-1 \mid N \setminus \{k\}) \in \{i_0, i_1\}$$

by part (b) of Lemma 4 and our universal assumption $|N| \geq 5$. Let $w_1 := \sigma_j(n-1 \mid N \setminus \{j\})$ for some (and hence for all) $j \in N \setminus \{i_0, i_1\}$; w_1 is the common worst agent with respect to $\{>_j\}_{j \in N \setminus \{i_0, i_1\}}$. Let $A_0 := N$.

Let $A_1 := A_0 \setminus \{w_1\}$. We show that $A_1 \in \mathcal{A}$. The unique agent i satisfying $i \notin A_1$ is w_1 . It is immediate to see

$$\{\sigma_{w_1}(m \mid N \setminus \{w_1\}) \mid m \in \{1, 2, \dots, |A_1|\}\} = A_1.$$

For any other agent i with $i \in A_1$, the $(n-1)$ th priority agent with respect to $>_i$

is w_1 . So,

$$\{\sigma_i(m \mid N \setminus \{i\}) \mid m \in \{1, 2, \dots, |A_1| - 1\}\} = A_1 \setminus \{i\}.$$

Assume that we are given $\{A_\ell\}_{\ell \in \{1, 2, \dots, k\}} \subset \mathcal{A}$ with $|A_\ell| \geq 5$ and $\{w_\ell\}_{\ell \in \{1, 2, \dots, k\}}$ with w_ℓ being the $(n - \ell)$ th priority agent with respect to $>_i$ for all $i \in A_\ell$ and the $(n - \ell + 1)$ th priority agent with respect to $>_i$ for all $i \in N \setminus A_{\ell-1}$. Pick $i_0 \in A_k$ arbitrarily and let $i_1 := \sigma_{i_0}(n - k \mid N \setminus \{i_0\})$. Since $|A_k| \geq 5$, for each pair $j, k \in A_k \setminus \{i_0, i_1\}$,

$$\sigma_j(n - k \mid N \setminus \{j\}) = \sigma_k(n - k \mid N \setminus \{k\}) \in \{i_0, i_1\}$$

by part (b) of Lemma 4. Let $w_{k+1} := \sigma_j(n - k \mid N \setminus \{j\})$ for some (and hence for all) $j \in A_k \setminus \{i_0, i_1\}$. w_{k+1} is $(n - (k + 1))$ th priority agent with respect to $>_i$ for all $i \in A_{k+1} := A_k \setminus \{w_{k+1}\}$; and w_{k+1} is $(n - k)$ th priority agent with respect to $>_i$ for all $i \notin A_k$ by part (b) of Lemma 4. Therefore, $A_{k+1} \in \mathcal{A}$. This inductive construction works up to $k = n - 4$ and we have $A_{n-4} \in \mathcal{A}$ with $|A_{n-4}| = 4$. Renaming agents if necessary, let $A_{n-4} = \{1, 2, 3, 4\}$ and $w_k = k$ for each $k = 5, 6, \dots, n$.

For any distinct pair $i, j \in N \setminus A_{n-4}$ and for each $m \in \{3, 4\}$, $\sigma_i(m \mid A_{n-4}) = \sigma_j(m \mid A_{n-4})$ by part (b) of Lemma 4. Therefore, without loss of generality, we can express priority order $>_i$ for each $i \in N \setminus A_{n-4}$ as follows:

$$>_i: 1, 2, 3, 4, \dots, i - 1, i + 1, \dots, n \text{ or } >_i: 2, 1, 3, 4, \dots, i - 1, i + 1, \dots, n.$$

For each $j \in \{1, 2, 3\}$, applying part (b) of Lemma 4 to pair $\{j, 5\}$, we have $\sigma_j(3 \mid A_{n-4} \setminus \{j\}) = 4$. This yields $D := \{1, 2, 3\} \in \mathcal{A}$ as desired. $\square$

In the rest of Appendix A, without loss of generality, we assume that $D = \{1, 2, 3\}$ the set of agents occupying the top positions in all priority orderings, except in their own. The remaining agents are ordered according to the natural ordering $>$. The following lemma completes the characterization of all admissible agent-specific priority profiles, $\succ$.

Lemma 6. $\exists i \neq j \in \{1, 2, 3\}$ s.t. $>_i: j, k, 4, 5, \dots, n$ and $>_j: i, k, 4, 5, \dots, n$, where $k \in \{1, 2, 3\} \setminus \{i, j\}$.

Proof. Toward the contradiction, assume that there is an odd-length cycle in their

priorities: without loss of generality, assume that we have

$$2 \succ_1 3 \succ_2 1 \succ_3 2.$$

Consider submarket with $S = \{1, 2, 3\}$ in which every pair is mutually acceptable: $R_i^S = \{j, k\}$ for all distinct $i, j, k \in \{1, 2, 3\}$. By Pareto efficiency, the two of them are matched while one of them remains single under F . Without loss of generality, suppose 3 is the unmatched. Now, consider $Q_3^S := R_3^S \setminus \{1\}$. Then, (Q_3^S, R_{-3}^S) is in class $R^{(2, \{1, 3\})}$. Since $3 \succ_2 1$, 3 becomes matched with 2, which contradicts strategyproofness of F . $\square$

Lemma 7. *Let $k \notin N^F(R^S)$ with $A_k(R^S) \neq \emptyset$, and β be a maximum matching in $G(R^S)$ with $\beta(k) \neq k$. Then, there is $h \in N^F(R^S)$ such that $\beta(h) = h$ and $h \succ_{\beta(k)} k$.*

Proof. We prove this by contradiction. Assume that there are R^S , $k \notin N^F(R^S)$ with $A_k(R^S) \neq \emptyset$, and a maximum matching β in $G(R^S)$ with $\beta(k) \neq k$; and yet suppose that for all $h \in N^F(R^S)$,

$$\beta(h) \neq h \text{ or } k \succ_{\beta(k)} h.$$

Take $\alpha \in F(R^S)$. Let $\gamma := \alpha \Delta \beta$ be the symmetric difference of the two matchings, and let p be a path in γ whose one endpoint is k . Let j be the other endpoint of p (Figure 6). Let

$$N^p := \{j' \in S \mid \{j', j''\} \in p \text{ for some } j'' \in S\}$$

be the set of agents involved in this path.

Then, $N^p \setminus \{k\} \subset N^F(R^S)$. If $\beta(h) \neq h$ for all $h \in N^F(R^S)$, then β matches strictly more agents than α , contradicting the maximality of α . Therefore,

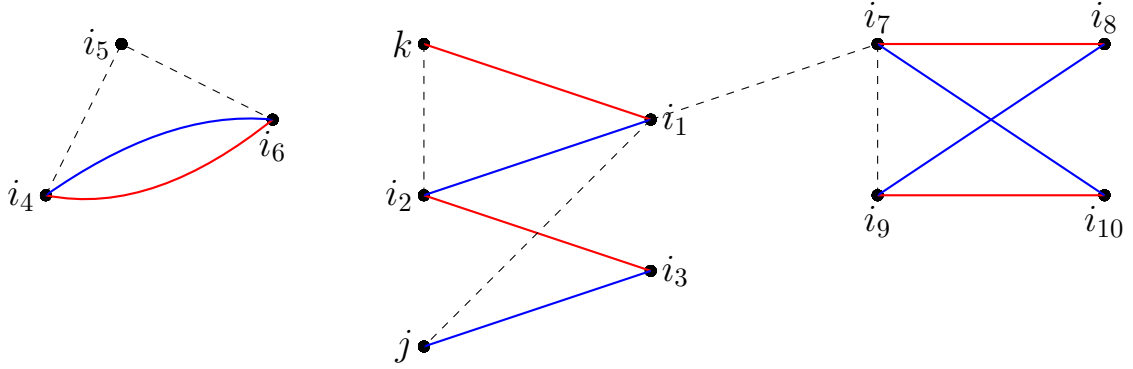
$$S^{<k} := \{h \in N^p \setminus \{k, \beta(k)\} \mid k \succ_{\beta(k)} h\} \neq \emptyset.$$

Since $\alpha(N^p) = N^p$, consistency implies that

$$\alpha|_{N^p} \in F(R^S|_{N^p}).$$

Next, consider a submarket with agents $S' := N^p \setminus \{k, \beta(k)\}$ and preference profile $R^S|_{S'}$. Matching $\alpha \cap p|_{S'}$ matches all but one agent, $S' \setminus \{\alpha(\beta(k))\}$; therefore, any maximum matching in this submarket must match all but one. Therefore,

Figure 6: An illustration of the symmetric difference $\gamma = \alpha \Delta \beta$



The edges corresponding to matching α are drawn in red, while those of matching β are drawn in blue; and dashed lines indicate edges that do not belong to either matching in $G(R^S)$. The edge $\{i_4, i_6\}$ belongs to both α and β . The symmetric difference of α and β is all edges used in exactly one of α and β . In this example,

$$\gamma = \alpha \Delta \beta = \alpha \cup \beta \setminus \{\{i_4, i_6\}\}.$$

The path p from k to j and N^p are

$$\begin{aligned} p &= \{ \{k, i_1\}, \{i_1, i_2\}, \{i_2, i_3\}, \{i_3, j\} \} \\ N^p &= \{k, i_1, i_2, i_3, j\}. \end{aligned}$$

there is the unique agent $i^* \in S'$ who is unmatched under each selected matching $\delta \in F(R^S|_{S'})$; in other words, $N^F(R^S|_{S'}) = S' \setminus \{i^*\}$.¹⁵ We will consider the manipulation incentive of i^* and arrive at a contradiction.

Assume for a moment $i^* \in S^{<k}$, which we will prove later. Recall that k is the only unmatched agent at $R^S|_{N^p}$ i.e.,

$$N^F(R^S|_{N^p}) = N^p \setminus \{k\}.$$

By strategyproofness of F , submitting $\{\beta(k)\}$ should not make k better off:

$$N^F(\{\beta(k)\}, R_{-k}^S|_{N^p}) = N^p \setminus \{k\}.$$

If

$$\{i^*, \beta(k)\} \in E(\{\beta(k)\}, R_{-k}^S|_{N^p}),$$

then

$$\delta \cup \{i^*, \beta(k)\} \in F(\{\beta(k)\}, R_{-k}^S|_{N^p})$$

because $N^F(\{\beta(k)\}, R_{-k}^S|_{N^p}) = N^p \setminus \{k\}$ and Pareto indifference of F . However, our assumption $i^* \in S^{<k}$ and Lemma 3 imply

$$k \in N^F(\{\beta(k)\}, R_{-k}^S|_{N^p}),$$

which is a contradiction. Hence, $\{i^*, \beta(k)\} \notin E(\{\beta(k)\}, R_{-k}^S|_{N^p})$. If $\beta(k) \in R_{i^*}^S$ and $i^* \notin R_{\beta(k)}^S$, then through the change

$$((\{\beta(k)\}, R_{\beta(k)}^S, R_{- \{k, \beta(k)\}}^S)|_{N^p}) \rightarrow ((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{- \{k, \beta(k)\}}^S)|_{N^p})$$

$\beta(k)$ additionally finds i^* acceptable and the two profiles are otherwise same. Hence, respect for improvements of F implies that i^* should not get strictly worse off. We will show that i^* actually gets strictly worse off. If k remains single at this new profile, then

$$\{i^*, \beta(k)\} \in E((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{- \{k, \beta(k)\}}^S)|_{N^p})$$

and Pareto indifference of F imply

$$\delta \cup \{i^*, \beta(k)\} \in F((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{- \{k, \beta(k)\}}^S)|_{N^p}).$$

¹⁵ i^* is not necessarily j .

But then Lemma 3 implies that

$$k \in N^F \left((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{-\{k, \beta(k)\}}^S) \mid_{N^p} \right),$$

a contradiction. Therefore, k must be matched at this new profile. k 's only acceptable partner is $\beta(k)$; thus, any selected matching in this profile

$$\delta' \in F \left((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{-\{k, \beta(k)\}}^S) \mid_{N^p} \right)$$

must match them: $\delta'(k) = \beta(k)$ by individual rationality of F . By consistency,

$$\delta' \mid_{S'} \in F \left((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{-\{k, \beta(k)\}}^S) \mid_{S'} \right).$$

By our construction, i^* is the unique unmatched agent at this submarket. Therefore, i^* is also the unique unmatched agent in the original market as well:

$$N^F \left((\{\beta(k)\}, R_{\beta(k)}^S \cup \{i^*\}, R_{-\{k, \beta(k)\}}^S) \mid_{N^p} \right) = N^p \setminus \{i^*\}.$$

Thus, i^* gets worse off, which contradicts respect for improvements of F .

The remaining possibilities for R^S are

$$(a) : \beta(k) \notin R_{i^*}^S \text{ and } i^* \notin R_{\beta(k)}^S$$

or

$$(b) : \beta(k) \notin R_{i^*}^S \text{ and } i^* \in R_{\beta(k)}^S.$$

If (a) is the case, since the change from (a) to (b) is the improvement of i^* and F respects for improvements,

$$N^F \left((R_{\beta(k)}^S \cup \{i^*\}, R_{-\beta(k)}^S) \mid_{N^p} \right) = N^p \setminus \{k\}.$$

Thus, the rest of the analysis is the same under (a) and (b). So, we suppose (b). Consider another preference profile

$$Q^S := (R_{i^*}^S \cup \{\beta(k)\}, R_{-i^*}^S).$$

When Q^S is the true preference profile, by the discussion so far, i^* must be the unique unmatched agent:

$$N^F (Q^S \mid_{N^p}) = N^p \setminus \{i^*\}.$$

However, submitting $R_{i^*}^S$ instead, i^* gets better off as

$$N^F(R^S|_{N^p}) = N^p \setminus \{k\},$$

contradicting strategyproofness of F . Therefore, as long as $i^* \in S^{<k}$, we get a contradiction.

As stated above, we now prove that $i^* \in S^{<k}$. Agents 1 and 2 need to be treated differently as they are special.

Case 1 $k = 1$ or 2: We consider only $k = 1$ because of the symmetry.

If $i^* \in S' \setminus S^{<1}$, then either (1) $i^* = 2$ or (2) $i^* = 3$ and $\beta(1) = 2$ with $3 >_2 1$.

First consider case (1). By the choice of i^* ,

$$N^F(R^S|_{S'}) = S' \setminus \{i^*\}.$$

Note that $\beta|_{S'}$ is a maximum matching in $G(R^S|_{S'})$ with $\beta|_{S'}(2) \neq 2$. Since $1 \notin S'$,

$$S^{<2} := \{h \in S' \setminus \{2, \beta(2)\} \mid 2 >_{\beta(2)} h\} = S' \setminus \{2, \beta(2)\}.$$

That is, for the submarket $P^S|_{S'}$, all the premises of the above discussion are true and all the candidates of i^* for this submarket is in $S^{<2}$. Therefore, the logic as above leads to a contradiction. Thus, (1) is not possible.

Next, consider case (2). In this case, $1, 2 \notin S'$ implies that all the candidates of i^* for the submarket $R^S|_{S'}$ is in $S^{<3}$. Thus, we can get a contradiction as above.

Case 2 $k \geq 3$: Let $k \geq 3$ be the smallest number (the highest priority agent) for whom $i^* \notin S^{<k}$ holds. But then,

$$i^* \in \{1, 2, \dots, k-1\}.$$

The minimality of k yields the desired contradiction from $N^F(R^S|_{S'}) = S' \setminus \{i^*\}$. $\square$

A.2 Proofs for The Main Result

Lemma 8. $\text{TPM}^{\bar{>}}$ satisfies consistency, Pareto efficiency, strategyproofness, and respect for improvements.

Proof. Since Pareto efficiency is not an inter-preference property, $\text{TPM}^{\bar{>}}$ inherits it from the associated priority mechanisms.

Before turning to inter-preference properties, observe that the five exceptional cases in the definition of the TPM that induce priority flips are mutually exclusive:

if one of them holds, none of the others can. This remains true even after some matched pairs leave the market; that is, the relative priorities among the remaining agents are unchanged when matched pairs exit. Let $N(R^S)$ denote the set of agents matched under $\text{TPM}^{\bar{>}}(R^S)$.

(1) $\text{TPM}^{\bar{>}}$ satisfies respect for improvements: In $\text{TPM}^{\bar{>}}$, any weak expansion of the set of compatible partners $A_i(R^S)$ will not hurt i 's welfare.

(2) $\text{TPM}^{\bar{>}}$ satisfies consistency: Note that in each of the exceptional cases with priority flips, there is a set of three agents $\{1, 2, j\}$, $j \in N \setminus \{1, 2\}$, and one of $\{1, 2\}$ remains single while the other two are matched with each other under $\text{TPM}^{\bar{>}}(R^S)$. The status (matched or not) of the three agents are not affected by the departures of other matched agents and, conversely, their presence will not affect the status of the rest of the participants. In this sense, the three agents $\{1, 2, j\}$ are essentially separated from the rest of the market. $\text{TPM}^{\bar{>}}(R^S)$ treats the rest of the agents $S \setminus \{1, 2, j\}$ in the same way as $\text{PM}^{\bar{>}}(R^S|_{S \setminus \{1, 2, j\}})$. Therefore, $\text{TPM}^{\bar{>}}$ inherits consistency from $\text{PM}^{\bar{>}}$.

(3) $\text{TPM}^{\bar{>}}$ satisfies strategyproofness: Assume not so. Then, there are R^S , $k \in S$ with $k \notin N(R^S)$, Q_k^S , $\alpha \in \text{TPM}^{\bar{>}}(Q_k^S, R_{-k}^S)$ such that $\alpha(k) \in A_k(R^S)$. Since the PM is strategyproof for any priority, any such manipulation must change the priority. There are many cases, all of which we will examine individually.

Case 1 $k = 1$ or 2 : We consider only $k = 1$ because 1 and 2 are symmetric. $1 \notin N(R^S)$ holds only if one of the followings hold:

1. $A_1(R^S) = \emptyset$
2. $A_1(R^S) = A_2(R^S) = \{i\}$ and $>_i: 2, 1$
3. $\{1, 2, 3\}$ is a connected component in $G(P^S)$ and $>_2: 3, 1, \dots$ and $>_3: 2, 1, \dots$

Case 1.1 $A_1(R^S) = \emptyset$: There is no way to manipulate.

Case 1.2 $A_1(R^S) = A_2(R^S) = \{i\}$ and $2 >_i 1$:

In this case, the priority used is $>_{1 \leftrightarrow 2}$ i.e., $\text{TPM}^{\bar{>}}(R^S) = \text{PM}^{\bar{>}_{1 \leftrightarrow 2}}(R^S)$. Any misreport Q_1^S with $A_2(Q_1^S, R_{-1}^S) = A_2(R^S) = \{i\}$ cannot have the mechanism choose matching α with $\alpha(1) = i$ even if the priority changes. Therefore, a possible misreport must be Q_1^S with $A_2(Q_1^S, R_{-1}^S) = \{1, i\}$. In this case, observe that the new priority used is either $>$ or $>_{1 \leftrightarrow 3}$. In either case, 2 is the 2nd priority agent. Since there are two acceptable partners $\{1, i\}$ to agent 2 , agent 2 must be matched at (Q_1^S, R_{-1}^S) . This in turn implies that 1 cannot be matched with i because 1 and i are the only acceptable partner for 2 and hence 2 must be matched with either 1 or i .

Case 1.3 $\{1, 2, 3\}$ is a connected component in $G(P^S)$ and $>_2: 3, 1, \dots$, and $>_3: 2, 1, \dots$

In this case, 1 cannot be matched with either 2 or 3 simply because 1's turn comes after 2 or 3 regardless of her preference revelation.

Case 2 $k = 3$: If $A_3(R^S) = \emptyset$, there is no way to manipulate. We consider the case with $A_3(R^S) \neq \emptyset$ and $3 \notin N(R^S)$. In this case, each $i \in A_3(R^S)$ must be matched with 1 or 2 for any matching β in $\text{TPM}^{\bar{}}(R^S)$. Therefore, $|A_3(R^S)| \leq 2$.

Case 2.1 First, suppose that $A_3(R^S) = \{i\}$.

Case 2.1.1 If $i \in \{1, 2\}$, then $\{\beta(i), i\} = \{1, 2\}$, $\beta(i) >_i 3$, and $A_{\beta(i)}(R^S) = \{i\}$. Therefore, the priority used at original profile R^S is $>$. 3 cannot match with i as long as the priority used is $>$ and $A_{\beta(i)}(R^S) = \{i\}$. The only situation in which agent 3 can change the priority is when a unilateral misreport by 3 results in an isolated odd cycle $\{1, 2, 3\}$ in the mutual acceptability graph and $3 >_{\beta(i)} i$. In this case, 3 and $\beta(i)$ will be matched as they are both prioritized over i . Thus, 3 ends up being matched with the unacceptable $\beta(i)$.

Case 2.1.2 If $i \notin \{1, 2\}$, then $\beta(i) \in \{1, 2\}$ and we are in one of the following two situations: $A_{\beta(i)}(R^S) = \{i\}$; or $A_1(R^S) \cup A_2(R^S) = \{i, j\}$ with some $j \notin \{1, 2, 3, i\}$ and both of $A_1(R^S)$ and $A_2(R^S)$ are nonempty. In either case, 3 cannot change the priority used.

Case 2.2 Suppose $A_3(R^S) = \{i, j\}$ with $i \neq j$. Then, $\beta(\{i, j\}) = \{1, 2\}$ and $A_1(R^S) \cup A_2(R^S) = \{i, j\}$ with both of them nonempty.

Case 2.2.1 If $\{i, j\} = \{1, 2\}$, then $\{1, 2, 3\}$ forms an isolated odd cycle with $1 >_2 3$ and $2 >_1 3$. Therefore, 3 has no way to change the priority used.

Case 2.2.2 Suppose $\{i, j\} \neq \{1, 2\}$. Note that $\{i, j\} \cap \{1, 2\} = \emptyset$. In this case, 3 cannot unilaterally cause the priority flip.

Case 3 $k \in \{4, 5, \dots, n\}$: The only way that k changes the priority is to pretend to find one of $\{1, 2\}$ acceptable. However, any such misreport will match k with an unacceptable partner for sure.

Thus, no agents can manipulate $\text{TPM}^{\bar{}}$. □

Theorem 1. *F satisfies consistency, Pareto efficiency, strategyproofness, and respect for improvements if and only if $F = \text{TPM}^{\bar{}}$ for some $\bar{}$.*

Proof. Suppose F satisfies the axioms. Let $\bar{ } = \{>_i\}_{i \in N}$ be the linear orders constructed from F as in the previous section. Also, by relabeling agents if needed, we assume $D = \{1, 2, 3\}$ and otherwise the agents are ordered according to $>$. Let $N^F(R^S)$ denote the set of matched agents under F at R^S , and let $N(R^S)$ denote

the set of matched agents under $\text{TPM}^{\bar{>}}$ at R^S . We prove $F = \text{TPM}^{\bar{>}}$ by showing

$$i \notin N^F(R^S) \text{ if and only if } i \notin N(R^S)$$

for each agent i . Fix an arbitrary agent i . If $A_i(R^S) = \emptyset$, the statement is true by individual rationality. So, assume $A_i(R^S) \neq \emptyset$.

Case 1 $i = 1$ or 2 : By symmetry, we focus on 1 . If $1 \notin N^F(R^S)$, then any maximum matching β that matches 1 leaves some $h \in N^F(R^S)$ with $h >_{\beta(1)} 1$ unmatched by Lemma 7. This is possible only when (1) $>_{\beta(1)} 2, 1, h = 2$, and $A_1(P^S) = A_2(P^S) = \{\beta(1)\}$; or (2) $\beta(1) = 2$ with $3 >_2 1, h = 3$, and $A_1(P^S) = A_3(P^S) = \{2\}$; or (3) $\beta(1) \in \{2, 3\}$ with $3 >_2 1$ or $2 >_3 1, h \in \{2, 3\} \setminus \{\beta(1)\}$, and $\{1, 2, 3\}$ constitute an isolated cycle in $G(P^S)$. These are exactly the three conditions under which 1 can be unmatched under $\text{TPM}^{\bar{>}}$ even if $A_1(R^S) \neq \emptyset$. Hence, $1 \notin N(R^S)$.

Conversely, if $1 \notin N(R^S)$, then one of the three conditions holds. 1 cannot be matched under F by Lemma 3 because matching 1 with one of her acceptable mates inevitably leaves another agent whose priority is higher than 1 unmatched.

Case 2 $i \geq 3$: Assume $i \in (N(R^S) \setminus N^F(R^S)) \cup (N^F(R^S) \setminus N(R^S))$. Take i to be the smallest number.

Let $\alpha \in \text{TPM}^{\bar{>}}(R^S)$. If $i \in N(R^S) \setminus N^F(R^S)$, by Lemma 7, there is $h \in N^F(R^S)$ such that $\alpha(h) = h$ and $h >_{\alpha(i)} i$. This means that $h \in \{1, 2, \dots, i-1\}$ and

$$h \in N^F(R^S) \setminus N(R^S),$$

which contradicts the choice of i . Hence, $i \in N^F(R^S) \setminus N(R^S)$. $i \notin N(R^S)$ implies that any maximum matching that matches i must leave another agent $h \in N(R^S)$ with $h > i$ unmatched. Take any $\beta \in F(R^S)$. Since β is a maximum matching with $\beta(i) \neq i$, there is h' such that $h' > i$ and $h' \notin N^F(R^S)$. This contradicts the choice of i because $h' \in N(R^S) \setminus N^F(R^S)$. $\square$

References

- Abdulkadiroğlu, Atila and Sönmez, Tayfun (2003). “School Choice: A Mechanism Design Approach”. *American Economic Review* 93.3, 729–747.
- Afacan, Mustafa Oğuz, Hu, Gaoji, and Li, Jiangtao (2024). “Housing markets since Shapley and Scarf”. *Journal of Mathematical Economics* 111.

- Alcalde-Unzu, Jorge and Molis, Elena (2011). “Exchange of indivisible goods and indifferences: The Top Trading Absorbing Sets mechanisms”. *Games and Economic Behavior* 73.1, 1–16.
- Andersson, Tommy, Cseh, Ágnes, Ehlers, Lars, and Erlanson, Albin (2021). “Organizing Time Exchanges: Lessons from Matching Markets”. *American Economic Journal: Microeconomics* 13.1, 338–73.
- Andersson, Tommy and Kratz, Jörgen (2020). “Pairwise Kidney Exchange over the Blood Group Barrier”. *The Review of Economic Studies* 87.3, 1091–1133.
- Balbuzanov, Ivan (2020). “Short Trading Cycles: Paired Kidney Exchange with Strict Ordinal Preferences”. *Mathematical Social Sciences* 104, 78–87.
- Biró, Péter, Klijn, Flip, Klimentova, Xenia, and Viana, Ana (2024). “Shapley-Scarf Housing Markets: Respecting Improvement, Integer Programming, and Kidney Exchange”. *Mathematics of Operations Research* 49.3, 1938–1972.
- Biró, Péter et al. (2021). “Modelling and optimisation in European Kidney Exchange Programmes”. *European Journal of Operational Research* 291.2, 447–456.
- Bogomolnaia, Anna, Deb, Rajat, and Ehlers, Lars (2005). “Strategy-proof assignment on the full preference domain”. *Journal of Economic Theory* 123.2, 161–186.
- Bogomolnaia, Anna and Moulin, Herve (2004). “Random Matching under Dichotomous Preferences”. *Econometrica* 72.1, 257–279.
- Chambers, Christopher P. and Echenique, Federico (2016). *Revealed Preference Theory*. Econometric Society Monographs. Cambridge: Cambridge University Press.
- Ehlers, Lars (2002). “Coalitional Strategy-Proof House Allocation”. *Journal of Economic Theory* 105.2, 298–317.
- Ehlers, Lars (2014). “Top trading with fixed tie-breaking in markets with indivisible goods”. *Journal of Economic Theory* 151, 64–87.
- Ergin, Haluk (2000). “Consistency in house allocation problems”. *Journal of Mathematical Economics* 34.1, 77–97.

- Ergin, Haluk, Sönmez, Tayfun, and Ünver, M. Utku (2020). “Efficient and Incentive-Compatible Liver Exchange”. *Econometrica* 88.3, 965–1005.
- Gale, D. and Shapley, L. S. (1962). “College Admissions and the Stability of Marriage”. *The American Mathematical Monthly* 69.1, 9–15.
- Han, Xiang, Kesten, Onur, and Ünver, M. Utku (2024). “Blood Allocation with Replacement Donors: Theory and Application”. *working paper*.
- Hatfield, John William (2005). “Pairwise kidney exchange: Comment”. *Journal of Economic Theory* 125.2, 189–193.
- Jaramillo, Paula and Manjunath, Vikram (2012). “The difference indifference makes in strategy-proof allocation of objects”. *Journal of Economic Theory* 147.5, 1913–1946.
- Klaassen, Mattheüs F., Klerk, Marry de, Dor, Frank J. M. F., Heidt, Sebastiaan, Laar, Stijn C. van de, Minnee, Robert C., Wetering, Jacqueline van de, Pengel, Liset H. M., and Weerd, Annelies E. de (2025). “Navigating a Quandary in Kidney Exchange Programs: A Review of Donor Travel versus Organ Shipment”. *Transplant International* 38, 14804.
- Klaus, Bettina (2011). “Competition and resource sensitivity in marriage and roommate markets”. *Games and Economic Behavior* 72.1, 172–186.
- Klaus, Bettina (2017). “Consistency and its converse for roommate markets”. *Games and Economic Behavior* 104, 43–58.
- Klaus, Bettina and Can, Burak (2013). “Consistency and population sensitivity properties in marriage and roommate markets”. *Social Choice and Welfare* 41, 835–862.
- Kratz, Jörgen (2024). “Conflicting objectives in kidney exchange”. *Journal of Economic Theory* 217, 105803.
- Kute et al. (2025). “Analysis of kidney and liver exchange transplantation in India (2000-2025): a multicentre, retrospective cohort study”. *Lancet Regional Health Southeast Asia* 37, 100597.
- Ma, Jinpeng (1994). “Strategy-proofness and the strict core in a market with indivisibilities”. *International Journal of Game Theory* 23, 75–83.

- Manjunath, Vikram and Westkamp, Alexander (2021). “Strategy-proof exchange under trichotomous preferences”. *Journal of Economic Theory* 193, 105197.
- National Kidney Registry (2018). *National Kidney Registry Quarterly Report, 2018 Q4*. Tech. rep.
- Nicolò, Antonio and Rodríguez-Álvarez, Carmelo (2012). “Transplant quality and patients’ preferences in paired kidney exchange”. *Games and Economic Behavior* 74.1, 299–310.
- Nicolò, Antonio and Rodríguez-Álvarez, Carmelo (2017). “Age-based preferences in paired kidney exchange”. *Games and Economic Behavior* 102, 508–524.
- Özkal-Sanver, İpek (2004). “A note on gender fairness in matching problems”. *Mathematical Social Sciences* 47.2, 211–217.
- Pápai, Szilvia (2000). “Strategyproof Assignment by Hierarchical Exchange”. *Econometrica* 68.6, 1403–1433.
- Pycia, Marek and Ünver, Utku (2017). “Incentive compatible allocation and exchange of discrete resources”. *Theoretical Economics* 12.1, 287–329.
- Root, Joseph and Ahn, David S. (2025). “Incentives and Efficiency in Constrained Allocation Mechanisms”. Working paper.
- Roth, Alvin E., Sönmez, Tayfun, and Ünver, M. Utku (2004). “Kidney Exchange”. *The Quarterly Journal of Economics* 119.2, 457–488.
- Roth, Alvin E., Sönmez, Tayfun, and Ünver, M. Utku (2005). “Pairwise kidney exchange”. *Journal of Economic Theory* 125.2, 151–188.
- Sasaki, Hiroo and Toda, Manabu (1992). “Consistency and characterization of the core of two-sided matching problems”. *Journal of Economic Theory* 56.1, 218–227.
- Shapley, Lloyd and Scarf, Herbert (1974). “On cores and indivisibility”. *Journal of Mathematical Economics* 1.1, 23–37.
- Sönmez, Tayfun and Ünver, M. Utku (2005). *Kidney Exchange with Good Samaritan Donors: A Characterization*. Working paper.
- Sönmez, Tayfun and Ünver, M. Utku (2010). “House allocation with existing tenants: A characterization”. *Games and Economic Behavior* 69.2, 425–445.

- Sönmez, Tayfun and Ünver, M. Utku (2025). “Matching under Non-transferable Utility: Applications”. *as Chapter 3 of Yeon-Koo Che, Pierre-Andre Chiaporri, and Bernard Salanie (eds) Handbook of the Economics of Matching, Volume 2 (Elsevier) 2025.*
- Svensson, Lars-Gunnar (1999). “Strategy-proof allocation of indivisible goods”. *Social Choice and Welfare* 16.4, 557–567.
- Takamiya, Koji (2001). “Coalition strategy-proofness and monotonicity in Shapley–Scarf housing markets”. *Mathematical Social Sciences* 41.2, 201–213.
- Thomson, William (1996). “Consistent Allocation Rules”. *University of Rochester Working Paper*.
- Thomson, William (2011). “Consistency and its converse: an introduction”. *Review of Economic Design* 15, 257–291.
- Toda, Manabu (2006). “Monotonicity and Consistency in Matching Markets”. *International Journal of Game Theory* 34, 13–31.
- Whitney, Hassler (1935). “On the Abstract Properties of Linear Dependence”. *American Journal of Mathematics* 57.3, 509–533.