



Contents lists available at ScienceDirect

## Games and Economic Behavior

journal homepage: [www.elsevier.com/locate/geb](http://www.elsevier.com/locate/geb)

## Rejection-proof mechanisms for multi-agent kidney exchange

Danny Blom<sup>\*</sup>, Bart Smeulders, Frits Spieksma

Department of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands

## ARTICLE INFO

## Keywords:

Multi-agent systems  
Kidney exchange  
Bilevel programming  
Computational complexity

## ABSTRACT

Kidney exchange programs (KEPs) increase kidney transplantation by facilitating the exchange of incompatible donors. Increasing the scale of KEPs leads to more opportunities for transplants. Collaboration between transplant organizations (agents) is thus desirable. As agents are primarily interested in providing transplants for their own patients, collaboration requires balancing individual and common objectives. In this paper, we consider ex-post strategic behavior, where agents can modify a proposed set of kidney exchanges. We introduce the class of rejection-proof mechanisms, which propose a set of exchanges such that agents have no incentive to reject them. We provide an exact mechanism and establish that the underlying optimization problem is  $\Sigma_2^P$ -hard; we also describe computationally less demanding heuristic mechanisms. We show rejection-proofness can be achieved at a limited cost for typical instances. Furthermore, our experiments show that the proposed rejection-proof mechanisms also remove incentives for strategic behavior in the ex-ante setting, where agents withhold information.

## 1. Introduction

For patients suffering from end-stage renal disease, a kidney transplant from a living donor is the preferred treatment option. A donor must be medically compatible with the recipient, which means some recipients cannot receive a transplant, even though they have a donor willing to undergo a transplant for them. We refer to such a combination of recipient and donor as an (*incompatible*) *pair*. Kidney exchange programs (KEPs), introduced by Rapaport (1986) and popularized by Roth et al. (2004, 2005), aim to match the recipients from a pair with compatible donors from other pairs. Donors in a pair only donate if their paired recipient receives a transplant in return, leading to *cycles* of transplants. Kidney exchange programs may also include so-called non-directed donors (NDDs). Non-directed donors are not associated with any recipient and are willing to donate to any recipient in the program, leading to donation *chains*. In a KEP, exchanges are based on a *compatibility graph* that reflects the medical compatibilities between donors and recipients, and a solution in a KEP typically consists of a disjoint set of cycles (of length at most  $K$ ) and chains (of length at most  $L$ ). Fig. 1 illustrates an example of a compatibility graph and a corresponding solution. Kidney exchange programs are widely used and are still expanding. We refer to Biró et al. (2019) for a recent overview of the practice in Europe.

Solutions in kidney exchange programs generally improve with an increase in the number of potentially successful donors, i.e., the size of the pool. For example, Agarwal et al. (2020) show that in the US, fragmentation of kidney exchanges over separate hospital, regional and national programs leads to a yearly loss of hundreds of transplants compared to a more centralized kidney exchange program, see also Melcher et al. (2013).

<sup>\*</sup> Corresponding author.

E-mail addresses: [d.a.m.p.blom@tue.nl](mailto:d.a.m.p.blom@tue.nl) (D. Blom), [b.smeulders@tue.nl](mailto:b.smeulders@tue.nl) (B. Smeulders), [f.c.r.spieksma@tue.nl](mailto:f.c.r.spieksma@tue.nl) (F. Spieksma).

<https://doi.org/10.1016/j.geb.2023.10.015>

Received 30 November 2022

Available online 23 November 2023

0899-8256/© 2023 The Author(s).

Published by Elsevier Inc.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

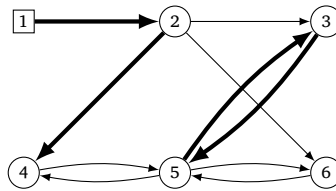


Fig. 1. Circles represent recipient-donor pairs, while squares represent non-directed donors. The presence of an arc  $(i, j)$  signifies compatibility of the donor of  $i$  and the recipient of  $j$ . The bold arcs represent a feasible KEP solution, which features a cycle between pairs 3 and 5 and a chain from NDD 1 to pair 2 and on to pair 4.

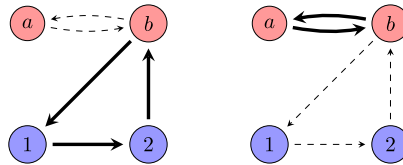


Fig. 2. The social optimum of this 2-agent kidney exchange instance is given by the cycle  $(b, 1, 2)$ , which provides 3 transplants. However, the red agent prefers the cycle  $(a, b)$ , as both transplants in this cycle are for red recipients.

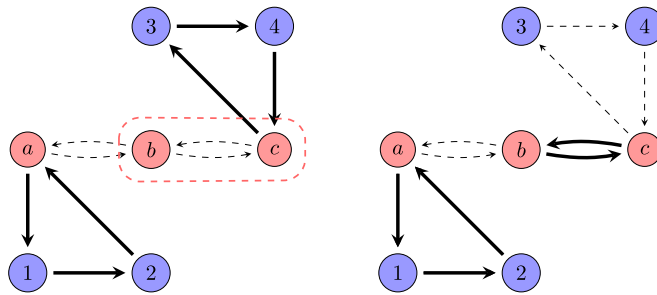


Fig. 3. The social optimum (left) consists of two 3-cycles  $(a, 1, 2)$  and  $(c, 3, 4)$ . If this is the proposed solution, the red agent has the incentive to reject the cycle  $(c, 3, 4)$  and internally match  $(b, c)$ , resulting in the solution on the right, thereby increasing the number of transplants for the red agent.

Within Europe, several countries have started international cooperations known as *multi-agent kidney exchange programs* (m-KEP) to increase pool size (Böhmgig et al., 2017; Valentín et al., 2019).

While multi-agent kidney exchanges aim to increase the total number of transplants, the competing goals of countries or individual transplant centers (whom we will refer to as *agents*) raise issues. Indeed, single agents are primarily interested in transplanting their patients, rather than achieving the overall social optimum, e.g., the maximum number of transplants over all agents. For example, we may consider the example in Fig. 2 taken from Ashlagi and Roth (2014), in which the red agent prefers the solution that follows from deviating from cooperation, over the solution that is a social optimum. We also note agent objectives can differ from simple transplant maximization, possibly leading to further misalignment of objectives (see Biró et al. (2019) for an overview of kidney exchange objectives in Europe). Furthermore, different agents might face different legislative constraints. Indeed, the maximum allowed length of a cycle might differ between countries. For example, in the Austrian-Czech collaboration, Austria allows only short (simultaneous) cycles, whereas the Czech Republic allows for larger and non-simultaneous cycles as well, see, e.g., Mincu et al. (2021).

Thus, when designing mechanisms for m-KEPs, we must consider the strategic behavior of agents that is aimed at increasing transplants for their own patients. Typically, m-KEP is modeled as a private information game, where agents first decide on their *withholding strategy*, a term coined by Ashlagi and Roth (2014). A withholding strategy is a decision, to be made by an agent, on which of their pairs and NDDs to report to, and which to withhold from, the *collaboration*. This collaboration can be regarded as an independent decision-making entity, whose goal is to propose a kidney exchange solution that is optimal with respect to some predefined (hierarchy) of objectives, using the pairs and NDDs revealed to it. We will refer to the set of pairs and non-directed donors revealed to the collaboration as the *common pool*. Finally, each agent can then also select transplants for its unmatched pairs, which can either be pairs that were hidden from the collaboration, or pairs that were revealed to the collaboration yet not part of the collaboration’s proposed transplant plan (Ashlagi and Roth, 2014). We refer to this setting as the *ex-ante setting*, as each agent makes their decision before the solution is proposed by the collaboration.

In this paper, we introduce the *ex-post setting*. Here the collaboration first proposes a solution, and then agents can use so-called *rejection strategies* to decide whether to reject (part of) this proposed solution. This entails withdrawing pairs and NDDs from cycles and chains from the proposed solution. In addition, the agents can identify new internal cycles and chains with pairs or NDDs that were either withdrawn or unmatched in the proposed solution. An example is given in Fig. 3.

The ex-post setting differs from the more traditional ex-ante setting in two aspects. First, in the ex-post setting, we assume that the collaboration has full information on the individual agents' pools and the corresponding compatibilities. Secondly, in the ex-post setting agents can evaluate the solution determined by the collaboration, withdraw pairs from proposed exchanges, and use alternative internal exchanges.

To support the first assumption, we note that, in practice, a collaboration between hospitals often implies the disclosure of information about each other's patients. We also point out that in the ex-ante setting it is a computational challenge for an individual agent to determine an optimal withholding strategy, even when given full information. Indeed, Smeulders et al. (2022) show that identifying an optimal withholding strategy is already  $\Sigma_2^P$ -hard in such a setting. Uncertainty about the pools of other agents and the strategic interaction further complicate the problem. These difficulties form an incentive for agents to provide full information. Of course, an agent may still resort to heuristic strategies. Ashlagi and Roth (2014) identify a heuristic strategy that involves an agent withholding a subset of its pairs for internal exchanges. Although such a strategy can lead to transplants for more of the agents' patients for some specific m-KEP mechanisms, we will show empirically, using both established and state-of-the-art instance generators, that this is typically not the case for the mechanisms proposed in this paper.

The second assumption is justified by the setting of organ transplantation. Ethical and legal considerations make binding agreements to execute proposed transplant plans impossible, as donors and recipients can never be forced or pressured to undergo transplantation. Agents thus have the option of rejecting proposed solutions. This possibility for agents to reject exchanges also diminishes the risk of agents of contributing all their pairs to the collaboration, thus providing more support for the first assumption as well.

All this motivates a concept that we study in this contribution: solutions that are rejection-proof (see Section 2 for formal definitions). The goal of this paper is to study mechanisms that lead to solutions such that no agent has the incentive to withdraw a pair or NDD from the proposed solution.

### 1.1. Structure of the paper

The paper unfolds as follows. We briefly survey related literature in Subsection 1.2. Next, in Section 2, we introduce the necessary notation as well as the definitions of withholding strategies relevant in an ex-ante setting, and rejection strategies relevant in an ex-post setting. This allows us to introduce the *rejection-proofness* property. We also link rejection-proof solutions to Nash equilibria in a related kidney exchange game. In Section 3 we further motivate our choice to study the ex-post setting by experimentally comparing withholding strategies and rejection strategies. We find that rejection strategies lead to better solutions for individual agents who employ such strategies than withholding strategies. Thus, an ex-post setting may offer more potential for finding good solutions than an ex-ante setting. Sections 4 and 5 introduce our specific rejection-proof mechanisms. The first mechanism, Maximum Rejection-proof KEP (MAXRP-KEP), always finds, among the set of rejection-proof solutions, one for which social welfare value is maximized. Notice, however, that the problem of finding such a solution is shown to be  $\Sigma_2^P$ -hard, and the mechanism thus needs to solve a computationally challenging optimization problem.

The second mechanism, Maximum Internal Transplants (MAXINT), is inspired by current practice and is easy to compute. However, the solution it finds, while rejection-proof, may not be a social optimum. The final mechanism, the Reject-and-Rebuild (RR) mechanism, mimics the process where each agent employs rejection strategies until stability is reached. Whenever stability is attained, the mechanism considers if additional exchanges can be selected involving currently unmatched pairs and NDDs, and adds these to the current solution. Each of these three mechanisms returns a rejection-proof solution which is also a maximal feasible kidney exchange solution, i.e., there exists no alternative kidney exchange solution with a strict superset of exchanges. In Section 6, we describe the outcomes of all mechanisms. Firstly, we study the impact of each of the three mechanisms on the social welfare value of the returned solutions. Furthermore, we consider computation times for the exact mechanism, for different choices of exchange weights. We finish the paper with a discussion in Section 7.

### 1.2. Related literature

This paper is part of a growing literature on multi-agent kidney exchange programs. Starting with Sönmez and Ünver (2013) and Ashlagi and Roth (2014), the main focus has been on withholding strategies and mechanisms to incentivize agents to reveal their complete pools. In multi-period settings, a number of credit mechanisms have been proposed. Hajaj et al. (2015) proposes awarding credits based on the differences between the expected and reported numbers of pairs by an agent, this mechanism is strategy-proof assuming the expected number of pairs known. Agarwal et al. (2020) presents a system where expected marginal transplants per pair, based on blood type and PRA, are used to assign credits. Klimentova et al. (2020) propose two methods of assigning credits, the first based on the maximum number of transplants achievable for a patient, the second based on the marginal benefit an agent adds to overall transplants by joining the collaboration. Finally, Biró et al. (2020) proposes a mechanism based on Shapley values. In single-period settings, and limiting the maximum cycle length to 2, Ashlagi et al. (2015) present a strategy-proof randomized mechanism. For the same cycle length limits, Carvalho et al. (2017) and Carvalho and Lodi (2023) study a non-cooperative game where agents' strategies consist of selecting internal exchanges, and shared exchanges on remaining pairs are determined by an independent agent. They show social welfare Nash equilibria for these games can be efficiently computed. Finally, in Carvalho et al. (2022), a game is presented where agents' strategies consist of internal and shared exchanges. Exchanges proceed to transplant if all involved agents have this exchange in their strategy.

Settings of strategic interactions between the agents and the collaboration are typically modeled by a *Stackelberg game*, i.e., a type of sequential game in which a *leader* selects an initial strategy, after which a *follower* selects an optimal response strategy (von Stackelberg, 1952). The computational complexity of determining an optimal strategy is typically different for the leader and the follower. While the follower needs to solve an optimization problem parameterized by the leader’s strategy, the leader also needs to anticipate the follower’s optimal response to their decision. In particular, this means that the leader faces a *bilevel optimization problem*.

Clearly, bilevel optimization problems are intimately related to the ex-post setting in m-KEPs where the collaboration (the leader) first generates a solution after which the agents (the followers) act. In fact, our mechanism MAXRP-KEP falls into this paradigm. Therefore, we now compactly discuss some of the existing literature in this field; for surveys, we refer to Kleinert et al. (2021) and Beck et al. (2023).

When solving a bilevel optimization problem, its computational complexity becomes relevant. An important notion when solving multi-level optimization problems is the *polynomial hierarchy* (Stockmeyer, 1976). The polynomial hierarchy is a chain of complexity classes  $\Sigma_0^P \subseteq \Sigma_1^P \subseteq \dots$ , generalizing the notion of the classes P and NP. In particular, we have that  $\Sigma_0^P = P$  and  $\Sigma_1^P = NP$ . As a rule of thumb, whenever the decision version of follower’s best-response problem is contained in  $\Sigma_{k-1}^P$ , the decision version of the problem faced by the leader is in  $\Sigma_k^P$  ( $k \geq 1$ ). In this paper, we will specifically consider the complexity class  $\Sigma_2^P$ .

In recent years, research has focused on bilevel optimization problems and their connections to  $\Sigma_2^P$ . In particular, there exists a natural connection between problems in the class NP and their  $\Sigma_2^P$  counterparts. Such a  $\Sigma_2^P$  counterpart can often be characterized by a bipartition of decision variables into sets  $X$  and  $Y$ , with a question of the form “does there exist  $x \in X$  such that for all  $y \in Y$ ,  $P(x, y)$  is true”, where  $P(x, y)$  is a proposition that can be evaluated in polynomial time, see e.g. Woeginger (2021).

An important subclass of such problems is the class of *adversarial problems*. In an adversarial problem, the leader acts as an adversary, meaning that they aim at blocking the follower from attaining their (lower-level) objective as much as possible. Johannes (2011) proved that adversarial problems are  $\Sigma_2^P$ -hard, provided the non-adversarial counterpart is NP-hard and satisfies some other weak conditions. In this paper, although the problems we consider are not necessarily adversarial, the leader may aim at disincentivizing the follower from performing strategic behavior.

Very often, bilevel optimization problems can be formulated as (*mixed-integer bilevel linear programming* (MIBLP) (Jeroslow, 1985). As mixed-integer linear programming (MILP) is NP-hard in general, MIBLPs cannot be reformulated as a compact mixed-integer linear program unless  $NP = \Sigma_2^P$  (Lodi et al., 2014).

When attempting to solve an MIBLP, an intuitive starting point is to first consider a relaxation of the problem, and next to iteratively solve the associated relaxed problem and cut away infeasible solutions (a strategy similar to strategies used for solving MILPs). A natural relaxation of bilevel optimization problems is the *high-point relaxation* (HPR), which is obtained by relaxing optimality of the lower-level solution to lower-level feasibility (Bialas and Karwan, 1984; Moore and Bard, 1990). One can further consider the linear relaxation of this high-point relaxation. As the high-point relaxation is typically a mixed-integer linear program, one can further consider the linear relaxation of HPR. Many HPR-based methods are inspired by conventional MILP techniques, such as branch-and-cut. One of the first branch-and-cut methods for MIBLP was proposed by DeNegre and Ralphs (2009), who use Chvátal-Gomory-like no-good cuts to separate a single bilevel infeasible point from the inducible region. Tahernejad et al. (2020) generalize this approach in the case of binary linking variables by introducing cuts separating all solutions for which the set of linking variables have some fixed value.

Another useful approach is to reformulate the bilevel problem as a single-level problem by means of a *value function reformulation*. In most settings, the value function relates to a highly nonconvex problem. Lozano and Smith (2017) propose a sampling-based reformulation based on the value-function using big-M constraints to guarantee bilevel feasibility of its solutions. Furthermore, Bolusani and Ralphs (2022) consider the connections of value functions and duality, and introduce a solution method based on approximating the value function with dual bounding functions.

## 2. Preliminaries and notation

The basic model for kidney exchange programs considers a directed graph  $G = (V, A)$  called the *compatibility graph*. Each vertex  $v \in V$  corresponds to either an incompatible pair or a non-directed donor. We represent the set of pairs by  $V_P$ , and the set of non-directed donors by  $V_N$ , where  $V = V_P \cup V_N$ . Each arc  $(u, v) \in A$  indicates that the donor of  $u \in V$  and the patient of  $v \in V_P$  are medically compatible.

In multi-agent kidney exchange programs, hospitals and/or transplant centers (agents) are responsible for a set of pairs and NDDs. To model this setting, the vertex set  $V$  is partitioned into subsets  $V^p = V_P^p \cup V_N^p$  for each agent  $p \in \mathcal{P}$ , where  $\mathcal{P}$  denotes the set of agents. We have  $V = \bigcup_{p \in \mathcal{P}} V^p$  and  $V^p \cap V^q = \emptyset$  for  $p, q \in \mathcal{P}, p \neq q$ . We call an arc  $(u, v) \in A$  *internal (to agent  $p \in \mathcal{P}$ )* if  $\{u, v\} \subseteq V^p$ , and *shared* if there exists no agent to whom  $(u, v)$  is internal. This allows us to classify the arc set as  $A = (\bigcup_{p \in \mathcal{P}} A^p) \cup A^0$ , where for each  $p \in \mathcal{P}$ , we denote by  $A^p$  the set of all arcs internal to  $p$ , and by  $A^0$  the set of all shared arcs. Notice that this setting generalizes the basic model.

Let  $C$  denote the set of cycles in  $G$  that can be used for cyclic exchanges (typically such cycles have a length bounded by an integer  $K \geq 2$ ). We use  $V(c)$  and  $A(c)$  to denote the set of vertices and arcs respectively of cycle  $c \in C$ . For each agent  $p \in \mathcal{P}$ , the set  $C^p \subseteq C$  consists of the cycles  $c \in C$  for which all arcs are internal to  $p$ , i.e., the set of cycles  $c \in C$  such that  $A(c) \subseteq A^p$ . We refer to  $C^p$  as the set of *internal cycles* of  $p$ . The set of *shared cycles* is denoted by  $C^0 = C \setminus (\bigcup_{p \in \mathcal{P}} C^p)$ . The (social) value of cycle  $c \in C$  is given by  $w_c \in \mathbb{R}_{\geq 0}$ , and the value that  $c$  has to a certain agent  $p \in \mathcal{P}$  is denoted by  $w_c^p \in \mathbb{R}_{\geq 0}$ . Note that the social welfare value of

a cycle need not be strictly additive in the agent values, i.e., it need not be the case that  $w_c = \sum_{p \in \mathcal{P}} w_c^p$ . For instance, participating agents may have agreed that the collaboration should strive for transplant maximization, while privately wishing to prioritize certain transplants. Nevertheless, the social welfare value  $w_c$  of cycle  $c$  typically corresponds to the number of patients receiving a transplant according to cycle  $c$ , whereas the value  $w_c^p$  corresponds to the number of patients in cycle  $c$  that belong to the pool  $V^p$  of agent  $p \in \mathcal{P}$ . We refer to these social and agent values as *unit valuations*, as each transplant is considered to be equally valued.

Similar to the notation for cycles, we denote the set of all feasible chains by  $D = \bigcup_{p \in \mathcal{P}} D^p$ , where  $D^p$  refers to the set of all feasible internal chains of agent  $p$  (typically all acyclic paths of length at most  $L$  on  $G^p = G[V^p] = (V^p, A^p)$  originating from a vertex  $v \in V_N^p$ ). The value of a chain  $d \in D$  is denoted by  $w_d \in \mathbb{R}_{\geq 0}$ , and the value of a chain  $d$  to agent  $p$  by  $w_d^p \in \mathbb{R}_{\geq 0}$ . Typically, but not necessarily, these values correspond to the number of patients from the pool  $V^p$  receiving a transplant according to chain  $d$ . Note that we do not include shared chains. Since an NDD can initiate a chain of many transplants, and the donor of the final pair in a chain often becomes an NDD themselves, we consider that an agent never wants to risk chains ending with a pair of another agent. This also excludes intermediate pairs of other agents, as chains could fail when only partly executed for a variety of reasons. Going forward, we make no more distinction between cycles and chains, thus the inclusion of shared chains is possible in our approach.

In order to keep notation concise, we define the set of *exchanges* as  $\mathcal{E} = C \cup D$  in case distinguishing between cycles and chains is unnecessary. Similarly, for each agent  $p \in \mathcal{P}$ , we define  $\mathcal{E}^p = C^p \cup D^p$  as the set of *internal exchanges* of  $p$ , and  $\mathcal{E}^0 = C^0$  as the set of *shared exchanges*. For each vertex  $v \in V$ , we denote by  $\mathcal{E}(v)$  the set of exchanges involving  $v$ , with similar notions  $C(v)$  and  $D(v)$  for the sets of cycles and chains respectively. Finally, whenever we consider an exchange  $e \in \mathcal{E}$ , we adopt all other notions defined above for this exchange depending on whether it is a cycle or a chain.

The set of feasible (*kidney exchange*) solutions are captured by the following set

$$\mathcal{X} = \{X \subset \mathcal{E} \mid \forall e, e' \in X, e \neq e' : V(e) \cap V(e') = \emptyset\}.$$

By definition, for each  $X \in \mathcal{X}$ , the elements of  $X$  are pairwise vertex-disjoint exchanges, i.e., the elements of  $X$  form a packing of cycles and chains on  $G$ . This condition is required, as donors can donate at most one kidney and each paired or non-directed donor can thus be involved in at most one exchange. We encode each solution  $X \in \mathcal{X}$  with a vector  $\mathbf{x} = (x_e)_{e \in \mathcal{E}} \in \{0, 1\}^{\mathcal{E}}$  such that  $x_e = 1$  if and only if  $e \in X$ . The social welfare value of a solution  $X$  is denoted by  $z(X) = \sum_{e \in \mathcal{E}} w_e x_e$ , i.e., the sum of social welfare values of the exchanges of  $X$ . Similarly, we denote the value of solution  $X$  to agent  $p \in \mathcal{P}$  by  $z^p(X) = \sum_{e \in \mathcal{E}} w_e^p x_e$ .

The challenge in multi-agent kidney exchange problems is to design a method used by the collaboration to select a kidney exchange solution on the common pool that provides a satisfactory trade-off between the individual agents' objectives and social welfare maximization. In other words, the main question is: how to design a mechanism that arrives at a satisfactory solution for all parties involved?

We define mechanisms for multi-agent kidney exchange as follows:

**Definition 1.** A (multi-agent) kidney exchange mechanism is a function  $f$  mapping an instance  $I = (\mathcal{E}, \mathcal{P})$  of a multi-agent kidney exchange problem to a feasible kidney exchange solution  $X = f(I) \in \mathcal{X}$ .

In Subsections 2.1 and 2.2 we consider two ways in which agents are allowed to impact such a mechanism, namely in an *ex-ante* and *ex-post* fashion respectively.

### 2.1. The ex-ante setting: withholding strategies

In the ex-ante setting agents can choose not to reveal their pairs and NDDs to the collaboration. This gives rise to the class of withholding strategies, which can be described as follows.

1. Each agent  $p \in \mathcal{P}$  identifies a subset  $S^p \subseteq V^p$  of its pairs and NDDs to contribute to the collaboration. Throughout the rest of the paper, we will refer to this as agent  $p$  selecting the *withholding strategy*  $W^p = V^p \setminus S^p$ . We denote by  $G' = G[\bigcup_{p \in \mathcal{P}} S^p]$  the subgraph of  $G$  induced by the set of contributed pairs and NDDs. Let  $I' = (\mathcal{E}', \mathcal{P})$  where  $\mathcal{E}'$  denotes the exchanges that are feasible on  $G'$ .
2. A mechanism  $f$  computes a solution  $X = f(I')$ . We denote by  $U^p(X) \subseteq S^p$  the set of pairs and NDDs contributed by agent  $p \in \mathcal{P}$  that remain unmatched with respect to  $X$ .
3. Each agent  $p \in \mathcal{P}$  selects a solution  $X^p \in \mathcal{X}$  on the subgraph of  $G^p$  induced by  $W^p \cup U^p(X)$ , i.e.,

$$X^p \in \arg \max_{X' \in \mathcal{X}} \{z^p(X') \mid \forall e \in X' : V(e) \subseteq W^p \cup U^p(X)\}.$$

The final kidney exchange solution is given by

$$\tilde{X} = X \cup \bigcup_{p \in \mathcal{P}} X^p.$$

For each agent  $p \in \mathcal{P}$ , the value is given by  $z^p(\tilde{X})$ . In this game, it is theoretically hard for an agent to maximize their value in the final solution. Indeed, Smeulders et al. (2022) show that the problem of finding an exact withholding strategy for agent  $p \in \mathcal{P}$  is  $\Sigma_2^P$ -hard if  $f$  is a mechanism maximizing social welfare value on the contributed pool. This is the case even when the withholding

strategies  $W^q \subseteq V^q$  for each other agent  $q \in \mathcal{P} \setminus \{p\}$  are fixed and known to agent  $p$ . In practice, even the vertex set  $V^q$  of another agent  $q$  is not known to an agent, further complicating the problem.

Due to the complexity of computing exact withholding strategies, we consider an intuitive withholding strategy proposed in Ashlagi and Roth (2014), which we call the *greedy withholding strategy*.

**Definition 2 (Greedy withholding strategy).** An agent  $p \in \mathcal{P}$  applies a *greedy withholding strategy* when the agent computes an internal exchange solution  $Y^p$  that is maximum with respect to  $w_e^p$ , and decides to reveal  $V^p \setminus W^p$  to the collaboration, where  $W^p$  is the set of pairs and NDDs involved in exchanges of solution  $Y^p$ .

It was shown empirically by Ashlagi and Roth (2014) that, under a mechanism that maximizes total transplants, an agent can on average increase the number of transplants for their own patients by employing this strategy at the expense of the social optimum.

## 2.2. The ex-post setting: rejection strategies

In this paper we consider a setting where agents can perform strategic behavior in an ex-post fashion, i.e., after the mechanism  $f$  has proposed a kidney exchange solution. Specifically, we consider the following setting.

1. Given an m-KEP instance  $I = (\mathcal{E}, \mathcal{P})$ , the mechanism  $f$  proposes the solution  $X = f(I)$ .
2. Simultaneously, each agent  $p \in \mathcal{P}$  selects a *rejection strategy*, i.e., a solution  $Y^p$  that can be obtained from the proposed solution  $X$  using the following two operations:

(O1) for each exchange  $e \in X$  such that  $V(e) \cap V^p \neq \emptyset$ , agent  $p$  can choose to remove  $e$  from  $X$ , i.e., to *reject* exchange  $e$ .

(O2) for each exchange  $e \in \mathcal{E}^p \setminus X$ , agent  $p$  can choose to add  $e$  to  $X$ .

For each agent  $p \in \mathcal{P}$ , we define the set of feasible rejection strategies as

$$\mathcal{R}^p(X) = \{Y^p \in \mathcal{X} \mid Y^p \text{ can be obtained from } X \text{ through (O1)-(O2)}\}$$

3. The final kidney exchange solution is given by

$$Y = \bigcup_{p \in \mathcal{P}} (\mathcal{E}^p \cap Y^p) \cup \bigcap_{p \in \mathcal{P}} (\mathcal{E}^0 \cap Y^p).$$

Here, the first term consists of the set of internal exchanges selected by each agent individually, whereas the second term consists of the set of shared exchanges that are not rejected by any agent.

From now on, with a slight abuse of notation, a rejection strategy can be represented by both a kidney exchange solution  $Y^p \in \mathcal{R}^p(X)$  and its corresponding encoding  $y^p = (y_e^p)_{e \in \mathcal{E}}$ . We say that agent  $p \in \mathcal{P}$  *accepts* the solution  $X$  proposed by the mechanism if  $p$  selects the rejection strategy  $Y^p = X$ . Generally speaking, we say agent  $p$  *rejects* solution  $X$  whenever  $p$  selects a rejection strategy  $Y^p \neq X$ . The goal of agent  $p \in \mathcal{P}$  is to select a rejection strategy  $Y^p \in \mathcal{R}^p(X)$  such that the resulting solution  $Y$  after Step 3 is of maximum value  $z^p(Y)$  to  $p$ . However, this optimization problem for agent  $p$  is subject to strategic interaction with other agents, as  $Y$  is determined by the rejection strategies  $Y^{p'}$  of all agents  $p' \in \mathcal{P}$ . In particular, a shared exchange  $e \in X \cap \mathcal{E}^0$  is only contained in  $Y$  if all of the agents with at least one pair involved in  $e$  include it in their solution.

Due to these interactions based on the rejection of shared exchanges, we can characterize the decision on rejection strategies for each agent  $p$  as a noncooperative game called the *rejection game*.

**Definition 3 (Rejection game).** Suppose the collaboration has proposed a feasible kidney exchange solution  $X \in \mathcal{X}$ . Then, the *rejection game* (with respect to  $X$ ) is the game with the set of agents  $\mathcal{P} = \{1, \dots, \ell\}$ , the strategy sets  $\mathcal{R}^p(X)$  for each agent  $p \in \mathcal{P}$ , and the utilities defined as

$$u_p(Y^1, \dots, Y^\ell) = z^p(Y^p \cap \mathcal{E}^p) + z^p\left(\bigcap_{p' \in \mathcal{P}} Y^{p'} \cap \mathcal{E}^0\right),$$

for each agent  $p \in \mathcal{P}$  and each strategy profile  $(Y^1, \dots, Y^\ell) \in \mathcal{R}^1(X) \times \dots \times \mathcal{R}^\ell(X)$ .

Here, the utilities coincide with the (agent) values of the solution obtained in Step 3 provided the agents select the rejection strategies  $Y^p$  for  $p \in \mathcal{P}$ . In essence, the rejection game with respect to  $X$  is equivalent to the kidney exchange game described in Carvalho et al. (2022), with the only exception that the choice of shared exchanges is limited to  $X \cap \mathcal{E}^0$ . Let us denote by  $Y^{-p} = (Y^1, \dots, Y^{p-1}, Y^{p+1}, \dots, Y^\ell)$  a strategy profile played by all agents except agent  $p$ .

**Definition 4 (Best-response rejection strategy).** Let  $X \in \mathcal{X}$  be the feasible kidney exchange solution proposed by the collaboration. For some  $p \in \mathcal{P}$ , let  $Y^{-p}$  be the strategy profile played by all agents except  $p$ . Then,  $p$  plays a best-response rejection strategy  $Y^p$  if

$$Y^p \in \arg \max_{Y^* \in \mathcal{R}^p(X)} u_p(Y^*, Y^{-p}) \tag{1}$$

In all further experiments in this work, we consider a Nash equilibrium as the outcome for this game, i.e. the solution is a strategy  $Y^p$  for each agent, such that the best-response rejection strategy for each agent is this same strategy  $Y^p$ . We obtain one such Nash equilibrium by repeated applications of best-response. Initially, we consider the strategy  $X$  for each agent and compute the best-responses for every agent simultaneously. If the strategy of at least one agent changes, we recompute the best-responses. We repeat this process until a Nash equilibrium is reached. Note that since each change in the best-response strategies removes at least one shared exchange, so this process will eventually converge to an equilibrium. In this work, we will consider mechanisms that, given an m-KEP instance  $I = (\mathcal{E}, \mathcal{P})$ , output a feasible kidney exchange solution  $X \in \mathcal{X}$  such that accepting  $X$  is a best-response strategy assuming all other agents will also accept  $X$ , i.e., the strategy profile  $(X, \dots, X)$  is a pure Nash equilibrium in the rejection game with respect to  $X$ . This gives rise to the definition of *rejection-proofness*.

**Definition 5 (Rejection-proofness).** A mechanism  $f$  is called rejection-proof if for any m-KEP instance  $I = (\mathcal{E}, \mathcal{P})$ , and any agent  $p \in \mathcal{P}$ , the rejection strategy  $X \in \mathcal{R}^p(X)$  is a best-response strategy with respect to the strategy profile corresponding to accepting  $X$  for all agents  $p' \in \mathcal{P} \setminus \{p\}$ :

$$X \in \arg \max_{Y^* \in \mathcal{R}^p(X)} u_p(X, \dots, X, Y^*, X, \dots, X) \tag{2}$$

We call Problem (2) the Reject Kidney Exchange Problem (RKEP). Formulation (3) describes an integer linear programming formulation of the concrete instance RKEP( $\mathbf{x}, p$ ) of the Reject Kidney Exchange Problem faced by agent  $p$  given a feasible kidney exchange solution  $X$ , encoded by  $\mathbf{x} = (x_e)_{e \in \mathcal{E}}$ :

$$\text{RKEP}(\mathbf{x}, p) = \max_y \sum_{e \in \mathcal{E}} w_e^p y_e^p \tag{3a}$$

$$\text{s. t. } \sum_{e \in \mathcal{E}(v)} y_e^p \leq 1, \quad v \in V \tag{3b}$$

$$y_e^p = x_e, \quad e \in \mathcal{E}^0 : V(e) \cap V^p = \emptyset \tag{3c}$$

$$y_e^p \leq x_e, \quad e \in \mathcal{E}^0 : V(e) \cap V^p \neq \emptyset \tag{3d}$$

$$y_e^p \in \{0, 1\}. \quad e \in \mathcal{E} \tag{3e}$$

In this formulation, a rejection strategy  $Y^p \in \mathcal{R}^p(X)$  is encoded by a vector  $\mathbf{y}^p$  of binary variables  $y_e^p$  for each exchange  $e \in \mathcal{E}$  such that  $y_e^p = 1$  if and only if  $e \in Y^p$ . The operations (O1) and (O2) do not allow an agent  $p$  to reject exchanges that do not involve any of its pairs or NDDs; this is enforced by Constraints (3c), whereas the remaining exchanges could be rejected as modeled by Constraints (3d). Furthermore, the agent is free to select any internal exchange  $e \in \mathcal{E}^p$ , provided that the final solution is vertex-disjoint as is ensured by Constraints (3b).

As accepting the proposed solution is a feasible rejection strategy for each agent  $p \in \mathcal{P}$ , i.e.,  $X \in \mathcal{R}^p(X)$ , we have that  $\text{RKEP}(\mathbf{x}, p) \geq z^p(X)$ . Now, if  $\text{RKEP}(\mathbf{x}, p) = z^p(X)$ , accepting the proposed solution  $X$  is an optimal rejection strategy. If all agents accept the solution  $X$ , the value of the final solution for  $p$  is  $z^p(X)$ . Notice that we will also refer to a solution  $X$  satisfying  $\text{RKEP}(\mathbf{x}, p) = z^p(X)$  for each agent  $p \in \mathcal{P}$  as rejection-proof.

On the other hand, whenever, for some agent  $p \in \mathcal{P}$ ,  $\text{RKEP}(\mathbf{x}, p) > z^p(X)$ , then there exists a rejection strategy  $Y^p \in \mathcal{R}^p(X) \setminus \{X\}$  with which agent  $p$  can increase its value assuming all other agents accept  $X$ .

### 3. Comparing withholding and rejection strategies

Having laid out the ex-ante and the ex-post setting in Section 2, it is a natural question to wonder how strategic behavior affects the outcomes in a particular setting, or more concretely, which strategies lead to better solutions for the individual agents. This section is devoted to answering that question by running some computational experiments. The outcomes will serve as further motivation for analyzing the ex-post setting and the corresponding rejection-proof mechanisms. These experiments are conducted in order to pinpoint the benefits that agents can achieve as a result of strategic behavior, and the impact that this strategic behavior has on the overall performance of the multi-agent KEP. This will show that incentives for strategic behavior are present in realistic settings and that the resulting overall loss is large enough to motivate the development of mechanisms preventing strategic behavior.

Since the outcomes of strategic behavior will strongly depend on the graph structure, we consider kidney exchange instances generated through two different sets of characteristic distributions for donor-recipient pairs. Considering multiple distributions lowers bias resulting from any particular graph structure. One set is generated using the traditional Saidman et al. (2006) distributions, which were also used in Ashlagi and Roth (2014). The second set is based on recent work, which aims to more closely resemble instances of

the British kidney exchange program, as described by Delorme et al. (2022); we use their SplitPRA-BandXMatch-PRA0 distributions.<sup>1</sup> We consider the same instance sizes as Ashlagi and Roth (2014). Specifically, we consider a select number of combinations with  $|\mathcal{P}| \in \{2, 4, 6, 8, 10, 12, 14, 16\}$  and the number of pairs per pool  $|V^p| \in \{10, 15, 30, 50, 80\}$  such that in each combination, there are at most 180 pairs in total (notice that we do not consider non-directed donors in these experiments). See Table 1 for a list of all considered combinations. For each such combination, a set of 50 instances is generated.

As we generate these instances using both distributions, there is a total of  $2 \times 18 \times 50 = 1800$  instances. Following Ashlagi and Roth (2014), we assume that each agent  $p \in \mathcal{P}$  values each transplant equally, i.e.,  $w_e^p = |V_p^p \cap V(e)|$  for each exchange  $e \in \mathcal{E}$ , and that the cycle lengths are bounded by  $K = 3$ .

To establish the gains or losses of strategic behavior, we follow the experimental set-up of Ashlagi and Roth (2014), and consider the following scenarios, where “the strategy” refers to either a greedy withholding strategy, or a best-response rejection strategy.

1. All agents are truthful (i.e., each agent contributes its complete pool to the collaboration, and accepts any solution proposed by the collaboration respectively).
2. A single agent employs the strategy, all others are truthful.
3. All agents employ the strategy.

In all cases, the collaboration uses a mechanism to maximize the number of transplants for the common pool. To establish the gain or loss an agent achieves by employing a strategy, we compare the number of transplants for this agent between scenarios 1 and 2. In Table 1, this comparison is shown in the columns “Withhold Agent” and “Reject Agent” respectively. A positive number shows the percentage-wise increase in transplants, whereas a negative number indicates that the agent loses transplants by employing the strategy. We also compare scenarios 1 and 3 to evaluate the total loss in transplants if all agents employ the same type of strategy. The outcome of this comparison is shown in the columns “Withhold Total” and “Reject Total”. Each number or percentage in these tables represents an average over the 50 instances of the respective instance set. The names of the instance sets are of the form  $a \times b$ , where  $a$  is the number of agents in the m-KEP, and  $b$  is the number of pairs in each agent’s pool. As a basis, we compute for each instance a kidney exchange solution with a maximum number of transplants, i.e., a socially optimal solution. The corresponding average values are reported in the column called “#Transplants”.

From the columns “Withhold Agent” and “Reject Agent”, we find that both withholding and rejection strategies provide comparable benefits whenever agents control small pools (with at most 15 pairs per pool); however, agents controlling larger pools benefit significantly more from rejection strategies than from greedy withholding. This effect is even more pronounced in the instances from Saidman et al. (2006). Furthermore, even though in almost all cases the average gain of greedy withholding is positive, nearly every instance set contains instances where withholding led to transplant losses (only the instances sets  $8 \times 10$  and  $14 \times 10$  based on the pool distributions from Delorme et al. (2022) did not). By contrast, rejection strategies are risk-free to the agent, as an agent only plays a rejection strategy if they can strictly improve their outcome. The reason for this is that rejection strategies are a form of ex-post manipulation, meaning that agents have an information benefit compared to when they need to select a withholding strategy ex-ante. Finally, from the columns “Withhold Total” and “Reject Total”, it is clear that the negative effect on the total number of transplants is much larger in the case where all agents use greedy withholding compared to when all agents use a rejection strategy. The combined result of higher average benefits and reduced risk motivates studying rejection strategies, and the corresponding ex-post setting, in more depth. We do note it could be of interest to validate these results in different settings, e.g., by including NDDs, cases where agents have different sizes within the same instance, etc. Furthermore, Kidney Exchange problems often allow many optimal solutions, and biases due to the solver, representation of the data, etc. could impact these results.

#### 4. An exact rejection-proof mechanism

In this section, we will formulate an exact rejection-proof mechanism, i.e., a mechanism that outputs a rejection-proof solution maximizing social welfare value. For that purpose, we first provide a description of the underlying optimization problem OPT-RP-KEP as follows:

##### OPT-RP-KEP

*Given:* a multi-agent KEP instance  $I = (\mathcal{E}, \mathcal{P})$  consisting of a set of exchanges  $\mathcal{E}$ , a set of agents  $\mathcal{P} = \{1, \dots, \ell = |\mathcal{P}|\}$ .

*Find:* a rejection-proof kidney exchange solution  $X$  of maximum social welfare value, i.e.,

$$X \in \arg \max_{Y \in \mathcal{X}} z(Y) = \arg \max_{Y \in \mathcal{X}} \sum_{e \in Y} w_e.$$

The natural decision variant of OPT-RP-KEP, which uses an additional input parameter  $t$  and asks whether there exists a rejection-proof kidney exchange solution  $X$  with  $z(X) \geq t$ , will be referred to as DEC-RP-KEP. We now define the *Maximum Rejection-Proof KEP* mechanism, or simply MAXRP-KEP, and also propose an algorithm for solving the resulting bilevel optimization problem.

<sup>1</sup> Both sets of instances were generated using the generator from <https://wpettersson.github.io/kidney-webapp/#/generator>, making use of the pre-built configurations.



**Table 1**

Impact of strategic behavior in both the ex-ante setting (withholding) and the ex-post setting (rejection). The total number of transplants in an optimal kidney exchange solution, i.e., whenever each agent would truthfully contribute all of their pairs, is compared with percentage gains/losses on individual and social welfare whenever strategic behavior would be employed, for instances generated using the distributions proposed by Delorme et al. (2022) and Saidman et al. (2006) respectively.

Instance set	Delorme et al. (2022)				
	#Transplants	Withhold Agent	Reject Agent	Withhold Total	Reject Total
2 × 10	2.14	0.0%	0.0%	-6.5%	0.0%
4 × 10	9.02	4.5%	5.5%	-11.8%	-3.8%
6 × 10	18.56	1.4%	1.4%	-7.4%	-1.4%
8 × 10	28.62	0.6%	0.6%	-7.3%	-2.0%
10 × 10	42.46	0.5%	0.5%	-5.7%	-1.5%
12 × 10	57.08	1.2%	0.8%	-6.9%	-1.5%
14 × 10	72.16	1.9%	1.2%	-5.6%	-1.5%
16 × 10	87.12	1.9%	0.4%	-6.2%	-1.2%
2 × 15	4.56	0.0%	0.8%	-6.1%	-0.9%
4 × 15	19.56	0.9%	1.7%	-12.7%	-3.0%
6 × 15	35.80	4.6%	2.5%	-12.3%	-3.0%
8 × 15	54.62	2.4%	3.8%	-9.3%	-2.6%
10 × 15	79.40	1.5%	1.0%	-9.5%	-2.5%
2 × 30	19.88	0.4%	5.5%	-14.6%	-1.9%
4 × 30	56.92	-0.8%	3.9%	-18.7%	-4.7%
6 × 30	100.64	1.2%	3.6%	-16.1%	-4.5%
2 × 50	44.82	0.3%	7.4%	-18.0%	-2.9%
2 × 80	84.96	-1.6%	5.9%	-19.3%	-4.5%

Instance set	Saidman et al. (2006)				
	#Transplants	Withhold Agent	Reject Agent	Withhold Total	Reject Total
2 × 10	9.38	-0.8%	4.2%	-14.5%	-1.3%
4 × 10	23.96	3.8%	6.4%	-15.4%	-4.7%
6 × 10	38.06	6.8%	6.5%	-15.6%	-6.8%
8 × 10	51.42	6.6%	6.6%	-13.1%	-8.9%
10 × 10	68.06	0.3%	3.3%	-13.3%	-7.7%
12 × 10	80.12	4.0%	3.7%	-11.0%	-7.3%
14 × 10	96.82	2.1%	5.0%	-10.0%	-8.4%
16 × 10	109.60	3.1%	4.6%	-10.1%	-8.5%
2 × 15	16.82	2.8%	7.7%	-13.8%	-2.1%
4 × 15	37.24	7.6%	10.0%	-15.4%	-6.8%
6 × 15	58.62	4.7%	8.7%	-14.0%	-7.0%
8 × 15	79.72	2.0%	4.7%	-13.3%	-8.3%
10 × 15	102.06	5.1%	6.0%	-11.4%	-7.7%
2 × 30	36.54	4.0%	8.0%	-9.9%	-2.1%
4 × 30	80.26	4.8%	8.1%	-12.9%	-5.7%
6 × 30	122.12	4.8%	8.7%	-12.0%	-6.4%
2 × 50	67.66	3.3%	5.6%	-7.2%	-1.7%
2 × 80	105.74	3.4%	5.7%	-4.2%	-1.3%

**Definition 6.** Maximum Rejection-Proof KEP (MAXRP-KEP)

The MAXRP-KEP mechanism is the mechanism  $f$  that, for any m-KEP instance  $I = (\mathcal{E}, \mathcal{P})$ , selects a feasible kidney exchange solution  $X = f(I) \in \mathcal{X}$  optimal to OPT-RP-KEP. This solution  $X$ , encoded by  $\mathbf{x} = (x_e)_{e \in \mathcal{E}}$ , can be found by solving the following integer bilevel program:

$$z^{\text{MAXRP-KEP}}(I) = \max_{\mathbf{x}} \sum_{e \in \mathcal{E}} w_e x_e \tag{4a}$$

$$\text{s. t. } \sum_{e \in \mathcal{E}(v)} x_e \leq 1, \quad v \in V^p \tag{4b}$$

$$\sum_{e \in \mathcal{E}} w_e^p x_e \geq \text{RKEP}(\mathbf{x}, p), \quad p \in \mathcal{P} \tag{4c}$$

$$x_e \in \{0, 1\}. \quad e \in \mathcal{E} \tag{4d}$$

Observe that (4) is indeed an integer bilevel program. The right-hand side of Constraints (4c) is given by the optimal values of the lower-level problems of finding the best rejection strategy  $Y^p \in \mathcal{R}^p(X)$  to solution  $X$ , for each agent  $p \in \mathcal{P}$ . In other words, this bilevel optimization problem implicitly considers lower-level variables  $\mathbf{y}^p$  encoding the selected rejection strategies  $Y^p \in \mathcal{R}^p(X)$  for each  $p \in \mathcal{P}$  respectively. If we omit Constraints (4c), we have a standard weighted kidney exchange problem. Constraints (4c) are added to impose rejection-proofness of the solution. Intuitively, a kidney exchange solution  $X$  can only be rejection-proof if its value

$z^p(X)$  to agent  $p$  is at least the value of any optimal rejection strategy  $Y^p \in \mathcal{R}^p(X)$ . We show that for cases in which standard kidney exchange problems are NP-hard, the problem of finding maximum cardinality rejection-proof kidney exchange solutions is complete in the complexity class  $\Sigma_2^P$ . We refer to Woeginger (2021) for a more elaborate background on the polynomial hierarchy.

**Theorem 1.** DEC-RP-KEP is  $\Sigma_2^P$ -complete for any fixed maximum cycle length  $K \geq 3$  (even in the case of  $|\mathcal{P}| = 2$ , transplants being valued equally and no non-directed donors).

The proof of this theorem can be found in Appendix A. Informally, this proof is based on a polynomial reduction from the  $\Sigma_2^P$ -complete problem ADVERSARIAL (2,2)-SAT (Johannes, 2011). In this problem, we are given a boolean CNF expression  $E$  consisting of a set of clauses  $C$ , based on two sets of variables  $X$  and  $Y$ . The question is whether there exists a truth setting for  $X$  such that there exists no truth setting for  $Y$  such that  $E$  is satisfied. In our reduction, we construct a compatibility graph corresponding to an m-KEP with 2 agents.

Given the NP-hardness of the basic kidney exchange problem, which appears in the lower level problem, the complexity of the bilevel problem is perhaps not surprising. It does raise the question of the complexity of the special cases in which the standard kidney exchange problem is solvable in polynomial time. This is the case for  $K = 2$ , and here the problem of finding an optimal rejection-proof solution follows from results on a related game, studied in Carvalho and Lodi (2023). They show one can find a socially optimal Nash equilibrium in polynomial time, for any number of agents. Since the exchanges in this Nash equilibrium also form a rejection-proof solution, this gives a polynomial time method for finding an optimal rejection-proof solution. The case for  $K = \infty$  is, to the best of our knowledge, still open.

In the remainder of this section, we discuss strategies to solve Problem (4), based on the connection between rejection strategies and vertex subsets. In Section 4.1, we propose a single-level reformulation of the problem with exponentially many constraints. Next, Section 4.2 discusses an iterative constraint generation procedure.

#### 4.1. A single-level reformulation

In this section, we identify a set of constraints that can replace Constraints (4c) in Problem (4), thus reducing it to a single-level optimization problem, albeit being exponential in the size of the input. First, for each agent  $p \in \mathcal{P}$  and each subset  $U^p \subseteq V^p$ , we define

$$\beta(U^p) = \max_{X \in \mathcal{X}} \left\{ \sum_{e \in X} w_e^p \mid \forall e \in X, V(e) \subseteq U^p \right\} \tag{5}$$

as the maximum value of a feasible kidney exchange solution to  $p$  on the subgraph induced by  $U^p$ . Note that calculating  $\beta(U^p)$  requires solving a standard weighted kidney exchange problem, which is an NP-hard problem in general. We now claim the following:

**Theorem 2.** A kidney exchange solution  $X \in \mathcal{X}$  encoded by  $x = (x_e)_{e \in \mathcal{E}}$  is feasible to Problem (4), that is rejection-proof, if and only if it satisfies

$$\sum_{e \in \mathcal{E} : V(e) \cap U^p \neq \emptyset} w_e^p x_e \geq \beta(U^p), \quad \forall p \in \mathcal{P}, U^p \subseteq V^p. \tag{6}$$

**Proof.** ( $\Rightarrow$ ) Consider a rejection-proof solution  $X \in \mathcal{X}$ . We argue by contradiction. Suppose there exists some agent  $p \in \mathcal{P}$  and a subset  $U^p \subseteq V^p$  for which the solution  $X$  does not satisfy (6). Let  $Y^p \in \mathcal{R}^p(X)$  be obtained from  $X$  by first rejecting all exchanges  $e \in X$  such that  $V(e) \cap U^p \neq \emptyset$  and secondly adding the exchanges of a solution on  $G[U^p]$  of maximum value to  $p$ .

As  $Y^p \in \mathcal{R}^p(X)$ , we have that  $z^p(X) < z^p(Y^p) \leq \text{RKEP}(x, p)$ . Here, the first inequality follows from the assumption that constraint (6) is violated with respect to  $U^p$ , as this means that  $z^p(X \setminus Y^p) < z^p(Y^p \setminus X)$ . The second inequality follows from the fact that  $y^p$ , which encodes the solution  $Y^p$ , is feasible to  $\text{RKEP}(x, p)$ . Hence, it follows that  $X$  is not rejection-proof.

( $\Leftarrow$ ) Suppose the solution  $X = (x_e)_{e \in \mathcal{E}}$  is not rejection-proof. Then there exists an agent  $p \in \mathcal{P}$  and a solution  $Y^p = (y_e^p)_{e \in \mathcal{E}}$  feasible to  $\text{RKEP}(x, p)$  such that

$$\sum_{e \in \mathcal{E}} w_e^p y_e^p > \sum_{e \in \mathcal{E}} w_e^p x_e. \tag{7}$$

We now construct the set  $U^p$  as follows:

$$U^p := \{v \in V^p \mid \exists e \in \mathcal{E}(v) : y_e^p = 1, x_e = 0\}. \tag{8}$$

Thus, the set  $U^p$  is the set of vertices that are part of an exchange in  $Y^p$ , and not part of an exchange in  $X$ . By construction of the set  $U^p$ , we have that  $y_e^p \leq x_e$  if  $V(e) \cap U^p = \emptyset$ . Thus, we have:

$$\sum_{e \in \mathcal{E} : V(e) \cap U^p = \emptyset} w_e^p y_e^p \leq \sum_{e \in \mathcal{E} : V(e) \cap U^p = \emptyset} w_e^p x_e. \tag{9}$$

Combining (7) and (9) gives

$$\sum_{e \in \mathcal{E} : V(e) \cap U^p \neq \emptyset} w_e^p y_e^p > \sum_{e \in \mathcal{E} : V(e) \cap U^p \neq \emptyset} w_e^p x_e. \tag{10}$$

Finally, note that by construction of  $U^p$ , if for an exchange  $e \in Y^p$  it holds that  $V(e) \cap U^p \neq \emptyset$ , then actually  $V(e) \subseteq U^p$ . These exchanges thus describe a feasible kidney exchange solution on  $G[U^p]$ . We now have

$$\beta(U^p) > \sum_{e \in \mathcal{E} : V(e) \cap U^p \neq \emptyset} w_e^p y_e^p. \tag{11}$$

From combining (10) and (11), we can conclude that (6) is violated for the set  $U^p$ .  $\square$

**Remark 1.** Notice that the subset rejection constraint corresponding to agent  $p \in \mathcal{P}$  and subset  $U^p = V^p$  imposes individual rationality for  $p$ , i.e., the proposed solution should have a higher total value to agent  $p$  than any internal solution on  $V^p$ . In particular, the leftmost solution in Fig. 3 satisfies individual rationality for all agents, but violates rejection-proofness (due to the subset rejection constraint for  $U^r = \{b, c\}$ ).

We can now replace Constraints (4c) with Constraints (6), reducing the problem to a single-level integer program. We refer to Constraints (6) as the *subset rejection constraints (SRC)*. We denote by  $\mathcal{J}$  the index set of the subset rejection constraints, i.e.,

$$\mathcal{J} = \{(p, U^p) \mid p \in \mathcal{P}, U^p \subseteq V^p\}.$$

Violation of the subset rejection constraint for some  $p \in \mathcal{P}, U^p \subseteq V^p$  means that agent  $p$  has the incentive to select the rejection strategy obtained from rejecting all exchanges intersecting this subset and replacing it with a kidney exchange solution on  $G[U^p]$  of maximum value to  $p$ .

We now obtain the following formulation:

$$z^{\text{MAXRP-KEP}}(\mathcal{I}; \mathcal{J}) = \max_{\mathbf{x}} \sum_{e \in \mathcal{E}} w_e x_e \tag{12a}$$

$$\text{s. t. } \sum_{e \in \mathcal{E}(v)} x_e \leq 1, \quad v \in V \tag{12b}$$

$$\sum_{e \in \mathcal{E} : V(e) \cap U^p \neq \emptyset} w_e^p x_e \geq \beta(U^p), \quad (p, U^p) \in \mathcal{J} \tag{12c}$$

$$x_e^p \in \{0, 1\}. \quad e \in \mathcal{E} \tag{12d}$$

Observe that, given the weight vector  $\mathbf{w}$ , the quantity  $\beta(U^p)$  is a constant number for each agent  $p \in \mathcal{P}$  and subset  $U^p \subseteq V^p$ . Hence, this is indeed a single-level reformulation of the problem. This comes at the cost of introducing an exponential number of constraints (12c) due to the number of subsets  $U^p \subseteq V^p$  for each agent  $p \in \mathcal{P}$ . However, as finding a rejection-proof solution involving the largest number of pairs and NDDs is  $\Sigma_2^P$ -hard due to Theorem 1, we cannot hope to find a compact formulation unless the polynomial hierarchy collapses.

#### 4.2. Solving MAXRP-KEP

In this section, we discuss our solution approach for MAXRP-KEP. Problem (12) has an exponential number of constraints due to  $|\mathcal{J}|$  being exponential in the number of vertices. In our solution method, we initially consider a relaxation in which the subset rejection constraints (12c) are restricted to some small subset  $\mathcal{J}' \subset \mathcal{J}$ .

Iteratively, we solve this relaxation and apply a separation method to identify for each agent  $p \in \mathcal{P}$  whether there exists a subset  $U^p \subseteq V^p$  such that the subset rejection constraint with respect to  $(p, U^p) \in \mathcal{J} \setminus \mathcal{J}'$  is violated. This iterative algorithm terminates when there do not exist any more violated constraints.

Observe that, viewing formulation (12) through the lens of bilevel optimization, we can consider the lower level variables ( $y^p$  for each  $p \in \mathcal{P}$ ) to be included implicitly in this single-level reformulation. In order to see this, for any feasible kidney exchange (leader) solution  $\hat{\mathbf{x}}$ , note that the lower-level solution  $\hat{\mathbf{y}}^p$  defined by  $\hat{y}_e^p = \hat{x}_e$  for all  $e \in \mathcal{E}$  is feasible to the lower-level problem RKEP( $\hat{\mathbf{x}}, p$ ). Constraints (12c) thus cut off bilevel infeasible solutions.

Next, we will discuss how to separate violated subset rejection constraints and tiebreaking approaches for the relaxed master problem in more detail.

##### 4.2.1. Separating the subset rejection constraints

Consider a solution  $X$  that is not rejection-proof. In the proof of Theorem 2, we show that if there exists an agent  $p \in \mathcal{P}$  and solution  $Y^p$  feasible to RKEP( $\mathbf{x}, p$ ) with  $z^p(Y^p) > z^p(X)$ , then at least one inequality from the class of subset rejection constraints (6) must be violated. In other words, then there is an agent  $p$  and a subset

$$U^p = \{v \in V^p \mid \exists e \in \mathcal{E}(v) : y_e^p = 1, x_e = 0\}$$

for which an inequality from (6) is violated. The right-hand side of the violated subset rejection constraint can be computed as follows.

$$\beta(U^p) = \sum_{e \in \mathcal{E} : V(e) \subseteq U^p} w_e^p y_e^p. \tag{13}$$

In other words, the separation problem can be decoupled into a separation problem for each agent  $p \in \mathcal{P}$ , which is just the problem  $\text{RKEP}(\mathbf{x}, p)$ . For each  $p \in \mathcal{P}$ , we can deduce from its optimal solution  $Y^p \in \mathcal{R}^p(X)$  a violated subset rejection constraint if  $p$  is incentivized to reject  $X$ .

#### 4.2.2. Tiebreaking in MAXRP-KEP

It has been observed that kidney exchange problems usually have many optimal solutions (Farnadi et al., 2021). This carries over to MAXRP-KEP, in particular when the number of rejection constraints  $|J'|$  is still small, as the problem is then almost similar to a regular kidney exchange problem. Selecting a solution among distinct optima is a delicate matter, as some solutions may be rejection-proof while others are not. It is desirable to identify an optimal rejection-proof solution if one exists, as we can then restrict the number of iterations of the (relatively costly) procedure of separating violated subset rejection constraints and resolving of the resulting, strengthened,  $z^{\text{MAXRP-KEP}}(I; J')$ . The following lemma gives us insight into how to compare solutions in terms of rejection-proofness.

**Lemma 1.** Consider two solutions  $X, \tilde{X} \in \mathcal{X}$  such that

- (a)  $(\tilde{X} \cap \mathcal{E}^0) \subseteq (X \cap \mathcal{E}^0)$ ,
- (b)  $z^p(\tilde{X}) \geq z^p(X)$  for each agent  $p \in \mathcal{P}$ .

If  $X$  is rejection-proof, then  $\tilde{X}$  is rejection-proof.

**Proof.** We argue by contradiction, suppose that  $\tilde{X}$  is not rejection-proof. Then, given encoding  $\tilde{\mathbf{x}}$  for  $\tilde{X}$ , for some agent  $p \in \mathcal{P}$  there exists a solution  $\tilde{Y}^p \in \mathcal{R}^p(\tilde{X})$  optimal to  $\text{RKEP}(\tilde{\mathbf{x}}, p)$  such that  $z^p(\tilde{Y}^p) > z^p(\tilde{X})$ . Note that any solution feasible to  $\text{RKEP}(\tilde{\mathbf{x}}, p)$  is also feasible to  $\text{RKEP}(\mathbf{x}, p)$ , i.e.,  $\mathcal{R}^p(\tilde{X}) \subseteq \mathcal{R}^p(X)$ , because of Condition (a). Finally,  $z^p(\tilde{Y}^p) > z^p(\tilde{X}) \geq z^p(X)$  means the agent has incentive to reject  $X$  as well, meaning that  $X$  is also not rejection-proof.  $\square$

This lemma shows that the number of shared exchanges in a solution and agent value are two factors influencing whether or not a solution is rejection-proof. Indeed, when the number of shared exchanges in a solution  $X$  is small, this implies that the number of feasible rejection strategies, i.e. the number of feasible solutions of  $\text{RKEP}(\mathbf{x}, p)$ , is more restricted compared to when there are many shared exchanges. This feasible region is also more restricted if the value  $z^p(X)$  for solution  $X$  to agent  $p$  is larger, which is consistent with Constraints (4c).

We use this knowledge to propose as a tiebreaker the internal preference objective

$$t_1(X) := \sum_{p \in \mathcal{P}} \sum_{e \in \mathcal{E}^p} w_e^p x_e. \tag{14}$$

By counting only the agent value of internal exchanges, it immediately takes into account both conditions of Lemma 1. Clearly, this tiebreaker prioritizes solutions with large agent values, whereas solutions with many shared exchanges are disadvantaged. In fact, we can show that given two solutions as in Lemma 1, mechanisms deploying the tiebreaking rule (14) will always (weakly) prioritize  $\tilde{X}$  over  $X$ .

**Lemma 2.** Consider two solutions,  $X$  and  $\tilde{X}$ , such that  $\tilde{X} \cap \mathcal{E}^0 \subseteq X \cap \mathcal{E}^0$  and  $z^p(\tilde{X}) \geq z^p(X)$  for all agents  $p \in \mathcal{P}$ . Then, we have  $t_1(\tilde{X}) \geq t_1(X)$ , i.e., the tiebreak value of  $\tilde{X}$  is at least as high as the tiebreak value of  $X$ .

**Proof.** Let  $\tilde{X} \cap \mathcal{E}^0 \subseteq X \cap \mathcal{E}^0$ . Then, using the definitions of  $z^p(\tilde{X})$  and  $z^p(X)$ , we get

$$\sum_{e \in \mathcal{E}^0} w_e^p \tilde{x}_e \leq \sum_{e \in \mathcal{E}^0} w_e^p x_e \Rightarrow \sum_{e \in \mathcal{E}^p} w_e^p \tilde{x}_e \geq \sum_{e \in \mathcal{E}^p} w_e^p x_e. \tag{15}$$

Now, it follows from (14) and (15) that  $t_1(\tilde{X}) \geq t_1(X)$ .  $\square$

The tiebreaker is implemented as follows. Let  $Z$  be an upper bound to the agent value of internal exchanges in any solution ( $Z \geq \max_{X \in \mathcal{X}} \sum_{p \in \mathcal{P}} \sum_{e \in \mathcal{E}^p} w_e^p x_e$ ). We then replace the objective (4a) of MAXRP-KEP, by

$$\sum_{e \in \mathcal{E}} Z w_e x_e + \sum_{p \in \mathcal{P}} \sum_{e \in \mathcal{E}^p} w_e^p x_e. \tag{16}$$

This objective takes into account both the social welfare value of all transplants ( $\sum_{e \in \mathcal{E}} Z w_e x_e$ ) and the agent value of internal exchanges ( $\sum_{e \in \mathcal{E}^p} w_e^p x_e$ ). Due to the heavy weighting of social welfare value, the agent value will only function as a tiebreaker between solutions with equal social welfare value.

To summarize, we can characterize the steps performed by the MAXRP-KEP mechanism with the pseudocode described in Algorithm 1.

---

**Algorithm 1:** Description of the MAXRP-KEP mechanism.

---

```

Data:  $m$ -KEP instance  $I = (\mathcal{E}, \mathcal{P})$ 
Result: Rejection-proof solution  $X \in \mathcal{X}$  of maximum social welfare value
 $J' = \{\emptyset\}$ ;
 $\mathbf{x}^* \in \arg \max_{\mathbf{x}} z^{\text{MAXRP-KEP}}(I; J')$ ; /* objective function as in (16) */
 $X^* = \{e \in \mathcal{E} : x_e^* = 1\}$ ;
while  $X^*$  is not rejection-proof do
  forall  $p \in \mathcal{P}$  do
    Identify  $U^p \subseteq V^p$  such that (12c) violated for  $(p, U^p) \in \mathcal{J} \setminus J'$  w.r.t.  $\mathbf{x}^*$ ;
     $J' = J' \cup (p, U^p)$ ;
  end
   $\mathbf{x}^* \in \arg \max_{\mathbf{x}} z^{\text{MAXRP-KEP}}(I; J')$ ;
   $X^* = \{e \in \mathcal{E} : x_e^* = 1\}$ ;
end
return  $X^*$ 

```

---

### 5. Heuristic mechanisms

In Section 4, we have discussed a mechanism for identifying optimal solutions that are rejection-proof by solving a bilevel optimization problem. In this section, we consider two heuristic mechanisms for finding rejection-proof kidney exchange solutions. On the one hand, these mechanisms have reduced computational complexity. On the other hand, they do not guarantee that the found solutions are indeed optimal. In Section 5.1, we describe a heuristic mechanism called *Maximum Weight Internal KEP*, which proposes solutions that maximize the sum of the agents' values for internal exchanges. This ensures that there are no incentives for the agents to deviate from accepting the proposed solution. In Section 5.2, we propose a mechanism that first identifies a rejection-proof solution that results from the best responses to a proposed solution. In the second step, local improvements preserving rejection-proofness are made.

#### 5.1. Maximum weight internal matching

We now describe our first heuristic mechanism, which we call the *Maximum Weight Internal KEP* mechanism, shortened to the MAXINT mechanism or simply MAXINT. This mechanism proposes a solution where the social transplant value is maximized, under the restriction that the agent value of internal exchanges is maximum.

**Definition 7.** Maximum Weight Internal Transplants (MAXINT)

For any  $m$ -KEP instance  $I = (\mathcal{E}, \mathcal{P})$ , the MAXINT mechanism selects a feasible kidney exchange solution  $X \in \mathcal{X}$  encoded by  $\mathbf{x} = (x_e)_{e \in \mathcal{E}}$  optimal to the following integer linear program (MAXINT-KEP):

$$z^{\text{MAXINT}} = \max_{\mathbf{x}} \sum_{e \in \mathcal{E}} w_e x_e \tag{17a}$$

$$\text{s. t. } \sum_{e \in \mathcal{E}(v)} x_e \leq 1, \quad v \in V \tag{17b}$$

$$\sum_{e \in \mathcal{E}^p} w_e^p x_e = \beta(V^p), \quad p \in \mathcal{P} \tag{17c}$$

$$x_e \in \{0, 1\}. \quad e \in \mathcal{E} \tag{17d}$$

Constraints (17b) again impose that  $\mathbf{x}$  should induce a set of pairwise vertex-disjoint exchanges. Moreover, Constraints (17c) ensure that  $X^p = \{e \in \mathcal{E}^p \mid x_e = 1\}$  forms a maximum (agent) value internal kidney exchange solution for each agent  $p \in \mathcal{P}$ .

Notice that the feasible region of Problem (17) is equivalent to the set of kidney exchange solutions that can be obtained if each agent selects a greedy withholding strategy. As MAXINT aims to maximize the social welfare value of the kidney exchange solution, it can be regarded as the kidney exchange solution for which all agents coordinate their greedy withholding strategies in order to maximize social welfare. We note that there are existing transnational collaborations where greedily withholding within countries is the official policy (Valentín et al., 2019). While simple, the MAXINT mechanism guarantees at least equal social welfare compared to this policy, and can improve it.

**Theorem 3.** *The MAXINT mechanism is rejection-proof.*

**Proof.** Let  $X \in \mathcal{X}$  be a kidney exchange solution proposed by MAXINT, encoded by binary variables  $(x_e)_{e \in \mathcal{E}}$ , and let  $p \in \mathcal{P}$  be an arbitrary agent. Then

$$z^p(X) = \sum_{e \in \mathcal{E}^p} w_e^p x_e + \sum_{e \in \mathcal{E}^0} w_e^p x_e.$$

For any optimal solution  $y^p$  to RKEP( $x, p$ ), encoding solution  $Y^p \in \mathcal{R}^p(X)$ , we have

$$\text{RKEP}(x, p) = \sum_{e \in \mathcal{E}^p} w_e^p y_e^p + \sum_{e \in \mathcal{E}^0} w_e^p y_e^p.$$

As  $X \cap \mathcal{E}^p$  is a maximum value internal solution on  $V^p$ , we have

$$\sum_{e \in \mathcal{E}^p} w_e^p x_e \geq \sum_{e \in \mathcal{E}^p} w_e^p y_e^p. \tag{18}$$

Next, notice that a shared exchange  $e \in \mathcal{E}^0$  can only be contained in  $Y^p$  if it was in the proposed solution  $X$  in the first place, hence

$$\sum_{e \in \mathcal{E}^0} w_e^p x_e \geq \sum_{e \in \mathcal{E}^0} w_e^p y_e^p. \tag{19}$$

It follows from adding (18) and (19) that  $z^p(X) \geq \text{RKEP}(x, p)$ , showing that MAXINT is rejection-proof.  $\square$

We illustrate the workings of this mechanism in Section 5.3.

### 5.2. Reject-and-Rebuild

We now consider a second heuristic mechanism for rejection-proof kidney exchange solutions, which we call the REJECT-AND-REBUILD (RR) mechanism. Informally, the mechanism computes a solution by initially considering a socially optimal solution and mimicking best-response strategies of agents to this solution (REJECT). This process iterates until rejection-proofness is reached. At this point, the mechanism adds local improvements preserving rejection-proofness (REBUILD). Formally, the mechanism takes an instance  $I = (\mathcal{E}, \mathcal{P})$  and outputs a rejection-proof kidney exchange solution  $X = f(I) \in \mathcal{X}$  after performing the following steps:

1. An arbitrary kidney exchange solution  $\tilde{X} \in \mathcal{X}$  of maximum social welfare value is computed. This solution  $\tilde{X}$  is encoded by  $\tilde{x} = (x_e)_{e \in \mathcal{E}}$ .
2. (REJECT) For each agent  $p \in \mathcal{P}$ , solve the rejection problem RKEP( $\tilde{x}, p$ ) with optimal rejection strategy  $Y^p \in \mathcal{R}^p(\tilde{X})$ . If  $\tilde{X}$  is rejection-proof, i.e.,  $Y^p = \tilde{X}$  for all  $p \in \mathcal{P}$ , go to Step 3. Otherwise, let  $\tilde{X}^I = \bigcup_{p \in \mathcal{P}} (Y^p \cap \mathcal{E}^p)$  be the set of internal exchanges obtained from the rejection strategies  $Y^p$  of all agents  $p \in \mathcal{P}$ . Furthermore, let  $\tilde{X}^0 = \bigcap_{p \in \mathcal{P}} (Y^p \cap \mathcal{E}^0)$  be the set of non-rejected shared exchanges of  $\tilde{X}$ . We set  $\tilde{X} = \tilde{X}^I \cup \tilde{X}^0$  and repeat Step 2.
3. (REBUILD) For each agent  $p \in \mathcal{P}$ , let  $U^p(\tilde{X}) \subseteq V^p$  be the set of pairs and NDDs of  $p$  that are unmatched in  $\tilde{X}$ , i.e.,

$$U^p(\tilde{X}) := \{v \in V^p \mid \nexists e \in \tilde{X} : v \in V(e)\}.$$

Let now  $X' \in \mathcal{X}$  be some kidney exchange solution on  $G[\bigcup_{p \in \mathcal{P}} U^p(\tilde{X})]$  maximizing social welfare value. Output  $X = \tilde{X} \cup X'$ .

We now show that the RR mechanism indeed terminates and outputs rejection-proof mechanisms.

**Theorem 4.** *The REJECT-AND-REBUILD mechanism is rejection-proof.*

**Proof.** Let  $\tilde{X} \in \mathcal{X}$  be a kidney exchange solution proposed by the RR mechanism in Step 1, encoded by  $(\tilde{x}_e)_{e \in \mathcal{E}}$ . To show that the mechanism terminates, we need to argue that the iterative approach in Step 2 gives us a rejection-proof solution after a finite number of iterations. Observe that whenever an agent  $p$  rejects the solution using a nontrivial rejection strategy  $Y^p \in \mathcal{R}^p(\tilde{X}) \setminus \{\tilde{X}\}$ , the number of pairs involved in an internal exchange strictly increases by at least one. Hence, for any m-KEP with finite agent pools, this means that the number of iterations is also finite. Furthermore, once the number of pairs involved in an internal exchange is maximum, the corresponding solution  $\tilde{X}$  will be rejection-proof, and we go to Step 3. As a solution does not lose the rejection-proofness property after adding exchanges to it, the final solution is also rejection-proof, hence the Reject-and-Rebuild mechanism is rejection-proof.  $\square$

### 5.3. Illustration of MAXINT and REJECT-AND-REBUILD

We give an illustration of the two mechanisms MAXINT and REJECT-AND-REBUILD using two different instances. First, consider the instance depicted in Fig. 4.

We provide the MAXRP-KEP solution for this instance in Fig. 5a; this solution features six transplants, is a social optimum and is rejection-proof as neither the blue nor the red agent has an incentive to reject this solution. The solution obtained from the MAXINT mechanism is depicted in Fig. 5b; notice that it is also rejection-proof. The solution obtained from the REJECT-AND-REBUILD mechanism is depicted in Fig. 5c, and found in two steps. Given the first solution (which is a social optimum), the blue agent rejects one cycle (the rightmost 3-cycle) in favor of the blue (internal) 2-cycle. The resulting solution is both rejection-proof and maximal, hence no shared exchanges are added in the REBUILD step.

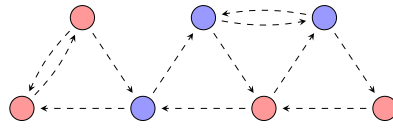


Fig. 4. Example of an m-KEP instance with two agents.

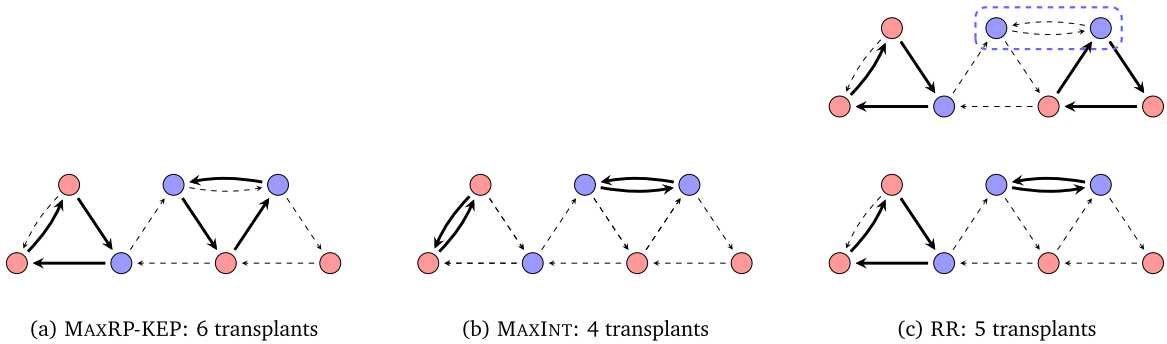


Fig. 5. Example of an m-KEP instance  $I$  for which the three rejection-proof mechanisms output three different kidney exchange solutions, with  $z^{\text{MAXRP-KEP}}(I) > z^{\text{RR}}(I) > z^{\text{MAXINT}}(I)$ .

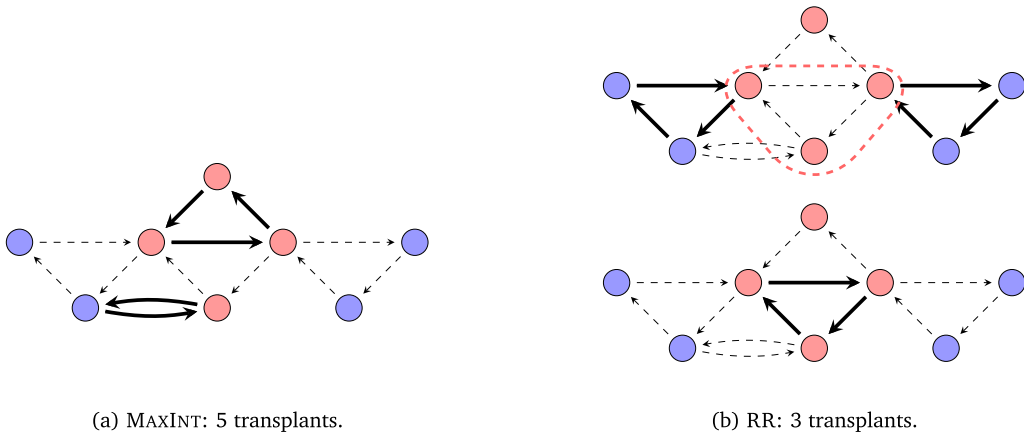


Fig. 6. Example of an m-KEP instance  $I$  with  $z^{\text{MAXINT}}(I) > z^{\text{RR}}(I)$ .

Let us now consider MAXINT and REJECT-AND-REBUILD, when applied to the instance depicted in Fig. 6. MAXINT gives the solution depicted in Fig. 6a, consisting of five transplants. When considering the REJECT-AND-REBUILD mechanism in Fig. 6b, notice that the red agent has two optimal rejection strategies, given by rejecting all existing cycles and adding either of the two 3-cycles internal to the red agent. As the red agent selects an arbitrary optimal rejection strategy, it can be the case that they pick this strategy in an unfortunate fashion, leading to a strictly worse solution compared to the MAXINT solution in terms of social welfare value. Furthermore, this instance also shows that the REBUILD step is not redundant. If the red agent would select the topmost 3-cycle as their rejection strategy, the 2-cycle (on the bottom) would become completely unmatched, meaning it could be added to the solution (obtaining the same solution as MAXINT). We finally point out that these two instances show that REJECT-AND-REBUILD does not dominate MAXINT and vice versa.

In the next section, we conduct experiments to compare the optimal (MAXRP-KEP) and the two heuristic mechanisms (MAXINT and RR) against each other in terms of computational complexity and solution quality, tested on instances obtained from two state-of-the-art instance generators for general kidney exchange programs.

### 6. Computational results

In this section, we describe our computational experiments dedicated to comparing the mechanisms proposed in Sections 4 and 5.

First, in Section 6.1 we look at the mechanism outcomes in two ways. We compare the quality of solutions of the instance sets against the maximum value of (unrestricted) kidney exchange solutions, i.e., the maximum value of kidney exchange solutions for which rejection-proofness is not necessarily satisfied. Furthermore, we repeat our experiments from Section 3, testing whether

**Table 2**

Comparison of the solution quality (average over 50 instances per instance set) of the rejection-proof mechanisms (MAXRP-KEP / MAXINT / RR) with the setting in which all agents perform greedy withholding (Withhold Total) or rejection strategies (Reject Total). The numbers in the columns show the average percentage loss in the total number of transplants realized in comparison with the maximum value kidney exchange solution, for instances generated using the distributions proposed by Delorme et al. (2022) and Saidman et al. (2006) respectively.

Instance set	Delorme et al. (2022)					
	#Transplants	MAXRP-KEP	MAXINT	RR	Withhold Total	Reject Total
2 × 10	2.14	0.0%	-6.5%	0.0%	-6.5%	0.0%
4 × 10	9.02	-0.9%	-11.8%	-3.8%	-11.8%	-3.8%
6 × 10	18.56	-0.4%	-6.9%	-1.4%	-7.4%	-1.4%
8 × 10	28.62	-0.6%	-7.2%	-2.0%	-7.3%	-2.0%
10 × 10	42.46	-0.6%	-5.6%	-1.3%	-5.7%	-1.5%
12 × 10	57.08	-0.4%	-6.6%	-1.5%	-6.9%	-1.5%
14 × 10	72.16	-0.4%	-5.6%	-1.4%	-5.6%	-1.5%
16 × 10	87.12	-0.3%	-6.0%	-1.1%	-6.2%	-1.2%
2 × 15	4.56	-0.4%	-5.3%	-0.9%	-6.1%	-0.9%
4 × 15	19.56	-1.0%	-12.1%	-3.0%	-12.7%	-3.0%
6 × 15	35.80	-0.4%	-10.8%	-2.9%	-12.3%	-3.0%
8 × 15	54.62	-0.4%	-9.0%	-2.5%	-9.3%	-2.6%
10 × 15	79.40	-0.7%	-9.2%	-2.4%	-9.5%	-2.5%
2 × 30	19.88	-0.8%	-11.5%	-1.9%	-14.6%	-1.9%
4 × 30	56.92	-1.0%	-15.7%	-4.5%	-18.7%	-4.7%
6 × 30	100.64	-0.5%	-13.4%	-4.1%	-16.1%	-4.5%
2 × 50	44.82	-0.8%	-14.2%	-2.9%	-18.0%	-2.9%
2 × 80	84.96	-0.8%	-15.5%	-4.4%	-19.3%	-4.5%
Instance set	Saidman et al. (2006)					
	#Transplants	MAXRP-KEP	MAXINT	RR	Withhold Total	Reject Total
2 × 10	9.38	-0.2%	-10.9%	-1.3%	-14.5%	-1.3%
4 × 10	23.96	-0.6%	-13.4%	-4.7%	-15.4%	-4.7%
6 × 10	38.06	-0.4%	-13.2%	-6.1%	-15.6%	-6.8%
8 × 10	51.42	-0.5%	-11.5%	-7.8%	-13.1%	-8.9%
10 × 10	68.06	-0.1%	-11.5%	-6.7%	-13.3%	-7.7%
12 × 10	80.12	-0.0%	-8.9%	-5.9%	-11.0%	-7.3%
14 × 10	96.82	-0.0%	-7.8%	-6.2%	-10.0%	-8.4%
16 × 10	109.60	-0.1%	-8.0%	-6.2%	-10.1%	-8.5%
2 × 15	16.82	-0.4%	-11.2%	-2.1%	-13.8%	-2.1%
4 × 15	37.24	-0.4%	-12.9%	-6.2%	-15.4%	-6.8%
6 × 15	58.62	-0.2%	-11.2%	-6.3%	-14.0%	-7.0%
8 × 15	79.72	-0.1%	-10.9%	-6.6%	-13.3%	-8.3%
10 × 15	102.06	0.0%	-9.0%	-6.2%	-11.4%	-7.7%
2 × 30	36.54	-0.3%	-8.5%	-2.0%	-9.9%	-2.1%
4 × 30	80.26	-0.0%	-10.4%	-5.3%	-12.9%	-5.7%
6 × 30	122.12	-0.0%	-8.8%	-5.8%	-12.0%	-6.4%
2 × 50	67.66	-0.2%	-6.0%	-1.7%	-7.2%	-1.7%
2 × 80	105.74	0.0%	-4.1%	-1.3%	-4.2%	-1.3%

individual agents can benefit from employing a greedy withholding strategy. This time, we assume the collaboration uses one of the rejection-proof mechanisms, rather than computing a socially optimal solution.

Second, in Section 6.2 we discuss the performance of the MAXRP-KEP mechanism in terms of computation times, showing that despite the computation complexity the problem of computing optimal rejection-proof solutions is tractable for many types of instances.

### 6.1. Quality of rejection-proof solutions

We compare the quality of the solutions obtained from the exact mechanism (MAXRP-KEP) and the two heuristic mechanisms (MAXINT and REJECT-AND-REBUILD (RR)) against the solution value of basic kidney exchange solutions without additional restrictions. To this end, we consider the same instances generated for the discussion in Section 3. Our goal is to compare the maximum number of transplants that can be attained under the three different rejection-proof regimes.

Table 2 shows that MAXRP-KEP achieves rejection-proofness at a small cost on average. For all instance sets, the loss in the number of transplants to the social optimum is at most 1%, which is significantly smaller than the losses that can occur if agents employ withholding (Withhold Total) or rejection strategies (Reject Total). Through the rebuild step, Reject-and-Rebuild also strictly improves over agents employing rejection strategies, but only slightly. The MAXINT mechanism comes at a higher cost, with losses up to 15% of total transplants and consistently higher than losses from agents using rejection strategies. However, it still improves upon the situation where agents use greedy withholding strategies without coordination.



**Table 3**

Average percentage gain in the number of transplants achieved by a single agent selecting a greedy withholding strategy compared to the setting where this agent is truthful, assuming all other agents are truthful, whenever any of the three mechanisms MAXRP-KEP, MAXINT or RR is employed. Results are shown for instances generated using the distributions proposed by Delorme et al. (2022) and Saidman et al. (2006) respectively.

Instance set	Delorme et al. (2022)		
	Withhold Agent MAXRP-KEP	Withhold Agent MAXINT	Withhold Agent RR
2 × 10	0.0%	0.0%	0.0%
4 × 10	-0.9%	1.1%	-1.0%
6 × 10	0.7%	0.0%	0.0%
8 × 10	0.6%	0.0%	0.0%
10 × 10	0.5%	0.0%	0.5%
12 × 10	0.0%	0.0%	0.4%
14 × 10	0.8%	0.4%	0.8%
16 × 10	2.0%	0.0%	1.5%
2 × 15	0.0%	0.0%	0.0%
4 × 15	-1.4%	0.5%	-1.4%
6 × 15	-0.4%	-2.0%	1.4%
8 × 15	-0.6%	-0.6%	-1.8%
10 × 15	0.8%	0.0%	0.5%
2 × 30	-3.0%	-2.4%	-3.2%
4 × 30	-3.6%	-0.8%	-4.2%
6 × 30	-1.2%	-1.8%	-3.1%
2 × 50	-5.5%	-2.6%	-6.0%
2 × 80	-8.3%	-2.9%	-6.3%
Instance set	Saidman et al. (2006)		
	Withhold Agent MAXRP-KEP	Withhold Agent MAXINT	Withhold Agent RR
2 × 10	-0.4%	-1.5%	0.0%
4 × 10	-1.3%	-0.4%	-2.8%
6 × 10	-2.4%	0.4%	2.0%
8 × 10	-0.9%	0.0%	0.0%
10 × 10	0.8%	-1.3%	-3.8%
12 × 10	0.0%	-2.1%	-3.9%
14 × 10	0.8%	-1.5%	-1.9%
16 × 10	-0.3%	-0.9%	-0.3%
2 × 15	-2.8%	-1.0%	-2.4%
4 × 15	-0.2%	-0.7%	0.2%
6 × 15	0.2%	-2.4%	-1.7%
8 × 15	-2.1%	-0.7%	-3.2%
10 × 15	-1.1%	-2.4%	-4.0%
2 × 30	-1.6%	-0.9%	-0.6%
4 × 30	-1.5%	-1.3%	-1.2%
6 × 30	-1.1%	-1.7%	-2.5%
2 × 50	-1.9%	-0.8%	-1.2%
2 × 80	-0.7%	-0.0%	-0.6%

Analogous to the experiments in Section 3, we also investigate the gains an agent can achieve by using the greedy withholding strategy. Specifically, we consider the difference in transplants for an agent between being truthful and using the greedy withholding strategy, assuming all other agents are truthful. In contrast to Section 3, we now consider the case where the collaboration uses one of the rejection-proof mechanisms.

Table 3 shows the average in- or decrease in transplants for an agent  $p$  for all three mechanisms. Here, we consider the case in which  $p$  would unilaterally decide not to report its entire pool (like all other agents), but instead to select some greedy withholding strategy before the rejection-proof mechanism is employed. For most instance sets, greedy withholding leads to a loss of transplants for the agent on average. This is in contrast to the situation where a solution on the common pool is selected of maximum value (where agents gained by using greedy withholding, as shown in Table 1). These losses are generally small (1 or 2%) and small gains are possible, in particular for the Delorme instances with smaller pool sizes. However, for larger pool sizes, agents may lose up to 8% of transplants by using greedy withholding. These results show that while the rejection-proof mechanisms only provide guaranteed protection against agents employing rejection strategies, in practice they may also disincentivize (greedy) withholding strategies.

## 6.2. Computation times of the MaxRP-KEP mechanism

In this section we further investigate computational aspects of the MaxRP-KEP mechanism. We have previously established the theoretical complexity of the underlying optimization problem, now we consider computational experiments relating to the bilevel optimization problem central to the mechanism. In order to compare computation times for different distributions of agent pool sizes,

**Table 4**

Comparison of computational results of MAXRP-KEP for all instance sets. All results were obtained using no initial constraints, weighted tiebreaking, and maximum violation constraints added. Instances solved is the number of instances solved to optimality in less than 3 hours (10 800s), out of 30. The reported numbers on computation times, constraints added and the number of iterations are averages over all instances solved to optimality within the given time limit.

Instance Set	Agent Sizes	Total Time (s)	Constraints Added	Iterations	Instances Solved
DelBXM-PRA0Score_96_4_112m	1 × 50, 2 × 25	0.46	5.50	4.77	30
DelBXM-PRA0Score_96_4_115s	1 × 50, 5 × 10	0.17	3.37	3.80	30
DelBXM-PRA0Score_96_4_2equal	2 × 50	1.15	9.67	7.63	30
DelBXM-PRA0Score_96_4_4equal	4 × 25	0.06	3.13	2.70	30
DelBXM-PRA0_96_4_112m	1 × 50, 2 × 25	0.09	1.60	2.47	30
DelBXM-PRA0_96_4_115s	1 × 50, 5 × 10	0.08	1.07	1.93	30
DelBXM-PRA0_96_4_2equal	2 × 50	0.09	1.40	2.17	30
DelBXM-PRA0_96_4_4equal	4 × 25	0.04	1.20	1.83	30
Instance Set	Agent Sizes	Total Time (s)	Constraints Added	Iterations	Instances Solved
DelBXM-PRA0Score_192_8_112m	1 × 100, 2 × 50	1809.19	63.21	35.29	14
DelBXM-PRA0Score_192_8_115s	1 × 100, 5 × 20	2063.00	74.21	67.37	19
DelBXM-PRA0Score_192_8_2equal	2 × 100	4537.86	81.50	43.50	2
DelBXM-PRA0Score_192_8_4equal	4 × 50	1332.89	56.31	24.92	26
DelBXM-PRA0_192_8_112m	1 × 100, 2 × 50	256.94	43.50	31.13	30
DelBXM-PRA0_192_8_115s	1 × 100, 5 × 20	166.61	36.25	35.39	28
DelBXM-PRA0_192_8_2equal	2 × 100	589.97	73.18	48.50	28
DelBXM-PRA0_192_8_4equal	4 × 50	24.01	14.20	8.70	30
Instance Set	Agent Sizes	Total Time (s)	Constraints Added	Iterations	Instances Solved
SaidmanScore_96_4_112m	1 × 50, 2 × 25	308.04	36.42	22.83	24
SaidmanScore_96_4_115s	1 × 50, 5 × 10	303.07	33.26	28.67	27
SaidmanScore_96_4_2equal	2 × 50	732.42	41.83	28.04	24
SaidmanScore_96_4_4equal	4 × 25	518.13	26.64	13.89	28
Saidman_96_4_112m	1 × 50, 2 × 25	2.97	3.40	3.37	30
Saidman_96_4_115s	1 × 50, 5 × 10	5.75	7.80	7.40	30
Saidman_96_4_2equal	2 × 50	4.70	2.50	2.77	30
Saidman_96_4_4equal	4 × 25	2.50	9.07	6.87	30

we consider a new set of multi-agent kidney exchange instances which include NDDs. Typically, the existence of NDDs increases the computational challenge in kidney exchange problems, which is our reason for adding these. As before, we consider instances generated using Saidman and Delorme distributions (with the BXM-PRA-0 configuration). This configuration uses a classification of PRA values based on score bins (“bands”) in order to determine whether or not donors and recipients are compatible. Furthermore, additional considerations are made to establish compatibilities for patients with PRA score 0. Each instance contains either 96 pairs and 4 non-directed donors, or 192 pairs and 8 non-directed donors, divided at random among the agents. The division over agents differs, with either 1 large and 2 medium size agents (“112m”: 50% + 2 × 25%), 1 large and 5 small (“115s”: 50% + 5 × 10%), and 2 or 4 equal-sized agents (“2equal” and “4equal”, respectively). For each instance set of 30 instances, we set the social welfare value equal to the number of transplants. In all instances, agents only value transplants to their own patients. In the “Score” instances, the agent valuation for each transplant is drawn independently at random from a uniform distribution between 0 and 99, e.g., to indicate recipient priorities or quality of matches.

These experiments were run using CPLEX 12.10, using 4 threads on a 2.8GHz Intel Xeon Platinum 8260 Processor. The computation times for the MAXRP-KEP mechanism can be seen in Table 4. Although the basic characteristics of these instances such as the proportion of pairs vs. NDDs are fixed, we note massive differences in computation time required for the different instance sets. The differences were most pronounced when we compare the distributions used for instance generation (Saidman vs. Delorme instances), with the Saidman instances considerably more difficult. Another factor that significantly impacted mechanism tractability is the set of preferences of individual agents, more concretely the valuations that agents had for transplants. Instances where agents follow unit valuations for each transplant were significantly easier to solve. Finally, less pronounced differences exist depending on the distribution of agent pool sizes, where harder-to-solve instances were typically among the set of instances with larger pool sizes per agent.

- **Saidman vs. Delorme:** It is clear that there exist huge differences based on the instance generation procedure used. All instance sets generated using the Delorme distributions with pool size 100 are solved completely, with average computation times exceeding one second for only one instance set. Instances of the same size generated using the Saidman distributions on the other hand are much more difficult, with some instances not solved optimally within the time limit. For the solved instances, the average time to optimality is also much longer, up 12 minutes for one instance set. The larger Delorme instances were most difficult, but are double the size of the Saidman instances.

As was shown in Delorme et al. (2022), instances generated using their distributions are significantly sparser than those generated using the Saidman configuration, i.e., the compatibility graphs in Delorme instances consistently have significantly fewer arcs than the Saidman instances. For denser compatibility graphs the computational cost of solving the integer programs

**Table 5**

Comparison of SaidmanScore instances based on whether or not tiebreaks are used. Results are obtained using no initial constraints. Instances solved is the number of instances solved to optimality within 10800 seconds (out of 30). Reported computation times, master times, and numbers of iterations are averages restricted to the instances solved to optimality within the time limit.

Instance Set	Tiebreak	Total Time (s)	Master Time (s)	Iterations	Instances Solved
1l2m	Used	308.04	298.82	22.83	24
	Not Used	942.85	911.45	44.48	25
1l5s	Used	303.07	290.88	28.67	27
	Not Used	726.00	672.74	43.71	28
2equal	Used	732.42	720.24	28.04	24
	Not Used	1176.52	1156.51	39.17	24
4equal	Used	518.13	516.44	13.89	28
	Not Used	378.73	376.41	18.71	28

increases as there exist more potential exchanges that could be performed. Furthermore, denser graphs increase the set of possible rejection strategies. This gives the agents more freedom and they are thus more likely to reject proposed solutions. This is also reflected in the number of iterations required to find the optimal rejection-proof solution.

- **Unit vs. Score:** Instances in which the agents have non-unit valuations (denoted by “Score” in the instance set name) for transplants are much more difficult. This is most apparent for the Saidman and larger Delorme instances, where almost all instances with unit valuations are solved within the time limit. For the non-unit values, instances in each set were unsolved, and average times for solved instances are considerably larger.

The reason for the additional difficulty is reflected in the number of iterations and the number of constraints. In the case of unit values, the overall value and agent values of a given solution are closely aligned. For non-unit values, however, solutions with many low-value transplants are often rejected by the agents, adding more constraints to the master problem. This leads to more, and more costly, iterations as the additional constraints add difficulty to the master problem.

- **Distribution over agents:** We focus on the difficult Saidman Score and larger Delorme instances, and those instances with larger agents are more difficult to solve. The instance set with the suffix “4equal”, has the most solved instances, and while the average time per instance is the second highest. For all other instances, the largest agent has half of the total pool, and we see that more instances are solved, for the instances where the remaining pairs are divided over more agents. Since larger agents have more opportunities to identify profitable rejections, the increased difficulty is as expected. This is also reflected in the number of iterations required and constraints added. In instance sets with larger agents, more constraints were needed to identify rejection-proof solutions.

The instance sets in this study cannot capture the variety of possibilities that could be encountered in real life, and the feasibility of using MAXRP-KEP in practice will depend on the specific properties of the application. This said we believe the above results are positive. Using our algorithms, MAXRP-KEP is solvable on large Delorme instances with unit valuations, most medium-size instances for the dense Saidman instances, and trivial for the medium-size Delorme instances. The strong results on Delorme instances are particularly encouraging, as they more strongly resemble real-life kidney exchange pools. The increased complexity of increased agent pool sizes might become a serious concern, especially in cooperations between large countries. Still, pools with around 50 pairs are manageable, a size which is currently only exceeded by three European Kidney Exchanges in practice (UK, Spain, and The Netherlands (Biró et al., 2019)).

### 6.3. Tiebreaker

We now discuss the tiebreaker and its effect on the computation time for the challenging SaidmanScore instances. Its effect depends on instance characteristics. Fig. 7 shows the number of instances solved optimally at each point in time and gives the clearest comparison. For the instances with one large and 2 medium or 5 small agents, the tiebreaker generally speeds up computation for easier instances, but the gap disappears for harder instances. For both instance sets, not using the tiebreaker allows for one more instance to be solved within the time limit. For the 2equal and 4equal instance sets, little difference can be seen. We do note that using the tiebreaker strongly reduces the number of iterations required to identify optimal solutions, as can be seen in Table 5. As before, this is offset by the increased difficulty of solving the master problem, now due to the harder objective function, leading to similar overall results.

## 7. Discussion

In this paper, we address the issue of cooperation in kidney exchange from a new angle. Previous work focused on incentivizing agents to contribute all of their patients and donors to a common pool. The underlying assumption of this work is that agents will strategically choose to reveal or withhold their pairs from cooperation, to maximize the value of transplants to their patients (withholding strategies). One of our contributions is to recognize a second option for agents, the rejection strategies. In Section 3, we show that such strategies provide better outcomes for the agent than existing withholding strategies with reasonable computational and

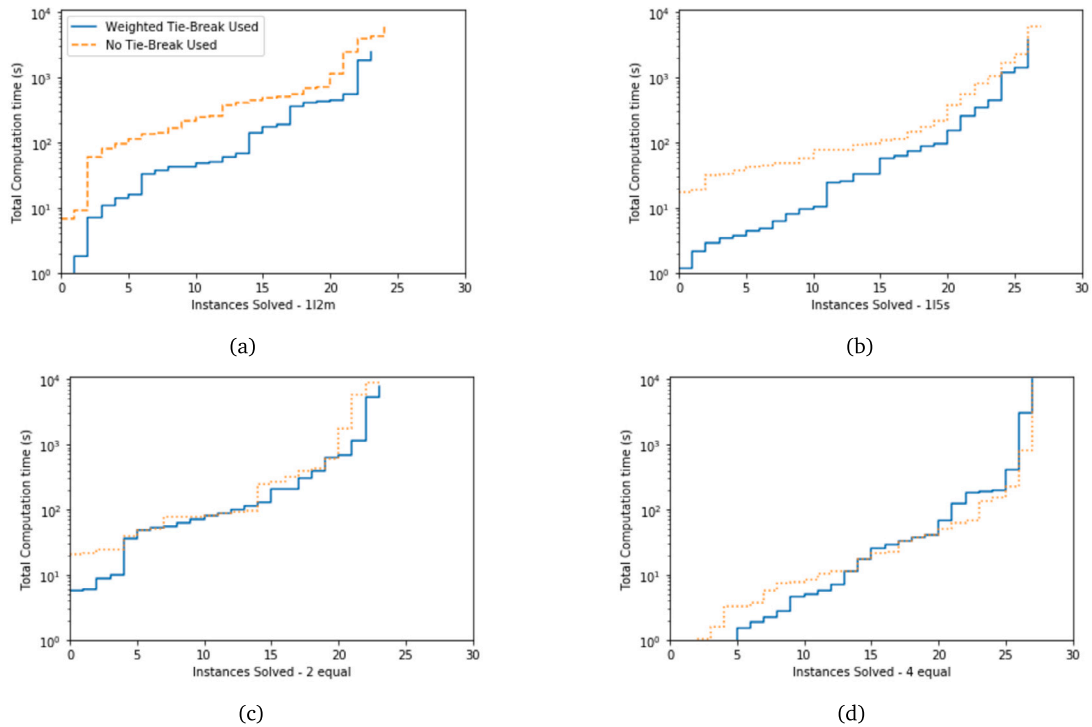


Fig. 7. Comparison of the computation times of MAXRP-KEP for different distributions of agent pool sizes, depending on the use of the tiebreaker.

information requirements. Furthermore, we show that such rejection strategies can cause an overall loss of transplants. As a result, we propose that rejection-proofness is a valuable property for cooperation mechanisms. We note that some of our computational observations are potentially influenced by solver bias, as mentioned in Section 3. Investigating whether this is actually the case, and to what extent this can be corrected is an area for future research.

We provide three different rejection-proof mechanisms, MAXRP-KEP, MAXINT, and RR. Our main mechanism, MAXRP-KEP, guarantees rejection-proofness at a very small cost in terms of number of transplants. This performance does require solving a complex bilevel problem which comes at a computational cost; however, we show this is manageable for instance sets of reasonable size. Further research into algorithms for this problem may be necessary to allow for cooperation with multiple agents with large pool sizes. One potential avenue is to make use of existing approaches to scale up the basic kidney exchange problem, such as branch-and-price approaches (Riascos-Álvarez et al., 2020; Omer et al., 2022), if they are adjusted for the subset rejection constraints. The second mechanism, MAXINT, is simple to implement, easy to understand, and close to the policy of existing international cooperations. However, it performs worst in terms of transplants lost for realistic instances. The REJECT-AND-REBUILD mechanism provides a modest improvement over agents using rejection strategies.

Common to all mechanisms is that besides their designed outcome of nullifying, for individual agents, the benefit of using rejection strategies, these mechanisms also eliminate expected benefits provided by greedy withholding strategies. We do note that they are not strategy-proof, as evidenced by the existence of instances where the greedy withholding strategy does provide a benefit.

As a final remark about the benefits of our mechanisms, we note that they provide a natural way of reconciling differences in transplant valuation in cooperations. These mechanisms make use of agent values ( $w_e^p$ ) and social welfare values ( $w_e$ ). Through the constraints ensuring rejection-proofness, agents are assured that their private valuations will be taken into account. This allows agents in a cooperation to set their priorities.

Looking forward, we believe that combinations of rejection-proof and credit mechanisms are interesting to study. Current credit mechanisms do not handle failure or rejection of proposed transplants (see, e.g., Hajaj et al. (2015); Klimentova et al. (2020); Benedek et al. (2022)). By removing incentives to reject, this issue is ameliorated. Rejection-proof mechanisms provide guarantees that agents will not lose out by cooperating, but can still allow for an unequal division of the benefits of cooperation. Furthermore, in our experiments, rejection-proof mechanisms still allow situations where greedy withholding strategies are profitable. Depending on the graph structure, there may be situations where they are consistently profitable. Credit mechanisms can reduce these issues in multi-period settings. By combining rejection-proof and credit mechanisms, the risk of strategic behavior can be further reduced.

### Declaration of competing interest

We have no conflicts of interests to disclose.

## Acknowledgments

The authors thank Péter Biró for helpful discussions. The research of Frits Spieksma is supported by Dutch Research Council (NWO) Gravitation Project NETWORKS [Grant 024.002.003].

## Appendix A. Proof of Theorem 1

We consider the computational complexity of the underlying optimization problem OPT-RP-KEP of the MAXRP-KEP mechanism. Theorem 1 claims that DEC-RP-KEP is  $\Sigma_2^P$ -complete for fixed maximum cycle length  $K \geq 3$ . This result remains to hold even for  $|\mathcal{P}| = 2$ , no non-directed donors and unit transplant valuations, i.e., for all exchanges  $e \in \mathcal{E}$ , we have

$$w_e^p = |V(e) \cap V^p| \quad \text{and} \quad w_e = \sum_{p \in \mathcal{P}} w_e^p.$$

We prove this statement by providing a reduction from ADVERSARIAL (2,2)-SAT, which is proven to be  $\Sigma_2^P$ -complete in Smeulders et al. (2022). This problem is closely related to the foundational  $\Sigma_2^P$ -complete problem,  $B_2^{\text{CNF}}$ .

### Adversarial (2,2)-SAT

*Given:* a boolean CNF expression  $E = (X, Y, C)$ , consisting of two sets  $X$  and  $Y$  of variables, and a set  $C$  of disjunctive clauses  $C$  over  $X$  and  $Y$ , such that each literal of the set  $\{z, \bar{z} \mid z \in X \cup Y\}$  occurs in exactly two clauses of  $E$ .

*Question:* does there exist a truth setting  $\theta_X : X \rightarrow \{\text{True}, \text{False}\}$  such that  $E$  is unsatisfiable, i.e., there exists no truth setting  $\theta_Y : Y \rightarrow \{\text{True}, \text{False}\}$  such that all clauses  $C$  are satisfied given  $(\theta_X, \theta_Y)$ .

We will say that a variable  $z \in X \cup Y$  occurs in sense  $\psi = t$  or  $\psi = f$  in clause  $c \in C$ , if  $c$  contains the literal  $z$  or  $\bar{z}$  respectively. Furthermore, we denote by  $c(z, i)$  and  $c(\bar{z}, i)$  the clause in which  $z$  or  $\bar{z}$  respectively occur for the  $i$ -th time, where  $i \in \{1, 2\}$ . In the following sections, we prove that

$$\text{ADVERSARIAL (2,2)-SAT} \leq_p \text{DEC-RP-KEP}.$$

First, we describe a mapping  $h$  of an ADVERSARIAL (2,2)-SAT instance  $E$  to an instance  $h(E) = \mathcal{I} = (\mathcal{E}, \mathcal{P}, t)$  of DEC-RP-KEP. Next, we construct a kidney exchange solution on the instance  $h(E)$ . Finally, we show that it is indeed rejection-proof and attains value  $t$  if and only if  $E$  is a yes-instance to ADVERSARIAL (2,2)-SAT, hence proving our polynomial reduction is valid.

### A.1. Instance mapping

Let us be given an instance  $E = (X, Y, C)$  to ADVERSARIAL (2,2)-SAT. We construct an instance  $\mathcal{I} = h(E) = (\mathcal{E}, \mathcal{P}, t)$  to DEC-RP-KEP forming a 2-agent kidney exchange program, i.e.,  $\mathcal{P} = \{b, g\}$ . Informally, we will refer to these agents as the blue and the green agent for illustrative purposes.

**$X$ -gadgets:** for each variable  $x \in X$ , we define the vertex set  $V_x := V_x^g \cup V_x^b$ , with vertex sets  $V_x^g$  and  $V_x^b$  belonging to agent  $g$  and  $b$  respectively, as follows.

$$V_x^g := \{x_0^g\} \cup \{x_{\psi,i}^g : \psi \in \{t, f\}, i \in \{1, 2\}\},$$

$$V_x^b := \{x_i^b : i \in \{1, 2\}\} \cup \{x_{c(x,\psi,i)}^b : \psi \in \{t, f\}, i \in \{1, 2\}\}.$$

The arc set  $A_x \subseteq \{(i, j) : i, j \in V_x\}$  is defined as

$$A_x := \left\{ (x_0^g, x_{\psi,1}^g), (x_{\psi,1}^g, x_{\psi,2}^g), (x_{\psi,2}^g, x_0^g) : \psi \in \{t, f\} \right\}$$

$$\cup \left\{ (x_{\psi,i}^g, x_i^b), (x_i^b, x_{c(x,\psi,i)}^b), (x_{c(x,\psi,i)}^b, x_{\psi,i}^g) : \psi \in \{t, f\}, i \in \{1, 2\} \right\}$$

The graph  $G_x = (V_x, A_x)$  will be referred to as the  $x$ -gadget or the *variable gadget* of  $x$ , and is depicted as the leftmost gadget in Fig. A.1 (within the rectangle).

**$Y$ -gadgets:** for each variable  $y \in Y$ , we obtain the vertex set  $V_y := V_y^b$ , defined by

$$V_y^b := \{y_i^b : i \in \{0, 1, 2\}\} \cup \{y_{\psi,i}^b : \psi \in \{t, f\}, i \in \{1, 2\}\} \cup \{y_{c(y,\psi,i)}^b : \psi \in \{t, f\}, i \in \{1, 2\}\}$$

Similar to the arc set  $A_x$  for each  $x$ -gadget, the arc set  $A_y \subseteq \{(i, j) : i, j \in V_y\}$  is defined as

$$A_y := \left\{ (y_0^b, y_{\psi,1}^b), (y_{\psi,1}^b, y_{\psi,2}^b), (y_{\psi,2}^b, y_0^b) : \psi \in \{t, f\} \right\}$$

$$\cup \left\{ (y_{\psi,i}^b, y_i^b), (y_i^b, y_{c(y,\psi,i)}^b), (y_{c(y,\psi,i)}^b, y_{\psi,i}^b) : \psi \in \{t, f\}, i \in \{1, 2\} \right\}$$

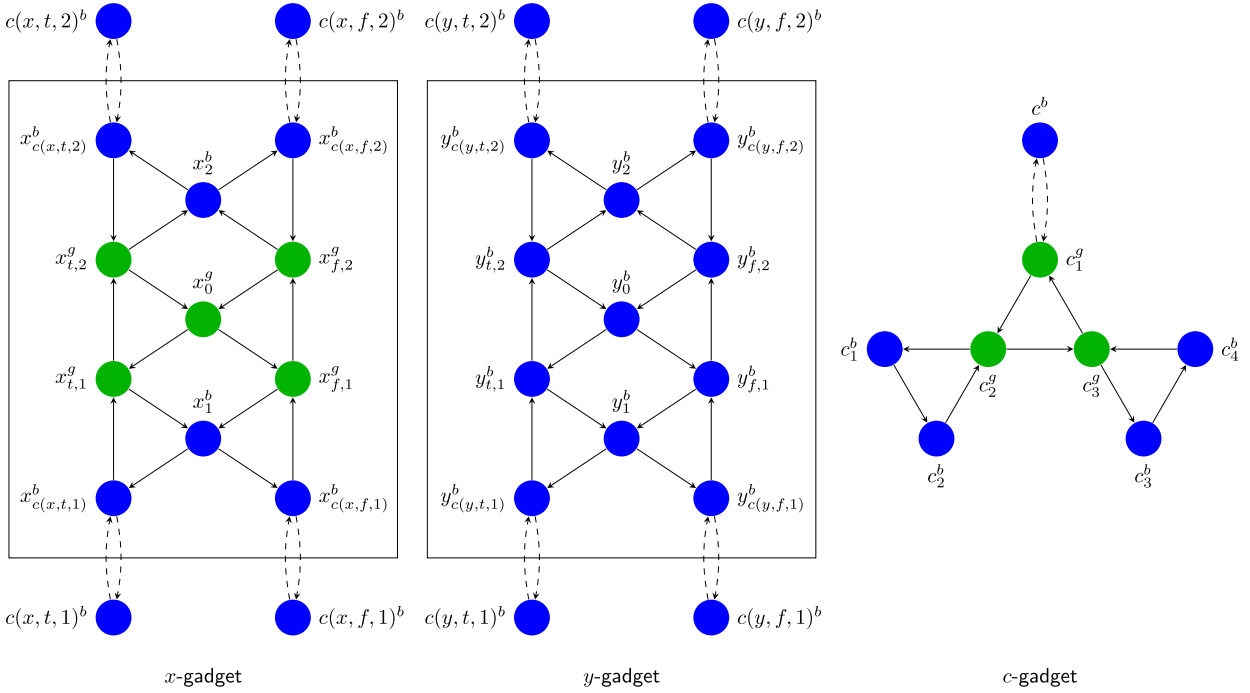


Fig. A.1. Gadgets used as building blocks for the transformed instance  $h(E)$ .

The graph  $G_y = (V_y, A_y)$  will be referred to as the  $y$ -gadget or *variable gadget* of  $y$ , and is illustrated as the middle gadget in Fig. A.1 (within the rectangle).

**Clause gadgets:** for each clause  $c \in C$ , we define the vertex set  $W_c := W_c^g \cup W_c^b$ , where the vertex sets  $W_c^g$  and  $W_c^b$  are respectively defined by

$$W_c^g := \{c_1^g, c_2^g, c_3^g\},$$

$$W_c^b := \{c^b, c_1^b, c_2^b, c_3^b, c_4^b\}.$$

The arc set  $A_c$  within the  $c$ -gadget is defined as

$$A_c := \left\{ (c_1^g, c_2^g), (c_2^g, c_3^g), (c_3^g, c_1^g) \right\} \cup \left\{ (c^b, c_1^b), (c_1^b, c^b) \right\}$$

$$\cup \left\{ (c_2^g, c_1^b), (c_1^b, c_2^b), (c_2^b, c_2^g) \right\} \cup \left\{ (c_3^g, c_3^b), (c_3^b, c_4^b), (c_4^b, c_3^g) \right\}$$

The graph  $G_c = (W_c, A_c)$  will be referred to as the  $c$ -gadget or the *clause gadget* of  $c$ . The  $c$ -gadget is illustrated as the rightmost gadget in Fig. A.1.

**Arcs connecting variable and clause gadgets:** finally, we also add arcs connecting  $x$ - and  $y$ -gadgets to  $c$ -gadgets. In particular, we define these arcs as

$$A_{X \cup Y, C} = \left\{ \left( z_{c(z, \psi, i)}^b, c(z, \psi, i)^b \right), \left( c(z, \psi, i)^b, z_{c(z, \psi, i)}^b \right) : z \in X \cup Y, \psi \in \{t, f\}, i \in \{1, 2\} \right\}$$

These arcs are depicted in Fig. A.1 as dashed arcs. The arcs  $A_{X \cup Y, C}$  induce a set of 2-cycles, which we will refer to as *clause satisfaction cycles*. Summarizing the above, we have constructed a compatibility graph  $G = (V, A)$  on which we will base our transformed instance  $h(E)$  to DEC-RP-KEP, defined by

$$V = \left( \bigcup_{x \in X} V_x \right) \cup \left( \bigcup_{y \in Y} V_y \right) \cup \left( \bigcup_{c \in C} W_c \right), \tag{20}$$

$$A = \left( \bigcup_{x \in X} A_x \right) \cup \left( \bigcup_{y \in Y} A_y \right) \cup \left( \bigcup_{c \in C} A_c \right) \cup A_{X \cup Y, C}. \tag{21}$$

The instance  $h(E) = (\mathcal{E}, \mathcal{P}, t)$  can then be formalized by selecting  $\mathcal{E}$  as the set of cycles on  $G$  of length at most 3,  $\mathcal{P} = \{b, g\}$  and  $t = 9|X| + 9|Y| + 5|C| + 3$ .

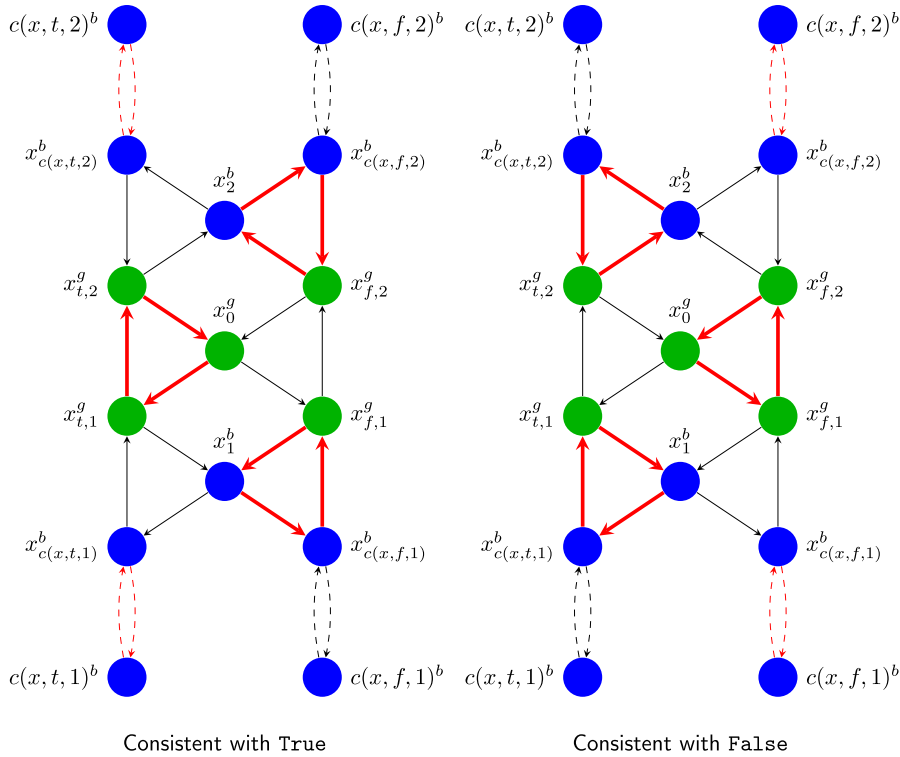


Fig. A.2. Gadget packings  $S_{x,t}$  for variable  $x \in X$  consistent with True (left) and  $S_{x,f}$  consistent with False (right). Consistent packings work similarly for variables  $y \in Y$ , where the gadget consists of blue vertices only. Notice that each red dashed clause satisfaction cycle can be associated with satisfying a clause as the truth setting of this variable implies that this clause is satisfied.

**Remark 2.** We consider the proof for the case of  $K = 3$ . The proof can be extended to all fixed  $K \geq 3$  by modifying each of the variable gadgets in the following way: subdivide the arcs  $(z_{c(z,\psi,i)}^b, z_{\psi,i}^b) \in A_z$  into a directed path of  $K - 2$  arcs, such that each of the intermediate vertices is blue, to obtain compatibility graph  $G'$ . Now, we need to consider the set  $\mathcal{E}'$  of cycles of length at most  $K$  on the graph  $G'$  and an alternative target integer  $t' = (2K + 3)(|X| + |Y|) + 5|C| + 3$ .

A.2. Constructing a rejection-proof solution

We construct a rejection-proof solution  $S$  – which is initially empty – feasible with respect to the constructed instance  $h(E)$  as follows:

1. Let  $\theta_X$  be a truth setting for  $X$  such that  $E$  becomes unsatisfiable given  $\theta_X$ . For each variable  $x \in X$ , add to  $S$  the variable gadget packing  $S_{x,\psi}$  which we refer to as being consistent with setting  $x$  to  $\psi = \theta_X(x)$ . We define the packing consistent with  $\psi$  as

$$S_{x,\psi} := \left\{ \left\{ x_0^g, x_{\psi,1}^g, x_{\psi,2}^g \right\}, \left\{ x_{\bar{\psi},1}^g, x_1^b, x_{c(x,\bar{\psi},1)}^b \right\}, \left\{ x_{\bar{\psi},2}^g, x_2^b, x_{c(x,\bar{\psi},2)}^b \right\} \right\}$$

where  $\bar{\psi}$  denotes the negation of  $\psi$ , i.e., if  $\psi = t$ , then  $\bar{\psi} = f$  and vice versa. In that regard, we can associate with each variable  $x \in X$  the packings  $S_{x,t}$  and  $S_{x,f}$ , both of which are illustrated in Fig. A.2. Similar notions can be regarded for the variable gadgets for  $y \in Y$ . This covers  $9|X|$  vertices, as each consistent packing includes three 3-cycles.

2. Select a maximum cardinality kidney exchange solution on the subgraph  $G[U^b]$  induced by the set  $U^b \subset V^b$  of blue vertices that are not yet covered by a cycle after Step 1. Notice that an example of a feasible kidney exchange solution on  $G[U^b]$  can be constructed by selecting  $\{S_{y,\theta_Y(y)} \mid y \in Y\}$  to  $S$ , where  $\theta_Y$  is a truth setting for  $Y$  satisfying the maximum number of clauses given  $\theta_X$ . We denote this quantity by  $k_{\max}(\theta_X)$ . Given such  $\theta_X$  and  $\theta_Y$ , we can select for each satisfied clause  $c \in C$  a clause satisfaction cycle associated with some variable  $z \in X \cup Y$  occurring in the clause whose truth setting implies that  $c$  is satisfied. Hence, a maximum cardinality kidney exchange solution on  $G[U^b]$  covers at least  $9|Y| + 2k_{\max}(\theta_X)$  vertices.
3. For each clause  $c \in C$ , we add to  $S$  the satisfied clause packing

$$S_{c,t} := \left\{ \left\{ c_1^g, c_2^g, c_3^g \right\} \right\}$$

if  $c^b$  is already covered by a clause satisfaction cycle (which is the case for  $k_{\max}(\theta_X)$  clauses), and the unsatisfied clause packing

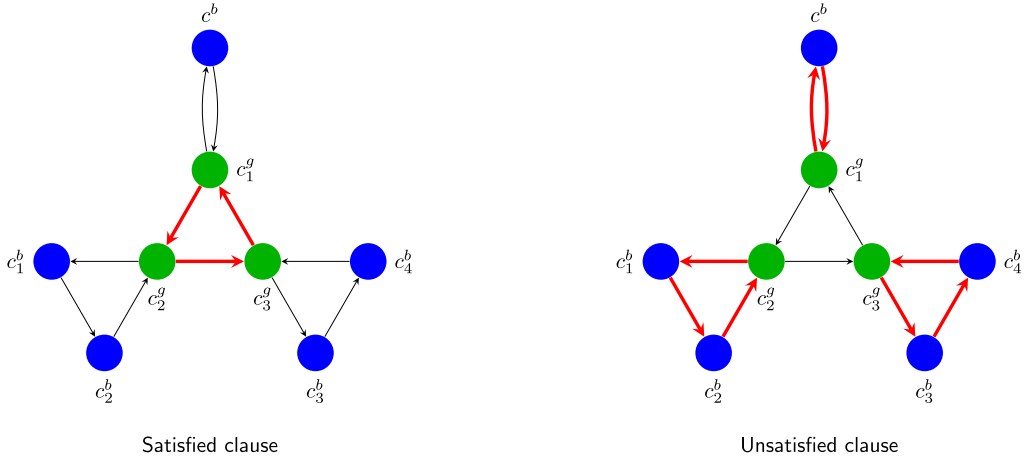


Fig. A.3. As rejection-proofness requires all three green vertices are covered, these are the only possible packings in a feasible rejection-proof solution.

$$S_{c,f} := \{ \{c_2^g, c_1^b, c_2^b\}, \{c_3^g, c_3^b, c_4^b\}, \{c_1^g, c^b\} \}$$

otherwise (which is the case for the remaining  $|C| - k_{\max}(\theta_X)$  clauses). Fig. A.3 illustrates the satisfied and unsatisfied clause packings respectively. Hence, this step covers  $3k_{\max}(\theta_X) + 8(|C| - k_{\max}(\theta_X))$  vertices.

**Lemma 3.** *An instance  $E$  is a yes-instance for ADVERSARIAL (2,2)-SAT if and only if the transformed instance  $h(E)$  is a yes-instance for DEC-RP-KEP.*

**Proof.**  $\Rightarrow$ ) Let  $E = (X, Y, C)$  be a yes-instance to ADVERSARIAL (2,2)-SAT, i.e., there exists a truth setting  $\theta_X : X \rightarrow \{\text{True}, \text{False}\}$  such that for any truth setting  $\theta_Y : Y \rightarrow \{\text{True}, \text{False}\}$  we have that  $E$  is not satisfied. Let  $\theta_X$  be such a truth setting for  $X$ , and let  $k_{\max}(\theta_X) \leq |C| - 1$  be the maximum number of clauses that can be satisfied simultaneously by a truth setting  $\theta_Y$  for  $Y$ .

It is now easy to check that we have constructed a solution  $S$  that covers at least

$$\begin{aligned} 9|X| + 9|Y| + 2k_{\max}(\theta_X) + 3k_{\max}(\theta_X) + 8(|C| - k_{\max}(\theta_X)) &= 9|X| + 9|Y| + 8|C| - 3k_{\max}(\theta_X) \\ &\geq 9|X| + 9|Y| + 5|C| + 3, \end{aligned}$$

vertices, where the last inequality follows from  $E$  being a yes-instance, i.e.,  $k_{\max}(\theta_X) \leq |C| - 1$ . In the remainder of this proof, we will denote by  $s$  a binary encoding of the solution  $S$ .

It remains to show that  $S$  is indeed rejection-proof. The transformed instance  $h(E)$  of DEC-RP-KEP contains a green and blue agent. Observe that the green agent has no incentive to reject the solution, as all of their vertices are covered by  $S$ . Hence, we can restrict ourselves to proving that the blue agent has no incentive to reject  $S$ , i.e.,

$$\forall S' \in \mathcal{R}^b(S) : z^b(S') \leq z^b(S).$$

Suppose on the contrary that there exists a rejection strategy  $S^b \in \mathcal{R}^b(S)$  feasible to RKEP( $s, b$ ) such that  $z^b(S^b) > z^b(S)$ , which rejects the minimal number of proposed shared cycles. This minimality principle will play an important role for us to obtain a contradiction. Notice that the only shared exchanges worth rejecting are either those of the form

- (i)  $\{x_{\psi,i}^g, x_i^b, x_{c(x,\psi,i)}^b\} \in S \setminus S^b$  for some  $x \in X, \psi \in \{t, f\}, i \in \{1, 2\}$  (internal to some  $x$ -gadget), or
- (ii)  $\{c^b, c_i^g\}$  for some  $c \in C$ .

However, there cannot be any such cycles of type (ii), as in that case the blue agent would not have selected a maximum cardinality internal solution in Step 2 of the construction algorithm. Hence, any profitable rejection strategy must reject some cycle of the form  $\{x_{\psi,i}^g, x_i^b, x_{c(x,\psi,i)}^b\} \in S \setminus S^b$  where  $\psi \in \{t, f\}$  and  $i \in \{1, 2\}$ , i.e., the blue agent rejects some proposed shared cycle within a gadget of some variable  $x \in X$ .

Without loss of generality, suppose  $S^b$  rejects a shared exchange of type (i) for some  $x \in X, \psi = t$ , and  $i = 1$ . Then,  $S^b$  must contain some clause satisfaction cycle

$$\{x_{c(x,t,1)}^b, c(x, t, 1)^b\}.$$

However, the alternative solution  $\hat{S}^b$  obtained from  $S^b$  from removing the 2-cycle and adding the rejected type (i) exchange is still feasible to RKEP( $s, b$ ), i.e.,  $\hat{S}^b \in \mathcal{R}^b(S)$  and  $z^b(\hat{S}^b) = z^b(S^b) > z^b(S)$ . As  $\hat{S}^b$  rejects fewer shared exchanges than  $S^b$ , this contradicts



the assumed minimality of  $S^b$ . Hence, also the blue agent has no incentive to reject the constructed solution  $S$ , which is therefore a rejection-proof kidney exchange solution covering at least  $t = 9|X| + 9|Y| + 5|C| + 3$  vertices. Therefore,  $S$  is a valid certificate showing that  $h(E)$  is a yes-instance to DEC-RP-KEP.

⇐) Let us consider a feasible solution  $S$  to DEC-RP-KEP for the transformed instance that covers at least  $9|X| + 9|Y| + 5|C| + 3$  vertices. Without loss of generality, we can suppose that in  $S$ , all gadgets are packed consistent with some truth settings  $\theta_X$  and  $\theta_Y$  (observe that for each variable gadget, switching from a clause satisfaction cycle to its unique intersecting 3-cycle within the variable gadget preserves rejection-proofness). Remark that there must exist a clause  $c \in C$  for which the  $c$ -gadget is packed with  $S_{c,f}$ , since any rejection-proof kidney exchange solution would otherwise cover at most  $9|X| + 9|Y| + 2|C| + 3|C| = 9|X| + 9|Y| + 5|C|$  vertices. If  $E$  would now be a no-instance to ADVERSARIAL (2,2)-SAT, then  $k_{\max}(\theta_X) = |C|$  for any truth setting  $\theta_X : X \rightarrow \{\text{True}, \text{False}\}$ . Hence, for each  $\theta_X$  there must exist some truth setting  $\theta_Y$  satisfying all clauses. However, this truth setting would imply a profitable rejection strategy  $S^b \in \mathcal{R}^b(S)$  for the blue agent. This rejection strategy  $S^b$  is obtained from  $S$  by switching to packings consistent with  $\theta_Y(y)$  for each  $y \in Y$  and adding exactly  $|C|$  disjoint clause satisfaction cycles. As  $S^b$  contains at least one more clause satisfaction cycle compared to  $S$ , we have that  $z^b(S^b) \geq z^b(S) + 1$ , contradicting rejection-proofness of  $S$ . We can conclude from this that the truth setting  $\theta_X$  is a valid certificate for  $E = (X, Y, C)$  being a yes-instance to ADVERSARIAL (2,2)-SAT.  $\square$

**Theorem 1.** DEC-RP-KEP is  $\Sigma_2^P$ -complete for any fixed maximum cycle length  $K \geq 3$ , even in the case of  $|\mathcal{P}| = 2$ , transplants being valued equally and no non-directed donors.

**Proof.** Hardness was proven in Lemma 3. We continue by showing membership of DEC-RP-KEP in the class  $\Sigma_2^P$ . Following the definition of  $\Sigma_2^P$  in Arora and Barak (2009), we consider DEC-RP-KEP as the language given by

$$\{ \langle \mathcal{E}, \mathcal{P}, t \rangle : \exists \text{ RP kidney exchange solution } X \in \mathcal{X} \subset 2^{\mathcal{E}} \text{ with } z(X) = \sum_{e \in X} w(e) \geq t \}$$

Remark that this language is in  $\Sigma_2^P$  since a tuple  $\langle \mathcal{E}, \mathcal{P}, t \rangle \in \text{DEC-RP-KEP}$  if and only if

$$\exists X \in \mathcal{X} : z(X) \geq t$$

$$\forall X' \in \mathcal{X} : X' \notin \mathcal{R}^p(X) \text{ or } z^p(X') \leq z^p(X), \text{ for all } p \in \mathcal{P}.$$

Remark that there exist compact encodings of  $X, X'$ , i.e. polynomial in the number of vertices and arcs of the graph  $G$ , such as arc-based formulations of kidney exchange programs. Furthermore, it can be checked in polynomial time whether such encodings indeed imply feasible cycle-chain packings. This also holds for checking if  $X' \in \mathcal{R}^p(X)$  and checking the value  $z^p(X')$ . Therefore, given a rejection-proof solution  $X$  and some alternative vector  $X'$ , we can check in polynomial time that  $X'$  does not prove a violation of rejection-proofness of  $X$ , hence the language DEC-RP-KEP is in  $\Sigma_2^P$ .  $\square$

## References

- Agarwal, Nikhil, Ashlagi, Itai, Azevedo, Eduardo, Featherstone, Clayton R., Karaduman, Ömer, 2020. Market failure in kidney exchange. *Am. Econ. Rev.* 109 (11), 4026–4070.
- Arora, Sanjeev, Barak, Boaz, 2009. *Computational Complexity: A Modern Approach*. Cambridge University Press.
- Ashlagi, Itai, Roth, Alvin E., 2014. Free riding and participation in large scale, multi-hospital kidney exchange. *Theor. Econ.* 9 (3), 817–863.
- Ashlagi, Itai, Fischer, Felix, Kash, Ian A., Procaccia, Ariel D., 2015. Mix and match: a strategyproof mechanism for multi-hospital kidney exchange. *Games Econ. Behav.* 91, 284–296.
- Beck, Yasmine, Ljubić, Ivana, Schmidt, Martin, 2023. A survey on bilevel optimization under uncertainty. *Eur. J. Oper. Res.*
- Benedek, Márton, Biró, Péter, Kern, Walter, Paulusma, Daniël, 2022. Computing balanced solutions for large international kidney exchange schemes. In: *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS '22*. Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450392136, pp. 82–90.
- Bialas, Wayne F., Karwan, Mark H., 1984. Two-level linear programming. *Manag. Sci.* 30 (8), 1004–1020. <https://EconPapers.repec.org/RePEc:nm:ormnsc:v:30:y:1984:i:8:p:1004-1020>.
- Biró, Péter, Haase-Kromwijk, Bernadette, Andersson, Tommy, Ásgeirsson, Eyjólfur Ingi, Baltessová, Tatiana, Boletis, Ioannis, Bolotinha, Catarina, Bond, Gregor, Böhmig, Georg, Burnapp, Lisa, et al., 2019. Building kidney exchange programmes in Europe—an overview of exchange practice and activities. *Transplantation* 103 (7), 1514.
- Biró, Péter, Gyetvai, Márton, Klimentova, Xenia, Pedro Pedroso, João, Pettersson, William, Viana, Ana, 2020. Compensation scheme with Shapley value for multi-country kidney exchange programmes. (Accessed: 06 December 2018).
- Böhmig, Georg A., Fronek, Jiří, Slavcev, Antonij, Fischer, Gottfried F., Berlakovich, Gabriela, Viklicky, Ondrej, 2017. Czech-Austrian kidney paired donation: first European cross-border living donor kidney exchange. *Transpl. Int.* 30 (6), 638–639.
- Bolusani, Suresh, Ralphs, Ted K., 2022. A framework for generalized Benders' decomposition and its application to multilevel optimization. *Math. Program.*, 1–38.
- Carvalho, Margarida, Lodi, Andrea, 2023. A theoretical and computational equilibria analysis of a multi-player kidney exchange program. *Eur. J. Oper. Res.* (ISSN 0377-2217) 305 (1), 373–385. <https://doi.org/10.1016/j.ejor.2022.05.027>.
- Carvalho, Margarida, Lodi, Andrea, Pedroso, Joao Pedro, Viana, Ana, 2017. Nash equilibria in the two-player kidney exchange game. *Math. Program.* 161 (1), 389–417.
- Carvalho, Margarida, Lodi, Andrea, Pedroso, João P., 2022. Computing equilibria for integer programming games. *Eur. J. Oper. Res.* 303 (3), 1057–1070.
- Delorme, Maxence, Garcia, Sergio, Gondzio, Jacek, Kalcsics, Jörg, Manlove, David F., Pettersson, William, Trimble, James, 2022. A new matheuristic and improved instance generation for kidney exchange programmes. *Comput. Oper. Res.* 141.
- DeNegre, Scott T., Ralphs, Ted K., 2009. A branch-and-cut algorithm for integer bilevel linear programs. In: Chinneck, John W., Kristjansson, Bjarni, Saltzman, Matthew J. (Eds.), *Operations Research and Cyber-Infrastructure*. Springer US, Boston, MA. ISBN 978-0-387-88843-9, pp. 65–78.

- Farnadi, Golnoosh, St-Arnaud, William, Babaki, Behrouz, Carvalho, Margarida, 2021. Individual fairness in kidney exchange programs. In: Proceedings of the AAAI Conference on Artificial Intelligence, 35 (13), 11496–11505.
- Hajaj, Chen, Dickerson, John P., Hassidim, Avinatan, Sandholm, Tuomas, Sarne, David, 2015. Strategy-proof and efficient kidney exchange using a credit mechanism. In: Twenty-Ninth AAAI Conference on Artificial Intelligence.
- Jeroslow, Robert G., 1985. The polynomial hierarchy and a simple model for competitive analysis. *Math. Program.* 32 (2), 146–164.
- Johannes, Berit, 2011. New Classes of Complete Problems for the Second Level of the Polynomial Hierarchy. Technical report, PhD thesis. TU Berlin.
- Kleinert, Thomas, Labbé, Martine, Ljubić, Ivana, Schmidt, Martin, 2021. A survey on mixed-integer programming techniques in bilevel optimization. *EURO J. Comput. Optim.* 9, 100007.
- Klimentova, Xenia, Viana, Ana, Pedroso, João Pedro, Santos, Nicolau, 2020. Fairness models for multi-agent kidney exchange programmes. *Omega*, 102333.
- Lodi, Andrea, Ralphs, Ted, Woeginger, Gerhard, 2014. Bilevel programming and the separation problem. *Math. Program.* 146 (1–2), 437–458.
- Lozano, Leonardo, Smith, J. Cole, 2017. A value-function-based exact approach for the bilevel mixed-integer programming problem. *Oper. Res.* 65 (3), 768–786.
- Melcher, Marc L., Blosser, Christopher D., Baxter-Lowe, Lee Ann, Delmonico, Francis L., Gentry, Sommer E., Leishman, Ruthanne, Knoll, Gregory A., Leffell, Mary Sue, Leichtman, Alan B., Mast, Debbie A., et al., 2013. Dynamic challenges inhibiting optimal adoption of kidney paired donation: findings of a consensus conference. *Am. J. Transplant.* 13 (4), 851–860.
- Mincu, Radu-Stefan, Biró, Péter, Gyetvai, Márton, Popa, Alexandru, Verma, Utkarsh, 2021. IP solutions for international kidney exchange programmes. *Cent. Eur. J. Oper. Res.* 29, 403–423.
- Moore, James T., Bard, Jonathan F., 1990. The mixed integer linear bilevel programming problem. *Oper. Res.* 38 (5), 911–921. ISSN 0030364X, 15265463. <http://www.jstor.org/stable/171050>.
- Omer, Jérémy, Arslan, Ayse N., Yan, Fulin, 2022. Kidneyexchange.jl: a Julia package for solving the kidney exchange problem with branch-and-price.
- Rapaport, Felix, 1986. The case for a living emotionally related international kidney donor exchange registry. *Transplant. Proc.* (ISSN 0041-1345) 18 (3, Suppl. 2), 5–9. <http://europepmc.org/abstract/MED/11649919>.
- Riascos-Álvarez, Lizeth Carolina, Bodur, Merve, Aleman, Dionne M., 2020. A branch-and-price algorithm enhanced by decision diagrams for the kidney exchange problem. Preprint. [arXiv:2009.13715](https://arxiv.org/abs/2009.13715).
- Roth, Alvin E., Sönmez, Tayfun, Ünver, M. Utku, 2004. Kidney exchange. *Q. J. Econ.* 119 (2), 457–488.
- Roth, Alvin E., Sönmez, Tayfun, Ünver, M. Utku, 2005. Pairwise kidney exchange. *J. Econ. Theory* 125 (2), 151–188.
- Saidman, Susan L., Roth, Alvin E., Sönmez, Tayfun, Ünver, M. Utku, Delmonico, Francis L., 2006. Increasing the opportunity of live kidney donation by matching for two- and three-way exchanges. *Transplantation* 81 (5), 773–782.
- Smeulders, Bart, Blom, Danny A.M.P., Spieksma, Frits C.R., 2022. Identifying optimal strategies in kidney exchange games is  $\Sigma_2^P$ -complete. *Math. Program.*, 1–22.
- Sönmez, Tayfun, Ünver, M. Utku, 2013. Market design for kidney exchange. In: Vulkan, Nir, Roth, Alvin E., Neeman, Zvika (Eds.), *The Handbook of Market Design*. Oxford University Press, Oxford.
- Stockmeyer, Larry J., 1976. The polynomial-time hierarchy. *Theor. Comput. Sci.* (ISSN 0304-3975) 3 (1), 1–22. [https://doi.org/10.1016/0304-3975\(76\)90061-X](https://doi.org/10.1016/0304-3975(76)90061-X). <https://www.sciencedirect.com/science/article/pii/030439757690061X>.
- Tahernejad, Sahar, Ralphs, Ted K., DeNegre, Scott T., 2020. A branch-and-cut algorithm for mixed integer bilevel linear optimization problems and its implementation. *Math. Program. Comput.* 12 (4), 529–568.
- Valentín, María Oliva, García, Marta, Costa, Alessandro Nanni, Bolotinha, Catarina, Guirado, Lluís, Vistoli, Fabio, Breda, Alberto, Fiaschetti, Pamela, Dominguez-Gil, Beatriz, 2019. International cooperation for kidney exchange success. *Transplantation* 103 (6), e180–e181.
- von Stackelberg, Heinrich, 1952. *The Theory of the Market Economy*. William Hodge.
- Woeginger, Gerhard J., 2021. The trouble with the second quantifier. *4OR*, 1–25.