Short Trading Cycles: Kidney Exchange with Strict Ordinal Preferences

Ivan Balbuzanov*

JOB MARKET PAPER

November 12, 2014

Abstract

I study the problem of kidney exchange under strict ordinal preferences and with constraints on the length of the trading cycles. The assumption of strict ordinal preferences, which is a departure from the traditional assumption that all compatible kidneys are perfect substitutes for each other, allows the mechanism I propose to take advantage of the welfare-relevant information that strict preferences carry. Additionally, individual rationality in this setting incentivizes patient-donor pairs who are compatible with each other to participate in the kidney exchange, thus increasing the match rate for incompatible pairs. I show that deterministic mechanisms have poor properties in this environment. Instead, I explicitly define an individually rational, ordinally efficient and anonymous random mechanism for the case of pairwise kidney exchange. I then extend the idea behind this mechanism to arrive at a constrained ordinally efficient mechanism no matter what the ex-post constraints on the outcome are, including individual rationality, limits on the cycle lengths, maximizing the number of proposed transplantations etc. Several mechanisms from the existing literature are special cases of this mechanism. Finally, I show that individual rationality, ex-post efficiency and weak strategyproofness are incompatible for the cycle-constrained case making the proposed mechanism a second-best mechanism.

Keywords: random assignment; kidney exchange; probabilistic serial mechanism; individual rationality; ordinal efficiency; anonymity; housing market.

*University of California, Berkeley, Department of Economics, 530 Evans Hall #3880, Berkeley, CA 94720-3880. Email: `isbalbuz@econ.berkeley.edu`. Website: `https://sites.google.com/site/ibalbuzanov/` (the most recent version of this paper is available at my website). I am grateful to my advisor, Haluk Ergin, for his continual encouragement, support and guidance throughout this project. I also thank David Ahn, Yuichiro Kamada and Chris Shannon for excellent discussions that have helped improve the paper. This paper has also benefited from comments and suggestions by Itai Ashlagi, Eric Auerbach, Aaron Bodoh-Creed, Eric Budish, Christopher P. Chambers, Yeon-Koo Che, Loren Chen, Julien Combe, Satoshi Fukuda, Axel Gottfries, Fuhito Kojima, C. Matthew Leister, Antonio Miralles, John Mondragon, Michèle Müller-Itten, Takeshi Murooka, Paulo Natenzon, Omar Nayeem, Aniko Öry, Alex Teytelboym, Emilia Tjernström, M. Utku Ünver, Xavier Vives, and seminar participants at UC Berkeley. The research was partly supported by NSF grant SES-1227707, which I gratefully acknowledge.

1 Introduction

As of September 2014, the number of patients with end-stage kidney disease and in need of a kidney transplant exceeded 100,000 on the US waiting list alone. The number has been growing for years. Annually, more than 4,000 patients die while waiting and thousands more become too sick to receive a transplant and have to withdraw from the list (see, for example, The New York Times Editorial Board 2014). One of the methods attempted to alleviate the severe kidney shortage, has been the creation of living-donor kidney-exchange clearinghouses (Roth et al. 2004, 2005a, 2007), which allow otherwise incompatible patient-donor pairs to trade kidneys amongst themselves.[1]

The United States National Organ Transplantation Act (NOTA) of 1984 forbids the transfer of human organs for "valuable considerations."[2] This has two important consequences for kidney exchange. First, it is illegal to use the price mechanism to guarantee that the post-exchange allocation is socially optimal. Instead, it falls to the existing clearinghouses, which are in charge of kidney exchanges, to organize efficient trading. Second, NOTA's provisions make it impossible for donors to contractually commit to donate a kidney so kidney exchanges need to be performed simultaneously to avoid donors backing out of a promised donation. Thus any exchange with k patient-donor pairs requires k donor nephrectomies (i.e., kidney removals) and k transplantations, each of which requires an operating theatre and a surgical team working simultaneously. This constraint creates significant logistical challenges which in practice limit the number of pairs who could participate in each exchange. Since each exchange takes the form of a trading cycle, where the first donor donates a kidney to the second patient, whose donor donates to the third patient and so on, until a donor closes the cycle by donating a kidney to the first patient, the length of these trading cycles, as measured by the number of pairs in them, cannot be too large.

With this motivation, I study the problem of object exchange without monetary transfers (Shapley and Scarf 1974), with strict ordinal preferences and with constraints on the length of the trading cycles. The assumption of strict ordinal preferences is the main departure from the existing kidney-exchange literature which is based on the assumption of binary preferences, as initially postulated by Roth et al. (2005a), so that all compatible kidneys are viewed as perfect substitutes from the point of view of the transplant patient.[3] There is mounting evidence in the transplantation literature, however, that a variety of factors beyond simple compatibility can impact the short- and long-term survival rates of kidney grafts, including, for example, age and sex.[4] Thus reducing the problem to simple dichotomous compatibility-based preferences disposes of some welfare-relevant information. Additionally, an individually rational mechanism that takes strict preferences into account can also induce the participation of patients who are compatible with their related donor. This would greatly increase the transplantation rates for incompatible pairs: for example, Gentry et al. (2007) estimate that the rate would almost double. Finding a suitable mechanism in this environment of great real-life interest has been an open problem until now.

After I show that deterministic mechanisms in this setting have poor properties, my first main result is to propose a random mechanism that satisfies the following three properties. Firstly, it is individually rational. This guarantees compatible transplantations and the participation of compatible pairs. Secondly,

[1]Kidney-exchange clearinghouses have been organized in the US (Wallis et al. 2011), UK (Manlove and O'Malley 2012), the Netherlands (Keizer et al. 2005; De Klerk et al. 2005), South Korea (Park et al. 1999, 2004), Romania (Lucan 2007), Portugal, Australia, New Zealand, Canada, Spain (Constantino et al. 2013) and others.

[2]The buying and selling of kidneys is forbidden almost everywhere in the world (Roth 2007).

[3]Efficiency in this case reduces to maximizing the number of exchanges performed. While the main efficiency criterion in this paper is Pareto efficiency defined with respect to the patients' preferences over kidneys, my most general result, Proposition 8, allows the addition of other criteria that are ex-post desirable, including maximum-cardinality matchings.

[4]I discuss this further in Section 2.

it is ordinally efficient, where ordinal efficiency is the natural form of efficiency for random environments with ordinal preferences (Bogomolnaia and Moulin 2001). Thirdly, it is fair where fairness is represented by anonymity/name-invariance. I explicitly define the mechanism for the kidney-exchange setting, where each trade can involve no more than two pairs. I call it the *2-cycle probabilistic serial* (2CPS) mechanism, as it is based on Bogomolnaia and Moulin's (2001) probabilistic serial (PS) mechanism.[5] The mechanism can be extended to a setting including a social endowment of kidneys such as kidneys coming from deceased or altruistic living donors. I also show that no mechanism can simultaneously satisfy an arbitrary cycle constraint, individual rationality, ordinal efficiency, and also guarantee that agents truthfully report their preferences, which, I argue, is the least important desideratum in the setting of kidney exchange.[6]

For any preference-profile, the 2CPS mechanism selects a lottery over deterministic allocations, in each of which all trades satisfy individual rationality and the cycle-length cap. The cycle-length constraint and individual rationality limit the set of deterministic allocations on which a random allocation could place positive probability. I then consider the general case of arbitrary ex-post constraints. My second main result is to show how the PS mechanism can be adapted to account for any possible such constraints while maintaining (constrained) ordinal efficiency and (if the constraints allow it) anonymity. This result is striking since it includes not only the setting of object exchange with caps on the cycle length but also more general object-exchange and object-allocation problems, as well as two-sided matching markets such as school-choice problems or the assignment of medical interns to hospitals. Moreover, any of these markets can include arbitrary constraints, including not only cycle-length caps and individual rationality but also requiring maximum-cardinality matchings, quotas, regional caps, the assignment of couples in close proximity to each other in job-placement services, and other constraints motivated by geographical or diversity considerations.

Another reason to consider strict preferences is related to the participation of compatible patient-donor couples in kidney exchange programs. Even if the differences in the graft survival and rejection rates between different kidneys are of secondary importance to actually receiving a kidney, current mechanisms based on the binary-preference assumption do not provide incentives for compatible pairs to enroll in kidney-exchange programs. Under these mechanisms, the patient from a compatible pair would always be guaranteed to receive a compatible kidney but the kidney she might end up receiving can have a worse expected outcome than the one from her donor. Adding to that the waiting time and other extra costs associated with enrolling in a kidney exchange, compatible pairs are unlikely to want to participate. However, their involvement can increase the matching rate among incompatible pairs, significantly improving efficiency: simulations in Gentry et al. (2007) suggest that the match rate for incompatible pairs would double if known compatible pairs participate.[7] Imposing individual rationality in my setting provides incentives for compatible patient-donor pairs to enroll by guaranteeing the patient a kidney that is at least as good for her as her donor's kidney. Notice that it's individually rational for compatible patient-donor to participate in an ex-post way: compatible pairs would never want to back out of a proposed kidney exchange.

I begin by showing that deterministic mechanisms do not perform very well in my environment. Namely, the mechanism desiderata of basic fairness (represented by anonymity/name-invariance) and efficiency are

[5] It is notable that this setting is equivalent to the well-known roommate problem (Gale and Shapley 1962). Thus the 2CPS mechanism is also a general roommate-problem solution with desirable properties.

[6] Previous models of kidney exchange have been concerned with strategyproofness but blood testing for kidney exchange has become more standardized and centralized over time, which has made misreporting of compatibility-induced preferences harder. See Ashlagi and Roth (2014).

[7] See also Roth et al. (2005b) for a similar estimate. In line with these recommendations, the National Kidney Registry in the US is actively trying to recruit compatible pairs for participation in their program. There is still, however, a debate in the medical literature on the ethics of allowing compatible pairs to participate in kidney exchange. See Sönmez and Ünver (2014) for relevant references and a discussion.

incompatible with one another, as are individual rationality, efficiency and strategyproofness. I then examine the performance of Gale's Top Trading Cycles (TTC) in my setting since TTC is the solution with the best properties in the absence of cycle constraints.[8] If we assume that patients' preferences are drawn from a uniform distribution over the space of preference profiles, TTC fails to satisfy the cycle constraints with probability approaching 1 as the number of patient-donor pairs diverges to infinity. If this weren't the case, TTC would have made a good solution: it would have selected a matching that is cycle-constraint compliant with positive probability and, due to TTC's desirable properties, the cost of the occasional long cycle might have been acceptable. However, the result suggests that TTC would not work, even approximately, in my setting.

Shifting attention to random mechanisms, I show that a couple of likely candidates for a suitable mechanism fail to satisfy ordinal efficiency. Instead, I propose the 2CPS mechanism, which is individually rational (ensuring compatible transplantations), ordinally efficient, and anonymous, which I argue are the most important desiderata in this setting.

The 2CPS mechanism is based on a simultaneous-eating algorithm. The algorithm treats all kidneys *as if* they are infinitely divisible and all agents *as if* they are claiming larger and larger shares from the donors' kidneys in continuous time starting with their most preferred kidney. The algorithm ends when all the kidneys have been completely claimed or, equivalently, when all patients have one unit of kidney shares. Then, for any patient i and kidney from donor j, we treat the share that patient i has claimed from kidney j as the probability with which patient i receives kidney j. Thus, at the end of the algorithm, there is an associated probability-share matrix M, where $M(i, j)$ denotes the probability that patient i receives kidney j. If $i = j$, we interpret the corresponding diagonal matrix entries to denote the probability that patient i does not participate in an exchange with other pairs. Since the kidney of any donor i has to be assigned to exactly one patient (potentially patient i) and each patient i receives exactly one kidney (potentially kidney i), these relationships have to be true in expectation so we must have

$$\sum_j M(i,j) = 1 \text{ and } \sum_i M(i,j) = 1$$

for each i and j. Matrices that satisfy these conditions are called *bistochastic*.

Not all bistochastic matrices represent valid lotteries over deterministic matchings, however. To see that, consider a simple problem with three patient-donor pairs, numbered 1 through 3, participating in a paired kidney exchange. That is to say, all exchanges are limited to including no more than two patient-donor pairs. First, consider the matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

P is not permissible in my setting since it represents a probability-one three-way exchange where donor 1 donates to patient 3, whose willing donor donates to patient 2, whose donor donates to patient 1.[9] It is not hard to see, however, that, with two-pair trading cycles, the probability with which patient i receives kidney j must equal the probability that patient j receives kidney i. Thus any matrix that represents a lottery over permissible trades must be symmetric. So for the case of paired kidney exchange we need to consider only

[8] For example, TTC selects the unique allocation in the core (Roth and Postlewaite 1977) and is the unique mechanism that is individually rational, Pareto efficient, and strategyproof (Ma 1994).

[9] Furthermore, it can be shown that P is a possible outcome of Bogomolnaia and Moulin's (2001) Probabilistic Serial mechanism, which means that that mechanism is not directly applicable in my setting.

symmetric bistochastic matrices. But restricting our attention to symmetric bistochastic matrices is also not enough. To see this, consider the matrix

$$Q = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

While Q is a symmetric bistochastic matrix, it also does not represent a lottery over two-pair exchanges. To see that, note that the probability that a patient is left unmatched under Q is zero since the trace of Q is zero but in any exchange at least one patient is left unmatched. So this is impossible. In contrast, the matrix R, defined by

$$R = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

represents the lottery that places probability $1/3$ on each two pairs trading. The adaptation of the PS mechanism in my setting needs to output a matrix that represents a permissible lottery like R. I accomplish this by using a corollary of the celebrated Edmonds' characterization of the matching polytope (Edmonds 1965), which gives sufficient and necessary conditions for a bistochastic matrix to be the representation of a lottery over two-way exchanges. The 2CPS is constructed to guarantee that each of the conditions from Edmonds' Theorem are satisfied for each interim probability-share matrix, which in turn guarantees that those conditions are satisfied for the final matrix as well.

I next consider the the case of general ex-post restrictions on the mechanism, including, but not limited to, having a cap on the number of pairs in each exchange and requiring individual rationality. I show how to construct a set of constraints on the probability-share matrices that guarantees the following two things. First, the adaptation of the PS mechanism under those constraints selects a bistochastic matrix that represents a lottery over allowable deterministic allocations. Second, that lottery is constrained ordinally efficient. The conditions in question are the minimal set of constraints defining a naturally-defined bounded convex polytope and are in theory computable given a list of allowable ex-post allocations.

Note that the ex-post restrictions can include requiring the maximum possible number of recommended transplantations. Adding this restriction can be viewed as a refinement of the efficiency criterion used in this paper since it also ensures the ex-post satisfaction of the binary-preference efficiency criterion. The reason for this is that, as Roth et al. (2005a) show, efficiency for binary preferences reduces to maximizing the number of proposed exchanges. A simple kidney-exchange problem with three patient-donor pairs, numbered 1 through 3, illustrates this point. Assume that the maximum cycle length is at least 3 so it does not represent a binding constraint. Let's assume that all three patients are compatible with the other two patients' donors. Also, let patients 1 and 2 find each other's donors to be most preferable. Then the kidney exchange which matches pairs 1 and 2 is Pareto optimal. However, it might be preferable to match all three patient-donor pairs in a cycle of length 3 to avoid leaving patient 3 without a kidney. It is not hard to verify that the two possible three-way exchanges are also Pareto optimal.

This is the first paper that proposes a variant of the PS mechanism for an object-exchange or an agent-matching setting, such as the kidney exchange and the roommate problems.[10] Additionally, my proof of the

[10] Yılmaz (2010) considers a modification of the PS mechanism, where the objects to be allocated are a mix of social and private endowment. If all objects are part of the private endowment, his model can be viewed as a model of pure exchange.

ordinal efficiency of the mechanisms here is simpler than the proofs used in the preceding literature (e.g. Bogomolnaia and Moulin 2001; Budish et al. 2013), which are indirect and rely on the characterization of ordinal efficiency via the acyclicity of a certain relation. The proof I offer is more direct, can be used to simplify the analogous proofs in the existing literature, and allows me to prove the ordinal efficiency of the most general class of mechanisms in my setting.

Even though I adopt the language of kidney exchange for the rest of this paper with patient-donor pairs being the main agents, the theory developed here can find applications in other settings. For example, the setup in my paper permits a patient to have multiple willing donors. A salient example from a different sphere of life can be found in housing exchanges: be it public housing, on-campus housing, offices within a company or an academic department, and prison cells and other rooms in institutional living facilities. For example, in the case of public housing, all trades should be performed simultaneously to avoid inefficiencies associated with some families remaining homeless, being forced into a short-term rental, having to move twice for the same trade, or occupying the same unit simultaneously with another family. Since the difficulty of finding a moving date that works for everyone involved in the trade increases with the number of agents involved, it might be infeasible to perform trades with long exchange cycles.[11]

In a similar vein, there are online platforms offering the possibilities of members exchanging real estate (GoSwap.org, DaytonaHomeTrader.com) or vacation rentals (Intervac-HomeExchange.com, HomeExchange.com, ExchangeHolidayHomes.com). In addition to homes and land, GoSwap.org also offers the possibilities of exchanging various vehicles (including planes and boats), as well as businesses. Thus far exchanges are restricted to two-way swaps. Other platforms with similar constraints include barter exchanges for shoes (The National Odd Shoe Exchange; oddshoe.org), books (ReadItSwapIt, www.readitswapit.co.uk), and used goods (Netcycler). More generally, the structure in this paper can be applied to any object-exchange setting where trades are difficult or expensive to carry through and thus might require significant coordination. Essentially, in any setting where trades cannot be executed by a centralized clearinghouse by collecting all the objects and then redistributing them among the recipients, high coordination costs could limit maximum trading-cycle length. Of course, as noted above, the most general class of mechanisms I propose can apply to virtually any matching problem, be it one- or two-sided ones with arbitrary constraints on the possible outcomes.

1.1 Literature Review

This paper is part of a burgeoning matching literature.[12] More precisely, it is situated at the intersection of kidney exchange[13] and random matching. The first economic study of kidney exchange (Roth et al. 2004) proposed a modification of Gale's Top Trading Cycles mechanism in a setting with strict ordinal preferences but without cycle constraints. Subsequent work (starting with Roth et al. (2005a)) has accounted for the cycle constraints but assumes dichotomous preferences. As noted above, my work combines the two approaches: I study the kidney-exchange problem with strict ordinal preferences subject to cycle constraints. The only

However, his version of the PS mechanism offers little benefit over Gale's Top Trading Cycles in that setting.

[11] A similar feature occurs in commercial real-estate markets. There is anecdotal evidence that some cities' tradition that the majority of rental leases should expire on the same date improves the quality of the renter-housing match. This might explain the longevity of these customs. Examples include New York City's *Moving Day* (until WWII), Quebec's *fête du déménagement*, Boston's *Allston Christmas*, and Madison's *Hippie Christmas*. Conversely, the tight residential real-estate market in the UK has caused the appearance of so-called "upward chains", where a home sale transaction would be delayed until the current owners complete their own purchase of a new home, which could further be delayed for the same reason, etc.

[12] See Sönmez and Ünver (2011) and Abdulkadiroğlu and Sönmez (2013) for a pair of recent reviews.

[13] See Sönmez and Ünver (2013) for a recent review of the kidney-exchange literature.

similar work I am aware of is a trio of papers by Nicoló and Rodríguez-Álvarez (2011, 2012, 2013), who consider a model identical to the one presented here. Two of the papers (Nicoló and Rodríguez-Álvarez 2012, 2013) present impossibility results, while Nicoló and Rodríguez-Álvarez (2011) proposes a solution for the kidney-exchange problem but on a very restricted preference domain: namely, the authors assume that all patients rank all kidneys in the same way, barring incompatibilities.

Some recent work on general random matching mechanisms include Budish et al. (2013); Pycia and Ünver (2014); Akbarpour and Nikzad (2014); Kesten and Ünver (forthcoming). Budish et al. (2013) is the one that is closest to my paper. The authors study the feasibility of a class of exogenously-existing constraints that need to be respected for some assignment problem, such as school choice. If we interpret agents as students and objects as school seats, these constraints can represent, for example, maximum quotas for a certain type of students. The focus of their paper is finding a class of constraints such that whenever a bistochastic matrix satisfies them, one can decompose the matrix as a convex combination of permutation matrices, each one of which satisfies the same constraints, regardless of the desired lower and upper bound of each constraint. The authors provide a sufficient condition for such universal implementability, as they call this property. Namely, if the constraints satisfy a bihierarchical property, then universal implementability obtains. The main difference with my approach is that I care about a different class of constraints: for example, satisfaction of individual rationality and the cycle constraints. Furthermore, one of the mechanisms that Budish et al. (2013) propose is a special case of the most general mechanism defined in Section 7.

Another similarity between their approach and mine is that the main mechanisms proposed in each paper are based on the Probabilistic Serial mechanism, initially defined by Bogomolnaia and Moulin (2001) in the simple object-assignment setting. Since their seminal contribution, their work has been generalized for ordinal preferences allowing indifferences (Katta and Sethuraman 2006), for multi-unit demand (Kojima 2009), for property rights necessitating individual rationality (Yılmaz 2009, 2010), for fractional endowments (Athanassoglou and Sethuraman 2011), and for combinatorial demand (Nguyen et al. 2014).[14]

More broadly speaking, the most general result of my paper proposes a desirable mechanism for matching under arbitrary constraints, which includes two-sided matching markets. Some work related to that includes three recent papers (Kamada and Kojima forthcoming, 2014a,b) studying two-sided matching markets under relatively general constraints. Their main concerns, however, are stability concepts under these constraints, while I do not address stability in this work, in line with the remainder of the kidney-exchange literature.

The paper is organized as follows. Section 2 presents a background on the practice of kidney exchange. Section 3 presents the model. Section 4 and 5 go over some preliminary observations and results with those in Section 5 relating to deterministic mechanisms. Section 6 presents the 2-Cycle Probabilistic Serial mechanism and Section 7 proposes the general mechanism. Section 8 discusses incentives in my setting and Section 9 concludes.

2 Background on Kidney Exchange

Kidney transplantation is generally the only long-term treatment for end-stage chronic renal disease. Not only is transplantation associated with longer expected survival rate as compared to dialysis but the quality of life of renal patients is higher after a kidney graft (Wolfe et al. 1999). Most transplanted kidneys originate

[14]Other work spurred on by Bogomolnaia and Moulin (2001) include characterizations of ordinal efficiency (McLennan 2002; Abdulkadiroğlu and Sönmez 2003; Manea 2008; Carroll 2010), the study of the behavior of the PS mechanism in large markets (Che and Kojima 2010; Kojima and Manea 2010; Liu and Pycia 2013), and axiomatic characterizations of the PS mechanism and its extensions (Bogomolnaia and Heo 2012; Heo and Yılmaz 2013; Hashimoto et al. 2014; Heo 2014a,b).

from deceased donors but, since the functional capacity of a single kidney is sufficient for most people, living-donor transplantations are also possible. Most living-donor transplantations come about when a patient in need of a kidney transplant finds a donor (often a relative or a friend), who is willing to donate one of her kidneys to the patient. If the pair is blood-type and tissue-type compatible, the transplantation can take place. If they are incompatible, however, they still have the opportunity to effect a transplantation if there exists another mutually incompatible patient-donor pair, such that the donor of each pair is compatible with the patient of the other pair. In such a case, the patient in the first pair receives a kidney from the donor in the second pair and vice versa. This is referred to as *paired kidney exchange*.

Larger exchanges, involving three or more pairs, are also possible. However, the near universal ban on the buying and selling of kidneys[15] means that donors cannot contractually commit to donate a kidney. Thus if all the surgeries in a kidney exchange are not performed simultaneously, the last donor in the exchange might back out of the trade since her patient has already received a kidney. That would be extremely damaging to the designated recipient of her kidney since that patient's donor has already donated a kidney: that patient, while still in need of a kidney, would not be able to participate in another kidney exchange unless she finds a new donor. Simultaneous kidney exchange places significant logistical burden on the participating hospital (or hospitals), however: a kidney exchange with k pairs requires the availability of $2k$ operating rooms and surgical teams working at the same time. This availability is particularly hard to guarantee since living-donor kidney transplantations are considered elective surgeries which are of secondary priority to emergency surgeries.

In order to maximize the possible benefit of such kidney exchanges, a number of regional and national clearing houses have been organized in the US, the UK, the Netherlands and a number of other countries. Following the seminal work of Roth et al. (2005a), models of kidney exchange assume that all patients have binary preferences over the available kidneys: each kidney is either compatible (acceptable) or incompatible (unacceptable) and, furthermore, patients find all compatible kidneys to be perfect substitutes for each other. In this simplified setting, ex-ante, ordinal and ex-post efficiency coincide and, in fact, any efficiency criterion reduces to the maximization of the number of potential transplantations. However, there is substantial medical evidence that a variety of interactions between the donor and patient characteristics may significantly impact the graft long-term survival rates. For example, the age and gender of living donors have been shown to affect graft failure rates (Gjertson 2003; Øien et al. 2007).[16]

Tissue incompatibility stems from the human leukocyte antigen (HLA) system. The HLA antigens are proteins with important roles in the immune system. There are six major HLA antigens and different people have different sets of them. It is possible for an individual to develop antibodies for antigens that she does not possess if she is exposed to them during pregnancy or after blood transfusion, organ or tissue transplantation. If a potential kidney transplant recipient has an antibody for a HLA antigen present in the kidney donor, that would cause incompatibility. ABO blood-type incompatibility works in a similar fashion: people who are of blood type A, for example, have type A antigens and type B antibodies. Thus they can receive blood or organ donations from donors who were themselves type O (who have neither of the two possible antigens) or type A.

Recent medical research, however, has made possible the transplantation of organs, including kidneys, even when there is blood- or tissue-type incompatibility. The process allowing that, called *desensitization*,

[15] The Islamic Republic of Iran is the only exception (Fatemi 2012).

[16] The literature on the effect of deceased donor characteristics on graft survival rates is much more extensive. Some statistically important factors include donor's age, sex, race, pre-decease health status, cause of death, as well as the mass of the transplanted kidney (Chertow et al. 1996; Koning et al. 1997; Ojo et al. 2000; Pessione et al. 2003).

has been described as "risky, technically demanding and costly," however (Wallis et al. 2011). Additionally, kidneys transplanted after such a treatment have slightly lower survival rates than similar compatible kidneys. For more details see Tobian et al. (2008) and Montgomery et al. (2011), for example. Furthermore, there is evidence in the medical literature that similarity in patient and donor's HLA antigens affects transplantation success (Opelz 1997, 1998; Opelz and Döhler 2007; Sasaki and Idica 2010).[17]

Other factors contributing to the view that patients' preferences over kidneys aren't binary are considerations of logistical feasibility (for example, all else held equal, a patient-donor pair prefers being matched with a pair that is geographically closer) or fairness (the UK national kidney exchange prefers that the age difference of the two donors in each two-pair kidney exchange is not too great and uses the actual age difference as a final tiebreaker when determining the proposed matching).

As noted in the introduction, another reason for why studying the model with strict preferences is valuable has to do with the provision of sufficient incentives to patient-donor pairs who are compatible with each other to enroll in the kidney exchange program. This would be guaranteed by individual rationality, the most important desideratum for mechanisms in my setting. In addition to providing incentives for compatible pairs to participate, it is of paramount importance since it ensures compatible transplantations. I define individual rationality (together with the other mechanism properties) in Section 3 and discuss its implications at the beginning of Section 4.

Conditional on recommending only compatible transplantations, I would like my mechanism to be efficient. This is captured by Pareto efficiency in the case of deterministic matchings and by ordinal efficiency for random matchings. I discuss the connection between these two concepts in Section 6. The third most important condition is some notion of fairness. Fairness, in addition to efficiency, is one of the fundamental requirements in the setting of matching without monetary transfers.[18] I use the notion of *equal treatment of equals*, which in my setting reduces to *anonymity*. Essentially, it requires that the outcome of a mechanism depends only on the profile of preferences and not on the identity of the agents. I discuss other possible justice criteria in Section B.1.

Finally, the least important criterion in my setting is strategyproofness. Early economic models of kidney exchange were concerned with providing incentives to the patients and their doctors to report their preference truthfully. Recent work, however, has moved away from this paradigm. Ashlagi and Roth (2014) note:

> During the initial startup period [of kidney exchange in the US], attention to the incentives of patients and their surgeons to reveal information was important. But as infrastructure has developed, the information contained in blood tests has come to be conducted and reported in a more standard manner (sometimes at a centralized testing facility), reducing some of the choice about what information to report and with what accuracy. So some strategic issues have become less important over time (and indeed current practice does not deal with the provision of information that derives from blood tests as an incentive issue).

In line with this, the mechanism I propose will satisfy the first three desiderata but will have poor incentive properties. I discuss this issue in Section 8.

[17] For a contradictory view, see Delmonico (2004).

[18] See Roth et al. (2005a); Yılmaz (2011) for some work aimed to achieve fairness in kidney exchanges with binary preferences.

3 Set-Up

Let $A = \{1, \ldots, n\}$ be a set of n patient-donor pairs. I assume that each patient i has a preference order $\succ_i$ over the set of kidneys. I will identify each of the kidneys via its donor so each patient can be said to have a preference order over the set A.[19] I will assume that $\succ_i$ represents a strict order so that for any $i, j, k \in A$, $j \neq k$ implies either $j \succ_i k$ or $k \succ_i j$ but not both. I write $j \succsim_i k$ whenever either $j = k$ or $j \succ_i k$. Denote all possible strict preference orders over A by $\mathbb{P}$ and let $\mathcal{P} = \mathbb{P}^n$ be all possible preference profiles for the n patients. I will denote a generic element of $\mathcal{P}$ by $\succ$. For any $A' \subsetneq A$, I will denote the preference profile of all patients not in A' by $(\succ_{-A'})$. For simplicity, when $A' = \{i\}$ is a singleton, I will denote the preference profile by $(\succ_{-i})$.

The goal is to study the ways in which the agents can organize a kidney exchange among themselves so that each donor donates a kidney if and only if her patient has received a kidney in order to avoid a situation where a donor has given a kidney, but her patient has not received one. Any exchange among the agents resulting in a final **(deterministic) matching** can be represented by a bijective function $m : A \to A$, where $m(i) = j$ indicates that patient i receives donor j's kidney. I interpret $m(i) = i$ to mean that patient i either receives her donor i's kidney or, equivalently for our purposes, is left unmatched and does not participate in the kidney exchange. I will denote the set of all such matchings by $\mathcal{M}$. Given a preference profile $\succ$, I will say that a matching $m \in \mathcal{M}$ is **efficient** if it is not Pareto dominated: i.e., if there does not exist a matching $m' \in \mathcal{M}$ such that $m'(i) \succsim_i m(i)$ for all $i \in A$ and $m'(i) \succ_i m(i)$ for some $i \in A$.

For any $k \in \mathbb{N}$, I will say that m **satisfies the k-cycle constraint** if there do not exist $k+1$ distinct elements of A denoted by $a_1, \ldots, a_{k+1}$ such that for $i = 1, \ldots, k$ we have $m(a_i) = a_{i+1}$. I will denote the set of all matchings that satisfy the k-cycle constraint by $\mathcal{M}_k$. Given a preference profile $\succ$, I will say that a matching $m \in \mathcal{M}_k$ is **k-constrained efficient** if there does not exist a matching $m' \in \mathcal{M}_k$ such that $m'(i) \succsim_i m(i)$ for all $i \in A$ and $m'(i) \succ_i m(i)$ for some $i \in A$.

Note that each $m \in \mathcal{M}$ can be represented as a matrix P_m with a generic entry $P_m(i, j)$ defined by

$$P_m(i,j) = \begin{cases} 1 & \text{if } m(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

So $P_m(i, j)$ equals 1 if and only if patient i receives donor j's kidney under matching m. In all other cases, $P_m(i, j) = 0$. Thus the matrix P_m for each $m \in \mathcal{M}$ has exactly one entry equal to 1 in each row and in each column. Matrices of this kind are called **permutation matrices**.[20] In fact, every permutation matrix represents some deterministic matching so there is a bijective relationship between the set of permutation matrices and the set $\mathcal{M}$.

I denote the space of lotteries over deterministic matchings as $\Delta\mathcal{M}$ and refer to its elements as **random matchings**. Each random matching $\mu \in \Delta\mathcal{M}$ can in turn be represented as a convex combination of the matrices corresponding to the elements of $\mathcal{M}$ which form the support of μ. It is then easy to see that μ can be represented as a bistochastic matrix P_μ, where, as usual, I use the term **bistochastic matrix** to refer to any non-negative matrix, such that the sum of its entries along any given row or column is 1. Note then that $P_\mu(i, j)$ denotes the probability that patient i receives donor j's kidney. Note that row i of P_μ denotes the probabilities with which patient i receives each of the n kidneys. I will refer to this as patient i's

[19] In fact, I use each $i \in A$ to refer to the patient from patient-donor pair i, to the donor of that pair, or to that donor's kidney. The corresponding context makes it clear in case I have not specified what i refers to.

[20] Note that one can view each such matrix P_m as the adjacency matrix of a directed graph. Then m satisfies the k-cycle constraint if and only if that directed graph does not have directed cycles of length greater than k.

probability-share allocation and will denote it by $P_\mu(i)$. I similarly define a **sub-bistochastic matrix** to be any non-negative matrix, such that the sum of its entries along any given row or column is no more than 1.

Assume that $\succ_i$ ranks the kidneys in A in the order $(a_1, a_2, \ldots, a_n)$ from best to worst. Consider p and q to be two probability-share allocations and let p_{a_j} and q_{a_j} denote the probability with which patient i receives donor a_j's kidney. Then I say that p **first-order stochastically dominates** q **with respect to** $\succ_i$ if

$$\sum_{l=1}^{j} p_{a_l} \geq \sum_{l=1}^{j} q_{a_l} \tag{1}$$

for each $j = 1, \ldots, n$. I will say that p **strictly first-order stochastically dominates** q **with respect to** $\succ_i$ if (1) holds and $p \neq q$.

I refer to functions $f : \mathcal{P} \to \mathcal{M}$ and $g : \mathcal{P} \to \Delta\mathcal{M}$ as a **deterministic mechanism** and a **random mechanism**, respectively. Abusing terminology, I say that a deterministic mechanism f is **efficient** (resp. k**-constrained efficient**) if for all $\succ$, $f(\succ)$ is efficient (k-constrained efficient) with respect to $\succ$. A random mechanism f is **ex-post efficient** (k**-constrained ex-post efficient**) if for all $\succ$, $f(\succ)$ places positive probability only on deterministic matchings that are efficient (k-constrained efficient) with respect to $\succ$. Note that any efficient (k-constrained efficient) deterministic mechanism can be viewed as an ex-post efficient (k-constrained ex-post efficient) random mechanism.

A random mechanism f is **ordinally efficient** if for all $\succ$ there does not exist an element μ in $\Delta\mathcal{M}$ such that $P_\mu(i)$ first-order stochastically dominates $P_{f(\succ)}(i)$ with respect to $\succ_i$ for all $i \in A$ and strictly so for some $i \in A$. A random mechanism f is k**-constrained ordinally efficient** if for all $\succ$ we have $f(\succ) \in \Delta\mathcal{M}_k$ and there does not exist $\mu \in \Delta\mathcal{M}_k$ such that $P_\mu(i)$ first-order stochastically dominates $P_{f(\succ)}(i)$ with respect to $\succ_i$ for all $i \in A$ and strictly so for some $i \in A$.

If $i \succ_i j$ for some $i, j \in A$, I will say that i finds kidney j **unacceptable**. Consequently, I say that a matching $m \in \mathcal{M}$ is **individually rational** if $m(i) \succsim_i i$ for all $i \in A$. Analogously, a random matching μ is **individually rational** if the deterministic matchings in its support are all individually rational themselves. Equivalently, $P_\mu(i, j) = 0$ whenever i finds j unacceptable. A deterministic mechanism (random mechanism) f is **individually rational** if for all $\succ$, $f(\succ)$ is an individually rational deterministic matching (random matching).

I move to defining the incentive and fairness properties of mechanisms. I will define them only for random mechanisms but the same definitions apply to deterministic mechanisms when viewed as a subset of the random ones. A random mechanism f is **strategyproof** if for all preference profiles $\succ$, all patients i and all preference orders $\succ_i'$, $P_{f(\succ)}(i)$ first order stochastically dominates $P_{f(\succ_i', \succ_{-i})}(i)$ with respect to $\succ_i$. I say that a random mechanism f is **weakly strategyproof** if for all preference profiles $\succ$ and all patients i, there does not exist an alternative preference order $\succ_i'$ such that $P_{f(\succ_i', \succ_{-i})}(i)$ strictly first-order stochastically dominates $P_{f(\succ)}(i)$ with respect to $\succ_i$.[21] It is easy to see that the two notions of strategyproofness are equivalent for deterministic mechanisms since first-order stochastic dominance is a total order over deterministic allocations.

A random mechanism f is said to be anonymous if the patient-donor pairs' names are irrelevant for the outcome of the mechanism. Formally, start by fixing an arbitrary bijective function $\pi : A \to A$ and a preference profile $\succ$. If $\succ_i$ for some $i \in A$ ranks the kidneys from A in the order $(a_1, \ldots, a_n)$ (from

[21] This is not a misnomer. Weak strategyproofness is a very weak property indeed. Balbuzanov (2014) shows that weak strategyproofness fails to satisfy even a simple intuitive incentive property. Namely, weak strategyproofness is not enough to guarantee that there exists a von Neumann-Morgenstern utility vector, under which the agent would prefer reporting truthfully.

best to worst), construct the preference relation $\succ^{\pi}_{\pi(i)}$ as the preference relation corresponding to the order $(\pi(a_1), \ldots, \pi(a_n))$. I.e.,

$$\pi(a_1) \succ^{\pi}_{\pi(i)} \cdots \succ^{\pi}_{\pi(i)} \pi(a_n).$$

I say that μ is **anonymous** if for all $\succ$ the matrices $P_{f(\succ)}$ and $P_{f(\succ^{\pi})}$ are identical up to the permutation π. In other words, $P_{f(\succ)}(i,j) = P_{f(\succ^{\pi})}(\pi(i), \pi(j))$ for all $i, j \in A$.

I end by noting some relationships between the properties outlined in this section. It is clear that strategyproofness implies weak strategyproofness. In addition, ordinal efficiency is stronger than ex-post efficiency. See Bogomolnaia and Moulin (2001) for the full proof but the intuition is clear: if a random matching μ placed positive probability on an inefficient deterministic matching m then one can shift some of that probability to a matching that Pareto dominates m and thus improve μ in the ordinal-efficiency sense. Analogously, k-constrained ordinal efficiency implies k-constrained ex-post efficiency.

4 Preliminary Observations and Results

I start this section with a remark about the connection between transplantation incompatibility and individual rationality in my setting from Section 3. Individual rationality can be defined in one of two ways. First, the method adopted here is to assume that a matching is individually rational only if each patient i receives a compatible kidney that she ranks higher than kidney i or, failing that, she is left unmatched.[22] This guarantees that no patient will receive a kidney that has worse graft survival expectations than her own donor's kidney. Under this interpretation individual rationality is a stronger condition than simply ensuring compatible transplantations. The alternative way to think about individual rationality, however, is to define it to be equivalent to compatibility. Thus patient i finds a kidney acceptable if and only if that kidney is compatible for her. In this case it might be possible for an agent to receive a kidney that she ranks lower than her donor's kidney. However that might deter compatible pairs from being part of the kidney exchange. This might hurt overall efficiency as the participation of compatible pairs may significantly improve the outcome's efficiency (Roth et al. 2005b; Gentry et al. 2007; Sönmez and Ünver 2014). However, while this is not the approach I take here, what follows can be modified in a straightforward manner to accommodate it.

One of the strongest assumptions in my set-up is that each donor-patient pair either receives *and* donates a kidney, or *neither* receives *nor* donates a kidney at the end of the mechanism. This assumption is obviously justified from the point of view of the bioethics of kidney exchange: it would be unfair for a patient's donor to donate a kidney without the patient receiving one since that destroys the patient's "bargaining chip" in any future kidney exchanges. And, since each donor can donate at most one kidney, there are n kidneys to be donated and received and so, conversely, there cannot exist a donor-patient pair which receives a kidney but does not donate one. This symmetry allows representing all exchanges as permutation matrices and lotteries over exchanges (what I call random matchings) as bistochastic matrices, which simplifies the analysis.[23]

I noted above that any random matching can be represented as a bistochastic matrix. The following celebrated result, shown by Birkhoff (1946) and von Neumann (1953), also provides the converse.

Theorem 1 (Birkhoff-von Neumann Theorem). *The convex hull of all $n \times n$ permutation matrices equals the set of all $n \times n$ bistochastic matrices.*

[22] Again, being left unmatched could mean either that the patient does not undergo a kidney transplantation or that she receives her donor's kidney. In either case, the patient does not participate in any exchanges.

[23] I allow the existence of "unattached" kidneys (such as kidneys donated by deceased or undirected altruistic donors) in an extension of the main model below.

I continue by explicitly considering the case $k = 2$. Note that for any matching in $\mathcal{M}_2$ in which patient i receives donor j's kidney, patient j must receive donor i's kidney in order for the matching to satisfy the 2-cycle constraint. This implies that any permutation matrix representing a matching in $\mathcal{M}_2$, and thus any bistochastic matrix representing a matching in $\Delta\mathcal{M}_2$, is symmetric: the probability that any patient i receives donor j's kidney must equal the probability that patient j receives donor i's kidney. The converse is also true for the permutation matrices: it is easy to see that any symmetric permutation matrix represents a matching in $\mathcal{M}_2$. Consider the converse in the case of bistochastic matrices. In particular, consider the following bistochastic matrix for $n = 3$:

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \tag{2}$$

I claim that this matrix cannot be decomposed into a convex combination of symmetric permutation matrices and, therefore, it cannot represent a random matching in $\Delta\mathcal{M}_2$. To see that, observe that in each 3×3 symmetric permutation matrix at least one of the diagonal elements is equal to 1. Thus in all 3×3 symmetric permutation matrices, the trace is at least 1 and, since this property is preserved under convex combinations, this must also be the case for any convex combination of 3×3 symmetric permutation matrices. Clearly, matrix (2) does not satisfy that property and so it does not represent an element of $\Delta\mathcal{M}_2$. In the language of kidney exchange, if the planner can organize only two-way exchanges and if there are three patient-donor pairs, at least one of them must not be a part of the exchange. At the same time, the probability that any given patient is not part of the exchange is zero for any random matching represented by (2). This, however, cannot happen in any lottery over deterministic matchings in which there is always a patient who is not in the exchange. Consequently, it need not be the case that every symmetric bistochastic matrix is a convex combination of symmetric permutation matrices (Katz 1970; Cruse 1975). In fact, as pointed out by Schrijver (2003), Edmonds' (1965) celebrated matching polytope characterization implies the following result.

Theorem 2 (Edmonds' Theorem)**.** *A symmetric bistochastic matrix P can be represented as a convex combination of symmetric permutation matrices if and only if*

$$\sum_{i \in E} \sum_{j \in E \setminus \{i\}} P(i,j) \leq 2p$$

for all $p \in \mathbb{N}$, for all $E \subseteq A$ with $|E| = 2p + 1$.

Each of the possible sets E defines a separate constraint that the matrix P must satisfy. I call such sets **Edmonds sets** and the corresponding constraints **Edmonds constraints**.[24] I will speak about Edmonds constraints containing a pair $\{i, j\}$ or an entry $P(i,j)$ if the corresponding Edmonds set contains both i and j. Analogously, I will refer to the sum on the left-hand side of the Edmonds constraint as a **Edmonds sum**. To illustrate the intuition behind the result, I provide a brief proof of the necessity condition. For the more involved sufficiency condition, see Edmonds (1965), Balinski (1972) or Cruse (1975).[25]

Proof of necessity: Assume that a given symmetric bistochastic matrix $P(i,j)$ can be represented as a convex combination of symmetric permutation matrices. Equivalently, P represents the matching probabilities

[24] It is worth noting that the Edmonds constraints do not satisfy the bihierarchy condition of Budish et al. (2013). So the results of this paper are logically independent.

[25] Schrijver (2003) has a list of additional alternative proofs proposed in the literature.

induced by a lottery over the deterministic matchings associated with the permutation matrices. Take some $k \in \mathbb{N}$ and an Edmonds set E with $|E| = 2k + 1$. Note that in any deterministic matching, at least one patient $i \in E$ does not receive a kidney from the set $E \setminus \{i\}$ since E has odd cardinality. Therefore, since the Edmonds sum for the associated permutation matrix equals the number of patients in E who receive a kidney from E other than their own donor's kidney, the Edmonds constraint is satisfied for all permutation matrices. Since the bistochastic matrix is a convex combination of the permutation matrices, it also satisfies the constraint. $\square$

Finally, I will call symmetric bistochastic matrices that satisfy the Edmonds constraints **2-implementable**. Generally, I call any bistochastic matrix P that satisfies $P = P_\mu$ for some $\mu \in \Delta\mathcal{M}_k$ k**-implementable**.

5 Deterministic Mechanisms with Cycle Constraints

In the setting of kidney exchange, the main desideratum for the mechanism should be individual rationality in order to guarantee compatible transplantations and, potentially, provide incentives for compatible couples to participate in the exchange. The other desiderata, in order of importance, are efficiency, anonymity (the fairness criterion[26]), and, if possible, good incentive properties. In this section, I consider to what extent these properties are compatible for deterministic mechanisms when we have cycle constraints.

First, note that without (binding) cycle constraints, Gale's Top Trading Cycles (TTC), introduced by Shapley and Scarf (1974), is individually rational, efficient, anonymous and strategyproof (Ma 1994; Miyagawa 2002). In addition, TTC is quite simple: at any stage, each patient i "points" to the patient, whose donor's kidney is the highest-ranked for i among the kidneys remaining. Note that i could point to herself. This forms a directed graph and, by finiteness of A, it must have at least one directed cycle. We perform the trades implied by that cycle and remove the patient-donor pairs involved in that trade. We iterate this process until all agents have been involved in a trade or have dropped out in cycles of length 1. It should be noted that the order of elimination of the cycles does not change the outcome of the mechanism; that is to say, the same cycles are selected, regardless of that order. In fact, the first mechanism proposed for the kidney-exchange problem was an adaptation of Top Trading Cycle (Roth et al. 2004).

As long as $k < n$, however, it is possible that TTC's outcome might fail to satisfy the k-cycle constraint. Even if $n = k + 1$, it is conceivable that the only cycle in the mechanism is a cycle that includes all n agents. A natural question to ask at this stage is indeed how often TTC's outcome fails to satisfy the k-cycle constraint. The approach I take to answer this question is inspired by Pittel (1989), who studies the core in stable-marriage problems by looking at randomly-drawn preferences.[27] I look at a random iteration of the problem by drawing a preference profile from a uniform distribution over $\mathcal{P}$.

Proposition 3. *For any k, if the preferences are drawn from a uniform distribution over $\mathcal{P}$, the probability that TTC selects at least one cycle of length greater than k goes to one as n goes to infinity.*[28]

Of course, preferences in many problems such as kidney-exchange would be correlated and not independently drawn from a uniform distribution but, nevertheless, this result suggests that, especially for problems with large n, TTC would not make for a satisfactory mechanism even approximately since it is vanishingly unlikely that all the cycles it selects will be sufficiently short. This motivates me to turn to other potential

[26] I later consider the compatibility of the other desiderata with other possible fairness properties.

[27] See also Knuth et al. (1990); Pittel (1992) and, more recently, Ashlagi et al. (2014) for papers studying the same question with similar methods.

[28] The proofs of all propositions are in the Appendix.

deterministic mechanisms. However, it turns out that deterministic mechanisms do not have good properties either. First, there does not exist a deterministic mechanism that is anonymous and efficient.

Proposition 4. *If cycle length cannot exceed $k > 1$, there does not exist an anonymous k-constrained efficient deterministic mechanism for all $n \geq k + 1$.*

Proposition 4 suggests a question. Since I view efficiency as being more important than anonymity, if one insists on using a deterministic mechanism, they must dispense with my fairness criterion. Ma (1994) characterizes the TTC mechanism as being the unique mechanism that is individually rational, Pareto-efficient, and strategyproof. One can then ask whether a mechanism similar to TTC can be found that is individually rational, k-constrained efficient, and strategyproof; i.e., a mechanism that satisfies the three remaining desiderata. It turns out, however, that these properties are also incompatible.

Proposition 5. *If cycle length cannot exceed $k > 1$, there does not exist an individually rational, k-constrained efficient, and strategyproof deterministic mechanism for all $n \geq k + 1$.*[29]

Since without cycle-length constraints TTC satisfies anonymity, Pareto optimality, individual rationality and strategyproofness, the results in this section suggest that imposing cycle-length constraints significantly limits the range of desirable properties of all deterministic mechanisms. Indeed, the best one can do within the class of deterministic mechanisms is to construct a mechanism that satisfies individual rationality and k-constrained efficiency. Indeed, a suitably redefined version of the serial dictatorship satisfies these two conditions.[30] One can achieve anonymity by using the analogously modified random serial dictatorship. This observation and the results above provide a compelling motivation for my consideration of the performance of random mechanisms in this setting.

6 The 2-Cycle Probabilistic Serial Mechanism

The proposed mechanism is based on the simultaneous eating mechanism (Bogomolnaia and Moulin 2001) and its extension (Budish et al. 2013). The intuition behind the mechanism is simple: all kidneys are viewed as infinitely divisible and the patients, each endowed with a claiming-speed function, continuously claim shares of their most preferred kidney in continuous time. Once all the kidneys have been completely claimed, we interpret the share each patient has claimed of each kidney as the probability with which that patient receives the kidney. For the simpler object-assignment setting, the Birkhoff-von Neumann theorem and its generalization proved by Budish et al. (2013) guarantee that the resulting matrix of probability shares would be implementable as a lottery over deterministic matchings. In my setting, I need to make sure that the resulting bistochastic matrix is symmetric and satisfies the Edmonds' conditions.

Budish et al. (2013) allow the existence of additional exogenously-imposed constraints that need to be respected for each ex-post assignment at the conclusion of the mechanism. In the school-choice problems, these constraints can be interpreted as quotas related to affirmative action, for example. In my setting, I similarly need to respect a set of constraints but they arise in order to guarantee individual rationality and k-implementability. I note here that the Edmonds constraints are not a special case of the constraints considered in Budish et al. (2013). However, the approaches of the two mechanisms in the way that they

[29] The same result was independently discovered by Nicoló and Rodríguez-Álvarez (2012). I am indebted to Antonio Miralles for pointing that out to me.

[30] For the case $k = 2$, such a mechanism would, at each step, match the remaining pair that is highest in a given priority order with their most-preferred mutually compatible remaining pair. This approach can be extended to $k > 2$. I do not formally study this mechanism in this paper. See, however, Example 2.

guarantee the desired constraints is similar: each patient is allowed to claim probability shares of her highest-ranked kidney among the ones available to her *as long as none of the Edmonds constraints corresponding to that patient and that kidney bind*. To make that clear, I proceed by describing the algorithm defining the outcome of the 2-*cycle simultaneous eating* (2CSE) mechanism given a preference profile $\succ$.

Definition 1. The 2-Cycle Simultaneous Eating mechanism. Time runs continuously starting at $t = 0$. For each point in time, there is an associated sub-bistochastic matrix M^t, where M^0 is the initial zero matrix. Each patient i has an associated claiming-speed function $e_i : [0, \infty) \to \mathbb{R}_+$ with $\int_0^\infty e_i(t)dt \geq 1$. I say that kidney j is available to patient i at time $t \geq 0$ if the following three conditions are satisfied: first, neither of the patients i and j finds the other one's donor's kidney unacceptable; second, the row sums corresponding to patients i and j are strictly less than 1 at time t (i.e., both i and j have remaining probability shares and remaining demand); third, none of the Edmonds constraints containing i and j bind at time t. Note that due to the symmetry, if j is available to i at time t, then i is available to j at the same time. For simplicity, I then say that the pair (i, j) is available.

At each instance of time t, each patient i claims with speed $e_i(t)$ the available remaining probability shares of her highest-ranked reported kidney j (possibly $j = i$) among the kidneys that are available to i at that instance. That increases the probability that patient i receives kidney j but, since I am restricted to cycles of length 2, it must also increase the probability that patient j receives kidney i. In other words, this action increases both $M^t(i, j)$ and $M^t(j, i)$. Note that j claiming kidney i's probability shares would increase the same two matrix entries. This activity also decreases the remaining probability shares of both i and j. Since the Edmonds constraints do not depend on any of the values of $M^t(i, i)$, the "pair" (i, i) is available if and only if i's row constraint doesn't bind. This guarantees that any point of time in the mechanism, each agent has available kidneys whose probability shares she is allowed to claim. Each patient (together with her associated donor) exits when her demand is met or, equivalently, when her donor's kidney's probability shares are depleted. The algorithm ends when all agents have exited.

The final output of this procedure is a probability-shares matrix M that is symmetric (since whenever $M^t(i, j)$ increases, so does $M^t(j, i)$ for all $i, j \in A$), bistochastic (since the procedure ends only when all patients' row sums equal 1) and satisfies all the Edmonds constraints (since M^t satisfies them for any t). Hence, M is implementable as a lottery over deterministic matchings that satisfy the 2-cycle constraint. That lottery is set to be the outcome of the 2CSE mechanism. Since, for our purposes, we are indifferent between all lotteries represented by a given bistochastic matrix, I write:

$$2CSE(\succ, e) = M,$$

where e denotes the profile of claiming-speed functions. Additionally, M is individually rational with respect to the reported preferences—this is guaranteed by the first condition defining availability above: no patient i is allowed to claim probability shares from kidney j if patient j finds kidney i unacceptable. Thus patient i is prevented from increasing the probability that patient-donor pairs i and j form a paired kidney exchange. Hence $M(i, j) = 0$. Following Bogomolnaia and Moulin (2001), whenever all the patients have the same claiming speeds (which, without loss of generality, I can assume to satisfy $e_i(t) = 1$ for all i and t), I will call this mechanism the 2-*cycle probabilistic serial* (2CPS) mechanism.

As an illustration, consider the following example of the 2CPS mechanism in action. Let $A = \{1, 2, 3, 4\}$

and let the preferences be defined by:

$$\succ_1: 2 \succ_1 3 \succ_1 4 \succ_1 1,$$
$$\succ_2: 1 \succ_2 3 \succ_2 4 \succ_2 2,$$
$$\succ_3: 1 \succ_3 2 \succ_3 4 \succ_3 3,$$
$$\succ_4: 1 \succ_4 2 \succ_4 3 \succ_4 4.$$

Since all kidneys are initially available to all agents, between $t = 0$ and $t = 1/4$ during the mechanism's implementation, patient $2, 3$, and 4 claim kidney 1's probability shares, while 1 claims kidney 2's probability shares. This increases $M^t(1,3)$ and $M^t(1,4)$ from 0 to $1/4$, while $M^t(1,2)$ increases to $1/2$ since 1 and 2 are claiming each other's donors' kidneys' probability shares. The matrix entries in its lower-diagonal part change correspondingly to preserve its symmetry. Thus at $t = 1/4$, the probability-share matrix looks like this:

$$M^{1/4} = \begin{pmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 \end{pmatrix}.$$

At this point of time, patient 1 exits since her her unit-demand has been met or, equivalently, kidney 1's probability shares have been completely claimed. Afterward, 3 and 4's highest-ranked available kidney becomes 2, while patient 2's is 3. So at time $t = 3/8$, the probability-share matrix looks like this:

$$M^{3/8} = \begin{pmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/4 & 1/8 \\ 1/4 & 1/4 & 0 & 0 \\ 1/4 & 1/8 & 0 & 0 \end{pmatrix}.$$

Notice that at this point of time the Edmonds constraint corresponding to the set $E = \{1, 2, 3\}$ starts binding. Hence the pair $\{2, 3\}$ becomes unavailable, in addition to all the pairs containing 1. In what follows, 4's highest-ranked kidney remains 2, while the only kidney available to 2 and 3 is 4. At $t = 7/16$ the probability-share matrix has taken the following form:

$$M^{7/16} = \begin{pmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/16 \\ 1/4 & 1/4 & 1/16 & 0 \end{pmatrix}.$$

At this time, 2 exits. With only 3 and 4 remaining, it is easy to verify that the final probability-share matrix and the outcome of the 2CPS mechanism is

$$M = \begin{pmatrix} 0 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 & 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

The definition of the 2CSE mechanism guarantees that it is individually rational. *A fortiori*, so is the

2CPS mechanism. Additionally, since all agents have the same claiming-speed function, it is easy to see that the 2CPS mechanism is also anonymous. In this section, I will show that the 2CPS mechanism also satisfies the last desideratum in my setting. Namely, I demonstrate that the 2CSE mechanism (including 2CPS) is 2-constrained ordinally efficient. I start with a brief illustration of the concept of ordinal efficiency and a discussion of why it is preferable over the weaker ex-post efficiency condition.

The current efficiency criterion used by kidney-exchange clearinghouses involves maximizing the number of transplantations performed by finding a maximum-cardinality matching on the compatibility graph. Anonymity could be attained by uniformly randomizing over all maximum-cardinality matchings. Since Bogomolnaia and Moulin (2001) show that ex-post efficiency is strictly weaker than ordinal efficiency in the object-assignment setting, it is also worth asking to what extent ex-post and ordinal efficiency are mismatched in the setting of kidney exchange. For example, I observe in Proposition 4 that there are no anonymous 2-constrained efficient deterministic mechanisms but anonymity can be easily attained by mixing uniformly over all Pareto optimal matchings for a given preference profile. We consider both of these two approaches in the following example.

Example 1. Let $A = \{1, 2, 3, 4, 5\}$ and let the preference profile $\succ$ be defined by

$$\begin{aligned}
\succ_1 &: 2 \succ_1 3 \succ_1 4 \succ_1 5 \succ_1 1, \\
\succ_2 &: 5 \succ_2 4 \succ_2 1 \succ_2 3 \succ_2 2, \\
\succ_3 &: 4 \succ_3 2 \succ_3 5 \succ_3 1 \succ_3 3, \\
\succ_4 &: 1 \succ_4 5 \succ_4 3 \succ_4 2 \succ_4 4, \\
\succ_5 &: 3 \succ_5 1 \succ_5 2 \succ_5 4 \succ_5 5.
\end{aligned}$$

It can be checked easily that every possible matching with cardinality 4 is Pareto optimal and every Pareto optimal matching has cardinality 4. There are fifteen such matchings so the maximal-cardinality ex-post efficient random mechanism assigns probability of $1/15$ to each one of them. The resulting ex-post efficient bistochastic matrix is

$$M = \begin{pmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}.$$

For example, one can calculate that the outcome of the 2CPS mechanism can be represented by

$$M' = \begin{pmatrix} 1/5 & 2/5 & 0 & 2/5 & 0 \\ 2/5 & 1/5 & 0 & 0 & 2/5 \\ 0 & 0 & 1/5 & 2/5 & 2/5 \\ 2/5 & 0 & 2/5 & 1/5 & 0 \\ 0 & 2/5 & 2/5 & 0 & 1/5 \end{pmatrix}.$$

Notice that each patient is unambitiously better off under the 2CPS outcome than under the random matching represented by M: each patient i prefers $M'(i)$ over $M(i)$ in first-order stochastic dominance sense. Thus the matrix M does not represent an ordinally efficient assignment as it is dominated by M'.

Example 1 demonstrates that maximum-matching efficiency and ex-post efficiency, while intuitive, turn out to not satisfy the simple and theoretically appealing property of ordinal efficiency. Ex-post efficiency is strictly weaker than ordinal-efficiency in my setting, as well. This is further related to the surprising observation that the same bistochastic matrix might represent both an ex-post efficient random matching and a random matching that is not ex-post efficient (Abdulkadiroğlu and Sönmez 2003). So, as Bogomolnaia and Moulin (2001) remark, ex-post efficiency is quite a subtle concept.

The following example is analogous. Consider the following version of the serial-dictatorship mechanism for the case $k = 2$ (cf. footnote 30). Given a priority order over A, at each step, the remaining pair that is highest in the priority order is matched with their most-preferred mutually-compatible remaining pair. The two pairs are removed from the problem and the iterative step is repeated. It is easy to see that the mechanism is Pareto efficient. Looking at a random version of this mechanism with the priority order drawn from a uniform distribution over all priority orders, I now show that it also fails to satisfy ordinal efficiency. This result echoes one of the main motivating observations of Bogomolnaia and Moulin (2001).

Example 2. Let $A = \{1, 2, 3, 4, 5, 6\}$ and let the preference profile $\succ$ be defined by

$$\begin{aligned}
&\succ_1 : 3 \succ_1 2 \succ_1 1,\\
&\succ_2 : 1 \succ_2 4 \succ_2 6 \succ_2 2,\\
&\succ_3 : 4 \succ_3 1 \succ_3 2 \succ_3 3,\\
&\succ_4 : 5 \succ_4 1 \succ_4 3 \succ_4 6 \succ_4 4,\\
&\succ_5 : 6 \succ_5 2 \succ_5 3 \succ_5 4 \succ_5 1 \succ_5 5,\\
&\succ_6 : 1 \succ_6 1 \succ_6 5 \succ_6 3 \succ_6 6.
\end{aligned}$$

The bistochastic matrix that represents the outcome of the random serial dictatorship here is

$$M = \begin{pmatrix} 1/12 & 11/24 & 11/24 & 0 & 0 & 0 \\ 11/24 & 1/12 & 0 & 0 & 0 & 11/24 \\ 11/24 & 0 & 1/12 & 11/24 & 0 & 0 \\ 0 & 0 & 11/24 & 1/12 & 11/24 & 0 \\ 0 & 0 & 0 & 11/24 & 1/12 & 11/24 \\ 0 & 11/24 & 0 & 0 & 11/24 & 1/12 \end{pmatrix},$$

which is dominated by the outcome of the 2CPS mechanism, which can be represented by

$$M' = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \end{pmatrix}$$

Proposition 6. *Every 2-cycle simultaneous eating mechanism is 2-constrained ordinally efficient.*

The proofs of ordinal efficiency for the PS and the generalized PS (Budish et al. 2013) mechanisms both require first characterizing ordinal efficiency as being equivalent to the acyclicity of a certain suitably specified relation (see also Katta and Sethuraman 2006; Kojima 2009; Yılmaz 2010). The idea of the proof above can be used to avoid the auxiliary characterization lemma for those settings as well, thus significantly simplifying those proofs. Conversely, the approach in the previous literature cannot be applied here: the existence of a cycle in the mentioned relation implies the existence of a Pareto-improving trade in probability shares. However, the existence of such a trade in my setting does not guarantee that the new, Pareto-improving profile of probability-share allocations would form a 2-implementable matrix.

6.1 Multiple Donors, Social Endowment and Chains

While, for reason of simplicity of the exposition in this paper, I talk about patients having a single donor, the model here applies without modification to settings where patients may have multiple willing donors. Then in each kidney exchange each participating patient receives a kidney and one of her donors donates a kidney. The preferences over the set A for each patient can be inferred from the preferences over all donors' kidneys. Namely, if each patient i has a number of donors, denoted $i^1, i^2, \ldots$, then each patient j has strict preferences over all the donors denoted by $\succ'_j$. Then for each $i, j, l \in A$, I say

$$i \succ_j l \text{ if } \max_{\succ'_j}\{i^1, i^2, \ldots\} \succ'_j \max_{\succ'_j}\{l^1, l^2, \ldots\},$$

where $\max_{\succ} S$ for some set S and some strict preference order $\succ$ is the unique element satisfying $\max_{\succ} S \in S$ and $\max_{\succ} S \succ s'$ for all $s' \in S$. Note that individual rationality can be interpreted analogously: no patient receives a kidney that has worse prospects for her than any of her donors' kidneys.

In this section, I consider what happens if some of the kidneys come from deceased or undirected altruistic donors. Deceased donors have historically been the most common source of transplantable kidneys, but transplantation from living altruistic donors is becoming increasingly common. These "living" kidneys have two chief advantages. First, kidney grafts from living donors have better survival rates than those from deceased donors. Second, living donors can act as the first link in an altruistic chain of donations, wherein the first patient receives the altruistic donor's kidney, freeing her own donor to give her kidney to another compatible patient. This patient's willing donor can then continue the chain. Non-simultaneous chains avoid the main undesirable feature of non-simultaneous exchange cycles. Namely, if a donor backs out of the swap after a transplantation has already been performed in an exchange cycle, a patient would lose her donor's kidney without receiving one in return. In chains, on the other hand, even if a donor backs out after her patient receives a kidney, the anticipated next link in the chain is left no worse off than before, as her donor has not yet given up a kidney. Thus each successive transplantation in a chain can be delayed until a patient who is particularly compatible with the last transplantee's donor becomes available. As they need not be performed simultaneously, chains are particularly helpful in maximizing the number of compatible transplantations because the short-cycle feasibility constraints are relaxed for them. The longest chain ever completed included 60 people and 30 transplants (Sack 2012). The longest active chain as of November 2014 is coordinated by the University of Alabama at Birmingham School of Medicine and has so far included 56 people with 28 transplants. See Ashlagi et al. (2011, 2012) for more on transplantation chains.

The importance of chains motivates the study of what happens when some of the kidneys belong to an unattached donor in my framework. In terms of the model considered here, this translates into the following modification. While the set of patient-donor pairs remains $A = \{1, \ldots, n\}$, the set of altruistic donors is $A' = \{n+1, \ldots, p\}$. Each patient in A has a strict preference order over the set $A \cup A'$. The rest of the model remains the same. I am still interested in a mechanism that is invididually rational, ordinally efficient and anonymous. But before I define the modified 2CSE mechanism, I consider the question of implementability.

How does one need to modify Edmonds' theorem in order to guarantee that a family of probability-share allocations $\{P(i)\}_{i\in A}$ can be represented as a lottery over deterministic matchings where each patient-donor pair is either matched up with another such pair from the set A or receives a single kidney from A'?

To answer that, consider the following way to represent the profile of probability-share allocations as a symmetric bistochastic matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} & q_{1,n+1} & q_{1,n+2} & \cdots & q_{1p} \\ p_{21} & p_{22} & \cdots & p_{2n} & q_{2,n+1} & q_{2,n+2} & \cdots & q_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} & q_{n,n+1} & q_{n,n+2} & \cdots & q_{np} \\ q_{n+1,1} & q_{n+1,2} & \cdots & q_{n+1,n} & 1-\sum q_{i,n+1} & 0 & \cdots & 0 \\ q_{n+2,1} & q_{n+2,2} & \cdots & q_{n+2,n} & 0 & 1-\sum q_{i,n+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{p1} & q_{p2} & \cdots & q_{pn} & 0 & 0 & \cdots & 1-\sum q_{ip} \end{pmatrix},$$

Essentially, I treat each kidney i in A' as an artificial patient-donor pair which is matched with each patient-donor pair j in A with the same probability with which patient j receives kidney i. Thus $P(i,j) = P(j,i)$ represents the probability that patient j receives kidney i. At the same time, $P(i,l) = 0$ for all other $l \in A'\backslash\{i\}$ denotes that two kidneys in A' cannot be matched together and $P(i,i)$ is the probability with which kidney i is left unclaimed. In a similar fashion, one can think of the set of permutation matrices of dimension $p\times p$ as having a one-to-one and onto correspondence with all deterministic matchings with social endowment. Once one recasts the problem in this fashion, it becomes clear that if one applies the Edmonds' bounds to the bordered matrix, they get the same easy sufficient and necessary conditions for implementability in the social-endowment-augmented random assignment problem. Observe that the Edmonds constraint have no bite in the lower right quadrant of the representation matrix since all the off-diagonal entries are always zero.

The 2CSE mechanism is modified in a corresponding fashion. The main difference is that in terms of participation in the mechanism I treat all kidneys in A' as agents (i.e., patient-donor pairs) except for the fact that they have no preferences and can claim no probability shares from other patient-donor pairs. Thus a kidney $i \in A'$ is available to an agent $j \in A$ at time t if none of the Edmonds constraints containing i and j binds for M^t and the row corresponding to i is strictly less than 1.[31] Then j claiming probability shares from kidney i increases both $M^t(i,j)$ and $M^t(j,i)$, which correspond to q_{ij} and q_{ji} respectively.

As before, it is easy to see that the bistochastic matrix that is the outcome of this mechanism will satisfy individual rationality. Also, if I endow all patient-donor pairs in A with the same claiming speed (as in the 2CPS), the mechanism would also be anonymous in the sense that after permuting the names of the patient-donor pairs amongst themselves (i.e., within A) and the names of the other kidneys amongst themselves (i.e., within A') the new outcome of the mechanism would be represented by the same matrix with its rows and columns appropriately permuted. Finally, it is not hard to modify the proof of Proposition 6 to show that the mechanism is also ordinally efficient.

With this in mind, it is easy to see that if I add a number of kidneys from altruistic donors and all patient-donor pairs in A find each other unacceptable, the random allocation problem of Bogomolnaia and Moulin (2001) becomes a special case of the 2-cycle constrained setting from this section. The only difference

[31] One does not need to specifically worry about individual rationality since if patient j finds kidney i unacceptable, patient j would never want to claim probability shares from kidney i as she prefers kidney j, which is always available to her.

is that each agent now has an outside option: being left unmatched.

7 Longer Cycles and Other Constraints

In this section, I consider the feasibility of extending the main results of the previous section not only to the case $k > 2$ but also to cases with arbitrary ex-post constraints. More specifically, I look into the set of constraints that need to be imposed on the simultaneous-eating algorithm that would guarantee that the algorithm outputs a valid bistochastic assignment matrix that can be decomposed into a lottery over deterministic permutation matrices, each of which satisfies the desirable ex-post constraints. To fix ideas, I start by providing a few basic definitions and defining the simultaneous-eating algorithm subject to a set of constraints Ω before I continue.

We call any correspondence $C : \mathcal{P} \to 2^{\mathcal{M}} \setminus \{\varnothing\}$ a **constraint correspondence**. I interpret $C(\succ)$ as the set of allowable ex-post deterministic matchings for the preference profile $\succ$. I say that a constraint correspondence is anonymous if it is name-invariant. Formally, for an arbitrary bijective function $\pi : A \to A$ and a permutation matrix M, let $\Pi(M)$ be defined from M via $\Pi(M)(i,j) = M(\pi^{-1}(i), \pi^{-1}(j))$. Then I say that a constraint correspondence C satisfies **anonymity** if for all $\succ \in \mathcal{P}$ one has $\Pi(M(C(\succ))) = M(C(\succ^{\pi}))$, where $M(C(\succ))$ is the set of permutation matrices corresponding to each of the elements in $C(\succ)$. Given a set of ex-post deterministic matchings $C' \subset \mathcal{M}$ and a preference profile $\succ$, I say that a random matching $\mu \in \Delta C'$ is **C'-constrained ordinally efficient with respect to $\succ$** if there does not exist another random matching $\mu' \in \Delta C'$ such that $P_{\mu'}(i)$ first-order stochastically dominates $P_{\mu}(i)$ with respect to $\succ_i$ for all $i \in A$ and strictly so for some $i \in A$. Given a constraint correspondence C, I define a mechanism $f : \mathcal{P} \to \Delta\mathcal{M}$ to be **C-constrained ordinally efficient** if every $f(\succ)$ is $C(\succ)$-constrained ordinally efficient with respect to $\succ$.

Let Ω^0 be the collection of all ordered pairs (a, b) comprised of a function $a : A \times A \to \mathbb{R}_+$ and a scalar $b \in \mathbb{R}_+$. I interpret each one of these pairs as the representation of a constraint of the form

$$\sum_{(i,j) \in A \times A} a(i,j) M(i,j) \leq b,$$

where either $b \geq 0$ and $a(i,j) \geq 0$ for all (i,j). The simultaneous-eating algorithm will be subject to a subset Ω of Ω^0.

Definition 2. The generalized constrained simultaneous-eating algorithm subject to Ω. Each patient i has an associated claiming-speed function $e_i : [0,1] \to \mathbb{R}_+$ with $\int_0^1 e_i(t)dt = 1$. Time runs continuously starting at $t = 0$. For each point in time there is an associated sub-bistochastic matrix M^t where M^0 is the initial zero matrix. I say that kidney j is available to patient i at time $t \geq 0$ if none of the constraints in Ω for which $a(i,j) > 0$, bind at that time. Note that M^0 satisfies all the constraints in Ω and, in particular, all kidneys that patient i finds acceptable are available to her. At time t, each patient i claims with speed $e_i(t)$ the available remaining probability shares of her favorite reported kidney j among the kidneys that are available to i at that instance. That increases the probability that i receives j's kidney—i.e., it increases $M^t(i,j)$. Also note that $i = j$ can be true in this case. The algorithm ends at time $t = 1$ and the final matrix is M^1.

For simplicity, I refer to the outcome of the *generalized constrained simultaneous-eating algorithm subject to* Ω given a preference profile $\succ$ as $GCSE(\succ, e, \Omega)$. If the claiming-speed functions for all patients are the same (assumed, without loss of generality, to be $e_i(t) = 1$ for all $i \in A$ and $t \in [0,1]$), we will call the resulting mechanism the *generalized constrained probabilistic serial mechanism subject to* Ω (or GCPS for short) and will refer to its outcome by $GCPS(\succ, \Omega)$.

Note that the class of constraints Ω^0 contains "zero" constraints (such as $M(i,j) \leq 0$, which represents the case where patient i finds kidney j unacceptable and individual rationality is a requirement[32]), as well as constraints similar to the Edmonds constraints. Other types of constraints, such as $M(i,j) - M(h,l) \leq 1/2$ or $M(i,j) + M(h,l) \geq 1$ are not included. The first one of these has the potential to become slack after binding during the course of the algorithm. This is problematic since patient i would have stopped claiming shares from kidney j and moved on to her next-best available kidney. If that constraint becomes slack, however, patient i could benefit from coming back to kidney j and claiming more shares from it, which the algorithm does not allow. This would jeopardize the efficiency of the outcome. The second one of these constraints is not initially satisfied for the matrix M^0. It is crucial for each matrix M^t to satisfy all the constraints in Ω, however, since my goal is to define Ω in a way guaranteeing that M^1 is a valid bistochastic matrix that is implementable as a lottery over deterministic matchings that satisfy the ex-post constraints. Now I show that there exists a natural way to define the set of constraints so that they are all within Ω^0.

Definition 3. Lower contour set. Given a set $C' \subset \mathbb{R}_+^{n \times n}$, I say that the set defined by

$$\{M \in \mathbb{R}_+^{n \times n} | \exists M' \in C' : M' \geq M\}$$

is the *lower contour set of* C'. I denote it by $\mathrm{lcs}(C')$.

In what follows, I abuse notation in the following way. If $C' \subset \mathcal{M}$, I use $\mathrm{lcs}(C')$ to denote the lower contour set of the set of permutation matrices corresponding to each of the elements in C'. Similarly, $\mathrm{lcs}(\Delta C')$ denotes the lower contour set of the convex hull of those permutation matrices.

Proposition 7. *For any set $C' \subset \mathcal{M}$, there exists an essentially unique*[33] *minimal set of constraints $\Omega \subset \Omega^0$ such that*

$$\mathrm{lcs}(\Delta C') = \bigcap_{(a,b) \in \Omega} \left\{ M \in \mathbb{R}_+^{n \times n} \,\middle|\, \sum_{(i,j) \in A \times A} a(i,j) M(i,j) \leq b \right\}.$$

I denote these constraints by $\Omega^{C'}$.

Proposition 8. *For any non-empty set of allowable ex-post deterministic matchings $C' \subset \mathcal{M}$ and any preference profile $\succ$, the GCSE mechanism subject to $\Omega^{C'}$ terminates at an allowable bistochastic matrix that represents a C'-constrained ordinally efficient random matching in $\Delta C'$ with respect to $\succ$.*

Proposition 9. *If C is an anonymous constraint correspondence, the GCPS mechanism subject to $\Omega^{C(\succ)}$ for any given preference profile $\succ$ is C-constrained ordinally efficient and anonymous.*

In the following discussion, I attempt to unpack the intuition behind Propositions 7, 8 and 9. At first blush, finding a suitable set Ω guaranteeing that the outcome of the GCSE algorithm subject to Ω is C-constrained ordinally efficient appears quite difficult. The case $k = 2$ has a number of benefits. For one, the permutation matrices (representing the deterministic matchings) and bistochastic matrices (representing

[32] More formally, this means $P_\mu(i,j) = 0$ for all $\mu \in C(\succ)$.

[33] For the pairs (a,b) with $b > 0$, the constraints are unique up to rescaling by a positive scalar. For the case $b = 0$, this is not the case since multiple sets of constraints can be equivalent here: for example, $M(1,2) + M(1,3) \leq 0$ and $M(1,2) + 2M(1,3) \leq 0$ are equivalent to each other and also to the pair of constraints $M(1,2) \leq 0$ and $M(1,3) \leq 0$. They all denote the fact that $M(1,2) = M(1,3) = 0$ for all allowable allocations.

the random matchings) are all necessarily symmetric when $k = 2$ which simplifies the constraints that are sufficient and necessary for implementation.

One consequence of the simplicity of the Edmonds constraint is that any patient always has an acceptable available kidney at any point during the run of the mechanism's algorithm, thus vastly simplifying the individual rationality guarantee. Due to the fact that the Edmonds constraints do not pertain to the trace entries of the probability-shares matrix, as pointed out above, the only way that a patient would not find her donor's kidney available at any point during the run of the algorithm is if that kidney has had its entire unit mass of probability shares claimed already. But then, by the enforced symmetry of the procedure, this would imply that the patient has also fulfilled her unit demand. So each patient can always "retreat" to her donor's kidney even if all the other acceptable kidneys are unavailable to her. So the algorithm can never get "stuck."

So merely applying the Edmonds constraints, which are intended to characterize the convex hull of symmetric permutation matrices, to the higher-dimensional set of sub-bistochastic matrices, such as the interim probability-shares matrices at any instance during the running of the 2CSE's algorithm, is enough to guarantee that at no point does the algorithm get "stuck". There is no reason to believe, however, that constraints characterizing the convex hull of some other set of allowable deterministic matchings would have the same property, even if those constraints are simple. For example, the case of the object-allocation problem with unacceptabilities has a simple set of constraints guaranteeing individual rationality: $P_\mu(i, o) \leq 0$ if agent i finds object o unacceptable. However, naively attempting to run a simultaneous-eating mechanism with these constraints would quickly result in problems.

To be more specific, assume that there are two agents, 1 and 2, and two objects, o_1 and o_2. Agent 1 finds only object o_1 acceptable, while agent 2 finds both of them acceptable but prefers o_1. So the only constraint characterizing the individually rational polytope is $P_\mu(1, o_1) \leq 0$. If one attempts to run a simultaneous-eating algorithm with this as the only constraint, however, they would not be able to guarantee individual rationality. To see that, observe that if 2 starts claiming probability shares from her favorite object o_1, this would make the only individually rational allocation (where 1 gets o_1 with probability 1) impossible. The focus of Yılmaz (2010), who studies an individually-rational version of the PS mechanism, is characterizing the other constraints necessary for guaranteeing individual rationality. Those constraints are derived from Gale's Supply-Demand Theorem (Gale 1957) and essentially serve in the same manner as the Edmonds constraint serve to define the mechanism in the previous section.

This suggests, though, that any case other than the symmetric $k = 2$ would first require obtaining a characterization similar to Edmonds' theorem. Then one would need to find a set of additional constraints Ω (similar to those from Gale's theorem) that need to be imposed on the interim probability-shares matrices in order to guarantee that the bistochastic matrix obtained at the end of the algorithm satisfies the characterization from step one. This would still not be enough, however, as the resulting lottery also needs to be constrained ordinally efficient. It is easy to see that guaranteeing just implementability is easy: one can choose any allowable deterministic ex-post matching and set the constraints so that the patients can claim only shares of the object they receive in that matching. This need not be constrained-ordinally efficient though. So the constraints need to guarantee that the mechanism's algorithm always selects a constrained-ordinally efficient matching.

Consider an implementation of the simultaneous-eating algorithm under a certain set of constraints Ω. Let the preference profile be $\succ$ and let D' be the set of all matrices that represent lotteries over allowable deterministic matchings in $C(\succ)$. If at any time during the algorithm, there is an interim probability-share matrix M^t such that there does not exist $M' \in D'$ with $M' \geq M^t$, the algorithm would not output an allowable bistochastic matrix. This follows from the fact that the interim probability-share matrices are increasing in t. Thus, at the very least, we need the constraints to guarantee that $M^t \leq M'$ for some $M' \in D'$ for all t. This motivates Definition 2. It turns out that constraining the mechanism to the lower contour set of D' suffices in all cases: this guarantees that the algorithm terminates at an allowable matrix and that matrix represents a constrained ordinally efficient allocation.

The key observation summarized in Propositions 7 and 8 is that if D' is the convex hull of the allowable permutation matrices, the set $\text{lcs}(D')$ satisfies a handful of nice properties. First, if the GCSE algorithm is constrained within $\text{lcs}(D')$, the algorithm never gets "stuck" and always outputs a bistochastic matrix within D'. Second, $\text{lcs}(D')$ is a bounded convex polytope, so it equals the set of all the matrices whose entries satisfy certain constraints and, importantly, those constraints are well behaved in the sense that they are of the form

$$\sum_{(i,j)\in A\times A} a(i,j)M(i,j) \leq b$$

for some $b \geq 0$ and $a(i,j) \geq 0$ for all (i,j).[34] In other words, those constraints are from the set Ω^0.[35] As noted above, the constraints ensure that during the running of the GCSE mechanism's algorithm, once a constraint starts binding, it remains binding until the conclusion of the algorithm. Thus, whenever a constraint starts binding and a kidney becomes unavailable to a patient, that patient can move to the next highest-ranked available kidney and not worry about coming back to the one whose probability shares she was just claiming in case the constraint ever becomes slack again. Thus ordinal efficiency is not jeopardized. I note here that even though the objects in my setting have unit supply and their total supply equals the demand of the agents, the GCSE mechanism can be generalized in a straightforward manner to the multi-unit demand and/or supply, as well unequal total demand and supply.

Observe that in Proposition 9 I require that the constraint correspondence is anonymous in order to guarantee that GCPS is anaonymous. C-constrained ordinal efficiency would be preserved for any constraint correspondence C. Note that if the constraint correspondence C is generated from individual rationality and limiting cycle length to be at most k, C-constrained ordinal efficiency is equivalent to k-constrained ordinal efficiency.

So the main result, Proposition 8, together with Proposition 9 guarantee that GCPS is constrained ordinally efficient, *no matter what the set of allowable ex-post deterministic matchings is* and possibly anonymous if the constraint correspondence is anonymous itself. The constraint correspondence is anonymous if the constraints placed on the possible ex-post deterministic matchings do not depend on the names of the agents. It is easy to see that individual rationality and satisfaction of the k-cycle constraints jointly or separately define an anonymous constraint correspondence. Since individual rationality and the k-cycle constraints pertain only to the possible ex-post deterministic matchings, a consequence of that result is that for any $k \geq 2$ the GCPS mechanism is individually rational, k-constrained ordinally efficient and anonymous. I show later that, just as in the case $k = 2$, there does not exist a mechanism that satisfies k-constrained ordinal efficiency, individual rationality and weak strategyproofness.

[34] These exclude the non-negativity constraints but those are automatically satisfied since the interim probability-share matrix is initially the zero matrix and it is increasing in t.

[35] Computing these constraints given the set of extreme points is known as the "convex hull problem". This problem has been well studied in the computational-geometry literature. See Chazelle (1993); Clarkson et al. (1993); Burnikel et al. (1994); Barber et al. (1996), as well as Seidel (2004) for a recent survey. Since the polytopes of interest here are all 0/1-polytopes (Ziegler 2000), there is the potential to define specialized algorithms with good performance (potentially even running in polynomial time).

Corollary 10. *Fix $k \geq 2$. If the constraint correspondence C is such that $C(\succ)$ equals the set of individually rational matchings in $\mathcal{M}_k$, the GCPS subject to $\Omega^{C(\succ)}$ for each preference profile $\succ$ is k-constrained ordinally efficient, anonymous and individually rational.*

The power of the result is not just confined to considerations of individual rationality and cycle implementability. Any desired outcome that can be represented as a constraint on the final possible deterministic matchings can be accommodated here. I next consider a few examples as illustrations of the result's power. The first three examples, for instance, show that the ordinal efficiency results of a handful of preceding papers are implied by Proposition 8 but, furthermore, that the constraints that they use to guarantee implementability are special cases of the constraints defining the lower contour set of permissible matchings.

Example 3. (No constraints.) If $C(\succ) = \mathcal{M}$, we are back to the realm of pure object-allocation problems without any additional constraints as in Bogomolnaia and Moulin (2001). The constraints defining the set $\text{lcs}(\Delta C(\succ))$ are simply the sub-bistochasticity conditions: $M \in \text{lcs}(\Delta C(\succ))$ if and only if

$$\sum_j M(i,j) \leq 1 \text{ and } \sum_i M(i,j) \leq 1 \text{ for all } i,j \in A, \tag{3}$$

which are the constraints used in Bogomolnaia and Moulin (2001).

Example 4. (Individual rationality.) If

$$C(\succ) = \{m \in \mathcal{M} | i \succ_i j \Rightarrow m(i) = 0\},$$

the only ex-post constraint here is individual rationality. Yılmaz (2010) adapts the simultaneous-eating algorithm by imposing constraints to guarantee that it would output a matrix that satisfies individual rationality. In addition to the sub-bistochasticity constraints (3), the following must also be true for any interim matrix

$$|U_T| - |T| \geq \sum_{i \in A \setminus T, j \in U_T} M^t(i,j) \text{ for all } T \subset A, \tag{4}$$

where $U_T \subset A$ is the set of objects that at least one agent in T finds acceptable. Yılmaz' (2010) result implies that if a sub-bistochastic matrix satisfies these constraints, then that matrix is in $\text{lcs}(\Delta C(\succ))$ because, as noted above, if an interim matrix is ever not in this set, the outcome of the algorithm would not satisfy individual rationality. Here I show that the converse is also true. Let $M \in \text{lcs}(\Delta C(\succ))$ and let M' be a bistochastic matrix representing an individually rational lottery with $M' \geq M$. For any $T \subset A$, the agents in T must have received only probability shares in U_T by individual rationality. The remainder of the probability shares of objects in U_T must be distributed among agents outside T. Since agents and objects have unitary demand and supply, respectively, the following must be true:

$$\sum_{i \in A \setminus T, j \in U_T} M'(i,j) = |U_T| - |T|.$$

Since $M(i,j) \leq M'(i,j)$, we have

$$\sum_{i \in A \setminus T, j \in U_T} M(i,j) \leq |U_T| - |T|.$$

Thus, in addition to the non-negativity constraints, (3) and (4) characterize the set $\text{lcs}(\Delta C(\succ))$.

Example 5. (Individual rationality and constraint on the trading-cycle lengths: $k = 2$.) Now consider the case $C(\succ)$ equals all individually rational matchings in $\mathcal{M}_2$. The GCPS in this setting is individually rational, 2-constrained ordinally efficient and anonymous but differs from 2CPS. The set of constraints describing the polytope $\text{lcs}(\Delta C(\succ))$ for the case $n = 3, k = 2$ are $M(i,j) \leq 0$ if $i \succ_i j$ or $j \succ_j i$, the sub-bistochasticity constraints from (3), as well as

$$\begin{aligned}
M(1,2) + M(1,3) + M(2,3) &\leq 1, \\
M(1,2) + M(1,3) + M(3,2) &\leq 1, \\
M(1,2) + M(3,1) + M(2,3) &\leq 1, \\
M(1,2) + M(3,1) + M(3,2) &\leq 1, \\
M(2,1) + M(1,3) + M(2,3) &\leq 1, \\
M(2,1) + M(1,3) + M(3,2) &\leq 1, \\
M(2,1) + M(3,1) + M(2,3) &\leq 1, \\
M(2,1) + M(3,1) + M(3,2) &\leq 1.
\end{aligned}$$

If the preference profile $\succ$ is

$$\begin{aligned}
\succ_1 &: 2 \succ_1 3 \succ_1 1, \\
\succ_2 &: 1 \succ_2 3 \succ_2 2, \\
\succ_3 &: 1 \succ_3 2 \succ_3 3,
\end{aligned}$$

then

$$GCPS(\succ, \Omega^{C(\succ)}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix},$$

while the 2CPS selects

$$2CPS(\succ) = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 2/3 & 1/3 & 0 \\ 1/3 & 0 & 2/3 \end{pmatrix}.$$

Example 6. (Bihierarchical constraints.) Budish et al. (2013) define a class of constraints called bihierarchical constraints and generalize the probabilistic serial mechanism for those constraints. Even though their paper allows multi-unit supply of objects, as noted above, the GCPS mechanism readily generalizes to that case. A set of bihierarchical constraints includes maximum quotas placed on all rows (agent demand) and all columns (object supply). In the case of single-unit demand and supply, these reduce to (3). Additionally, the constraints may also include constraints placed on some subcolumns such that for any j the constraints

$$\sum_{i \in A'} M(i,j) \leq b' \text{ and } \sum_{i \in A''} M(i,j) \leq b''$$

must satisfy either $A' \subset A''$, $A'' \subset A'$ or $A' \cap A'' = \varnothing$.

In the case of school choice, for example, the subcolumnar constraints can be interpreted as not allowing too many students with certain domicile neighborhood or background characteristics into a given school. In my language, the constraint correspondence C here is constant. For any $\succ$, the set $C(\succ)$ equals all matchings in $\mathcal{M}$ that are represented by permutation matrices that satisfy all the bihierarchical constraints. Analogously to Example 4, the results of Budish et al. (2013) imply that if an interim sub-bistochastic matrix satisfies the bihierarchical constraints, then it is in the set $\mathrm{lcs}(\Delta C(\succ))$. The converse is straightforward as well: if $M \in \mathrm{lcs}(\Delta C(\succ))$, then there exists some assignment matrix $M' \geq M$ that satisfies the bihierarchical constraints. Since $M \leq M'$ and all constraints are from the set Ω^0, M must satisfy them as well.

Maximum-cardinality matchings are currently essentially, as noted above, the main efficiency criterion for kidney-exchange clearinghouses. Thus the GCPS mechanism allows generalizing the workhorse mechanism introduced by Roth et al. (2005a), whose main consideration is maximizing the number of patients receiving kidneys, by adding any other applicable constraints and, crucially, by respecting any strict preferences that the agents might have over the available kidneys.

For the case $k = 3$, another salient anonymous constraint correspondence requires all three-way exchanges to have a *back arc*. That is to say, if a cycle $1 \to 2 \to 3 \to 1$ is selected with positive probability, it is required that at least one of patients $1, 2,$ or 3 finds kidney $3, 1,$ or 2, respectively, acceptable. Then if one of the patients becomes too sick to undergo a transplantation and the proposed three-way exchange cannot go through, then there is a chance that the remaining two agents would be able to form a two-way exchange. This is known as failure-aware kidney exchange (Dickerson et al. 2013).

Another important class of problems that can be accommodated by the GCSE mechanism are two-sided matching problems. The constraints in that case divide the agents in two groups and make it impossible for any agent in either group to be matched with any other agent in that group. The simple marriage problem (or the one-to-one two-sided matching problem) can be accommodated under 2CPS as it is a special case of the roommate problem. However, since Proposition 8 can be extended to allow for multi-unit demand and supply, it holds for many-to-many two-sided markets as well.

8 Incentives and Impossibility Results

In this section, I discuss the incentive properties of the 2CPS and GCPS mechanisms.

Proposition 11. *The 2CPS mechanism is not weakly strategyproof for $n \geq 3$.*

Proof. Consider the following counterexample for $A = \{1, 2, 3\}$:

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

Under these preferences, the 2CPS mechanism matches couple 1 with couple 2 with probability $1/3$ and with couple 3 with probability $2/3$. However, if patient 1 instead reported kidney 2 as unacceptable, the 2CPS will match couples 1 and 3 with probability 1, thus strictly improving 1's probability-share allocation in FOSD manner. □

The 2CPS mechanism fails to be even weakly strategyproof. Since I allow the presence of unacceptabilities, these poor incentive properties are not surprising since most interesting mechanism will fail weak strategyproofness when one allows unacceptabilities (for example, in the object-allocation setting, the PS mechanism is not weakly strategyproofness if one allows for unacceptabilities).

I next present an impossibility result that shows that the 2CPS mechanism is indeed a second best mechanism in the sense that there does not exist a mechanism that is individually rational, 2-constrained ex-post efficient, and weakly strategy proof.

Proposition 12. *There does not exist a random mechanism that is individually rational,* 2*-constrained ex-post efficient and weakly strategy proof whenever* $n \geq 4$*.*

I end with the impossibility result for general cycle-length constraints.

Proposition 13. *For any* $k > 2$*, there does not exist a random mechanism that is individually rational,* k*-constrained ex-post efficient and weakly strategyproof whenever* $n \geq k+1$*.*

I next argue that Proposition 13 is tight in the sense that there are mechanisms that respect the k-cycle constraints and satisfy any two of the three axioms, as well as a mechanism (in this case, Top Trading Cycles) which satisfies the three axioms but does not satisfy the k-cycle constraints. Clearly the GCSE mechanism is individually rational and k-constrained ex-post efficient. The no-trade mechanism is individually rational and (weakly) strategyproof.

Finally, consider the following example of a mechanism that satisfies the k-cycle constraints, constrained ex-post efficiency and weak strategyproofness. Let there be some priority order over A and let each agent i point at the agent who has i's highest-ranked object. Starting with the agent, call her 1, who is highest in the priority order, one of three things can occur. One possibility is that 1 is part of a cycle of length no longer than k, in which case we perform the trades implied by that cycle, remove all agents and objects in that cycle, and move to the next step. Another possibility is that 1 is part of a cycle of length longer than k. In that case, considering the chain starting at 1, we take the k-th agent in that chain and close the cycle by giving her 1's object. So for example, if $k = 2$ and 1 points at 2 who points at 3 who points back at 1, the first selected trading cycle here would have 1 receiving 2's object and 2 receiving 1's object, even if that is individually irrational for her. The last possibility is that 1 is not part of a cycle. Then starting with the chain anchored at 1, we should reach some cycle. For example, in a very simple case we could have

$$1 \to 2 \to 3 \to 2.$$

If $k = 4$ here, we can "break the cycle" in a strategyproof way by considering who agent 3's second highest choice is. If that is 4 for example, we implement the trade implied by the cycle $1 \to 2 \to 3 \to 4 \to 1$. It is not hard to prove that this mechanism is k-constrained ex-post efficient and strategyproof. It is, however, obviously not individually rational.

9 Conclusion

In this paper, I propose a mechanism that is suitable for the setting of kidney exchange with strict ordinal preferences. The mechanism is individually rational so it recommends compatible transplantations and provides sufficient incentives for compatible patient-donor pairs to enter the kidney-exchange program. It is also ordinally efficient and anonymous. I also provide a method to generalize the mechanism to an arbitrary matching setting with arbitrary constraints: regardless of what the set of acceptable ex-post allocations is, the mechanism would select an ordinally efficient lottery over that set and would do that in an anonymous way if the constraints allow it. The settings include one-sided (such as object-exchange and object-assignment problems) and two-sided matching markets (such as school-choice problems and the assignment of workers to institutions). The constraints can include individual rationality, cycle constraints, maximal-cardinality matchings, and various quotas, caps etc.

One question I haven't addressed until now is the problem of considering ordinal versus cardinal preferences. For example, the United Network for Organ Sharing/Organ Procurement and Transplantation Network's (UNOS/OPTN) guidelines for deceased donors' kidney allocation changed recently to address considerations of differential survival rates of kidneys.[36] Their system computes a Kidney Donor Profile Index, which indicates how long a kidney is likely to function once transplanted. The goal is to maximize the net life-years benefit of transplanted kidneys over the life expectancy under dialysis treatment. So why shouldn't the matching algorithm simply select an allocation that maximizes the sum of expected gained life-years? The answer is multi-fold. First, these cardinal preferences would be hard to estimate. See Freeman (2007) for a discussion, for example. Any proposal for estimating these would necessarily involve ad-hoc assumptions: for example, Freeman (2007) mentions that for the purposes of quality-of-life adjustment, the social planner needs to decide how much is a year with a functioning renal graft worth in terms of years on dialysis. Furthermore, the expected life-year benefit for a given patient-donor pair is hard to determine with much precision, or at least harder than defining ordinal preferences for the patient over the available kidneys. Second, maximizing total gained life-years creates fairness issues. Specifically, younger and healthier patients receive preferential treatment since a kidney is likelier to extend their lives more than older patients'. One can account for that by making suitable adjustments. For example, the system adopted by UNOS/OPTN makes adjustments to give preferential treatment to patients who have been on the waiting list for a long time. But, again, any such adjustment would necessarily be ad-hoc. I believe that the mechanism proposed here does a better job of balancing utility and justice, as desired by UNOS (Wallis et al. 2011). Finally, a utility-maximizing system would introduce a new set of incentive-compatibility issues. For example, doctors could falsify their patients' medical records to make them appear healthier (to inflate the estimate of expected life-years they can gain from a kidney) or sicker (if that gives them a waitlist priority). For accounts of a series of similar cases in the liver- and heart-transplantation programs in the US and Germany, see Snyder (2010); Pondrom (2013); Roth (2014).

There is significant potential for future work extending and refining the results presented here. Extending the 2CPS mechanism or the GCPS mechanism to the case of non-strict ordinal preferences would be valuable. I note here that the techniques from network flow theory used by Katta and Sethuraman (2006) (and adapted in Yılmaz (2009) and Budish et al. (2013)) cannot be extended in a straightforward manner to my setting. The main issues is that the Edmonds constraints are not nested within each other so constructing an auxiliary network representing the constraints would not help here.

Understanding the properties of the general mechanism from Proposition 8 would also be valuable. Examples include finding a natural fairness property it satisfies. Also, that mechanism embeds Bogomolnaia and Moulin's (2001) PS and Budish et al.'s (2013) Generalized PS mechanisms, which are weakly strategyproof. However, it also includes those mechanisms with unacceptabilities, where weak strategyproofness fails (Yılmaz 2010). A natural question to ask is: what conditions guarantee weak strategyproofness. For that matter, what conditions guarantee stronger incentive criteria, such as convex strategyproofness as defined in Balbuzanov (2014)?[37] Finally, this paper assumes a hard cap on the possible length of trading cycles. It would be interesting to see if one can arrive at a better mechanism by relaxing the requirement that the mechanism satisfies the cycle constraints with probability 1. A mechanism that has better properties and

[36]See Roth (2013) for more details.

[37]Convex strategyproofness requires that for a given ordinal preference relation, there exists a compatible von Neumann-Morgenstern utility vector such that an agent with that utility vector prefers reporting truthfully.

fails to satisfy the cycle-length constraints "relatively seldom" could be acceptable in the sense that the extra cost of the occasional long trading cycles might be outweighed by the benefit of a better overall mechanism.[38]

As noted in Section B.2, the 2CSE mechanism does not fully characterize all possible ordinally efficient random allocations. It would be valuable to know what properties characterize the allocations that can be selected by that mechanism. Conversely, does there exist a mechanism parameterized by a certain vector that selects all possible ordinally efficient allocations by varying the parameter? Another open area is finding a "better" fairness condition that the 2CPS mechanism satisfies (see Section B.1). While anonymity implies that patients' names do not matter, there are anonymous mechanisms that are arguably unfair.[39] However, since the 2CPS affords all agents the same initial conditions and is procedurally fair, I expect that it would satisfy some stronger fairness conditions.

[38]For recent work on matching mechanisms that satisfy desirable properties in an approximate manner see Budish (2011) and Akbarpour and Nikzad (2014).

[39]For example, consider an object-allocation setting with two agents $\{1, 2\}$ with multi-unit demand and two objects $\{a, b\}$, where being assigned both objects is preferred to having either one of them. Let f be a mechanism that gives both objects to the agent whose preferences are $\{a, b\} \succ a \succ b$ whenever the preferences are different and flips a fair coin for who gets what if the preferences are the same. The mechanism is clearly anonymous but it can be argued that it is unfair in the sense that one agent receives an "unreasonably large" share of the endowment under certain preferences.

References

Abdulkadiroğlu, A. and T. Sönmez (1998): "Random serial dictatorship and the core from random endowments in house allocation problems," *Econometrica*, 66, 689–701.

——— (2003): "Ordinal efficiency and dominated sets of assignments," *Journal of Economic Theory*, 112, 157–172.

——— (2013): "Matching markets: Theory and practice," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arello, and E. Dekel, New York: Cambridge University Press, vol. 1, 3–47.

Akbarpour, M. and A. Nikzad (2014): "Approximate random allocation mechanisms," Mimeo.

Ashlagi, I., D. Gamarnik, M. A. Rees, and A. E. Roth (2012): "The need for (long) chains in kidney exchange," Mimeo.

Ashlagi, I., D. S. Gilchrist, A. E. Roth, and M. A. Rees (2011): "Nonsimultaneous chains and dominos in kidney-paired donation-revisited," *American Journal of Transplantation*, 11, 984–994.

Ashlagi, I., Y. Kanoria, and J. D. Leshno (2014): "Unbalanced random matching markets: The stark effect of competition," Mimeo.

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

Athanassoglou, S. and J. Sethuraman (2011): "House allocation with fractional endowments," *International Journal of Game Theory*, 40, 481–513.

Balbuzanov, I. (2014): "Convex strategyproofness with an application to the probabilistic serial mechanism," Mimeo.

Balinski, M. L. (1972): "Establishing the matching polytope," *Journal of Combinatorial Theory, Series B*, 13, 1–13.

Barber, C. B., D. P. Dobkin, and H. Huhdanpaa (1996): "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, 22, 469–483.

Birkhoff, G. (1946): "Tres observaciones sobre el algebra lineal," *Revista Facultad de Ciencias Exactas, Puras y Aplicadas Universidad Nacional de Tucumán, Serie A (Matemáticas y Física Teórica)*, 5, 147–151.

Bogomolnaia, A. and E. J. Heo (2012): "Probabilistic assignment of objects: Characterizing the serial rule," *Journal of Economic Theory*, 147, 2072–2082.

Bogomolnaia, A. and H. Moulin (2001): "A new solution to the random assignment problem," *Journal of Economic Theory*, 100, 295–328.

Budish, E. (2011): "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," *Journal of Political Economy*, 119, 1061–1103.

Budish, E., Y.-K. Che, F. Kojima, and P. Milgrom (2013): "Designing random allocation mechanisms: Theory and application," *American Economic Review*, 103, 585–623.

Burnikel, C., K. Mehlhorn, and S. Schirra (1994): "On degeneracy in geometric computations," in *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, SODA '94, 16–23.

Carroll, G. (2010): "An efficiency theorem for incompletely known preferences," *Journal of Economic Theory*, 145, 2463–2470.

Chazelle, B. (1993): "An optimal convex hull algorithm in any fixed dimension," *Discrete & Computational Geometry*, 10, 377–409.

Che, Y.-K. and F. Kojima (2010): "Asymptotic equivalence of probabilistic serial and random priority mechanisms," *Econometrica*, 78, 1625–1672.

Chertow, G. M., E. L. Milford, H. S. Mackenzie, and B. M. Brenner (1996): "Antigen-independent determinants of cadaveric kidney transplant failure," *The Journal of the American Medical Association*, 276, 1732–1736.

Clarkson, K. L., K. Mehlhorn, and R. Seidel (1993): "Four results on randomized incremental constructions," *Computational Geometry*, 3, 185–212.

Constantino, M., X. Klimentova, A. Viana, and A. Rais (2013): "New insights on integer-programming models for the kidney exchange problem," *European Journal of Operational Research*, 231, 57–68.

Cruse, A. B. (1975): "A note on symmetric doubly-stochastic matrices," *Discrete Mathematics*, 13, 109–119.

De Klerk, M., K. M. Keizer, F. H. J. Claas, M. Witvliet, B. J. J. M. Haase-Kromwijk, and W. Weimar (2005): "The Dutch national living donor kidney exchange program," *American Journal of Transplantation*, 5, 2302–2305.

Delmonico, F. L. (2004): "Exchanging kidneys–advances in living-donor transplantation," *New England Journal of Medicine*, 350, 1812–1813.

Dickerson, J. P., A. D. Procaccia, and T. Sandholm (2013): "Failure-aware kidney exchange," in *Proceedings of the fourteenth ACM conference on electronic commerce*, 323–340.

Edmonds, J. (1965): "Maximum matching and a polyhedron with 0,l-vertices," *Journal of Research of the National Bureau of Standards–B. Mathematics and Mathematical Physics*, 69B, 125–130.

Fatemi, F. (2012): "The regulated market for kidneys in Iran," in *Auctions, Market Mechanisms, and Their Applications*, ed. by P. Coles, S. Das, S. Lahaie, and B. Szymanski, Springer, 62–75.

Freeman, R. B. (2007): "Survival benefit: Quality versus quantity and trade-offs in developing new renal allocation systems," *American Journal of Transplantation*, 7, 1043–1046.

Gale, D. (1957): "A theorem on flows in networks," *Pacific Journal of Mathematics*, 7, 1073–1082.

Gale, D. and L. S. Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9–15.

Gentry, S. E., D. L. Segev, M. Simmerling, and R. A. Montgomery (2007): "Expanding kidney paired donation through participation by compatible pairs," *American Journal of Transplantation*, 7, 2361–2370.

Gjertson, D. W. (2003): "Explainable variation in renal transplant outcomes: A comparison of standard and expanded criteria donors," *Clinical Transplants*, 17, 303–314.

Hashimoto, T., D. Hirata, O. Kesten, M. Kurino, and M. U. Ünver (2014): "Two axiomatic approaches to the probabilistic serial mechanism," *Theoretical Economics*, 9, 253–277.

Heo, E. J. (2014a): "The extended serial correspondence on a rich preference domain," *International Journal of Game Theory*, 43, 439–454.

——— (2014b): "Probabilistic assignment problem with multi-unit demands: A generalization of the serial rule and its characterization," *Journal of Mathematical Economics*, 54, 40–47.

Heo, E. J. and O. Yilmaz (2013): "A characterization of the extended serial correspondence," Mimeo.

Kamada, Y. and F. Kojima (2014a): "General theory of matching under distributional constraints," Mimeo.

——— (2014b): "Stability concepts in matching under distributional constraints," Mimeo.

——— (forthcoming): "Efficient matching under distributional constraints: Theory and applications," *American Economic Review*.

Katta, A.-K. and J. Sethuraman (2006): "A Solution to the Random Assignment Problem on the Full Preference Domain," *Journal of Economic Theory*, 131, 231–250.

Katz, M. (1970): "On the extreme points of a certain convex polytope," *Journal of Combinatorial Theory*, 8, 417–423.

Keizer, K. M., M. De Klerk, B. J. J. M. Haase-Kromwijk, and W. Weimar (2005): "The Dutch algorithm for allocation in living donor kidney exchange," *Transplantation Proceedings*, 37, 589–591.

Kesten, O. and M. U. Ünver (forthcoming): "A theory of school-choice lotteries," *Theoretical Economics*.

Knuth, D. E., R. Motwani, and B. Pittel (1990): "Stable husbands," in *Proceedings of the first annual ACM-SIAM symposium on discrete algorithms*, 397–404.

Kojima, F. (2009): "Random assignment of multiple indivisible objects," *Mathematical Social Sciences*, 57, 134–142.

Kojima, F. and M. Manea (2010): "Incentives in the probabilistic serial mechanism," *Journal of Economic Theory*, 145, 106–123.

Koning, O. H. J., R. J. Ploeg, J. H. Van Bockel, M. Groenewegen, F. J. van der Woude, G. G. Persijn, and J. Hermans (1997): "Risk factors for delayed graft function in cadaveric kidney transplantation: A prospective study of renal function and graft survival after preservation with University of Wisconsin Solution in multi-organ donors," *Transplantation*, 63, 1620–1628.

Liu, Q. and M. Pycia (2013): "Ordinal efficiency, fairness and incentives in large markets," Mimeo.

Lucan, M. (2007): "Five years of single-center experience with paired kidney exchange transplantation," *Transplantation Proceedings*, 39, 1371–1375.

Ma, J. (1994): "Strategy-proofness and the strict core in a market with indivisibilities," *International Journal of Game Theory*, 23, 75–83.

Manea, M. (2008): "A constructive proof of the ordinal efficiency welfare theorem," *Journal of Economic Theory*, 141, 276–281.

Manlove, D. F. and G. O'Malley (2012): "Paired and altruistic kidney donation in the UK: Algorithms and experimentation," in *Experimental Algorithms*, ed. by R. Klasing, Springer, 271–282.

McLennan, A. (2002): "Ordinal efficiency and the polyhedral separating hyperplane theorem," *Journal of Economic Theory*, 105, 435–449.

Miyagawa, E. (2002): "Strategy-proofness and the core in house allocation problems," *Games and Economic Behavior*, 38, 347–361.

Montgomery, R. A., B. E. Lonze, K. E. King, E. S. Kraus, L. M. Kucirka, J. E. Locke, D. S. Warren, C. E. Simpkins, N. N. Dagher, A. L. Singer, A. A. Zachary, and D. L. Segev (2011): "Desensitization in HLA-incompatible kidney recipients and survival," *New England Journal of Medicine*, 365, 318–326.

Moulin, H. (1995): *Cooperative Microeconomics*, Prentice Hall.

Nguyen, T., A. Peivandi, and R. Vohra (2014): "One-sided matching with limited complementarities," Mimeo.

Nicoló, A. and C. Rodríguez-Álvarez (2011): "Age-Based Preferences: Incorporating Compatible Pairs into Paired Kidney Exchange," Mimeo.

——— (2012): "Transplant quality and patients' preferences in paired kidney exchange," *Games and Economic Behavior*, 74, 299–310.

——— (2013): "Incentive compatibility and feasibility constraints in housing markets," *Social Choice and Welfare*, 41, 625–635.

Øien, C. M., A. V. Reisæter, T. Leivestad, F. W. Dekker, P. D. Line, and I. Os (2007): "Living donor kidney transplantation: the effects of donor age and gender on short-and long-term outcomes," *Transplantation*, 83, 600–606.

Ojo, A. O., J. A. Hanson, R. A. Wolfe, A. B. Leichtman, L. Y. Agodoa, and F. K. Port (2000): "Long-term survival in renal transplant recipients with graft function," *Kidney International*, 57, 307–313.

Opelz, G. (1997): "Impact of HLA compatibility on survival of kidney transplants from unrelated live donors," *Transplantation*, 64, 1473–1475.

——— (1998): "HLA compatibility and kidney grafts from unrelated live donors," *Transplantation Proceedings*, 30, 704–705.

Opelz, G. and B. Döhler (2007): "Effect of human leukocyte antigen compatibility on kidney graft survival: comparative analysis of two decades," *Transplantation*, 84, 137–143.

Park, K., J. H. Lee, K. H. Huh, S. I. Kim, and Y. S. Kim (2004): "Exchange living-donor kidney transplantation: Diminution of donor organ shortage," *Transplantation Proceedings*, 36, 2949–2951.

Park, K., J. I. Moon, S. I. Kim, and Y. S. Kim (1999): "Exchange donor program in kidney transplantation," *Transplantation*, 67, 336–338.

Pessione, F., S. Cohen, D. Durand, M. Hourmant, M. Kessler, C. Legendre, G. Mourad, C. Noël, M.-N. Peraldi, C. Pouteil-Noble, P. Tuppin, and C. Hiesse (2003): "Multivariate analysis of donor risk factors for graft survival in kidney transplantation," *Transplantation*, 75, 361–367.

Pittel, B. (1989): "The average number of stable matchings," *SIAM Journal on Discrete Mathematics*, 2, 530–549.

——— (1992): "On likely solutions of a stable marriage problem," *The Annals of Applied Probability*, 2, 358–401.

Pondrom, S. (2013): "Trust is everything," *American Journal of Transplantation*, 13, 1115–1116.

Pycia, M. and M. U. Ünver (2014): "Decomposing random mechanisms," Mimeo.

Roth, A. E. (2007): "Repugnance as a constraint on markets," *Journal of Economic Perspectives*, 21, 37–58.

——— (2013): "UNOS Approves New Deceased Donor Allocation Rules," http://marketdesigner.blogspot.com/2013/06/optnunos-board-approves-significant.html.

——— (2014): "More Transplant Troubles in Germany," http://marketdesigner.blogspot.com/2014/10/more-transplant-troubles-in-germany.html.

Roth, A. E. and A. Postlewaite (1977): "Weak versus strong domination in a market with indivisible goods," *Journal of Mathematical Economics*, 4, 131–137.

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

——— (2005a): "Pairwise kidney exchange," *Journal of Economic Theory*, 125, 151–188.

——— (2005b): "A kidney exchange clearinghouse in New England," *American Economic Review Papers and Proceedings*, 95, 376–380.

——— (2007): "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *American Economic Review*, 97, 828–851.

Sack, K. (2012): "60 Lives, 30 Kidneys, All Linked," http://www.nytimes.com/2012/02/19/health/lives-forever-linked-through-kidney-transplant-chain-124.html.

Sasaki, N. and A. Idica (2010): "The HLA-matching effect in different cohorts of kidney transplant recipients: 10 years later," *Clinical Transplants*, 261–282.

Schrijver, A. (1986): *Theory of linear and integer programming*, John Wiley & Sons.

——— (2003): *Combinatorial optimization: Polyhedra and efficiency*, Springer.

Seidel, R. (2004): "Convex hull computations," in *Handbook of Discrete and Computational Geometry*, ed. by J. E. Goodman and J. O'Rourke, CRC Press, 495–512.

Shapley, L. S. and H. Scarf (1974): "On cores and indivisibility," *Journal of mathematical economics*, 1, 23–37.

Snyder, J. (2010): "Gaming the liver transplant market," *Journal of Law, Economics, and Organization*, 26, 546–568.

Sönmez, T. and M. U. Ünver (2011): "Matching, allocation, and exchange of discrete resources," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, North-Holland, vol. 1A, 781–852.

——— (2013): "Market design for kidney exchange," in *The Handbook of Market Design*, ed. by Z. Neeman, M. Niederle, A. E. Roth, and N. Vulkan, Oxford University Press, 93–137.

——— (2014): "Altruistically unbalanced kidney exchange," *Journal of Economic Theory*, 152, 105–129.

The New York Times Editorial Board (2014): "Ways to Reduce the Kidney Shortage," http://www.nytimes.com/2014/09/02/opinion/ways-to-reduce-the-kidney-shortage.html.

Tobian, A. A. R., R. S. Shirey, R. A. Montgomery, P. M. Ness, and K. E. King (2008): "The critical role of plasmapheresis in ABO-incompatible renal transplantation," *Transfusion*, 48, 2453–2460.

von Neumann, J. (1953): "A certain zero-sum two-person game equivalent to the optimal assignment problem," in *Contributions to the Theory of Games*, ed. by W. Kuhn and A. W. Tucker, Princeton: Princeton University Press, vol. 2, 5–12.

Wallis, C. B., K. P. Samy, A. E. Roth, and M. A. Rees (2011): "Kidney paired donation," *Nephrology Dialysis Transplantation*, 26, 2091–2099.

Wolfe, R. A., V. B. Ashby, E. L. Milford, A. O. Ojo, R. E. Ettenger, L. Y. Agodoa, P. J. Held, and F. K. Port (1999): "Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant," *New England Journal of Medicine*, 341, 1725–1730.

Yilmaz, Ö. (2009): "Random assignment under weak preferences," *Games and Economic Behavior*, 66, 546–558.

——— (2010): "The probabilistic serial mechanism with private endowments," *Games and Economic Behavior*, 69, 475–491.

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

Ziegler, G. M. (2000): "Lectures on 0/1-Polytopes," in *Polytopes Combinatorics and Computation*, ed. by G. Kalai and G. M. Ziegler, Birkhäuser Basel, vol. 29 of *DMV Seminar*, 1–41.

——— (2007): *Lectures on Polytopes*, Springer.

A Omitted Proofs

Proof of Proposition 3: I consider the following algorithm for the implementation of the TTC mechanism: at each step, we identify and remove only one trading cycle.

Fix the set of patient-donor pairs $A = \{1, 2, \ldots, n\}$ and the maximum allowed cycle length k, and consider the directed graph induced by each patient pointing at her highest-ranked kidney. I would like to estimate the probability that there exist no directed cycles of length less than or equal to k in that graph. The probability that patient 1 is pointing at a kidney other than kidney 1 is $\frac{n-1}{n}$. If this is the case, I assume without loss of generality that patient 1 is pointing at kidney 2. Then the probability that patient 2 is pointing at a kidney other than 1 and 2 is $\frac{n-2}{n}$. Again without loss of generality, I assume that patient 2 is pointing at kidney 3 if she is not pointing at 1 or 2.

We can continue in a similar fashion: at each stage, the probability that patient $m \leq k$ is not pointing at one of the kidneys $1, 2, \ldots, m$ (and thus being part of a cycle of length no greater than k) is $\frac{n-m}{n}$. The probability that patient $k+1$ points at kidney 1 (thus closing a cycle of length $k+1$) or at any of the kidneys $k+2, \ldots, n$ is $\frac{n-k}{n}$. Each subsequent patient $k+l$ may point at any of the kidneys $1, 2, \ldots, l, k+l+1, \ldots, n$ without being part of a cycle of length no longer than k. The probability that this happens is also $\frac{n-k}{n}$. Since these events are independent, the joint probability of this happening is:

$$p(n) := \begin{cases} \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\left(\frac{n-k}{n}\right)^{n-k+1} & \text{if } n > k, \\ 0 & \text{if } n \leq k. \end{cases}$$

Note that patient $k+l$ may not be a part of a short cycle even if she is pointing at kidney $k+l-1$ in case, for example, patient $k+l-1$ is part of a long cycle herself. So $p(n)$ underestimates the probability that there are no short cycles in the first stage of the algorithm. Therefore the probability that there is at least one cycle of length k or less is no greater than $1 - p(n)$.

Assume that there exists a short cycle. I remove all the kidneys and patients in a randomly chosen cycle of length no greater than k from the mechanism and from the remaining patients' preferences. Let's say there are n' remaining patient-donor pairs. Consider the next step in the algorithm. It is easy to see that the induced distribution over the reduced preferences satisfy the same properties as the original problem. Thus, recursively, the probability that the reduced problem has at least one short cycle is less than $1 - p(n')$. Denoting the remainder after dividing n by k by remainder(n, k), the overall probability that the TTC mechanism selects only short cycles is less than

$$q(n) = (1 - p(n))(1 - p(n-k)) \ldots (1 - p(\text{remainder}(n, k))),$$

where I have conservatively assumed that at each stage of the algorithm, one can find a cycle of length exactly k.

We use the well-known fact that

$$\lim_{n\to\infty} \left(\frac{n-k}{n}\right)^n = e^{-k}.$$

Thus all sufficiently large n satisfy $p(n) \geq \frac{1}{2}e^{-k}$, which implies $1 - p(n) \leq 1 - \frac{1}{2}e^{-k}$. Therefore

$$\lim_{n\to\infty} q(n) = \lim_{n\to\infty} (1 - \frac{1}{2}e^{-k})^n = 0,$$

which is what I wanted to show.

$\square$

Proof of Proposition 4: Assume that such a mechanism indeed exists and call it f and let $A = \{1, 2, \dots, n\}$.[40] I consider the cases $k = 2$ and $k \geq 3$ separately. I start with the case $k = 2$. Let the preference profile $\succ$ be

$$\begin{aligned} \succ_1 &: 2 \succ_1 n \succ_1 3 \succ_1 4 \dots \succ_1 n-1 \succ_1 1, \\ \succ_2 &: 3 \succ_2 1 \succ_2 4 \succ_2 5 \dots \succ_2 n \succ_2 2, \\ &\vdots \\ \succ_{n-1} &: n \succ_{n-1} n-2 \succ_{n-1} 1 \succ_{n-1} 2 \dots \succ_{n-1} n-3 \succ_{n-1} n-1, \\ \succ_n &: 1 \succ_n n-1 \succ_n 2 \succ_n 3 \dots \succ_n n-2 \succ_n n. \end{aligned}$$

Consider the following permutation: $\pi : A \to A$ defined by $\pi(i) = i + 1 (\bmod n)$. For simplicity, in what follows I omit the modulo notation. Note that the permuted preference profile $\succ^\pi$ is the same as the original preference profile $\succ$. Thus $P_{f(\succ)}(i,j) = P_{f(\succ^\pi)}(i,j)$. By anonymity, we also have $P_{f(\succ)}(i,j) = P_{f(\succ^\pi)}(\pi(i), \pi(j))$ and hence

$$P_{f(\succ)}(i,j) = P_{f(\succ)}(i+1, j+1). \tag{5}$$

Note that since f is 2-constrained efficient and there are no unacceptabilities, the matching $f(\succ)$ can have at most one patient who is unmatched. In fact, if n is odd, there is exactly one patient who is unmatched. Let that be i and so $P_{f(\succ)}(i,i) = 1$. But then, by (5), $P_{f(\succ)}(i+1, i+1) = 1$ as well, which implies that there are at least two people who are unmatched, which is a contradiction.

If n is even instead, all agents must be matched. If patient 1 does not receive either of her two highest-ranked kidneys 2 and n, by (5) none of the agents receives one of her two highest-ranked kidneys. But such a matching is dominated by the feasible matching where pairs 1 and 2, 3 and 4 etc. are matched together, since there each patient receives either her highest- or second highest-ranked kidney. This implies that patient 1 must receive either kidney 2 or kidney n. In the first case, (5) implies $P_{f(\succ)}(2,3) = 1$, which is impossible since cycles are of length no greater than 2. Similarly, if patient 1 receives kidney n, we must have $P_{f(\succ)}(1,2) = 1$, which is similarly impossible. This completes the case $k = 2$.

In the general case $k \geq 3$, I consider the cases $n-k$ being odd and even. If $n-k$ is even, let the preference profile $\succ$ be

$$\begin{aligned} \succ_1 &: 2 \succ_1 n \succ_1 n-k+2 \succ_1 3 \succ_1 4 \dots \succ_1 n-1 \succ_1 1, \\ \succ_2 &: 3 \succ_2 1 \succ_2 n-k+3 \succ_2 4 \succ_2 5 \dots \succ_2 n \succ_2 2, \\ \succ_3 &: 4 \succ_3 2 \succ_3 n-k+4 \succ_3 5 \succ_3 6 \dots \succ_3 1 \succ_3 3, \\ &\vdots \\ \succ_k &: k+1 \succ_k k-1 \succ_k 1 \succ_k k+2 \succ_k k+3 \dots \succ_k k-2 \succ_k k, \\ &\vdots \\ \succ_n &: 1 \succ_n n-1 \succ_n n-k+1 \succ_n 2 \succ_n 3 \dots \succ_n n-2 \succ_n n. \end{aligned}$$

In other words, patient i's preference order is $i+1, i-1, 1+i-k, i+2, i+3, i+4, \dots, i-2, i)$ from

[40]For maximum generality, the following proof assumes that there are no unacceptabilities in the patients' preferences.

most to least preferred, where all the operations are defined modulo n. Consider the same permutation as above: $\pi : A \to A$ defined by $\pi(i) = i + 1(\text{mod}\, n)$. By an analogous argument to above, (5) holds. As above, it is again impossible to have any agent left unmatched. Therefore $f(\succ)$ must match all the agents. Now assume that some cycle in the decomposition of $f(\succ)$ contains a segment $(\cdots - i - (i+1) - \cdots)$. This implies $P_{f(\succ)}(i, i+1) = 1$ and hence $P_{f(\succ)}(i+1, i+2) = 1$. So the cycle must contain the segment $(\cdots - i - (i+1) - (i+2) - \cdots)$. Inductive reasoning suggests that the cycle must be $(1 - 2 - \cdots - n - 1)$, which is impossible since cycles are constrained to be of length at most k. Thus no cycle contains the segment $(\cdots - i - (i+1) - \cdots)$. Analogously, one can show that no cycle contains the segment $(\cdots - (i+1) - i - \cdots)$. Examining the preferences, one can then see that in $f(\succ)$ no agent can do better than her third best option. But this is Pareto-dominated by the feasible matching

$$\{(1 - 2 - \cdots - k - 1), ((k+1) - (k+2) - (k+1)), ((k+3) - (k+4) - (k+3)), \ldots, ((n-1) - n - (n-1))\},$$

in which patients $1, 2, \ldots, k-1, k+1, k+3, \ldots, n-1$ receive their highest-ranked kidney, k receives her third choice and everyone else—their second choice. But this contradicts the fact that f is k-constrained efficient!

Finally, if $n - k$ is odd, let the preference profile $\succ$ be

$$\begin{aligned}
\succ_1 &: 2 \succ_1 n \succ_1 n-k+3 \succ_1 3 \succ_1 4 \ldots \succ_1 n-1 \succ_1 1,\\
\succ_2 &: 3 \succ_2 1 \succ_2 n-k+4 \succ_2 4 \succ_2 5 \ldots \succ_2 n \succ_2 2,\\
\succ_3 &: 4 \succ_3 2 \succ_3 n-k+5 \succ_3 5 \succ_3 6 \ldots \succ_3 1 \succ_3 3,\\
&\vdots\\
\succ_{k-1} &: k \succ_{k-1} k-2 \succ_{k-1} 1 \succ_{k-1} k+1 \succ_{k-1} k+2 \ldots \succ_{k-1} k-3 \succ_{k-1} k-1,\\
\succ_k &: k+1 \succ_k k-1 \succ_k 2 \succ_k k+2 \succ_k k+3 \ldots \succ_k k-2 \succ_k k,\\
&\vdots\\
\succ_n &: 1 \succ_n n-1 \succ_n n-k+2 \succ_n 2 \succ_n 3 \ldots \succ_n n-2 \succ_n n.
\end{aligned}$$

The preferences are the same as above, except patient i's third highest-ranked kidney is $2 + i - k(\text{mod}\, n)$ rather than $1 + i - k$ as above. The exact same arguments as above guarantee that all agents are matched and that in $f(\succ)$ no agent can be doing better than her third best option. But any such matching is Pareto dominated by the feasible matching

$$\{(1 - 2 - \cdots - (k-1) - 1), (k - (k+1) - k), ((k+2) - (k+3) - (k+2)), \ldots, ((n-1) - n - (n-1))\},$$

where, just like above, each agent receives her first, second or third best-choice. Contradiction! □

Proof of Proposition 5: Assume that such a mechanism indeed exists, call it f and let $A = \{1, 2, \ldots, k+1\}$. What follows holds for all $n > k$ since one can add additional patients who find no kidney or only each other's donors' kidneys acceptable. Consider the following preference profile $\succ$:

$$\begin{aligned}
&\succ_1 : 2 \succ_1 3 \succ_1 \ldots n \succ_1 1,\\
&\succ_2 : 3 \succ_2 4 \succ_2 \ldots 1 \succ_2 2,\\
&\quad\vdots\\
&\succ_n : 1 \succ_n 2 \succ_n \ldots n-1 \succ_n n.
\end{aligned}$$

For each patient i, construct the preference $\succ'_i$ by making only her top choice in $\succ_i$ acceptable. Using notation modulo n, this means:

$$\succ'_i : i+1 \succ'_i i \succ'_i i+2 \ldots$$

We first need to consider the case $k = 2$ and $n = 3$. There are three 2-efficient and individually rational matchings for the profile $\succ$. Due to the symmetry of the problem, I can assume without loss of generality that $f(\succ) = \{(1-2), (3)\}$, where a cycle of the form $(\alpha - \beta - \ldots - \omega)$ means that patient α receives kdiney β, and so on until patient ω receives kidney α. Note that by strategyproofness, $f(\succ'_1, \succ_2, \succ_3)$ must also be $\{(1-2), (3)\}$. Otherwise, patient 1 would be able to unilaterally deviate from that profile to $\succ$, which gives her her top choice. Finally, consider $f(\succ'_1, \succ'_2, \succ_3)$. The only possible individually rational matching here is $\{(1), (2-3)\}$. But that implies that patient 2 has a profitable deviation from $(\succ'_1, \succ_2, \succ_3)$ to $(\succ'_1, \succ'_2, \succ_3)$. Contradiction!

For the case of general k, consider the profile $\succ^0 := (\succ'_1, \ldots, \succ'_{n-1}, \succ_n)$. I argue that the only indivdually rational and k-efficient matching is the single cycle $(2-3-\ldots-n)$. First, it is clear that 1 remains unassigned by f. If she were assigned, she must receive kidney 2, who because of $f(\succ^0)$ being individually rational must receive kidney 3, and so on, until $n-1$ receives kidney n and, to close the cycle, n must receive kidney 1. This cycle is of length $n = k+1$, however, which is impossible. Thus the best the other patients can do is for 2 through $n-1$ to receive their top-choice kidneys, and n to receive her second highest-ranked kidney. Indeed, this is implementable in a cycle of length k: $(2-3-\ldots-n)$. Hence, since f is k-efficient, we must have

$$f(\succ^0) = \{(1), (2-3-\ldots-n)\}.$$

Now consider the profile obtainable from $\succ^0$ by changing the preferences of patient 2 from $\succ'_2$ to $\succ_2$. Call that profile $\succ^*$. I will argue that $f(\succ^*) = f(\succ^0)$. Note that due to f being strategyproof, agent 2 must receive her top choice (kidney 3) under $f(\succ^*)$. Otherwise she can unilaterally deviate back to $\succ^0$ and receive her top choice. Then, completely analogously to the above, one can show that 1 ends up unassigned under $f(\succ^*)$ and thus the unique k-efficient allocation is $\{(1), (2-3-\ldots-n)\}$.

Now define $\succ^1 := (\succ'_1, \succ_2, \succ'_3, \ldots, \succ'_n)$. Analogously to the case $\succ^0$, due to the symmetry of the problem, one can show that

$$f(\succ^1) = \{(3), (1-2-4-\ldots-n)\}.$$

Similarly to the above, one can show that by changing the preferences of patient n from $\succ'_n$ to $\succ_n$ in $\succ^1$, the matching that f selects for the resulting preference profile must remain unchanged. But that profile is $\succ^*$ and $f(\succ^0) \neq f(\succ^1)$. Contradiction! $\square$

Proof of Proposition 6: Fix a profile of preferences and claiming-speed functions $(\succ, e)$. Assume that the probability-share matrix M corresponding to $2CSE(\succ)$ is weakly Pareto dominated with respect to first-order stochastic dominance by a 2-implementable matrix M'. Note that this implies that M' is individually rational with respect to $\succ$. I will show that $M = M'$.

Note that a pair (i, j) can become unavailable during the mechanism in one of two ways: one (or both) of patients i and j exits the mechanism, or an Edmonds constraint containing $\{i, j\}$ starts binding. Let T^1 be the set of all times, at which a patient $i \in A$ exits the mechanism's procedure under $(\succ, e)$ and $M(i) \neq M'(i)$. Let T^2 be the set of all times, at which an Edmonds constraint corresponding to a set E starts binding during the procedure and there exists $\{i, j\} \subset E$ such that $M(i, j) \neq M'(i, j)$. Since there are finitely many such events, let $t^* = \min T^1 \cup T^2$. Assume toward contradiction that the set $T^1 \cup T^2$ is not empty and so $t^* > 0$ is well-defined. Consider the event that occurred at t^* (if there is more than one event occurring at that time, pick any one of them).

First, I consider the case in which the event is some patient i exiting. Since $M(i) \neq M'(i)$ and since, therefore, $M'(i)$ is strictly better than $M(i)$ in FOSD sense for i, there must exist kidneys j and l such that

$$M(i, j) > M'(i, j) \geq 0 \text{ and } 0 \leq M(i, l) < M'(i, l) \text{ and } l \succ_i j.$$

If $i = j$, this implies that i had claimed probability shares of kidney i when she exited even though she prefers kidney l. Note that l does not find i unacceptable since $M'(i, l) > 0$ and M' represents an individually rational matching with respect to the preference profile. So the pair (i, l) must have become unavailable at a time earlier than t^*. This is possible if an Edmonds constraint corresponding to an Edmonds set containing $\{i, l\}$ activated or l exited. Either way, though, this is impossible since the time of those events must be included in $T^1 \cup T^2$.

So $i \neq j$ must hold. Then, analogously to the above, there exists some patient h such that $h \succ_j i$ and $M(h, j) < M'(h, j)$. Since $M(i, j) > 0$, either i claimed some of kidney j's probability shares when she prefers l, or j claimed kidney i's probability shares when she prefers h. But then, analogously to above, one of the pairs (i, l) and (h, l) must have become unavailable earlier than. Using the same reasoning as above, this is impossible.

Therefore the first case is impossible. So I consider the case in which the event occurring at t^* is the activation of an Edmonds constraint. Let the corresponding Edmonds set be E with $i, j \in E$, $i \neq j$ and $M(i, j) \neq M'(i, j)$. Note that M' is 2-implementable. So we have

$$\sum_a \sum_{b \in E \setminus \{a\}} M'(a, b) \leq 2p = \sum_a \sum_{b \in E \setminus \{a\}} M(a, b),$$

where $|E| = 2p + 1$. I can assume that $M(i, j) > M'(i, j)$ since if $M(i, j) < M'(i, j)$ instead, the inequality above implies that there exists some other pair $i', j' \in E$ such that $M(i', j') > M'(i', j')$. As in the first case, this implies that there exist $h, l \in A$ such that

$$M(i, l) < M'(i, l), M(j, h) < M'(j, h), l \succ_i j \text{ and } h \succ_j i.$$

The rest of the analysis proceeds as above, leading to a contradiction. So the set $T_1 \cup T_2$ is empty and therefore $M = M'$. $\square$

Proof of Proposition 7: Start by fixing the number of agents to n and their preference profile to $\succ$. Let the set of permutation matrices that represent the permissible ex-post deterministic matchings in $C(\succ)$ be $C' := \{M_1, \ldots, M_p\}$. Their convex hull $\mathrm{co}(C')$ is the set of bistochastic matrices that are decomposable as a convex combination of matrices in C'. Define the set D to be the lower contour set of $\mathrm{co}(C')$.

It is clear that $\mathrm{co}(C') \subset D$. Next I will show that D is the convex hull of finitely many points and, hence, is bounded convex polytope in $\mathbb{R}^{n\times n}$. First, I claim that

$$D = \mathrm{co}\left(\bigcup_{i=1}^{p} \mathcal{E}_i\right), \tag{6}$$

where $\mathcal{E}_i$ is the set of all matrices each of whose entries equals either zero or the corresponding entry in $M_i \in C'$. In other words

$$\mathcal{E}_i = \{M \in \mathbb{R}_+^{n\times n} | \forall a, b \in A : M(a,b) = 0 \text{ or } M(a,b) = M_i(a,b)\}.$$

Note that if $M \in \mathcal{E}_i$, then $M \leq M_i$. It is easy to show that $\mathrm{co}(\mathcal{E}_i) = \{M \in \mathbb{R}_+^{n\times n} | M_i \geq M\}$. To start showing (6), take a matrix M that is in the convex hull on the right-hand side of (6). This means that I can represent M in the following way:

$$M = \sum_{i=1}^{q} \pi_i M^i,$$

where $\sum_{i=1}^{q} \pi_i = 1$ and $\pi_i \geq 0$ for all i. Also for each M^i, there exists $j \in \{1, \ldots, p\}$ such that $M^i \in \mathcal{E}_j$ and so there exists a corresponding matrix $\overline{M}^i \in C'$ such that $\overline{M}^i \geq M^i$. Then

$$M \leq \sum_{i=1}^{q} p_i \overline{M}^i \in \mathrm{co}(C').$$

Therefore $M \in D$. To show the other set inclusion for (6) assume that $M \in D$. Then there exists some $M' \in \mathrm{co}(C')$ with $M' = \sum_{i=1}^{p} \pi_i M_i$ such that $M' \geq M$. Define for each $\alpha, \beta \in \{1, \ldots, n\}$:

$$\gamma(\alpha,\beta) = \begin{cases} 0 & \text{if } M(\alpha,\beta) = 0, \\ M(\alpha,\beta)/M'(\alpha,\beta) & \text{otherwise.} \end{cases}$$

Define $\underline{\mathrm{M}}_i$ entry-by-entry via $\underline{\mathrm{M}}_i(\alpha,\beta) = \gamma(\alpha,\beta) M_i$. Clearly, $\underline{\mathrm{M}}_i \leq M_i$ and so

$$\underline{\mathrm{M}}_i \in \mathrm{co}(\mathcal{E}_i) \subset \mathrm{co}\left(\bigcup_{i=1}^{p} \mathcal{E}_i\right).$$

Also, it is easy to verify that $M = \sum_{i=1}^{p} \pi_i \underline{\mathrm{M}}_i$. Thus $M \in \mathrm{co}\left(\bigcup_{i=1}^{p} \mathcal{E}_i\right)$, which is the second set inclusion I wanted to show. Therefore, by (6), D is the convex hull of finitely many points and, hence, a bounded convex polytope in $\mathbb{R}^{n\times n}$. So D can be represented as the intersection of finitely many closed halfspaces in $\mathbb{R}^{n\times n}$.

Some matrix entries might be zero for all of the elements of D. Let there be q such entries. For the remainder of the proof, I will consider the elements of D as vectors with all entries which are zero for all elements from D removed. In other words, I will view D as a subset of $\mathbb{R}^{n^2-q}$ rather than $\mathbb{R}^{n\times n}$. I will show that the convex polytope D is fully dimensional when viewed in this space. To do that, it is enough to show that D has a non-empty interior.[41] Consider the following element of D:

$$N = \sum_{i=1}^{p} \frac{1}{p} M_i.$$

[41] See Section 2.3 in Ziegler (2007) for the relevant definitions and result.

Note that all of N's coordinates are positive because for all $n^2 - q$ coordinates, there exists some M_i for which that coordinate is equal to 1. Also it's clear that $N \in \text{co}(C') \subset D$. Consider then the element $\frac{1}{2}N$. It's clear that $\frac{1}{2}N < N$ and so $N \in D$. Also there exists some ε such that for every x in the open ball with radius ε around $\frac{1}{2}N$ (i.e., $B_\varepsilon(\frac{1}{2}N)$), $x \geq 0$ holds but also $x \leq N$ and hence $x \in D$. Thus $\frac{1}{2}N$ lies in the interior of D and D is fully dimensional in $\mathbb{R}^{n^2-q}$.

Since the convex polytope D is fully dimensional, it can be uniquely minimally defined (up to rescaling by a positive scalar) as all elements x in $\mathbb{R}^{n^2-q}$ that satisfy $Ax \leq b$ for some matrix A and a vector b. I will write each of the individual constraints as $A_i \cdot x \leq b_i$, where A_i is a row in A. Note that since D contains $\mathbf{0}$, $b_i \geq 0$ holds for all i. Furthermore, if some A_i satisfies $A_i \leq 0$, then the corresponding b_i must equal zero (otherwise, that would violate the minimality of the constraint set). Note that the corresponding constraint indicates that a coordinate or a sum of coordinates from each element of D must be non-negative. I will show that all the remaining constraints have positive coefficients. Indeed, take some A_i such that $A_i \cdot x \leq b_i$ is valid for all elements of D, and some coordinates of A_i are strictly positive and some strictly negative. By full dimensionality of D, there exists some $y \in D$ such that $A_i \cdot y = b_i$ but $A_j \cdot y < b_j$ for all $j \neq i$.[42] Then it is easy to see that one can find some $x \neq z$ distinct from y such that $z \geq x \geq y$ with $A_i \cdot z = b_i$ but $A_j \cdot z < b_j$ for all $j \neq i$, while $A_i \cdot x > b_i$. In other words, $z \in D$ while $x \notin D$. But by the way I defined D, there exists some $w \in \text{co}(C')$ such that $w \geq z$ and, since $z \geq x$, $w \geq x$ must hold and hence $x \in D$. Contradiction! Therefore all constraints, other than the ones guaranteeing non-negativity, are of the form

$$\sum_{(i,j) \in A \times A} a(i,j) M(i,j) \leq b$$

for some $b > 0$ and $a(i,j) \geq 0$ for all (i,j).

□

Proof of Propositions 8 and 9: Note that the notation from the proof of Proposition 7 carries through here. I start by revisiting the definition of the GCSE mechanism.

Each patient i has an associated claiming-speed function $e_i : [0,1] \to \mathbb{R}_+$ with $\int_0^1 e_i(t)dt = 1$. Time runs continuously starting at $t = 0$. For each point in time there is an associated sub-bistochastic matrix M^t where M^0 is the initial zero matrix. I say that kidney j is available to patient i at time $t \geq 0$ if none of the constraints defining the convex polytope D for which $a(i,j) > 0$, bind at that time. Note that M^0 is in D and satisfies all the constraints defining D and, in particular, all kidneys that could be available for patient i are initially available to her. At time t, each patient i claims with speed $e_i(t)$ the available remaining probability shares of her favorite reported kidney j among the kidneys that are available to i at that instance. That increases the probability that i receives j's kidney—i.e., it increases $M^t(i,j)$. Also note that $i = j$ can hold in this case. Note that the constraints defining D cannot be violated so $M^t \in D$ for all t and therefore there exists some implementable individually rational bistochastic matrix $M' \in \text{co}(C')$ such that $M' \geq M^t$ for all t. This means that for all agents who have not met their unit demand yet (i.e., for all agents for whom $M^t(i) \neq M'(i)$ yet), there exists a kidney j such that $M^t(i,j) < M'(i,j)$. Note that if patient i consumes some of kidney j, the new interim matrix would still be less than $M'(i,j)$ and thus no constraint defining D would be violated. So all constraints involving the pair (i,j) are slack at time t. Thus kidney j is available to patient i at time t. Therefore for each patient there exist at least one kidney which

[42] See, for example, Corollary 8.2a in Schrijver (1986) regarding fully dimensional polytopes. It states that given a constraint $A_i \cdot x \leq b_i$, for each other constraint $A_j \cdot x \leq b_j$, there exists some $x_j \in D$ such that $A_i \cdot x_j = b_i$ but $A_j \cdot x_j < b_j$. Then an equal-weight convex combination of all x_j's would satisfy the desired conditions.

is available to her at any point of time during the algorithm. In other words, at no point of time, does the algorithm become "stuck" without available kidneys' probability shares for one or more of the patients. Thus the algorithm ends at time $t = 1$ and the final matrix M^1 must satisfy all agents' unit demands. Therefore the sum of all the entries in the matrix is n and therefore M^t must be a bistochastic matrix. Finally, since each M^t is in D, the bistochastic matrix M^1 must be in $\text{co}(C')$.

Therefore M^1 can be decomposed into allowable permutation matrices. I implicitly define the mechanism f by setting $f(\succ)$ such that $P_{f(\succ)} = M^1$ for each possible preference profile $\succ$. By giving all agents the same claiming-speed function, f can be made anonymous since the constraint correspondence is anonymous itself. It remains to be shown that the f as defined is C-constrained ordinally efficient.

Assume that the bistochastic matrix M^1 is weakly Pareto dominated with respect to first-order stochastic dominance by some $M' \in \text{co}(C')$. I will show that $M^1 = M'$.

Let T be the set of times at which a constraint defining the polytope D starts binding and there exists some (i, j) such that $a(i, j)$ belonging to that constraint is strictly positive and $M^1(i, j) \neq M'(i, j)$. Note that the non-negativity of all the constraint coefficients guarantees that once a constraint starts binding, it will bind for the rest of the mechanism. This follows from the fact that all coefficients of the matrix M^t are non-decreasing in t by construction of the mechanism's algorithm. Since there are finitely many such events, set $t^* = \min T$.

So at time t^* in the mechanism's algorithm, a constraint defining the polytope D starts binding. Note that by a logic analogous to the one above, the all-positive coefficients and M^t being non-decreasing in t guarantee that the constraint must also bind for $t = 1$. In other words:

$$\sum_{(i',j') \in A \times A} a(i', j') M^1(i', j') = b.$$

Since $M' \in \text{co}(C')$, we also have

$$\sum_{(i',j') \in A \times A} a(i', j') M'(i', j') \leq b.$$

So

$$\sum_{(i',j') \in A \times A} a(i', j') M^1(i', j') \geq \sum_{(i',j') \in A \times A} a(i', j') M'(i', j'),$$

while $a(i, j) > 0$ and $M^1(i, j) \neq M'(i, j)$ hold. I can assume that $M^1(i, j) > M'(i, j)$ since if $M^1(i, j) < M'(i, j)$, the inequality above implies that there exists some other pair (i'', j'') with $a(i'', j'') > 0$ and $M^1(i'', j'') > M'(i'', j'')$.

Since patient i must strictly prefer $M'(i)$ over $M^1(i)$, there must exist kidney l such that

$$M^1(i, j) > M'(i, j) \geq 0 \text{ and } 0 \leq M^1(i, l) < M'(i, l) \text{ and } l \succ_i j.$$

But then since $M^1(i, j) > 0$, agent i must have consumed probability shares from kidney j during the mechanism even though she prefers kidney l. So when i was consuming j's probability shares, kidney l must have already become unavailable to i. But j becomes unavailable to i at time t^*. So l must have become unavailable to patient i strictly earlier than t^*. But $M^1(i, l) \neq M'(i, l)$, which contradicts the choice of t^*. $\square$

Proof of Proposition 12: It is enough to show the statement of the proposition for $n = 4$. If $n > 4$, I can simply consider the example that follows with all additional agents finding the original 4 agents unacceptable and vice versa. I start with the following simple Lemma.

Lemma 1. *Let f be some 2-constrained ex-post efficient, individually rational, and weakly strategyproof random mechanism. Then for all n, if*

$$\forall l \in A : j \succ_i l$$
$$\forall l \in A : i \succ_j l,$$

we must have

$$P_{f(\succ)}(i,j) = P_{f(\succ)}(j,i) = 1.$$

Proof. Let $\succ$ be such that

$$\forall l \in A : 2 \succ_1 l$$
$$\forall l \in A : 1 \succ_2 l,$$

Define $\succ_1'$ and $\succ_2'$ to be such that patient 1 finds only kidney 2 acceptable and vice versa. Formally:

$$\succ_1' : 2 \succ_1' 1 \dots$$
$$\succ_2' : 1 \succ_2' 2 \dots$$

Note that individual rationality and ex-post efficiency require that

$$P_{f(\succ_1', \succ_2', \succ_{-\{1,2\}})}(1,2) = 1.$$

By weak strategyproofness, agent 2 cannot improve her probability-share allocation in FOSD sense when deviating to a false report. Thus she must receive kidney 1 with probability one under $(\succ_1', \succ_{-1})$. Otherwise, she could unilaterally deviate to $\succ_2'$ and receive her first-best outcome. An analogous argument guarantees that pairs 1 and 2 are matched in a two-way exchange with probability 1 under $\succ$ as well, which is what I wanted to show. □

Toward contradiction, assume that such a mechanism exists. Call it f and consider the preference profile $\succ$ for $A = \{1,2,3,4\}$ defined by

$$\succ_1 : 2 \succ_1 4 \succ_1 3 \succ_1 1,$$
$$\succ_2 : 3 \succ_2 1 \succ_2 4 \succ_2 2,$$
$$\succ_3 : 4 \succ_3 2 \succ_3 1 \succ_3 3,$$
$$\succ_4 : 1 \succ_4 3 \succ_4 2 \succ_4 4.$$

Note that the set of 2-constrained Pareto optimal matchings are the two matchings $\{(1,2),(3,4)\}$ and $\{(1,4),(2,3)\}$, where a cycle of the form $(\alpha - \beta - \dots - \omega)$ means that patient α receives kdiney β, and so on until patient ω receives kidney α. Thus, since the f is 2-constrained ex-post efficient, $f(\succ)$ can place positive probability only on these two matchings. In particular, this implies that $P_{f(\succ)}(i,i) = 0$ for $i \in \{1,2,3,4\}$.

Now consider the preference profile $\succ^*$, which is the same as $\succ$ except for 1's preferences:

$$\succ_1^*: 2 \succ_1^* 1 \succ_1^* 4 \succ_1^* 3.$$

With this profile, there are only two 2-constrained efficient and individually rational matchings: $\{(1,2),(3,4)\}$ and $\{(1),(2,3),(4)\}$. (Note that $\{(1),(2,4),(3)\}$ is dominated by the first of these two matchings.) Thus, when applied to $\succ^*$, the mechanism must select some lottery over these two matchings. Consider a deviation by patient 4 to

$$\succ_4': 3 \succ_4 1 \succ_4 2 \succ_4 4$$

from the profile $\succ^*$. Note that by Lemma 1, this deviation guarantees her

$$P_{f(\succ_1^*,\succ_2,\succ_3,\succ_4')}(4,3) = 1.$$

But then, by weak strategyproofness, patient 4 must receive kidney 3 with probability 1 under $\succ^*$ as well (otherwise, she has a strictly first-order stochastically dominant deviation to $\succ_4'$). In other words, the mechanism must assign probability 1 to the matching $\{(1,2),(3,4)\}$ under $\succ^*$. But then 1 gets her top choice for sure and, in particular, by weak strategyproofness, $f(\succ)$ must place probability 1 on the matching $\{(1,2),(3,4)\}$.

By a completely analogous argument considering a potential deviation by patient 2, however, $f(\succ)$ must place probability 1 on the matching $\{(1,4),(2,3)\}$. Contradiction! □

Proof of Proposition 13: The proof is virtually identical to the proof of Proposition 5. I omit the details.

□

B Further discussion of the properties of the 2CSE/2CPS mechanisms

B.1 Justice

So far I have considered anonymity as a minimal fairness desideratum for any kidney-exchange mechanism. As far as fairness requirements go, anonymity is relatively weak. In this section I consider the compatibility of other justice criteria with the 2CSE (and 2CPS) mechanism, as well as with the other desiderata of the mechanism.

An important fairness criterion that is often considered in the literature (Moulin 1995) is non-envy. I say that a random mechanism f satisfies **no envy** if for every $i \in A$, the probability-share allocation $P_{f(\succ)}(i)$ first-order stochastically dominates $P_{f(\succ)}(j)$ for all $j \in A$ with respect to $\succ_i$. A random mechanism f satisfies **weak no envy** if for every $i \in A$, there does not exist $j \in A$ such that the probability-share allocation $P_{f(\succ)}(j)$ first-order stochastically dominates $P_{f(\succ)}(i)$ with respect to $\succ_i$. A random mechanism f satisfies **no justified envy** if for all pairs of patients $i, j \in A$ such that $P_{f(\succ)}(j)$ does not first-order stochastically dominate $P_{f(\succ)}(i)$ with respect to $\succ_j$ there exist some $l \in A$ such that $P_{f(\succ)}(j,l) > 0$ and $i \succ_i l$. It is clear that non-envy implies both weak non-envy and no justified envy.

As first noted by Yılmaz (2010), it is easy to see that non-envy is incompatible with individual rationality. A simple example would be an agent set $A = \{1,2\}$ where both patients rank kidney 1 the highest. Then, by individual rationality, both patients must be left unmatched with probability 1. But then patient

1's probability-share allocation strictly first-order stochastically dominates patient 2's with respect to 2's preferences. To get around this issue, Yılmaz (2010) proposes *no justified envy* as a refinement of non-envy that is suited for settings where individual rationality is of prime importance. Intuitively, no justified envy requires that if patient i has a cause to envy patient j's probability-share allocation, then i must with positive probability receive kidneys that j finds unacceptable. In other words, if i's probability-share allocation does not FOSD dominate another patient j's probability-share allocation with respect to i's preferences, then there is some kidney l such that j finds l unacceptable but i receives l with positive probability.

So does 2CPS satisfy no justified envy? It turns out that the answer is generally no. But, moreover, I show that no justified envy is incompatible with 2-constrained ordinal efficiency.

Example 7. Let $A = \{1, 2, 3, 4, 5\}$ and let the preference profile $\succ$ be defined by

$$\begin{aligned}
&\succ_1 : 2 \succ_1 3 \succ_1 4 \succ_1 5 \succ_1 1,\\
&\succ_2 : 5 \succ_2 4 \succ_2 1 \succ_2 3 \succ_2 2,\\
&\succ_3 : 4 \succ_3 2 \succ_3 5 \succ_3 1 \succ_3 3,\\
&\succ_4 : 1 \succ_4 5 \succ_4 3 \succ_4 2 \succ_4 4,\\
&\succ_5 : 3 \succ_5 1 \succ_5 2 \succ_5 4 \succ_5 5.
\end{aligned}$$

One can calculate that the outcome of the 2CPS mechanism is

$$\begin{pmatrix}
1/5 & 2/5 & 0 & 2/5 & 0\\
2/5 & 1/5 & 0 & 0 & 2/5\\
0 & 0 & 1/5 & 2/5 & 2/5\\
2/5 & 0 & 2/5 & 1/5 & 0\\
0 & 2/5 & 2/5 & 0 & 1/5
\end{pmatrix},$$

but this does not satisfy the no-justified-envy condition. More specifically, patient 1 envies patient 5's allocation.

We next show that 2-constrained ordinal efficiency and no justified envy are incompatible.

Proposition 14. *There is no 2-constrained ordinally efficient mechanism, whose outcomes always satisfy no justified envy.*

Proof. Consider the preferences from Example 7. Assume that f is a mechanism that is 2-constrained ordinally efficient, which satisfies no justified envy. Since there are no unacceptabilities in $\succ$, this implies that $P_{f(\succ)}(i)$ first-order stochastically dominates $P_{f(\succ)}(j)$ with respect to $\succ_i$ for all i and j. Then, denoting $M := P_{f(\succ)}$, the following hold:

$$\begin{aligned}
M(1,2) &\geq M(5,2) = M(2,5)\\
&\geq M(3,5) = M(5,3)\\
&\geq M(4,3) = M(3,4)\\
&\geq M(1,4) = M(4,1)\\
&\geq M(2,1) = M(1,2),
\end{aligned}$$

where the inequalities follow from the stochastic dominance between the rows, and the equalities follow from the symmetry of M due to its representing 2-constrained ordinally efficient matching. Therefore all these matrix entries are equal. Denote their value by a. Repeating the same logic, the following hold:

$$\begin{aligned} a + M(1,3) &\geq M(4,2) + a = a + M(2,4) \\ &\geq M(1,5) + a = a + M(5,1) \\ &\geq M(2,3) + a = a + M(3,2) \\ &\geq M(5,4) + a = a + M(4,5) \\ &\geq M(3,1) + a = a + M(1,3). \end{aligned}$$

Thus let

$$\begin{aligned} b :&= M(1,3) = M(3,1) = M(2,4) = M(4,2) = M(1,5) \\ &= M(5,1) = M(2,3) = M(3,2) = M(4,5) = M(5,4). \end{aligned}$$

So the matrix M equals

$$\begin{pmatrix} M(1,1) & a & b & a & b \\ a & M(2,2) & b & b & a \\ b & b & M(3,3) & a & a \\ a & b & a & M(4,4) & b \\ b & a & a & b & M(5,5) \end{pmatrix}.$$

Consider $M(1)$. By no justified envy

$$M(1,2) + M(1,3) \geq M(5,2) + M(5,3) \Rightarrow a + b \geq a + a \Rightarrow b \geq a \tag{7}$$

holds.

Note that since M must satisfy the Edmonds constraints, $a + b \leq 2/5$ (from the Edmonds constraint corresponding to the set $E = A$), $2a + b \leq 1$ and $a + 2b \leq 1$ (from the Edmonds constraint corresponding to the sets E with $|E| = 3$), where the first inequality clearly implies the other two. In fact, 2-constrained ordinal efficiency implies $a + b = 2/5$ since remaining unmatched is every patient's lowest-ranked outcome. Thus by (7), $b \geq 1/5$ but it is easily checked then that in such a case the matrix M is dominated in first-order stochastic dominance sense by the matrix

$$\begin{pmatrix} 1/5 & 2/5 & 0 & 2/5 & 0 \\ 2/5 & 1/5 & 0 & 0 & 2/5 \\ 0 & 0 & 1/5 & 2/5 & 2/5 \\ 2/5 & 0 & 2/5 & 1/5 & 0 \\ 0 & 2/5 & 2/5 & 0 & 1/5 \end{pmatrix},$$

which is clearly 2-implementable as it is, from Example 7, the outcome of the 2CPS mechanism for these preferences. Contradiction! $\square$

B.2 A note on welfare

Bogomolnaia and Moulin (2001) proposed their simultaneous eating mechanism as a way to characterize all ordinally efficient object-assignment allocations given a preference profile. They show that by varying the profile of claiming-speed functions of the agents, the mechanism can output any given ordinally efficient outcome. Yılmaz (2010) shows a similar result for his individually rational version of the PS mechanism. So a natural next question is to ask whether the 2CSE mechanism satisfies the same properties. In other words, for a given a preference profile $\succ$ and a random matching μ that is ordinally efficient with respect to $\succ$, does there exist some profile of claiming-speed functions e such that

$$P_\mu = 2CSE(\succ, e)?$$

We can answer that in the negative as the following example shows.

Example 8. Let $n = 3$ with preferences $\succ$ defined by

$$\begin{aligned}
&3 \succ_1 2 \succ_1 1,\\
&3 \succ_2 1 \succ_2 2 \text{ and}\\
&1 \succ_3 2 \succ_3 3.
\end{aligned}$$

Consider a random matching μ with

$$P_\mu = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is not hard to see that this random matching is 2-constrained ordinally efficient: the only way that $P_\mu(1)$ can be improved in first-order stochastic dominance fashion is if one increases the probability that patient-donor pairs 1 and 3 are matched together but that must increase the probability that 2 is left unmatched, which cannot be part of a first-order stochastic dominance improvement since being unmatched is patient 2's worst outcome. The same logic holds for $P_\mu(2)$ and $P_\mu(3)$ too. Thus μ is 2-ordinally efficient.

Note though that in any 2-cycle simultaneous eating mechanism, at least one of $P_{2CSE(\succ,e)}(1,3)$ and $P_{2CSE(\succ,e)}(2,3)$ must be positive since kidney 3 is patient 1 and 2's top choice, kidney 1 is patient 3's top choice, and initially all kidneys are available to all patients.

In addition to providing an answer to my question, example 8 also suggests the reason for why 2CSE fails to hit all ordinally efficient outcomes. More specifically, the example shows that there exist deterministic 2-efficient matchings in which none of the agents receives her top choice.[43] Since deterministic 2-efficient matchings are also ordinally efficient and, by construction of the 2CSE mechanism, at least one patient receives her highest-ranked kidney with positive probability (barring unacceptabilities), this means that it's impossible for all ordinally efficient random matchings to be outcomes of the 2CSE mechanism for some e. This raises two questions in turn. First, does there exist an ordinally efficient mechanism that is parameterized by some vector such that the mechanism selects all possible ordinally efficient allocations by varying the parameter? Second, can the bistochastic matrices that 2CSE selects be characterized?[44]

[43] This is in contrast to the fact that, in the object-allocation setting, any Pareto optimal allocation can be achieved via a serial dictatorship (Abdulkadiroğlu and Sönmez 1998), which means that at least one agent receives her favorite object.

[44] See Heo and Yılmaz (2013); Heo (2014a) for some results characterizing the possible matrix outcomes of an extension of

the PS mechanism.