

Fondazione Eni Enrico Mattei
Working Papers
Fondazione Eni Enrico Mattei

Year 2009

Paper 292

Feasibility Constraints and Protective
Behavior in Efficient Kidney Exchange

Antonio Nicolò*

Carmelo Rodríguez Álvarez†

*Università degli Studi di Padova

†Universidad Complutense de Madrid

This working paper site is hosted by The Berkeley Electronic Press (bepress).

<http://www.bepress.com/feem/paper292>

Copyright ©2009 by the authors.

Feasibility Constraints and Protective Behavior in Efficient Kidney Exchange

Abstract

We propose a model of Kidney-Exchange that incorporates the main European institutional features. We assume that patients do not consider all compatible kidneys homogeneous and patients are endowed with reservation values over the minimal quality of the kidney they may receive. Under feasibility constraints, patients' truthful revelation of reservation values is incompatible with constrained efficiency. In the light of this result, we introduce an alternative behavioral assumption on patients' incentives. Patients choose their revelation strategies as to "protect" themselves from bad outcomes and use a lexicographic refinement of maximin strategies. In this environment, if exchanges are pairwise, then priority rules or rules that maximize a fixed ordering provide incentives for the patients to report their true reservation values. The positive result vanishes if larger exchanges are admitted.

Feasibility Constraints and Protective Behavior in Efficient Kidney Exchange*

Antonio Nicoló[†]

Università degli Studi di Padova

Carmelo Rodríguez-Álvarez[‡]

Universidad Complutense de Madrid

May 7, 2009

Abstract

We propose a model of Kidney-Exchange that incorporates the main European institutional features. We assume that patients do not consider all compatible kidneys homogeneous and patients are endowed with reservation values over the minimal quality of the kidney they may receive. Under feasibility constraints, patients' truthful revelation of reservation values is incompatible with constrained efficiency.

*We are grateful to Bettina Klaus and Alvin Roth for insightful comments and discussions. We would also like to thank seminar audiences at the Universities of Vigo, York, and Warwick as well as the participants of the 27th Arne Ryde Symposium 2008 at Lund, the Asset Meeting 2008 at EUI, the BOMOPA Workshop at Bologna, and the 14th Coalition Theory Network Workshop at Maastricht. Nicoló thanks the Italian Ministry of University and Research for the financial support through grant 2005137858. Rodríguez-Álvarez gratefully acknowledges the financial support from the Ministerio de Educación y Ciencia through Programa Ramón y Cajal 2006, and Research Grant SEJ-2005-04805, the Fondo Social Europeo, the Consejería de Innovación Ciencia y Empresa (Junta de Andalucía), and the Fundación Ramón Areces.

[†]Dipartimento di Scienze Economiche "Marco Fanno". Università degli Studi di Padova. Via del Santo 33, 37123 PADOVA. Italy. antonio.nicolo@unipd.it.

[‡]Departamento de Fundamentos del Análisis Económico II. Facultad CC. Económicas. Campus de Somosaguas. Universidad Complutense de Madrid. 28223 MADRID. Spain. carmelorr@ccee.ucm.es.

In the light of this result, we introduce an alternative behavioral assumption on patients' incentives. Patients choose their revelation strategies as to "protect" themselves of bad outcomes and use a lexicographic refinement of maximin strategies. In this environment, if exchanges are pairwise, then priority rules or rules that maximize a fixed ordering provide incentives for the patients to report their true reservation values. The positive result vanishes if larger exchanges are admitted.

1 Introduction

In the past decade economists have become increasingly more involved in the design of markets/practical mechanisms (labor market clearinghouses, Roth 2002; power markets, Wilson 2002; school choice, Abdulkadirođlu and Sönmez 1999, 2003). With the recent seminal paper by Roth et al. (2004), the theory of mechanism design has found an important application in the design and implementation of matching mechanisms (rules) to allocate organs for transplantation. The complexity of institutional and feasibility constraints, the normative implications and the effects on patients' lives make the task of designing optimal rules a fascinating challenge.

The best treatment for end-stage kidney failure is kidney transplantation. Kidneys available for transplantation may be obtained from deceased donors or from willing living donors. In October 2008, more than 400,000 people in the US are being treated for end-stage kidney failure, and more than 76,000 are listed for a deceased donor kidney transplant. In 2007, there were 16628 kidney transplants in the US, 6041 of those transplants were from living donors.¹ Unfortunately, a donated kidney may be unsuitable for transplantation (incompatible) to a given patient because the mismatch between donor and patient blood types and tissues would lead to the immediate rejection and loss of the graft.

In recent years, there has been many proposals to alleviate the (universal) shortage of kidneys. Some authors as Becker and Elias (2002) present arguments favoring a market approach, but the medical community is firmly opposed to the application of monetary

¹The figures vary significantly in different developed countries. For instance, in the same period, there were 2211 transplants from deceased donor kidneys and only 137 from living donor for about 6000 patients on the waiting list in Spain. See Organ Procurement and Transplantation Network and Organización Nacional de Trasplantes webpages, www.optn.org and www.ont.es.

incentives to donation.² In their recent paper, Roth et al. (2004) show however, that monetary transfer between patients and donors may not be necessary to attenuate kidneys scarcity. The combination of dialysis (an alternative treatment for kidney–failure) and living donors generates interesting new protocols. Two incompatible donor–patient couples may be mutually compatible and a swap of donors between the two couples would result in two successful transplantations. (*Kidney Paired Exchange: KPE*).³ Analogously, the donor’s kidney may be transplanted to some patient in the deceased donor kidney waiting list and her initially intended patient may receive an absolute priority over kidneys in the cadaveric waiting list. (*List Paired Exchange: LPE*).⁴ Further gains can be obtained if kidney swaps involve more than two donor–patients couples. In fact, simulations carried out by Roth et al. (2004) suggest that the benefits of such an exchange could be very substantial, increasing live organ donations between unrelated donors from about 54% to as much as 91% if exchanges among multiple couples are feasible, and to as much as 75% even if only pairwise exchanges are feasible. In this environment a Central Transplant Coordinator (CTC) uses kidney assignment rules that taking into account the medical details of the patients and donors involved in the possible exchanges, propose compatible exchanges among the couples. A key issue when designing these rules is that they should not provide incentives for the patients to lie about their medical details in order to improve their chances of getting a match as good as possible.

In subsequent work, Roth et al. (2005a) have proposed a mechanism design approach to KPE that encompasses the specific features and institutional detail of New England.⁵ Roth et al. (2005a) assume that patients consider all compatible kidneys as homogeneous and patients’ sets of compatible kidneys are not known to the CTC. Since incentive constraints imply that all the operations involved in an exchange must be carried out

²In fact, the trade in human organs is a felony under the National Organ Transplant Act (NOTA) of 1984, and the Uniform Anatomical Gift Act of 1987. On the other hand, Iran has the highest living-donor rate in the world and it is the only country where monetary compensation for organs is officially sanctioned. See “The gap between supply and demand ” in *The Economist*, October 8th, 2008.

³See Delmonico (2004); Delmonico et al. (2004); Segev et al. (2005b); Spital (2004); Segev et al. (2005a).

⁴Patients remaining in the cadaveric waiting list would benefit as well because the patients who receive a kidney would drop from it. See Delmonico (2004); Kaplan et al. (2005); Zenios et al. (2001).

⁵Their proposal is actually been used by the New England Program of Paired Kidney Exchange since 2006. See www.nepke.org and Roth et al. (2005b) for additional details.

simultaneously, Roth et al. (2005a) consider that only pairwise exchanges between two donor–patient pairs are feasible. In fact, they prove that priority rules used in real life ensure that it is a dominant strategy for patients to truthfully reveal both the set of donors they can receive a kidney from and the set of patients that their donors can donate a kidney to.⁶

Roth et al. (2005a)’s focus on the institutional details of New England restricts the application of their model to other regions and countries. In Europe and in several areas of the US, it is generally accepted that patients and doctors do not consider all compatible kidneys as homogeneous. Many individual characteristics of the donor, like age, health status, as well as matching characteristics of the donor–patient pair, like HLA mismatches, are statistically significant in determining the probability of long–term graft survival and the quality of life of the patient after the operation.⁷ Thus, we propose an alternative approach to KPE that incorporates some important features of the European view on kidney exchange. First, as Roth et al. (2005a) we assume that there exist feasibility constraints on the number of simultaneous operations (even if we do not initially restrict our attention to pairwise exchanges). On the other hand, we depart from Roth et al. (2005a) in two important aspects.

- (i) Following Roth et al. (2004), *patients do not consider all compatible kidneys homogeneous*. Their preferences over available kidneys are based on the quality of the donor–patient match. *The quality of the match is determined by characteristics of patients and donors which are observable by doctors and verifiable by means of medical tests*. For instance, the quality of a match can be measured according to some objective criterion as the Lifetime Years From Transplant (LYFT) method that estimates the difference between the remaining lifetime with and without transplant for each candidate on the waiting list.⁸ Hence, *patients’ preferences are known for*

⁶Hatfield (2005) shows that the results are robust to arbitrary feasibility constraints. More recently, Saidman et al. (2006) and Roth et al. (2007) show that efficiency gains could be attained (and almost exhausted) if kidney exchanges among three donor–patient couples and LPE were admitted. Sönmez and Ünver (2005); Roth et al. (2006) also analyze the potential benefits of altruistic no–related –Samaritan–donors in LPE.

⁷See Duquesnoy et al. (2003), Merion et al. (2005), Keizer et al. (2005), Klerk et al. (2004), Opelz (1997), Kranenburg et al. (2004), and Schnitzler et al. (1999).

⁸See “Predicting the Life Years From Transplant (LYFT): Choosing a Metric”, Scientific Registry for Transplant Recipients working paper, May 16, 2007 at www.unos.org.

*the CTC and need not to be elicited.*⁹

- (ii) *Patients are endowed with a reservation value over the minimal quality of the donor-patient match required to accept a transplantation.* The choice of receiving a given kidney or continuing the dialysis treatment depends in fact on patient's eagerness to receive an organ instead of continuing the dialysis and waiting for a better kidney. Hence, each patient reservation value depends on how the patient subjectively evaluates the quality of life under dialysis, her expectations about the quality of future pools of kidneys available for exchange, and on her attitude towards risk and uncertainty.¹⁰ *Of course, patients reservation values are not observable and remain private information.*

In this scenario, we investigate whether we can construct rules such that revealing the true reservation value is a dominant strategy for the patients. Surprisingly, we show that in the presence of feasibility constraints, truthful revelation is not compatible with a weak version of efficiency.

In the light of the negative result that we obtain in the standard model, we propose an alternative “behavioral” approach. KPE programs normally involve the coordination of nephrology services and patients of several hospitals. In this environment, patients have little information about the remaining patients involved in the program. Patients may consider that by misreporting their own reservation values, they also could end up losing the possibility of a beneficial transplant. In fact, they could even receive an undesirable kidney. Therefore, it becomes natural to assume that patients might prefer to choose their strategies so as to “protect” themselves from the worst eventuality as far as possible. We capture formally this “lexicographic maximin” behavior assumption with the notion of “protective behavior” proposed by Barberà and Dutta (1982) and later axiomatized by Barberà and Jackson (1988). With his assumption on patients' strategic

⁹A possible problem is that, even doctors could be tempted to report strategically these information to the CTC in order to favor their own patients. This is the justification for the approach in Roth et al. (2004). The problem seems less relevant for European continental countries, where Transplant Services are normally public and have more information and less coordination problems than those in US.

¹⁰The existence of patients' reservation values is consistent with the existence of Extended Donor Criteria and the use of organs previously regarded as unsuitable, because improvements on immunosuppressive treatment imply that even those low quality organs have good probability of survival. See Su et al. (2004); Su and Zenios (2006).

behavior, we reconcile the notion that patients only care about obtaining a kidney in so far it is compatible with the evidence on the heterogeneity of compatible kidneys. In this scenario, we show that truthful revelation of patients reservation values can be attained despite the feasibility constraints. If kidney exchanges are restricted to involve only pairs of donor–patient couples, then a plethora of rules provide (strong) incentives for the patients to report their true reservation values. This is the case for priority rules or rules that maximize a fixed ordering over the set of feasible and individually rational kidney assignments. The positive result vanishes however, if larger exchanges are admitted. There are kidney allocation problems where if 3-way exchanges are feasible, then truth-telling is a protectively dominated strategy. Thus, in some sense, our results justify the possibility of introducing a pairwise kidney exchange in Europe, but also provide additional theoretical support beyond the logistic reasons for concentrating on the possibilities that arise in pairwise exchanges.

The remainder of the paper is organized as follows. In Section 2, we outline the model of kidney allocation problems and basic notation. In Section 3, we introduce the concept of kidney assignment rules and some desirable conditions. In Section 4 we present an introductory impossibility result. In Section 5, we define the protective behavior and present the positive results. In Section 6, we conclude and discuss lines of further research.

2 Kidney Assignment Problems

Consider a finite society consisting of a set $N = \{1, \dots, n\}$ of patients ($n > 3$) who need a kidney for transplantation. Each patient has a potential donor, and $\Omega = \{\omega_0, \omega_1, \dots, \omega_n\}$ denotes the set of kidneys available for transplantation. The kidney ω_0 refers to the situation in which a patient does not receive any kidney, while ω_i for each $i \neq 0$ refers to the kidney of patient i 's donor. We assume that each patient has only one potential donor and that there are not kidneys without living donor.¹¹

Each patient i is equipped with a complete and transitive preference relation \succsim_i on Ω . Patients' preferences are based on rankings expressed through objective, medical criteria that measure the fitness of each available kidney to each patient and that are observable by

¹¹Hence, we focus on KPE. We discuss in the concluding section the possibility of multiple donors.

the CTC.¹² We express patients' preferences by using numerical valuations over kidneys. For each $i, j \in N$, we denote by $v_i(\omega_j)$ i 's valuation of kidney ω_j . For each $i \in N$, $\omega, \omega' \in \Omega$, we say patient i considers kidney ω at least as good as kidney ω' , $\omega \succsim_i \omega'$, if and only if $v_i(\omega) \geq v_i(\omega')$. Of course, given \succsim_i the associated strict preference relation \succ_i and the indifference relation \sim_i are defined in the standard way. We normalize in such a way that for each $i \in N$, and each $\omega \in \Omega \setminus \{\omega_0\}$, $v_i(\omega) \in [0, 1)$. If for some $i \in N$ and $\omega \in \Omega$ $v_i(\omega) = 0$, we say that patient i and kidney ω are **incompatible**. This reflects the possibility that patient i 's body will reject the graft of kidney ω , because of blood-type incompatibility, positive crossmatch, or any other reason. We say that patient i and kidney ω are **compatible** if $v_i(\omega) > 0$. We assume that agents have strict preferences over compatible kidneys, and therefore for each $i \in N$ and each $\omega, \omega' \in \Omega$ if $v_i(\omega) \neq 0$ and $v_i(\omega') \neq 0$, then $v_i(\omega) \neq v_i(\omega')$. A **preference profile** is a matrix $\mathbf{P} \in \mathbb{M}_{n \times n}$ and it is defined by $P_{i,j} \equiv v_j(\omega_i)$ for each $i, j \in N$. Preference profiles contain all the information about patients' observable priority rankings.

Each patient i is also endowed with a reservation value $r_i \in (0, 1)$. We interpret r_i as the valuation that patient i assigns to receive ω_0 . Hence, $r_i \equiv v_i(\omega_0)$. Reservation values may incorporate patients' subjective valuation of being on dialysis and not receiving any kidney, as well as the endogenous expectation of receiving a new organ in the future from a new pool of donor-patient couples. We assume that for each patient i , $r_i > 0$. Thus, patients always prefer to stay on dialysis rather than receiving an incompatible kidney. In order to be consistent with our assumption on strict preferences for compatible kidneys, we assume that patients are never indifferent between receiving a kidney and not receiving a kidney and remaining on the waiting list. Thus, for each i and each $\omega \neq \omega_0$, $r_i \neq v_i(\omega)$. Given a patient i and a preference profile \mathbf{P} , we denote by $R_i \equiv \{r_i \in (0, 1) \mid \forall \omega \in \Omega \setminus \{\omega_0\}, r_i \neq v_i(\omega)\}$ the set of i 's reservation values that are consistent with \mathbf{P} .¹³ Let $\mathcal{R} \equiv \times_{i \in N} R_i$. We call $\mathbf{r} \in \mathcal{R}$ patients' reservation values profile. For each $i \in N$ and each $\mathbf{r} \in \mathcal{R}$, \mathbf{r}_{-i} denotes the restriction of \mathbf{r} to the patients in $N \setminus \{i\}$.

A (**kidney exchange**) **problem** \mathbf{K} is a pair $\mathbf{K} = (\mathbf{P}, \mathbf{r})$.

¹²These rankings may be based on the LYFT index –Life Years From Transplantation– or any other quality–efficiency criteria.

¹³The reader should keep in mind that R_i depends on \mathbf{P} . We are abusing notation but since \mathbf{P} is always a primitive of the analysis, there will not be room for confusion in the arguments.

An **assignment** a is an n -tuple of pairs $a = [(1, \omega), \dots, (n, \omega')]$ such that

- (i) for each $i, j \in N$, $i \neq j$ and each $\omega, \omega' \in \Omega \setminus \{\omega_0\}$, if $(i, \omega), (j, \omega') \in a$, then $\omega \neq \omega'$.
- (ii) if there are $i, j \in N$ such that $(i, \omega_j) \in a$, then $(j, \omega_0) \notin a$.

An assignment is an allocation of the available kidneys to the patients. By (i), an assignment never allocates the same kidney to two different patients, unless that kidney is the null kidney. By (ii), if the kidney of a patient's donor is assigned to another patient, then the initial patient is not assigned the null kidney. For each patient i and each assignment a , we denote by a_i the kidney that patient i receives at a .

In every assignment, kidneys are allocated by forming exchange cycles of patient–donors couples. In each exchange cycle, every patient receives a kidney from the donor of some patient in the cycle and simultaneously her donor's kidney is transplanted to another patient in the cycle. In an exchange cycle among k couples, all the kidneys must be reaped from the donors and transplanted to the patients simultaneously. If this constraint is not taken into account, once a donor's kidney is transplanted to another patient, the donor of the recipient may reject to donate her kidney in order to avoid any clinical complication involved in the operation. This fact implies that an assignment among k couples involves $2k$ simultaneous operations. Since hospitals face evident logistic restrictions, we incorporate such constraints in our analysis through a narrower definition of feasible assignments.

For each assignment a , let π_a be the finest partition of the set of patients such that for each $p \in \pi_a$ and each $i \in p$:

- (i) either there are $j, j' \in p$, with $a_i = \omega_j$ and $a_{j'} = \omega_i$,¹⁴
- (ii) or $a_i = \omega_0$.

Clearly, for each assignment a the partition π_a is unique and well-defined. We define the **cardinality of** a as the $\max_{p \in \pi_a} \#p$.

The cardinality of an assignment refers to the size of the largest exchange cycle formed in the assignment. Basically, it refers to the maximum number of simultaneous operations

¹⁴Note that $j = j'$ and $i = j = j'$ and then $a_i = \omega_i$ are allowed.

involved in an assignment. Of course, the concept of cardinality is crucial for our notion of feasibility.

For each $k \in \mathbb{N}$, $k \leq n$, we say that the assignment a is k -*feasible* if a 's cardinality is not larger than k . Let \mathcal{A}^k be the set of all k -feasible assignments.

An interesting case of feasibility restrictions appears when only immediate exchanges between two couples are admitted. An assignment a is a *pairwise-exchange* assignment ($a \in \mathcal{A}^2$) if a satisfies that if for some $i, j \in N$ $(i, \omega_j) \in a$, then $(j, \omega_i) \in a$.

3 Kidney Assignment Rules

In this paper, we are interested in rules that select a (kidney) assignment for each (kidney exchange) problem. An (*assignment*) *rule* is a mapping φ that selects an assignment a for each problem \mathbf{K} . For each patient i and each problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$, we denote by $\varphi_i(\mathbf{P}, \mathbf{r})$ the kidney assigned to i by φ at \mathbf{K} . As we take patients' preferences profile \mathbf{P} as given, whenever there is no room for confusion, we drop \mathbf{P} from the arguments and simply write $\varphi(\mathbf{r})$.

The assignment selected by a rule can be interpreted as an optimal recommendation that takes into account the preferences of the patients for the available kidneys and their reservation values and that tries to find a compromise between their (maybe conflicting) interests.

Every preference profile \mathbf{P} together with a rule φ define a revelation mechanism. The revelation mechanism (or game form) specifies a set of players (the patients), a set of strategies for each patient, the sets R_i that are consistent with \mathbf{P} ; and an outcome function, $\varphi(\mathbf{P}, \cdot)$. Note that the mechanism (\mathbf{P}, φ) fails short of defining a game in normal (strategic) form because \mathbf{P} does not introduce the information about patients' preferences about the possibility of receiving the null kidney ω_0 .

Next, we present formal definition of the standard conditions for desirable rules. The reader should keep in mind that all the conditions refer to a given observed preference profile \mathbf{P} .

Individual Rationality. For each $i \in N$ and each $\mathbf{r} \in \mathcal{R}$, $v_i(\varphi_i(\mathbf{r})) \geq \max \{r_i, v_i(\omega_i)\}$.

***k*-Efficiency.** For each $\mathbf{r} \in \mathcal{R}$, $\varphi(\mathbf{r}) \in \mathcal{A}^k$ and there is no $a \in \mathcal{A}^k$ such that for each $i \in N$ $a_i \succsim_i \varphi_i(\mathbf{r})$ and for some $j \in N$, $a_j \succ_j \varphi_j(\mathbf{P}, \mathbf{r})$.

Individual rationality is a minimal participation constraint which takes into account patient's right of refusing any transplant and receiving her donor's kidney. On the other hand, *k-efficiency* is the natural version of efficiency taking into account the feasibility restrictions on the cardinality of the assignments because at most $2k$ simultaneous operations can be carried out in the kidney exchange process. Of course, *n-efficiency* corresponds to the classical notion of (full) Pareto efficiency when there are not feasibility constraints.

4 Incentive-Compatibility and Feasibility Constraints

A central issue in the design of an optimal kidney exchange program is the use of all the relevant information in the assignment of the available kidneys. Although doctors may have all the information about the degree of compatibility and fitness among patients and kidneys that is imbedded in preference profiles, there is a key piece of information that remains private information for the patients and must be elicited for public use, their reservation values. The objective of this section is to analyze whether we can construct rules that provide incentives to the patients to reveal their true reservation values in the presence of feasibility constraints on the cardinality of the proposed assignments.

We are interested in rules that provide incentives for the patients to reveal their true reservation values in all the games induced by the revelation mechanism (\mathbf{P}, φ) .

Strategy-proofness. For each $i \in N$, each $\mathbf{r} \in \mathcal{R}$, and each $r'_i \in R_i$, $\varphi_i(\mathbf{r}) \succsim_i \varphi_i(r'_i, \mathbf{r}_{-i})$.

Strategy-proofness implies that reporting the true reservation value is a (weakly) dominant strategy for every patient in all the games compatible with the revelation mechanism \mathbf{P} and φ . Note that *strategy-proofness* is weak in our framework because only reservation values are private information. The justification for the requirements of *strategy-proofness* is two-fold. On the normative side, if patients do not provide the correct reservation values, then the assignment selected by the rule may be based on incorrect information, and

therefore it may represent a far from optimal recommendation to the society. On the positive side, in order to compute patients' best strategies in the revelation game induced by \mathbf{P} and φ , patients simply need to know their own reservation values.

The literature on the allocation of indivisible objects has extensively studied the problem of designing *strategy-proof* and *individually rational* assignment rules.¹⁵ When patients have strict preferences there is a natural way to assign the available kidneys among the patients. We can simply use the Gale's top trading cycle procedure to allocate objects in markets with individual property rights. According to this procedure, given a kidney allocation problem, let every patient point to the donor with her favorite kidney. A top trading cycle consists of patients such that each patient in the cycle points to the donor of the next patient in the cycle. (A single patient may constitute a cycle, by pointing to herself if her donor's kidney is her best preferred kidney or if she thinks that no available kidney is acceptable and she prefers not to perform any operation). Since there is a finite number of patients and kidneys, for each problem there is at least one top trading cycle. Give each patient in a top trading cycle her best preferred kidney, and remove them from the problem with her assigned kidney. Repeat the process until each patient receives a kidney (maybe the null kidney). The resulting assignment is unique given that preferences over compatible kidneys are strict.¹⁶ Moreover, the induced rule satisfies *individual rationality*, *n-efficiency*, and *strategy-proofness*.¹⁷ A top trading cycle however, may involve all the patients.

Example 1. Let $N = \{1, 2, 3, 4\}$. Consider the problem $\mathbf{K} = (\mathbf{P}, \mathbf{r})$ with

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0.99 \\ 0.99 & 0 & 0 & 0 \\ 0 & 0.99 & 0 & 0 \\ 0 & 0.95 & 0.99 & 0 \end{pmatrix},$$

and for each $i \in N$, $r_i = 0.9$. In this problem, the top trading cycle procedure selects the assignment

$$\bar{a} = [(1, \omega_2)(2, \omega_3)(3, \omega_4)(4, \omega_1)].$$

¹⁵See Gale and Shapley (1962); Shapley and Scarf (1974); Abdulkadiroğlu and Sönmez (1999).

¹⁶Patients will rather pick the null kidney rather than receiving an incompatible kidney.

¹⁷See Roth and Postlewaite (1977); Roth (1982).

Under \bar{a} , there is a top trading cycle that involves all the patients and every patient receives her best preferred kidney. Every rule that satisfies individual rationality and 3-efficiency however, selects the assignment

$$a' = [(1, \omega_2)(2, \omega_4)(3, \omega_0)(4, \omega_1)].$$

Clearly, for each $i \in N$, $\bar{a}_i \succsim_i a'_i$ and $\bar{a}_2 \succ_2 a'_2$ but \bar{a} is not 3-feasible.

Our first result shows that feasibility constraints may make it impossible to construct efficient rules that provide the right incentives to the patients at every preference profile.

Theorem 1. For each $2 \leq k \leq n - 1$, there are \mathbf{P} such that no rule satisfies individual rationality, k -efficiency, and strategy-proofness.

Proof. We study two cases. We analyze first the restriction to pairwise exchanges. Then, we provide the proof for $k \geq 3$. In both cases we exploit arguments similar to those employed in the literature of strategy-proof assignment rules in economies with indivisibilities where the core is empty ($k = 2$) or multi-valued ($k \geq 3$).¹⁸

Assume to the contrary there is a rule φ that satisfies *individual rationality*, *2-efficiency*, and *strategy-proofness* for every \mathbf{P} . Consider three patients $\{1, 2, 3\}$ and a preference profile \mathbf{P} such that its restriction to these patients and their donors' kidneys is:

$$\mathbf{P} = \begin{pmatrix} 0 & 0.75 & 0.99 \\ 0.99 & 0 & 0.75 \\ 0.75 & 0.99 & 0 \end{pmatrix},$$

and so that for each $i \in \{1, 2, 3\}$, and each $\omega \notin \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $v_i(\omega) = 0$. Thus, $\omega_2 \succ_1 \omega_3 \succ_1 \omega_1$, $\omega_3 \succ_2 \omega_1 \succ_2 \omega_2$, and $\omega_1 \succ_3 \omega_2 \succ_3 \omega_3$. In order to simplify notation, let $N = \{1, 2, 3\}$. (By *individual rationality*, this is without loss of generality.)

Let $\mathbf{r} = (r_1, r_2, r_3) = (0.6, 0.6, 0.6)$. By *individual rationality* and *2-efficiency*, φ selects an assignment in which two patients exchange their donors' kidneys while the remaining patient receives the null kidney. We assume without loss of generality that $\varphi(\mathbf{r}) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$.

Next, let $\mathbf{r}' = (r'_1, r'_2, r'_3) = (0.9, 0.6, 0.6)$. By *strategy-proofness*, $\varphi_1(\mathbf{r}') = \omega_2$. Finally, let $\mathbf{r}'' = (r''_1, r''_2, r''_3) = (0.9, 0.9, 0.6)$. By *individual rationality* and *strategy-proofness*, $\varphi_2(\mathbf{r}'') = \omega_0$. Then, $\varphi(\mathbf{r}'') = [(1, \omega_0), (2, \omega_0), (3, \omega_0)]$. Note that the assignment $a =$

¹⁸See Sönmez (1999).

$[(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ is 2-feasible, and $a_i \succsim_i \varphi_i(\mathbf{r}'')$ for each $i \in N$ and $a_2 \succ_2 \varphi_2(\mathbf{r}'')$. Then, φ violates 2-efficiency.

Next, we analyze the general case. Let $k \geq 3$. Remember that $k < n$ and then there are at least $k+1$ patients. Assume to the contrary there is a rule φ that satisfies *individual rationality*, *k-efficiency*, and *strategy-proofness* for every \mathbf{P} . Let the preference profile \mathbf{P} be such that for every $i = 1 \dots, k+1$:

$$v_i(\omega_{i+1}) > v_i(\omega_{i+2}) > v_i(\omega_i) > v_i(\omega) = 0, \\ \forall \omega \in \Omega \setminus \{\omega_0, \omega_i, \omega_{i+1}, \omega_{i+2}\}. \text{ (modulo } k+1 \text{)}.$$

Again, (by *individual rationality*, without loss of generality) simplify notation and let $N = \{1, \dots, k+1\}$.

| \succsim_1 | \succsim_2 | \dots | \succsim_{k-1} | \succsim_k | \succsim_{k+1} |
|--------------|--------------|---------|------------------|----------------|------------------|
| ω_2 | ω_3 | \dots | ω_k | ω_{k+1} | ω_1 |
| ω_3 | ω_4 | \dots | ω_{k+1} | ω_1 | ω_2 |
| ω_1 | ω_2 | \dots | ω_{k-1} | ω_k | ω_{k+1} |
| \dots | \dots | \dots | \dots | \dots | \dots |

Table 1: \mathbf{P} : Theorem 1, Case $k \geq 3$.

Let $\mathbf{r} \in \mathcal{R}$ be such that:

$$v_i(\omega_{i+2}) < r_i < v_i(\omega_{i+1}) \quad \text{for each } i \neq k+1 \\ v_{k+1}(\omega_{k+1}) < r_{k+1} < v_{k+1}(\omega_2)$$

Note that, by *individual rationality* either no object is assigned to any patient $1, \dots, k+1$, or patient $k+1$ receives ω_2 , patient 1 receives the null object, and every other patient i receives ω_{i+1} (the kidney of her next to the right neighbor). By *k-efficiency*:

$$\varphi(\mathbf{r}) = \left[\begin{array}{l} (1, \omega_0), \\ (i, \omega_{i+1}), \quad \forall i = 2, \dots, k \\ (k+1, \omega_2) \end{array} \right].$$

Let $\mathbf{r}' \in \mathcal{R}$ be such that for each $i \neq k-1$, $r_i = r'_i$ and $v_{k-1}(\omega_{k-1}) < r_{k-1} < v_{k-1}(\omega_{k+1})$. By *strategy-proofness*, $\varphi_{k-1}(\mathbf{r}') \succ_{k-1} \varphi_{k-1}(\mathbf{r}) = \omega_k$. Note that ω_k is patient $k-1$'s preferred kidney. Then, $\varphi_{k-1}(\mathbf{r}') = \omega_k$. By *k-efficiency* and *individual rationality*, $\varphi(\mathbf{r}) = \varphi(\mathbf{r}')$.

Let $\bar{\mathbf{r}} \in \mathcal{R}$ be such that for each $i \neq k + 1$, $r'_i = \bar{r}_i$ and $v_{k+1}(\omega_2) < \bar{r}_{k+1} < v_{k+1}(\omega_1)$. The same arguments we employed to determine $\varphi(\mathbf{r})$ apply here to obtain:

$$\varphi(\bar{\mathbf{r}}) = \begin{bmatrix} (i, \omega_{i+1}) \text{ (modulo } k+1), & \forall i \notin \{k, k-1\} \\ (k-1, \omega_{k+1}), \\ (k, \omega_0) \end{bmatrix}.$$

Note that $\omega_1 = \varphi_{k+1}(\bar{\mathbf{r}}) = \varphi(\bar{\mathbf{r}}'_{k+1}, \mathbf{r}'_{-(k+1)}) \succ_{k+1} \varphi_{k+1}(\mathbf{r}') = \omega_2$, which contradicts *strategy-proofness*. \square

The previous impossibility result is robust to the introduction of weak preferences over kidneys. All we require is to admit the existence of two indifference classes for acceptable kidneys. Hence, Theorem 1 contrasts with the positive results in dichotomous domains of preferences by Roth et al. (2005a) and Hatfield (2005). Moreover, the result can be extended to incomplete information settings where patients may have incomplete information (i.e. beliefs) about the reservation values of the remaining patients. In a result that parallels the results of Roth (1989), we can prove that there are preference profiles and sets of patients' beliefs about other patients' reservation values such that there is no rule that satisfies *individual rationality*, *k-efficiency*, and (Bayesian) *incentive compatibility*.¹⁹

5 Protective Behavior in Kidney Exchange Problems

The previous section presents a negative result for kidney assignment rules. This negative result is particularly discouraging since we balance the enrichment of preference domain with the increase in the information available to the CTC. Namely, we make the assumption that patients' preferences over available kidneys are known by the CTC because they depend on measurable and verifiable characteristics of patients and donors, like blood types of patients and donors, their age, health status, race, HLA mismatches, etc. The fact that reservation values depend on unobservable patients' characteristics however, introduces severe limitation on the properties that the kidney assignment rules

¹⁹A precise statement of this result can be found in Appendix A. In a recent paper, Villa and Patrone (2008) prove that the rule that maximizes the sum of the welfare of the patients is not incentive compatible.

may satisfy. Thus, information on the patients' reservation values have to be elicited and patients might be tempted to misreport such information to get better kidneys. In this scenario, if patients waiting for a transplant are strongly risk-averse, they might prefer to choose their strategies so as to “protect” themselves from the worst eventuality as far as possible. This “lexicographic maximin” behavior assumption is captured by the notion of “protective behavior”.

Consider a preference profile \mathbf{P} , a rule φ and a patient i who faces the revelation mechanism (game form) defined by (\mathbf{P}, φ) .²⁰

For each patient i , each pair of reservation values $r_i, s_i \in R_i$, and each real number $l \in \mathbb{R}$, let:

$$c^{r_i}(l, s_i) = \{\mathbf{s}_{-i} \in \mathcal{R}_{-i} \mid v_i(\varphi(\mathbf{P}, (s_i, \mathbf{s}_{-i}))) = l\}.$$

Then, $c^{r_i}(l, s_i)$ is the set of restricted profiles of reservation values of the remaining patients under which i receives a kidney ω with $v_i(\omega) = l$ when i announces s_i and her true reservation value is r_i .

Given \mathbf{P} and φ , for each patient i with reservation value $r_i \in \mathcal{R}_i$, $s_i, s'_i \in R_i$, s_i **protectively dominates** s'_i , denoted $s_i \text{ d}(r_i) s'_i$ if there exists $l \in \mathbb{R}$ such that:

(i) $c^{r_i}(t, s_i) \cap c^{r_i}(t', s'_i) = \emptyset$ for each $t \leq k$ and $t < t'$,

(ii) $c^{r_i}(l, s_i) \subsetneq c^{r_i}(l, s'_i)$.

For each patient i and each $r_i \in \mathcal{R}_i$, let $D(r_i) \equiv \{s_i \in \mathcal{R}_i \mid \text{there is no } s'_i \in \mathcal{R}_i \text{ with } s'_i \text{ d}(r_i) s_i\}$ be the set of **protective strategies** of patient i .

In order to compare two strategies according to this criterion an agent looks at the utility level of the worst outcome (say $t = \min_{\omega \in \Omega} v_i(\omega)$). Strategy s_i protectively dominates s'_i if two conditions hold. First, it never occurs that there exists a profile such that

²⁰Throughout this section, we assume that all the information available to the CTC, namely \mathbf{P} and φ is also available to the patients. The results are not altered however, if we only assume that patients have information about the set of mutually compatible exchanges and not the whole preference profile \mathbf{P} . We maintain the assumption on common knowledge of preference profiles just for the sake of clarity of exposition and to avoid the introduction of additional notation. We delay the discussion on the incomplete information case where patients only know her own preferences and reservation values to the end of this section.

strategy s_i induces this minimum utility level and strategy s'_i induces a higher level of utility. Second, there are some profiles such that s'_i induces the minimum level of utility, while s_i induces a larger payoff for patient i . If the first condition holds but not the second because $c^{ri}(t, s_i) = c^{ri}(t, s'_i)$, then patient i repeats the comparison with respect to the next to the worst utility level.

Clearly, the protective domination relation is not complete, but it is transitive. Thus, for each reservation value $r_i \in \mathcal{R}_i$, the set $D(r_i)$ is not empty. Moreover, if there is a unique protective strategy, $\{s_i\} = D(r_i)$, then $s_i \text{ d}(r_i) s'_i$ for each $s'_i \in R_i \setminus \{s_i\}$.

Truth-telling requires that reporting the true reservation value is a protective strategy for the player at every direct revelation game generated by \mathbf{P} and φ . Protective domination however, is not a complete relation. Thus, if there are several different protective strategies besides reporting the true value, then we could not say that truth-telling is an optimal strategy for the patients. This fact calls for a stronger implementability requirement.

Let $\mathbf{r}, \bar{\mathbf{r}} \in \mathcal{R}$. The reservation values profile $\bar{\mathbf{r}}$ is a **protective equilibrium at \mathbf{r}** iff for each patient i , $\bar{r}_i \in D(r_i)$.

A rule φ is **directly implementable via protective equilibria (DIPE)** iff for each $\mathbf{r}, \bar{\mathbf{r}} \in \mathcal{R}$ such that $\bar{\mathbf{r}}$ is a protective equilibrium at \mathbf{r} , $\varphi(\mathbf{r}) = \varphi(\bar{\mathbf{r}})$.

With our focus on DIPE rules, we emphasize that there is no *a priori* reason why implementation of rules should be achieved through equilibria involving truthful preference revelation. In fact, focusing on the implementability of rules reflects the consideration that correct revelation is not an objective *per se*, and what we really care about is the result of strategic behavior, rather than its correspondence to truth. The following result indicates, however, that the only rules which are implementable in our sense are those which in fact guarantee truthful revelation by all patients, under the behavioral and informational assumptions underlying the definition of protective equilibrium.

For each patient i , strategies $s_i, s'_i \in R_i$ are **equivalent** if for each $\mathbf{s}_{-i} \in \mathcal{R}_{-i}$, $\varphi_i(s_i, \mathbf{s}_{-i}) = \varphi_i(s'_i, \mathbf{s}_{-i})$.

Truth-telling is (essentially) the unique protective strategy for patient i
if for each $r_i \in \mathcal{R}_i$,

$$D(r_i) = \{r'_i \mid r_i \text{ and } r'_i \text{ are equivalent}\}.$$

Our first result in this section replicates Theorem 1 in Barberà and Dutta (1982).

Proposition 1. *Let $\mathbf{P} \in \mathcal{P}$. A rule φ is directly implementable via protective equilibria if and only if for each patient i , truth-telling is essentially the unique protective strategy.*

The proof of Proposition 1 mimics the proof Theorem 1 in Barberà and Dutta (1982), and it is relegated to the Appendix. The proof consists of three steps. First, we prove that the set of undominated strategies for non-equivalent reservation values (strategies) are disjoint. Then, we show reservation values with the same sets of admissible kidneys are equivalent. Finally, we check that revealing the true reservation value is indeed an undominated strategy.

The following example shows that the difference between truth-telling as a unique protective strategy and *strategy-proofness*. There are profiles for which no rule satisfies *individual rationality*, *2-efficiency*, and *strategy-proofness*, but there exist rules that satisfy *individual rationality*, *2-efficiency*, and such that truth-telling is the unique protective strategy.

Example 2. *Let $N = \{1, 2, 3\}$ and let ψ be the rule that maximizes the number of transplants obtained through pairwise exchanges, and ties are broken according to patient 1's preferences. Consider the preference profile presented in the first part of the proof of Theorem 1,*

$$\mathbf{P} = \begin{pmatrix} 0 & 0.75 & 0.75 \\ 0.99 & 0 & 0.99 \\ 0.75 & 0.99 & 0 \end{pmatrix}.$$

Let $r_2 \in \mathcal{R}_2$. If $s_2 < 0.75$, then

$$\begin{aligned} c^{r_2}(r_2, s_2) &= \{\mathbf{s}_{-2} \in \mathcal{R}_{-2} \mid s_1 > 0.99 \text{ and } s_3 > 0.75\}, \\ c^{r_2}(0.75, s_2) &= \{\mathbf{s}_{-2} \in \mathcal{R}_{-2} \mid s_1 < 0.99\}, \\ c^{r_2}(0.99, s_2) &= \{\mathbf{s}_{-2} \in \mathcal{R}_{-2} \mid s_1 > 0.99 \text{ and } s_3 < 0.75\}. \end{aligned}$$

If $0.75 < s'_2 < 0.99$, then

$$\begin{aligned} c^{r_2}(r_2, s'_2) &= \{\mathbf{s}_{-2} \in \mathcal{R}_{-2} \mid s_1 > 0.99 \text{ or } s_3 > 0.75\}, \\ c^{r_2}(0.99, s'_2) &= \{\mathbf{s}_{-2} \in \mathcal{R}_{-2} \mid s_1 > 0.75 \text{ and } s_3 > 0.75\}. \end{aligned}$$

Finally, if $s''_2 > 0.99$, then $c^{r_2}(r_2, s''_2) = \mathcal{R}_{-2}$.

Assume that $r_2 < 0.75$, then for each $s_2 < 0.75$, s_2 is strategically equivalent to r_2 . Moreover, for each $s'_2 > 0.75$, $c^{r_2}(r_2, r_2) \subsetneq c^{r_2}(r_2, s'_2)$. Using a similar argument for $r_2 > 0.75$, we can prove that truth-telling is (essentially) the unique protective strategy for patient 2. The same reasoning applies to the remaining patients, and therefore ψ is DIPE. On the other hand, It is immediate to check that ψ violates strategy-proofness. Let $\bar{\mathbf{r}} = (0.8, 0.1, 0.1)$ and $r'_2 = 0.8$, then $\psi(\bar{\mathbf{r}}) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$ and $\psi(r'_2 \bar{\mathbf{r}}_{-2}) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$. Because $\omega_3 \succ_2 \omega_1$, $\psi_2(r'_2 \bar{\mathbf{r}}_{-2}) \succ_2 \psi_2(\bar{\mathbf{r}})$. Notice that the violation of strategy-proofness appears for reservation values profile where patient 2 is not receiving her worst preferred outcome.

In order to follow with the analysis of truth-telling as a protective strategy, we need to introduce additional notation and definitions. Consider a patient i , (given \mathbf{P} and φ) the set of possible outcomes for patient i is defined by:

$$\Omega_i \equiv \{\omega \in \Omega \text{ such that there exists } \mathbf{r} \in \mathcal{R} \text{ with } \varphi_i(\mathbf{r}) = \omega\}.$$

With the definition of Ω_i at hand, for each problem reservation value $r_i \in \mathcal{R}_i$ the set of acceptable kidneys for i is defined as:

$$\Omega_i^+(r_i) \equiv \{\omega \in \Omega_i \text{ such that } v_i(\omega) \geq \max\{v_i(\omega_i), r_i\}\}.$$

Of course, i 's set of unacceptable kidneys is analogously defined:

$$\Omega_i^-(r_i) \equiv \{\omega \in \Omega_i \text{ such that } v_i(\omega) < \max\{v_i(\omega_i), r_i\}\}.$$

If φ satisfies *individual rationality*, then $\varphi_i(\mathbf{r}) \in \Omega_i^+(r_i) \neq \emptyset$. Finally let $\omega_i^+(r_i) \equiv \arg \min_{\omega \in \Omega_i^+(r_i) \setminus \{\omega_0\}} v_i(\omega)$ if $\Omega_i^+(r_i) \neq \emptyset$, $\omega_i^+(r_i) \equiv \emptyset$ otherwise; and $\omega_i^-(r_i) \equiv \arg \max_{\omega \in \Omega_i^-(r_i)} v_i(\omega)$ if $\Omega_i^-(r_i) \neq \emptyset$, $\omega_i^-(r_i) \equiv \emptyset$ otherwise. Thus, $\omega_i^+(r_i)$ is i 's worst acceptable kidney and $\omega_i^-(r_i)$ is i 's best unacceptable kidney. Note that if patient i 's reservation value is lower than her donor's kidney valuation, $r_i < v_i(\omega_i)$, then $\omega_i^+(r_i) = \omega_i$ and $\omega_i^-(r_i) = \emptyset$.

At this point, we introduce two conditions that turn out to be necessary for truth-telling being a unique protective strategy in our environment.

Invariance. For each patient i , and each pair $\mathbf{r}, \mathbf{r}' \in \mathcal{R}$ such that $\mathbf{r}_{-i} = \mathbf{r}'_{-i}$, if $\Omega_i^+(r_i) = \Omega_i^+(r'_i)$, then $\varphi_i(\mathbf{r}) = \varphi_i(\mathbf{r}')$.

Weak Consistency. For each $i \in N$, each $\mathbf{r} \in \mathcal{R}$, and each $r_i \in R_i$, if $\varphi_i(\mathbf{r}) = \omega_0$ and $r_i < r'_i$, then $\varphi_i(r'_i, \mathbf{r}_{-i}) = \omega_0$.

Invariance requires that if a patient changes her reported reservation value, but this change does not affect her set of acceptable kidneys, then the patient receives the same kidney. Note that a rule satisfying *invariance* may be responsive to the cardinal information of patients' reported reservation values. For instance, think of a serially dictatorial (priority) rule that always picks the best feasible allocation for a given patient, and then proceeds iteratively (serially breaking ties) according to a priority list that depends on the reservation value reported by that first patient on the list. *Weak Consistency* is a convenient weakening of the Axiom of Choice for single-valued choice functions.²¹ Simply, if a patient does not receive a kidney when she reports r_i , then she cannot be assigned to a compatible kidney when she raises her reservation value. Note that if a rule satisfies *individual rationality*, then *weak consistency* applies the logic behind the Axiom of Choice only at situations where the patients receive the worst possible outcome, the null kidney.

Proposition 2. For each preference profile \mathbf{P} and each rule φ , if φ satisfies individual rationality and for each patient i truth telling is the unique protective strategy, then φ satisfies invariance and weak consistency.

Proof. Let $i \in N$. We start with the proof of *invariance*. Assume first that $r_i < v_i(\omega_i)$. In this case, $\Omega_i^-(r_i) = \emptyset$ and we need to prove that every $s_i \in R_i$ such that $s_i < v_i(\omega_i^+(r_i))$, is equivalent to r_i . Let $s_i < v_i(\omega_i)$. By *individual rationality*, for each $\bar{\mathbf{s}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$, $v_i(\varphi_i(s_i, \bar{\mathbf{s}}_{-i})) \geq v_i(\omega_i)$. Let $\bar{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi(r_i, \bar{\mathbf{r}}_{-i}) = \omega_i$. Because truth-telling is the unique protective strategy for patient i , $c^{r_i}(v_i(\omega_i), r_i) \subseteq c^{r_i}(v_i(\omega_i), s_i)$ and $\varphi(s_i, \bar{\mathbf{r}}_{-i}) = \omega_i$. Next, assume that i 's true reservation value is s_i and let $\tilde{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi(s_i, \tilde{\mathbf{r}}_{-i}) = \omega_i$. Repeating the previous argument, *individual rationality* and the fact that truth-telling is the unique protective strategy for i , imply that $c^{s_i}(v_i(\omega_i), s_i) \subseteq$

²¹See Arrow (1959) and Sen (1971). We follow Hatfield (2005) in the terminology.

$c^{s_i}(v_i(\omega_i), r_i)$, and $\varphi(r_i, \tilde{\mathbf{r}}_{-i}) = \omega_i$. Applying the same argument iteratively for each level $k > r_i$, we conclude that r_i and s_i are equivalent.

Next, assume that $r_i > v_i(\omega_i)$ and that $\Omega_i^-(r_i) \neq \emptyset$.²² Hence, we need to check that every $s_i \in R_i$ such that

$$v_i(\omega^-(r_i)) < s_i < v_i(\omega^+(r_i))$$

is equivalent to r_i . Let $s_i \in (v_i(\omega^-(r_i)), v_i(\omega^+(r_i)))$. By *individual rationality* and the definitions of $\omega_i^-(r_i)$ and $\omega_i^+(r_i)$, for each $\bar{\mathbf{s}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$, $v_i(\varphi_i(s_i, \bar{\mathbf{s}}_{-i})) \geq r_i$. Let $\hat{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi_i(r_i, \hat{\mathbf{r}}_{-i}) = \omega_0$. Because truth-telling is the unique protective strategy for patient i , $c^{r_i}(r_i, r_i) \subseteq c^{r_i}(r_i, s_i)$, and $\varphi(s_i, \hat{\mathbf{r}}_{-i}) = \omega_0$. Next assume that i 's true reservation value is s_i and let $\tilde{\mathbf{r}}_{-i} \in \times_{j \neq i} \mathcal{R}_j$ be such that $\varphi(s_i, \tilde{\mathbf{r}}_{-i}) = \omega_0$. Repeating the previous argument, *individual rationality* and the fact that truth-telling is the unique protective strategy for i imply that $c^{s_i}(s_i, s_i) \subseteq c^{s_i}(s_i, r_i)$, and $\varphi(r_i, \tilde{\mathbf{r}}_{-i}) = \omega_0$. Applying the arguments of the previous paragraph iteratively for each level $k > r_i$, we conclude that r_i and s_i are equivalent. Finally note that the same argument suffices to prove that every $s_i < v_i(\omega_i^+(r_i))$ is equivalent to r_i if $r_i > v_i(\omega_i)$ and $\Omega_i^-(r_i) = \emptyset$, which completes the proof of *invariance*.

We conclude with the proof of *weak consistency*. Note that by *individual rationality*, for each $k < \max\{r_i, v_i(\omega_i)\}$, $c^{r_i}(k, r_i) = \emptyset$ and $c^{r_i}(\max\{r_i, v_i(\omega_i)\}, r_i) \neq \emptyset$. Assume that φ does not satisfy *weak consistency*, then there are $i \in N$, and $r_i < r'_i$ such that $\varphi_i(\mathbf{r}) = \omega_0$ and $\varphi_i(r'_i, \mathbf{r}_{-i}) \neq \omega_0$. Note that, by *individual rationality*, $r_i > v_i(\omega_i)$. Then, there is $\mathbf{r}' \in \mathcal{R}$ with $r'_i > r_i$ such that $\varphi_i(r_i, \mathbf{r}'_{-i}) = \omega_0$ but $\varphi_i(\mathbf{r}') \neq \omega_0$. By *individual rationality*, $\varphi_i(\mathbf{r}') \in \Omega_i^+(r_i) \setminus \{\omega_0\}$. Hence, by (i) of the definition of protective domination, r_i does not protectively dominates r'_i . Moreover, r_i and r'_i are not equivalent strategies. These facts contradict that truth-telling is essentially the unique protective strategy. \square

The next proposition shows that *invariance* and *weak consistency* are also sufficient if only pairwise exchanges are admitted.

Proposition 3. *For each preference profile \mathbf{P} , if the rule φ satisfies individual rationality, 2-efficiency, invariance, and weak consistency, then for each patient i truth-telling is the unique protective strategy.*

²²By *individual rationality*, this is always the case if $v_i(\omega_i) \neq 0$.

Proof. Fix a patient i and a reservation value $r_i \in \mathcal{R}_i$. By *invariance*, we need only prove that r_i protectively dominates every strategy $s_i \in R_i$ such that $s_i \notin [\omega_i^-(r_i), \omega_i^+(r_i)]$ if $\omega_i^-(r_i) \neq \emptyset$, and every $s'_i \notin (0, \omega_i^+(r_i)]$ if $\omega_i^-(r_i) = \emptyset$.

Assume first that $r_i < v_i(\omega_i)$. Then, $\omega_i^+(r_i) = \omega_i$. Note that, by *individual rationality*, for each $t < v_i(\omega_i)$, $c^{r_i}(t, r_i) = \emptyset$. Let $s_i \in R_i$ be such that $s_i > v_i(\omega_i) = \omega_i^+(r_i)$. Clearly, for each $t \leq r_i$ and each $t' > t$, $c^{r_i}(t, r_i) \cap c^{r_i}(t', s_i) = \emptyset$, which proves the (ii) of protective domination. Next, consider $\hat{\mathbf{r}} \in \mathcal{R}$ such that $\hat{r}_i = s_i$ and for each patient $j \neq i$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $\hat{r}_j > v_j(\omega)$. By *individual rationality*, for each $j \in N$ $\varphi_j(\hat{\mathbf{r}}) = \omega_0$. Then, $c^{r_i}(r_i, r_i) = \emptyset \subsetneq c^{r_i}(r_i, s_i) \neq \emptyset$, which proves condition (ii) of protective domination.

Next, assume that $r_i > v_i(\omega_i)$ and assume that $\omega_i^-(r_i) \neq \emptyset$. Let $s'_i \in R_i$ be such $s'_i < v_i(\omega_i^-(r_i))$. By *individual rationality*, $c^{r_i}(t, r_i) = \emptyset$ for all $t \leq v_i(\omega_i^-(r_i))$. Let $j \in N$ be such that $\omega_i^-(r_i) = \omega_j$.²³ Consider the reservation values profile $\mathbf{r}' \in \mathcal{R}$ such that $r'_l = s'_i$, for each $l \notin \{i, j\}$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $r'_l > v_l(\omega)$; and $v_j(\omega_i) > r'_j$. (Note that this is possible because $\omega_i^-(r_i) \in \Omega_i$.) By *individual rationality* and *k-efficiency*, $\varphi_i(\mathbf{r}') = \omega_j$ and $\varphi_i(r_i, \mathbf{r}'_{-i}) = \omega_0$. Clearly, $v_i(\varphi_i(r_i, \mathbf{r}'_{-i})) \geq r_i > v_i(\varphi_i(\mathbf{r}'))$. Hence $c^{r_i}(v_i(\varphi_i(\mathbf{r}')), r_i) \subsetneq c^{r_i}(v_i(\varphi_i(\mathbf{r}')), s'_i) \neq \emptyset$. This suffices to prove that r_i protectively dominates s'_i .

We conclude by checking that if $r_i > v_i(\omega_i)$, then r_i protectively dominates every $s''_i \in \mathcal{R}_i$ such that $s''_i > v_i(\omega_i^+(r_i)) \geq r_i$. By *individual rationality*, for each $t < r_i$, $c^{r_i}(t, r_i) = \emptyset$. Note that for each $\mathbf{r}^* \in \mathcal{R}$ such that $r_i^* = r_i$ and $\varphi_i(\mathbf{r}^*) = \omega_0$, and for each $s_i > r_i$, by *weak consistency*, $\varphi_i(s_i, \mathbf{r}^*_{-i}) = \omega_0$. Then, for each $t' > r_i$ we have $c^{r_i}(r_i, r_i) \cap c^{r_i}(t', s''_i) = \emptyset$ which proves condition (i) of the definition of protective domination. The previous argument implies that $c^{r_i}(r_i, r_i) \subseteq c^{r_i}(r_i, s''_i)$. In order to conclude the argument, we show that the previous inclusion is proper. Let $j \in N$ be such that $\omega^+(r_i) = \omega_j$. Consider the reservation values profile $\mathbf{r}'' \in \mathcal{R}$ such that $r''_i = s''_i$, $v_j(\omega_i) > r''_j$, and for each $m \notin \{i, j\}$ and each $\omega \in \Omega \setminus \{\omega_0\}$, $v_m(\omega) < r_m$. By *individual rationality* and *k-efficiency*, $\varphi_i(r_i, \mathbf{r}''_{-i}) = \omega_j$ and $\varphi_i(\mathbf{r}'') = \omega_0$. Hence $c^{r_i}(r_i, r_i) \subsetneq c^{r_i}(r_i, r''_i)$ which proves condition (ii) of the definition of protective domination. \square

The results in Propositions 2–3 are clearly positive. Under protective behavior, there is a large class of rules that provide incentives for the patients to reveal their true reservation

²³It is possible that $j = i$.

values. This family includes every rule that maximizes a strict order over the set of efficient and individually rational pairwise assignments. Hence, rules that maximize the number of exchanges, or serial priority rules satisfy our axioms. In addition, the previous results are in line with the results in Roth et al. (2005a) and Hatfield (2005) despite we start from different assumptions on patients' preferences and strategic behavior.

The previous result shows that when only pairwise exchanges are possible, under protective behavior *weak consistency* is almost sufficient for providing the right incentives to patients. In the light of the arguments of the proof of Proposition 3, the result extends to preference profiles and rules in which for every patient every kidney in the set of possible outcomes can be obtained through a pairwise exchange.²⁴ This observation however, does not hold generally. Our next result shows that the restriction on pairwise exchanges is essential for the positive result. If larger cycles are possible, then truth-telling may fail to be a protective strategy.

Proposition 4. *Let φ be a rule that satisfies individual rationality and k -efficiency for some $k \geq 3$. There are preference profiles \mathbf{P} such that if for some $\mathbf{r} \in \mathcal{R}$ $\#\varphi(\mathbf{r}) > 2$ then for some patient i $r_i \notin D(r_i)$.*

Proof. Consider the following counterexample. Let $N = \{1, 2, 3\}$ and let the preference profile \mathbf{P} be such that

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0.75 \\ 0.9 & 0 & 0.9 \\ 0 & 0.9 & 0 \end{pmatrix}.$$

Let the rule φ satisfy *individual rationality* and *3-efficiency*. *Individual rationality* and *3-efficiency* imply that φ is defined according to Table 2 and Table 3.

| $r_2 \setminus r_3$ | $r_3 > 0.9$ | $r_3 \in (0.75, 0.9)$ | $r_3 < 0.75$ |
|---------------------|---|---|---|
| $r_2 > 0.9$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ |
| $r_2 < 0.9$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ | $[(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ |

Table 2: $r_1 > 0.9$

²⁴This is the case for the preference profile presented in the proof of the case $k = 2$ of Theorem 1.

| $r_2 \setminus r_3$ | $r_3 > 0.9$ | $r_3 \in (0.75, 0.9)$ | $r_3 < 0.75$ |
|---------------------|---|---|--|
| $r_2 > 0.9$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ |
| $r_2 < 0.9$ | $[(1, \omega_0), (2, \omega_0), (3, \omega_0)]$ | $[(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ | either $[(1, \omega_0), (2, \omega_3), (3, \omega_2)]$, or $[(1, \omega_2), (2, \omega_3), (3, \omega_1)]$. |

Table 3: $r_1 < 0.9$

Assume now that there is $\mathbf{r} \in \mathcal{R}$ such that $\#\varphi(\mathbf{r}') = 3$. That is, there is \mathbf{r} such that $\varphi(\mathbf{r}) = [(1, \omega_2), (2, \omega_3), (3, \omega_1)]$. Necessarily, $r_1 < 0.9$, $r_2 < 0.9$, and $r_3 < 0.75$. It is immediate to compute: $c^{r_3}(r_3, r_3) = c^{r_3}(r_3, 0.8) = \{(s_1, s_2) \mid s_2 > 0.9\}$. On the other hand,

$$c^{r_3}(0.75, r_3) \neq \emptyset, \quad c^{r_3}(0.75, 0.8) = \emptyset,$$

Note that

$$\varphi_3(r_1, r_2, 0.8) = \omega_2 \succ_3 \omega_1 = \varphi_3(\mathbf{r}).$$

If patient 3 reservation value is r_3 , then $s_3 = 0.8$ protectively dominates r_3 . In this case, by reporting $s_3 = 0.8$, patient 3 is not taking any risk at the other patients' profiles for which she receives the null kidney. In the case of receiving an acceptable kidney however, by reporting $s_3 = 0.8$ she always gets her best preferred kidney. Interestingly, we have to move beyond the first round of comparisons between strategies to check domination. \square

Proposition 4 implies that at some preference profiles rules *individual rationality* and *k-efficiency*, it is necessary to introduce limits on the cardinality of the recommended exchanges in order to provide incentives for the patients to reveal their true reservation values. The problem affects reasonable rules like priority rules and every rule that maximizes the number of (individually rational) exchanges. Hence, Proposition 3 provides a strategic *rationale* for the focus on pairwise exchanges. Besides the logistic and direct incentives problems described by Roth et al. (2005a), the restriction to pairwise exchanges may be necessary to obtain the correct information from the patients in the protective behavior scenario.

Two final remarks are in order.

Proposition 4 depends crucially on the information available to the patients. In the light of the proof, it is clear that it is not necessary that the patients have perfect knowledge of the preference profile \mathbf{P} . It suffices that the patients know the sets of compatible

kidneys of the remaining patients. If each patient only has information about her own preferences, then the arguments in the proof of Proposition 3 would apply to prove a general version of Proposition 3. In such incomplete information framework, for each $k \leq n$, if a rule φ satisfies *individual rationality*, *k-efficiency*, *invariance*, and *weak consistency*, then φ is DIPE.

Finally, in this article we have focused on situations where each patient has only one possible donor. When patients can have multiple donors, it could be possible that a patient may have incentives to withdraw some of her possible donors if by doing so the assignment rule assigns her a better kidney. Again, with slight modifications of the arguments in the proof of Proposition 3, we can prove that rules that satisfy a version of Arrow's Axiom of Choice (defined over feasible assignments) are immune to such manipulations.

6 Concluding Remarks

In this paper, we have proposed a framework that departs from proposed by Roth et al. (2005a) in the design of the New England kidney exchange clearing-house in two relevant aspects. These departures try to convey some relevant institutional features of European approach to kidney exchange. On the one hand, we assume that patients may have heterogeneous preferences over the set of compatible kidneys. On the other hand, we assume that the CTC may avail with detailed information about patients' preferences. Our first result shows the difficulties to fulfill different forms of incentive compatibility if there are restrictions on the cardinality of feasible exchanges. The positive results are restored if patients follow the protective behavior and are strongly averse to the risk of refusing a transplant of a compatible kidney. Namely, if only pairwise kidney exchange are feasible, then the rules which satisfy *strategy-proofness* and (constrained) *efficiency* in the dichotomous domain provide incentives for truthful revelation in the protective behavior scenario. Interestingly, the difficulties return if larger exchanges are admitted. These results have policy implications. The efficiency gains of making possible cycles larger than pairwise exchanges can be overcome by the impossibility of eliciting the correct information from the patients. Since the cost of slackening the feasibility constraints are high, then our work puts some doubts on the economic advantage of these investments for the healthcare service.

In order to conclude, we devote a few lines to sketch some open venues of further research. Evidently, our assumption on patients' (protective) behavior deserves to be tested by means of controlled questionnaires on the population of patients in the waiting lists. On the other hand, incentive problems in kidney exchange environments have been studied on static model as ours. It is evident however, that kidney transplantation has a dynamic component. It seems a promising line of new research the analysis of dynamic and strategic models where patients and kidneys are available sequentially and simultaneously living donation and kidney exchange are feasible procedures.²⁵ In the light of the technical difficulties that appear in standard queue-management models, the analysis of protective behavior in such settings is a promising line of investigation.

References

- A. Abdulkadiroğlu and T. Sönmez. House allocation with existing tenants. *Journal of Economic Theory*, 88:233–260, 1999.
- A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93-3:729–747, 2003.
- K. J. Arrow. Rational choice functions and orderings. *Economica*, 26:121–127, 1959.
- S. Barberà and B. Dutta. Implementability via protective equilibria. *Journal of Mathematical Economics*, 10:49–65, 1982.
- S. Barberà and M. O. Jackson. Maximin, leximin, and the protective criterion: characterizations and comparisons. *Journal of Economic Theory*, 46:34–44, 1988.
- G. Becker and J. J. Elias. Introducing incentives in the market for living organ donations, 2002. Working Paper, University of Chicago.
- F. L. Delmonico. Exchanging kidneys – advances in living-donor transplantation. *The New England Journal of Medicine*, 350:1812–1814, 2004.

²⁵See ‘A Daisy Chain of Kidney Donations’ in *The Wall Street Journal*, September 23, 2008. Su and Zenios (2006) and Ünver (2009) are the first attempts to analyze the dynamic setting. Su and Zenios (2006) focus on self-selection mechanisms, where patients choose among different waiting list for different types of kidneys. Ünver (2009) study a dynamic setting that combines waiting lists, LPE, and KPE where patients only care about the absolute compatibility of prospective kidneys.

- F. L. Delmonico, G. S. Lipkowitz, P. E. Morrissey, J. S. Stoff, J. Himmelfarb, W. Harmon, M. Pavlakis, H. Mah, J. Goguen, R. Luskin, E. Milford, G. B. M. Chobanian, B. Bouthot, M. Lorber, and R. J. Rohrer. Donor kidney exchanges. *American Journal of Transplantation*, 4:1628–1634, 2004.
- R. J. Duquesnoy, S. Takemoto, P. de Lange, I. I. N. Doxiadis, G. M. T. Schreuder, G. G. Persijn, and F. J. Claas. HLAMatchmaker: A molecularly based algorithm for histocompatibility determination. III. effect of matching at the HLA–A,B amino acid triplet level o kidney transplant survival. *Transplantation*, 75(6):884–889, 2003.
- D. Gale and L. Shapley. College admission and stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.
- J. W. Hatfield. Pairwise kidney exchange: Comment. *Journal of Economic Theory*, 125:189–193, 2005.
- I. Kaplan, J. A. Houp, M. S. Leffell, J. M. Hart, and A. A. Zachary. A computer match program for paired and unconventional kidney exchanges. *American Journal of Transplantation*, 5:2306–2308, 2005.
- K. Keizer, M. de Klerk, B. Haase-Kromwijk, and W. Weimar. The dutch algorithm for allocation in living donor kidney exchange. *Transplantation Proceedings*, 37:589–591, 2005.
- M. d. Klerk, K. Keizer, and W. Weimar. Donor exchange for renal transplantation. *The New England Journal of Medicine*, 351:935, 2004.
- L. W. Kranenburg, T. Visak, W. Weimar, and et al. Startling a crossover kidney transplantation program in the Netherlands: ethical and psychological considerations. *Transplantation*, 78:194, 2004.
- R. M. Merion, V. B. Ashby, R. A. Wolfe, D. A. Distant, T. E. Hulbert-Shearon, R. A. Metzger, A. O. Ojo, and F. K. Port. Deceased–donor characteristics and the survival benefit of kidney transplantation. *Journal of American Medical Association*, 294(21):2726–2733, 2005.
- G. Opelz. Impact of HLA compatibility on survival of kidney transplants from unrelated live donors. *Transplantation*, 64:1473–1475, 1997.

- A. Roth. The economist as engineer: Game theory, experimental economics, and computation as tools of design economics. *Econometrica*, 70:1341–1378, 2002.
- A. E. Roth. Incentive compatibility in a market with indivisible goods. *Economics Letters*, 9:127–132, 1982.
- A. E. Roth. Two–sided matching with incomplete information about others’ preferences. *Games and Economic Behavior*, 1:191–209, 1989.
- A. E. Roth and A. Postlewaite. Weak versus strong domination in a market of indivisible goods. *Journal of Mathematical Economics*, 4:131–137, 1977.
- A. E. Roth, T. Sönmez, and M. U. Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.
- A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125:151–188, 2005a.
- A. E. Roth, T. Sönmez, and M. U. Ünver. A kidney exchange clearinghouse in New England. *American Economic Review, Papers and Proceedings*, 95:376–380, 2005b.
- A. E. Roth, T. Sönmez, M. U. Ünver, F. L. Delmonico, and S. L. Saidman. Utilizing list exchange and nondirected donations through “chain” paired kidney donations. *American Journal of Transplantation*, 6:2694–2705, 2006.
- A. E. Roth, T. Sönmez, and M. U. Ünver. Efficient kidney exchange: Coincidence of wants in a markets with compatibility preferences. *American Economic Review*, 97–3: 828–851, 2007.
- S. L. Saidman, A. E. Roth, T. Sönmez, M. U. Ünver, and F. L. Delmonico. Increasing the opportunity of live kidney donation by matching for two– and three–way exchanges. *Transplantation*, 81:773–782, 2006.
- M. A. Schnitzler, C. S. Hollenbeak, D. S. Cohen, R. S. Woodward, J. A. Lowell, G. G. Singer, R. J. Tesi, T. K. Howard, T. Mohanakumar, and D. C. Brennan. The economic implications of HLA matching in cadaveric renal transplantation. *The New England Journal of Medicine*, 341:1440–1446, 1999.

- D. L. Segev, S. E. Gentry, J. K. Melancon, and R. A. Montgomery. Characterization of waiting times in a simulation of kidney paired donation. *American Journal of Transplantation*, 5:2448–2455, 2005a.
- D. L. Segev, S. E. Gentry, D. S. Warren, B. Rech, and R. A. Montgomery. Kidney paired donation and optimizing the use of living donor organs. *Journal of American Medical Association*, 293:1883–1890, April 2005b.
- A. K. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38:307–317, 1971.
- L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.
- T. Sönmez. Strategy-proofness and essentially single-valued cores. *Econometrica*, 67:677–689, 1999.
- T. Sönmez and M. U. Ünver. Kidney exchange with good samaritan donors: A characterization, 2005. Unpublished Manuscript, Boston College and University of Pittsburgh.
- A. Spital. Donor exchange for renal transplantation. *The New England Journal of Medicine*, 351:936, 2004.
- X. Su and S. A. Zenios. Recipient choice can address the efficiency–equity trade–off in kidney transplantation: A mechanism design model. *Management Science*, 52:1647–1660, 2006.
- X. Su, S. A. Zenios, and G. M. Chertow. Incorporating recipient choice in kidney transplantation. *Journal of the American Society of Nephrology*, 15:1656–1663, 2004.
- M. U. Ünver. Dynamic exchange mechanisms. *Forthcoming, Review of Economic Studies*, 2009.
- S. Villa and F. Patrone. Incentive compatibility in kidney exchange problems. *Health Care Management Science*, 2008. doi: 10.1007/s10729-008-9089-0.
- R. Wilson. Architecture of power markets. *Econometrica*, 70:1299–1340, 2002.
- S. A. Zenios, E. Woodle, and L. Ross. *Primum non nocere*: avoiding harm to vulnerable waitlist candidates in an indirect kidney exchange. *Transplantation*, 72:648–654, 2001.

7 Appendices

7.1 Appendix A: Incomplete Information about Other Patients' Preferences

The results in Section 4 show the impossibility of obtaining the true values of the reservation values of the patient in fully private information environments. In this Appendix, we assume that patients do not know other patients' reservation values, but they do know the probability distribution from which they are drawn. Since we deal with probabilities we need to consider not only their preferences over kidneys (assignments) but also their preferences over lotteries on assignments. Hence, in this section, we assume that patients are expected utility maximizers.

A **expected utility function** is a function $u : \Omega \rightarrow \mathbb{R}$. For each $i \in N$, $\mathbf{P} \in \mathbb{M}_{n \times n}$, and each $r_i \in \mathcal{R}_i$, u is consistent with patient i preferences if for each $\omega, \omega' \in \Omega$, $v_i(\omega) \geq v_i(\omega')$, and for each probability $p \in [0, 1]$ and lottery $[p\omega, (1-p)\omega']$ which yields kidney ω with probability p and kidney ω' with probability $(1-p)$, the utility of the lottery is given by its expected utility $pu(\omega) + (1-p)u(\omega')$. Of course, a patient with expected utility u we assume that prefers a lottery $[p\omega, (1-p)\omega']$ to any other alternative α (which may or may not be a lottery itself) if and only if $pu(\omega) + (1-p)u(\omega') > u(\alpha)$. Clearly, v_i is consistent with patient i 's preferences, but there are many other consistent utility functions, because an expected utility function reflects not only the simple order of her preferences over kidneys, but also a measure of their intensity.

Let F denote a probability distribution over n -tuples of utility functions. We call such a probability distribution a *information structure*.

Given a preference profile \mathbf{P} and a information structure F , a rule φ satisfies **incentive compatibility** if for each $\mathbf{r} \in \mathcal{R}$ truth-telling for all the patients forms a Bayesian Nash equilibrium of the Bayesian game defined by $(\mathbf{P}, \mathbf{r}, \varphi, F)$. That is, for each patient i , each $\mathbf{r} \in \mathcal{R}$, each $r'_i \in \mathcal{R}_i$, and each u_i consistent with patient i preferences

$$\mathbb{E}[u_i(\varphi(\mathbf{r}) \mid F)] \geq \mathbb{E}[u_i(\varphi(r'_i, \mathbf{r}_{-i}) \mid F)].$$

Our next result is an impossibility theorem that extends the negative conclusions of Theorem 1. It says that in the incomplete information case, in equilibrium any revelation mechanism cannot satisfy the joint requirements of *k-feasibility*, *constrained efficiency*, and *individual rationality*.

Theorem 2. *Let $2 \geq k < n$. There exist preference profiles and states of information $[\mathbf{P}, F]$ for which no rule that satisfies individual rationality, k -efficiency, and incentive compatibility.*

Proof. Again, we consider the cases $k = 2$ and $k > 2$ separately.

Let $k = 2$ and consider the deterministic revelation mechanism φ that satisfies *individual rationality*, *k-efficiency*, and *incentive compatibility*. Consider three patients, 1, 2, and 3 with preference profile

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.75 & 0.99 \\ 0.99 & 0.5 & 0.75 \\ 0.75 & 0.99 & 0.5 \end{pmatrix},$$

and the $v_i(\omega_j) = 0$ for each $i \in \{1, 2, 3\}$, $j \notin \{1, 2, 3\}$. Let the type distribution F be such that $\text{Prob}(r_1 = 0.7) = p$, $\text{Prob}(r_1 = 0.8) = (1 - p)$, $\text{Prob}(r_2 = 0.7) = q$, $\text{Prob}(r_2 = 0.8) = (1 - q)$ for some $p, q \in (0, 1)$, and $\text{Prob}(r_3 = 0.7) = 1$. Assume without loss of generality that $\varphi(0.7, 0.7, 0.7) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$. By *individual rationality* and *k-efficiency*, $\varphi(0.8, 0.8, 0.7) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ and we have four possibilities:

- (i) $\varphi(0.8, 0.7, 0.7) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$ and $\varphi(0.7, 0.8, 0.7) = [(1, \omega_3), (2, \omega_0), (3, \omega_1)]$.
By reporting $r_2 = 0.7$, patient 2 receives kidney ω_1 . By reporting $r_2 = 0.8$, she receives ω_3 with probability q and the ω_0 with probability $(1 - q)$. If $u_2(\omega_0) = 0$ and $u_2(\omega_3) > \frac{1}{1-p}u_2(\omega_1)$, then patient 2 prefers to report $r_2 = 0.8$ even though her true reservation value is $r_2 = 0.7$, which contradicts *incentive compatibility*.
- (ii) $\varphi(0.8, 0.7, 0.7) = [(1, \omega_2), (2, \omega_1), (3, \omega_0)]$ and $\varphi(0.7, 0.8, 0.7) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$.
By reporting $r_2 = 0.7$, patient 2 receives kidney ω_1 . By reporting $r_2 = 0.8$, she receives ω_3 . Then, patient 2 always prefers to report $r_2 = 0.8$ independently of her true reservation value, which contradicts *incentive compatibility*.
- (iii) $\varphi(0.8, 0.7, 0.7) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ and $\varphi(0.7, 0.8, 0.7) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$.
By reporting $r_1 = 0.7$, patient 1 receives kidney ω_2 with probability q and ω_0 with probability $(1 - q)$ By reporting $r_2 = 0.8$, she receives ω_0 . Hence, patient 1, always has an incentive to report $r_1 = 0.7$, which contradicts *incentive compatibility*.
- (iv) $\varphi(0.8, 0.7, 0.7) = [(1, \omega_0), (2, \omega_3), (3, \omega_2)]$ and $\varphi(0.7, 0.8, 0.7) = [(1, \omega_3), (2, \omega_0), (3, \omega_1)]$.
By reporting $r_1 = 0.7$, patient 1 receives kidney ω_2 with probability q and ω_3 with

probability $(1 - q)$. By reporting $r_2 = 0.8$, she receives ω_0 . If $u_1(\omega_3)$ is close enough to $u_1(\omega_0)$ and $u_1(\omega_2)$ is large enough, then patient 1 has an incentive to report $r_1 = 0.7$ even though her true reservation is $r_1 = 0.8$, which leads to the desired contradiction.

Next, we analyze the general case. Let $k > 2$. Assume to the contrary there is a rule φ that satisfies *individual rationality*, *k-efficiency*, and *incentive compatibility*. Let \mathbf{P} be such that for every $i = 1 \dots, k$:

$$v_i(\omega_{i+1}) > v_i(\omega_{i+2}) > v_i(\omega_i) > v_i(\omega) = 0, \quad \forall \omega \in \Omega \setminus \{\omega_0, \omega_i, \omega_{i+1}, \omega_{i+2}\}. \quad (\text{modulo } k + 1),$$

and for patient $k + 1$, $v_{k+1}(\omega_2) > v_{k+1}(\omega_1) > v_{k+1}(\omega) = 0$ for each $\omega \in \Omega \setminus \{\omega_1, \omega_2\}$. By *individual rationality*, without any loss of generality, we can focus on the assignment restricted to the patients $1, \dots, k + 1$.

Consider type information distribution F be such that for each $i \in N \setminus \{k - 1, k + 1\}$, $\text{Prob}(v_i(\omega_{i+2} < r_i < \omega_{i+1}) = 1$, $\text{Prob}(v_{k-1}(\omega_{k-1} < r_{k-1} \geq \omega_k) = p$, $\text{Prob}(v_{k-1}(\omega_{k+1} < r_{k-1} < \omega_k) = (1 - p)$, $\text{Prob}(v_{k-1}(\omega_{k+1} < r_{k+1} < \omega_2) = q$, and $\text{Prob}(v_{k+1}(\omega_2 < r_{k+1} < \omega_1) = (1 - q)$.

Consider patient $k - 1$. By reporting $r_{k-1} \in (v_{k-1}(\omega_{k-1}, \omega_{k+1}))$, for $\lambda \in \{0, 1\}$, she obtains with probability q , $(\lambda\omega_k, (1 - \lambda)\omega_{k+1})$, and with probability $(1 - q)$ the kidney ω_{k+1} . On the other hand, by reporting $r_{k-1} \in (v_{k-1}(\omega_{k+1}, \omega_k))$, she receives with probability q the kidney ω_k and with probability $(1 - q)$ the null kidney ω_0 . Normalizing $u_{k-1}(\omega_0) = 0$, if $\lambda = 0$, for all $q \in (0, 1)$ there is u_{k-1} such that $u_{k-1}(\omega_k) > \frac{1}{q}u_{k-1}(\omega_{k+1})$, which contradicts *incentive compatibility*. (Note that patient $k - 1$ with such utility function would rather report a higher reservation value than her true one.) Thus $\lambda = 1$. Repeating the same argument with patient $k + 1$ yields the desired contradiction. \square

7.2 Appendix B: Omitted Proofs

Proof of Proposition 1. It is obvious that if for each patient truth-telling is the unique protective strategy then φ is DIPE. Hence, we focus on the converse result. Let φ be a DIPE rule. We proceed through a series of steps.

(a) If r_i and r'_i are not equivalent, then $D(r_i) \cap D(r'_i) = \emptyset$.

Let $r_i^* \in \mathcal{R}_i$. Assume to the contrary that $r_i^* \in D(r_i) \cap D(r'_i)$. Since r_i and r'_i are not equivalent, there is $\mathbf{r}_{-i} \in \mathcal{R}_{-i}$ such that $\varphi_i(r_i, \mathbf{r}_{-i}) \neq \varphi_i(r'_i, \mathbf{r}_{-i})$. For each $j \in N \setminus \{i\}$, let

$\hat{r}_j \in D(r_j)$. Since φ is DIPE, $\varphi_i(r_i^*, \hat{\mathbf{r}}_{-i}) = \varphi_i(r_i, \mathbf{r}_{-i})$ and $\varphi_i(r_i^*, \hat{\mathbf{r}}_{-i}) = \varphi_i(r_i', \mathbf{r}_{-i})$, which is not possible.

(b) *If r_i and r_i' are such that $\Omega_i^+(r_i) = \Omega_i^+(r_i')$, then r_i and r_i' are equivalent.*

Note first that both r_i and r_i' together with \mathbf{P} define the same preference relation \succsim_i . Let $\bar{r}_i \in D(r_i)$ and $\bar{r}_i' \in D(r_i')$. Let $\hat{\mathbf{r}}_{-i} \in \mathcal{R}_{-i}$ be such that for every $\mathbf{r}_{-i} \in \mathcal{R}_{-i}$, $\varphi_i(\bar{r}_i, \mathbf{r}_{-i}) \succsim_i \varphi_i(\bar{r}_i, \hat{\mathbf{r}}_{-i})$. Because $\bar{r}_i \in D(r_i)$, $\varphi_i(\bar{r}_i, \hat{\mathbf{r}}_{-i}) = \varphi_i(\bar{r}_i', \hat{\mathbf{r}}_{-i})$. Using the same argument with \bar{r}_i' , we conclude that the worst outcomes that i receives reporting \bar{r}_i and \bar{r}_i' are the same. Moreover, they receive such worst outcomes at the same reservation values profiles of the remaining patients. Repeating as many times as necessary the same argument with the following utility levels, we prove that \bar{r}_i and \bar{r}_i' are equivalent. Therefore, by step (a), r_i and r_i' are equivalent.

Remark 1. *Since there is a finite number of patients and kidneys, steps (a) and (b) imply that patients divide their strategy sets in a finite number of equivalence classes.*

(c) $D(r_i) = \{r_i' \mid r_i \text{ and } r_i' \text{ are equivalent}\}$.

By steps (a) and (b), it suffices to prove that $r_i \in D(r_i)$. Assume to the contrary that there are $i \in N$ and $r_i^0 \in \mathcal{R}_i$ such that $r_i^0 \notin D(r_i)$. Since $D(r_i^0) \neq \{\emptyset\}$, then there is $r_i^1 \in D(r_i^0)$. As r_i^1 protectively dominates r_i^0 , there is $r_{-i}^0 \in \mathcal{R}_{-i}$ such that $\varphi_i(r_i^1, r_{-i}^0) \succ_i \varphi_i(r_i^0, r_{-i}^0)$, and for each $r_{-i}^1 \in \mathcal{R}_{-i}$ such that $\varphi_i(r_i^1, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^1)$, $\varphi_i(r_i^0, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^0)$.

Since r_i^0 and r_i^1 are not equivalent, by step (a), $r_i^1 \notin D(r_i^1)$. Thus there exist $r_i^2 \in D(r_i^1)$ and r_i^2 is not equivalent to r_i^1 . By repeated application of this argument we can construct a sequence $r_i^0, r_i^1, r_i^2, \dots$ of i 's reservation values such that for all $t \in \mathbb{N}$, $r_i^t \notin D(r_i^t)$ and $r_i^t \in D(r_i^{t-1})$.

By step (b) there is a finite number of equivalence classes $D(r_i)$, there exist integers T and S such that r_i^T and r_i^{T+S} are equivalent. Then, by step (a), r_i^{T-h} and r_i^{T+S-h} are also equivalent for $h = 1, \dots, T$. In particular, r_i^0 and r_i^T are equivalent.

Let now define the sequence $\{r_{-i}^1, \dots, r_{-i}^S\}$ of elements of \mathcal{R}_{-i} such that for each $t = 1, \dots, S$, and each $j \in N \setminus \{i\}$, $r_{-i}^t \in D(r_{-i}^{t-1})$.

Since φ is DIPE, $\varphi_i(r_i^1, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^0)$, and because $r_i^1 \in D(r_i^0)$, $\varphi_i(r_i^0, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^0)$. Again, since φ is DIPE, $\varphi_i(r_i^0, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^0)$ implies $\varphi_i(r_i^1, r_{-i}^1) = \varphi_i(r_i^0, r_{-i}^1)$.

and $\varphi_i(r_i^0, r_{-i}^2) = \varphi_i(r_i^0, r_{-i}^0)$. Repeating the argument as many times as necessary, we obtain that $\varphi_i(r_i^0, r_{-i}^{S-1}) = \varphi_i(r_i^0, r_{-i}^0)$.

On the other hand, by repeated application of the fact that φ is DIPE, we obtain that for each $t = 1, \dots, S$; $\varphi_i(r_i^1, r_{-i}^0) = \varphi_i(r_i^t, r_{-i}^{t-1})$. In particular, $\varphi_i(r_i^S, r_{-i}^{S-1}) = \varphi_i(r_i^1, r_{-i}^0) \succ_i \varphi_i(r_i^0, r_{-i}^0)$, which contradicts that r_i^0 and r_i^S are equivalent. \square