

Matching, Allocation, and Exchange of Discrete Resources¹

Tayfun Sönmez²

M. Utku Ünver³

Boston College

Department of Economics, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA.

August 4, 2009

for the *Handbook of Social Economics* by Elsevier

edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson

¹We dedicate this survey to the memory of David Gale (1921-2008), one of the two founders of matching theory along with Lloyd Shapley. This survey is based on the lecture notes of Sönmez and Ünver for the graduate level “Matching Market Design” courses that they taught respectively at Boston College and the University of Pittsburgh from 2005 to 2008. We would like to thank Fuhito Kojima, Morimitsu Kurino, Alexey Kushnir, Muriel Niederle, and Parag Pathak for their comments.

²E-mail: sonmezt@bc.edu, url: <http://www2.bc.edu/~sonmezt>

³E-mail: unver@bc.edu, url: <http://www2.bc.edu/~unver>

Contents

1	Introduction	1
2	House Allocation and Exchange Models	5
2.1	House Allocation	5
2.2	The Housing Market	7
2.3	House Allocation with Existing Tenants	13
3	Kidney Exchange	19
3.1	Directed Live Donations and Donor Exchanges	20
3.2	The Designs	21
3.3	The Model	22
3.4	Multi-Way Kidney Exchanges with Strict Preferences	23
3.5	Two-Way Kidney Exchanges with 0-1 Preferences	27
3.5.1	Priority Mechanism	30
3.5.2	The Structure of Pareto-Efficient Matchings	30
3.5.3	Egalitarian Mechanism	32
3.6	Multi-Way Kidney Exchanges with 0-1 Preferences	36
3.7	Recent Developments and Related Literature	39
4	School Admissions	41
4.1	College Admissions	41
4.1.1	Differences between College Admissions, Student Placement, and School Choice Problems	44
4.2	Student Placement	44
4.2.1	Simple Case: One Skill Category	46
4.2.2	Current Mechanism in Turkish College Student Placement: Multi-Category Serial Dictatorship	46
4.2.3	Mechanisms via the Associated College Admissions Problem	49
4.2.4	Pareto Efficiency and Elimination of Justified Envy	49

4.2.5	Strategy-Proofness and Elimination of Justified Envy	50
4.2.6	Respecting Improvements	51
4.3	School Choice	52
4.3.1	The Boston School Choice Mechanism	53
4.3.2	Incentives, Pareto Efficiency, and Justified-Envy-Freeness with Strict and Weak Priorities	53
4.3.3	The School Choice TTC Mechanism	56
4.4	Recent Developments and Related Literature	57
5	Axiomatic Mechanisms and Market Design	61
5.1	House Allocation and Hierarchical Exchange	61
5.2	Trading Cycles with Brokers and Owners	64
5.3	Related Literature	68
6	Concluding Remarks	69
6.1	Lottery Mechanisms in Matching	69
6.2	Other Literature	71

Chapter 1

Introduction

Matching theory, a name referring to several loosely related research areas concerning matching, allocation, and exchange of indivisible resources, such as jobs, school seats, houses, etc., lies at the intersection of game theory, social choice theory, and mechanism design. Matching can involve the *allocation* or *exchange* of indivisible objects, such as dormitory rooms, transplant organs, courses, summer houses, etc. Or matching can involve *two-sided matching*, in markets with two sides, such as firms and workers, students and schools, or men and women, that need to be matched with each other. Auctions can be seen as special cases of matching models, in which there is a single seller. Recently, matching theory and its application to market design have emerged as one of the success stories of economic theory and applied mechanism design.

The seminal research paper on the subject was written by Gale and Shapley (1962), who introduced the two-sided matching model and a suitable solution concept called *stability*. They also showed that a stable matching always exists and proved this result through a simple iterative algorithm known as the *deferred acceptance algorithm*. Gale and Shapley were most likely unaware that this short note published in the *American Mathematical Monthly* would spark a new literature in game theory, which is now commonly referred to as *matching theory*.

Shapley and Shubik (1972) and Kelso and Crawford (1982) introduced variants of the two-sided matching model where monetary transfers are also possible between matching sides. However, Gale and Shapley's short note was almost forgotten until 1984, when Roth (1984) showed that the same algorithm was independently discovered by the National Residency Matching Program (NRMP)¹ in the United States (US), and since the 1950s, it had been used in matching medical interns with hospital residency positions (Roth 2008a also attributes the same discovery to David Gale). This discovery marked the start of the convergence of matching theory and game-theoretical field applications. In 1980s, several papers were written on the two-sided matching model and its variants exploring strategic and structural issues regarding stability.² Recently, new links between auctions,

¹See <http://www.nrmp.org>, retrieved on 10/16/2008.

²An excellent survey of these theoretical and practical developments from the 1950s to the 1990s is explored in

two-sided matching, and lattice theory were discovered (for example, see Hatfield and Milgrom 2005 for a summary of these discoveries and new results in a general two-sided matching domain).³

In this survey, we will focus on the other branch of matching theory, *allocation and exchange of indivisible goods*, which was also initiated by Shapley and (indirectly) Gale together with Scarf (Shapley and Scarf 1974).⁴ The basic model, referred to as the *housing market*, consists of agents each of whom owns an object, e.g. a house. They have preferences over all houses including their own. The agents are allowed to exchange the houses in an exchange economy. Shapley and Scarf showed that such a market always has a (*strict*) *core* matching, which is also a competitive equilibrium allocation. They also noted that a simple algorithm suggested by David Gale, now commonly referred to as *Gale's top trading cycles algorithm*, also finds this particular core outcome.

In the two-sided matching model, both sides of the market consist of agents, whereas in a housing market only one side of the market consists of agents. Subsequent research on the housing market showed that both competitive and core allocations are unique when preferences are strict (Roth and Postlewaite 1977). Moreover, when the core concept is used as a direct mechanism, it is strategy-proof (Roth 1982a). Subsequently, Ma (1994) showed that this is the only direct mechanism that is strategy-proof, Pareto-efficient, and individually rational. Although the core as a mechanism is the unique *nice* direct mechanism (unlike in most game-theoretical models including the two-sided matching model), the research on housing market model remained limited until recently with respect to the two sided-matching model. The links between the two models were later discovered and explored by Balinski and Sönmez (1999), Ergin (2002), Abdulkadiroğlu and Sönmez (2003a), Klaus and Ehlers (2006), and Kojima and Manea (2007b), among others.

The allocation model consists of objects and agents, each of whom has preferences over the objects. These objects will be allocated to the agents. Monetary transfers are not available. An exogenous control rights structure regarding the objects can be given in the definition of the problem. For example, each agent can have objects to begin with (as in the kidney exchange problem of Roth, Sönmez, and Ünver 2004, or the housing market), or some agents can have objects while others have none (as in the house allocation problem with existing tenants of Abdulkadiroğlu and Sönmez 1999). There can also be more complicated exogenous control rights structures, as in the school choice problem, where each school prioritizes the students (as defined by Abdulkadiroğlu and Sönmez 2003a). In the simplest of these models, there are no initial property rights, and objects

Roth and Sotomayor (1990). Also see Gusfield and Irving (1989) on the complementary work in operations research and computer science on algorithms regarding two-sided matching theory.

³For surveys on market design of the US Federal Communications Commission (FCC) auctions (see http://wireless.fcc.gov/auctions/default.htm?job=auctions_home, retrieved on 10/16/2008), electricity markets (e.g., for California market see <http://www.caiso.com>, retrieved on 10/16/2008), and other aspects of matching markets and their links to game theory and more specifically to auction and matching theory see Milgrom (2000, 2004, 2007), Klemperer (2004), Wilson (2002), and Roth (2002, 2008b), respectively.

⁴Nevertheless, we will also give basic results regarding Gale and Shapley's (1962) model and summarize important market design contributions on the subject in Chapter 4 under the "College Admissions" heading.

are socially endowed (as in the house allocation problem of Hylland and Zeckhauser 1979). Almost all of these models have real-life applications. In all of these applications, there exists a central planner (such as the housing office of a college allocating dorm rooms to students, a central health authority deciding which patients will receive kidneys, or a school board for assigning students to schools) that implements a direct mechanism by collecting preference information from the agents. The central authority uses a well-defined procedure that we will simply refer to as a *mechanism*. In this survey, we will inspect properties of different mechanisms proposed in the literature for these allocation problems. Most of the mechanisms we will introduce will be implemented by intuitive iterative algorithms.

In the models with initial property rights, various fairness and individual rights protection properties should be respected by any plausible mechanism for normative, institutional, or economic reasons. Some examples can be given as follows:

Normatively, one would expect there to be equal chances of assigning an object to agents who have identical rights over objects. In a school choice problem, students are the agents. Students who have the same priority at a school may be given the same chances of admission. Thus, from a fairness point of view, an even lottery can be used to order such students for tie-breaking purposes. On the other hand, if there exists a student who prefers a school to her assigned school and this more preferred school has admitted a student who has lower priority than her, then she has justified envy toward this student. Besides following certain normative criteria for institutional and legal reasons, adopted school choice mechanisms are expected to eliminate justified envy. For example, if there is justified envy regarding a student, her family can potentially take legal action against the school district.

In a kidney exchange problem, if a kidney transplant patient is not assigned a kidney as good as her live paired-donor's, she will not participate in the exchange in the first place. Under incomplete information, such possibilities may cause unnecessary efficiency loss. Thus, individual rationality is important for the kidney exchange problem.

Moreover, if possible, we would like the mechanisms to be incentive compatible: decision makers such as students, patients, and doctors should not be able to manipulate these systems by misreporting their preferences. This will be important not only in achieving allocations that satisfy the properties of the mechanisms under true preferences, but also for fairness reasons. For example, not all students are sophisticated enough to manipulate a mechanism successfully (see Pathak and Sönmez 2008 and also Vickrey 1961 for similar arguments in auction design). Moreover, one can expect that implementing a strategy-proof mechanism will minimize the informational burden of the agents. They will only need to form their (expected) preference ordering correctly and will not need to guess the preferences of other agents before submitting their preferences. Hence, in this survey, besides introducing several plausible mechanisms, we will explore what properties make these mechanisms plausible.

The survey will consist of four main chapters: In Chapter 2, we will introduce the *house allocation problem* and the *housing market* and explore mechanisms in this domain. As the market design application of these models, we will introduce one additional model and mechanism, inspired by dormitory room allocation at colleges. In Chapter 3, we will introduce the *kidney exchange* models under various institutional and modeling restrictions. We will draw parallels between some of these models and the house allocation and exchange models. We will also inspect real-life mechanisms designed by economists for these problems. In Chapter 4, we will explore the *school admissions problem*, and plausible mechanisms under different institutional restrictions. We will explore school admissions under three different models, the *college admissions problem*, the *student placement problem*, and the *school choice problem*. In Chapter 5, we will introduce general classes of mechanisms that can be used to characterize desirable house allocation mechanisms.

Chapter 2

House Allocation and Exchange Models

2.1 House Allocation

The simplest of the indivisible goods allocation models is known as the *house allocation problem* and is due to Hylland and Zeckhauser (1979). In this problem, there is a group of agents and houses (representing indivisible objects). Each agent shall be allocated a house by a central planner using their preferences over the houses. All houses are social endowments. Formally, a triple (A, H, \succ) is a **house allocation problem** if

- $A = \{a_1, a_2, \dots, a_n\}$ is a set of **agents**,
- $H = \{h_1, h_2, \dots, h_n\}$ is a set of **houses**,
- $\succ = (\succ_a)_{a \in A}$ is a strict preference profile such that for each agent $a \in A$, \succ_a is a **strict preference relation** over houses.¹ The induced weak preference relation of agent a is denoted by \succeq_a and for any $h, g \in H$, $h \succeq_a g \Leftrightarrow h \succ_a g$ or $h = g$ (i.e., a binary relation, which is a linear order).²

There are various applications of the house allocation problem, such as organ allocation for transplant patients waiting for deceased donor organs, dormitory room allocation at universities, and parking space and office allocation at workplaces.

¹For any subset of agents B , we will use \succ_{-B} to denote $(\succ_a)_{a \in A \setminus B}$ and \succ_B to denote $(\succ_a)_{a \in B}$.

²A binary relation β defined on a set X is a linear order if

- it is complete, i.e. for all $x, y \in X$, either $x\beta y$ or $y\beta x$,
- it is reflexive, i.e. for all $x \in X$, $x\beta x$,
- it is transitive, i.e. for all $x, y, z \in X$, $x\beta y$ and $y\beta z$ imply $x\beta z$, and
- it is anti-symmetric, i.e. for all $x, y \in X$, $x\beta y$ and $y\beta x$ imply $x = y$.

Throughout this subsection, we will fix A and H . A problem is denoted only through the preference profile \succ .

The outcome of a house allocation problem is a **matching**, which is a one-to-one and onto function $\mu : A \rightarrow H$ such that house $\mu(a)$ is the assigned house of agent a under matching μ . Let \mathcal{M} be the set of matchings.

We will inspect several desirable properties of matchings. A matching μ is **Pareto-efficient** if there is no other matching ν such that $\nu(a) \succ_a \mu(a)$ for all $a \in A$ and $\nu(a) \succ_a \mu(a)$ for some agent $a \in A$.

A (**deterministic direct**) **mechanism** is a procedure that assigns a matching for each house allocation problem. For any problem \succ , let $\phi[\succ] \in \mathcal{M}$ refer to the matching outcome of ϕ for problem \succ .

Next, we discuss several desirable properties of mechanisms. A mechanism ϕ is **strategy-proof** if for any problem \succ , any agent $a \in A$ and any preference relation \succ_a^*

$$\phi[\succ_a, \succ_{-a}](a) \succ_a \phi[\succ_a^*, \succ_{-a}](a).$$

That is, in a game induced by the direct mechanism ϕ , when agents reveal their preferences and the central planner implements a matching using ϕ according to the revealed preference profile, it is a weakly dominant strategy for each agent to truthfully report her preferences.

A mechanism is **Pareto-efficient** if it assigns a Pareto-efficient matching for each problem.

Next, we introduce a fundamental class of mechanisms, commonly referred to as *serial dictatorships* (or *priority mechanisms*) (for example, see Satterthwaite and Sonnenschein 1981 and Svensson 1994). A serial dictatorship is defined through a priority ordering of agents. A **priority ordering** is a one-to-one and onto function $f : \{1, 2, \dots, n\} \rightarrow A$. That is, for any $k \in \{1, \dots, n\}$, $f(k) \in A$ is the agent with the k^{th} highest priority agent under f . Let \mathcal{F} be the set of orderings. Each priority ordering induces a direct mechanism. We refer to the direct mechanism π^f as the **serial dictatorship induced by priority ordering** $f \in \mathcal{F}$, and its matching outcome $\pi^f[\succ]$ is found iteratively as follows:

Algorithm 1 *The serial dictatorship induced by f :*

Step 1: *The highest priority agent $f(1)$ is assigned her top choice house under $\succ_{f(1)}$*

\vdots

Step k : *The k^{th} highest priority agent $f(k)$ is assigned her top choice house under $\succ_{f(k)}$ among the remaining houses.*

We can summarize the desirable properties of serial dictatorships with the following theorem:

Theorem 1 *A serial dictatorship is strategy-proof and Pareto-efficient.*

Moreover, Abdulkadiroğlu and Sönmez (1998) show that for any Pareto-efficient matching of a given problem, there exists a serial dictatorship that achieves this matching.

Serial dictatorships can be easily implemented in real-life applications; therefore, they are very appealing. If it is not possible to distinguish between agents to determine the control rights of houses and order them as *serial dictators*, then a random ordering can be chosen and the induced serial dictatorship can be implemented to sustain fairness.

2.2 The Housing Market

The second model we consider is a variant of the house allocation problem and is known as a *housing market* (Shapley and Scarf, 1974). The only difference between this problem and the house allocation problem is that now each agent *owns* a house, i.e. has the initial property right of a house. Hence, a housing market is an exchange market (with indivisible objects) where agents have the option to trade their house in order to get a better one. On the other hand, a house allocation problem has no predefined control rights structure. The houses are social endowments, and the central planner allocates them.

Formally, a **housing market** is a list $(A, (a, h_a)_{a \in A}, \succ)$ such that

- $A = \{1, \dots, n\}$ is a **set of agents** and $\{h_1, \dots, h_n\}$ is a **set of houses** such that each agent a **occupies** house h_a satisfying $h_b \neq h_a$ for any $b \neq a$, and
- $\succ = (\succ_a)_{a \in A}$ is a strict preference profile such that for each agent $a \in A$, \succ_a is a **strict preference relation over houses**.

Throughout this subsection, we fix the set of agents A . We also fix the endowments of agents as above and denote the set as H . Thus, each market is denoted by a preference profile \succ .

There are several real-life applications of housing markets. We will focus on an important one in the next section. In this application, agents are end-stage kidney disease patients, are endowed with a live donor who would like to donate a kidney to them, and have the option to trade their donors to receive a better quality kidney.

Next, we define solution concepts for housing markets. The definitions of a matching, a mechanism, and their properties introduced for the housing allocation problem also apply to the housing market.

We also introduce a new concept about the additional structure of the housing market regarding initial property rights. A matching μ is **individually rational** if for each agent $a \in A$, $\mu(a) \succeq_a h_a$, that is, each agent is assigned a house at least as good as her own occupied house. A mechanism is **individually rational** if it always selects an individually rational matching for each market.

Although we focused on allocation through direct mechanisms, a decentralized solution may naturally exist for a housing market, which is an exchange economy with indivisible objects. A

competitive equilibrium may be achieved through decentralized trading. We define a price vector as a positive real vector assigning a price for each house, i.e. $p = (p_h)_{h \in H} \in \mathbb{R}_{++}^n$ such that p_h is the price of house h . A matching - price vector pair $(\mu, p) \in \mathcal{M} \times \mathbb{R}_+^n$ finds a **competitive equilibrium** if for each agent $a \in A$,

- $p_{\mu(a)} \leq p_{h_a}$ (budget constraint), and
- $\mu(a) \succsim_a h$ for all $h \in H$ such that $p_h \leq p_{h_a}$ (utility maximization).

Under a competitive equilibrium, each agent is assigned the best house that she can afford.

Another important concept for exchange economies is the *core*. With divisibilities, it is well known that any competitive equilibrium allocation is also in the core.

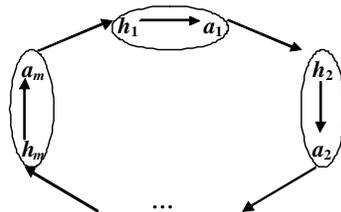
We formulate the core for a housing market as follows: A matching μ is in the **core** if there exists no coalition of agents $B \subseteq A$ such that for some matching $\nu \in \mathcal{M}$ such that for all $a \in B$, $\nu(a) = h_b$ for some $b \in B$, we have $\nu(a) \succsim_a \mu(a)$ for all $a \in B$ and $\nu(a) \succ_a \mu(a)$ for some $a \in B$. That is, the core is the collection of matchings such that no coalition could improve their assigned houses even if they traded their initial occupied houses only among each other.

Although competitive equilibrium and the core are very intuitive solution concepts with nice economic properties, it is not immediately clear that they exist and are related to each other for the housing market. Shapley and Scarf also proved that the core is non-empty and there exists a core matching that can be sustained under a competitive equilibrium.

Theorem 2 *The core of a housing market is non-empty and there exists a core matching that can be sustained as part of a competitive equilibrium.*

As an alternative proof to their initial proof, they introduced an iterative algorithm that is a core and competitive equilibrium matching. They attribute this algorithm to David Gale. This algorithm is a clearing algorithm that forms a directed graph in each iteration and assigns houses to a subset of agents. In order to define the algorithm, we define the following concept:

Consider a directed graph in which agents and houses are the vertices and edges are formed by each agent pointing to one house and each house pointing to one agent. We define a special subgraph of this graph. A **cycle** is a list of houses and agents $(h_1, a_1, h_2, a_2, \dots, h_m, a_m)$ such that each agent a_k points to house h_{k+1} for $k \in \{1, \dots, m-1\}$, a_m points to h_1 , and each house h_k points to agent a_k for $k \in \{1, \dots, m\}$. The figure below depicts such a cycle.



An interesting fact about any directed graph that is formed as explained above is the following:

Lemma 1 *Each directed graph formed by each agent pointing to a house and each house pointing to an agent has a cycle, and no two cycles intersect.*

This lemma will enable us to define the following algorithm properly:

Algorithm 2 Gale's top trading cycles (TTC) algorithm:

Step 1: *Let each agent point to her top choice house and each house point to its owner. In this graph there is necessarily a cycle and no two cycles intersect (by Lemma 1). Remove all cycles from the problem by assigning each agent the house that she is pointing to.*

⋮

Step k: *Let each remaining agent point to her top choice among the remaining houses and each remaining house point to its owner (note that houses leave with their owners and owners leave with their houses, so a house remaining in the problem implies that the owner is still in the problem and vice versa). There is necessarily a cycle and no two cycles intersect. Remove all cycles from the problem by assigning each agent the house that she is pointing to.*

The algorithm terminates when no agents and houses remain. The assignments formed during the execution of the algorithm is the matching outcome.

Shapley and Scarf also proved the following theorem:

Theorem 3 *Gale's TTC algorithm achieves a core matching that is also sustainable by a competitive equilibrium.*

A competitive equilibrium price vector supporting this core matching at the equilibrium can be formed as follows: Partition the set of agents as C_1, C_2, \dots, C_r where C_k is the set of agents removed in Step k of Gale's TTC algorithm. Price vector p is such that for any pair of houses h_a, h_b if the owners a and b were removed in the same step, i.e. $a, b \in C_k$ for some Step k , then we set $p_{h_a} = p_{h_b}$, if (without loss of generality) owner a is removed before agent b , i.e. $a \in C_k$ and $b \in C_\ell$ such that $k < \ell$, then we set $p_{h_a} > p_{h_b}$. That is, (1) the prices of the occupied houses whose owners are removed in the same step are set equal to each other and (2) the prices of those whose owners are removed in different steps are set such that the price of a house that leaves earlier is higher than the price of a house that leaves later.

Below, we demonstrate how Gale's TTC algorithm works with an involved example:

Example 1 *The execution of Gale's TTC algorithm*

Let

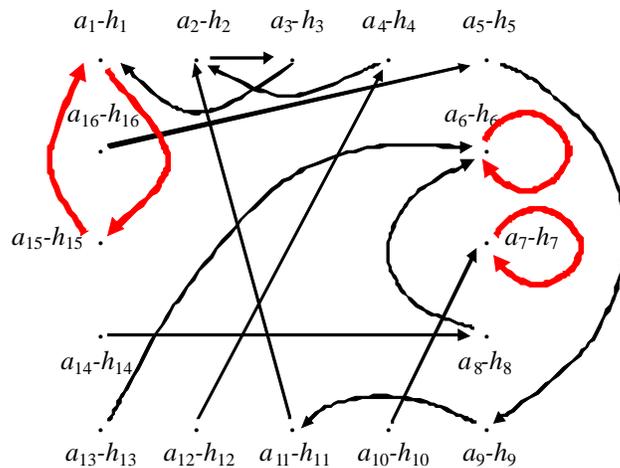
$$A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}\}.$$

Here h_i is the occupied house of agent a_i . Let the preference profile \succ be given as:

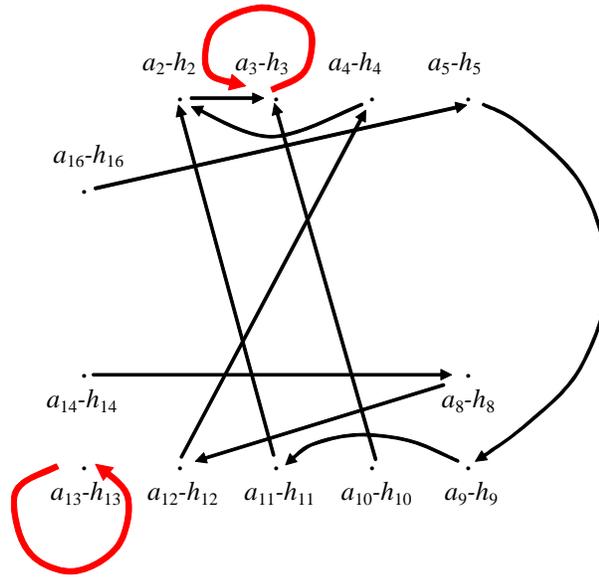
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
h_{15}	h_3	h_1	h_2	h_9	h_6	h_7	h_6	h_{11}	h_7	h_2	h_4	h_6	h_8	h_1	h_5
\vdots	h_4	h_3	\vdots	\vdots	\vdots	\vdots	h_{12}	\vdots	h_3	h_4	h_{14}	h_{13}	\vdots	\vdots	\vdots
	\vdots	\vdots					\vdots		h_{12}	h_{16}	\vdots	\vdots			
								h_{10}	\vdots	\vdots					
								\vdots							

We depict the directed graphs that are formed in each step of the algorithm below. The cycles are shown through bold arrows. Observe that we abbreviated in the graphs below the arrows through which each house points to its owner. When a cycle is removed, each agent in the cycle is assigned the house she is pointing to.

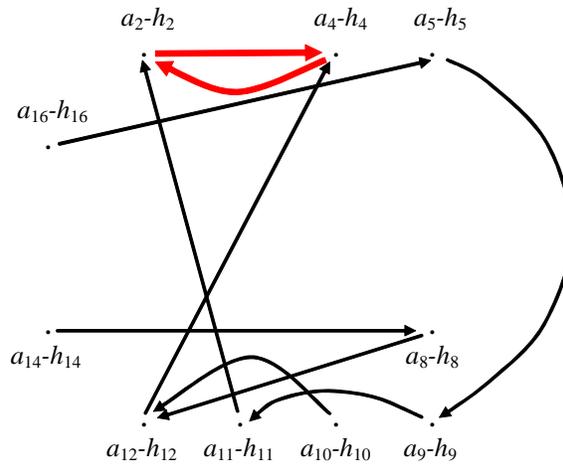
Step 1:



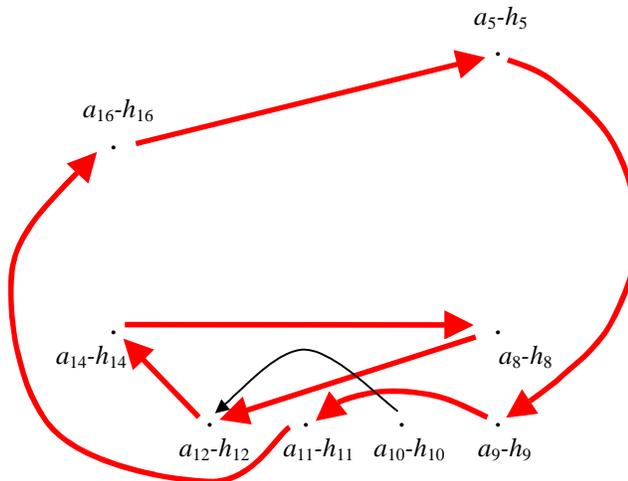
Step 2:



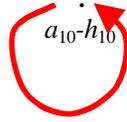
Step 3:



Step 4:



Step 5:



The outcome is:

$$\mu = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ h_{15} & h_4 & h_3 & h_2 & h_9 & h_6 & h_7 & h_{12} & h_{11} & h_{10} & h_{16} & h_{14} & h_{13} & h_8 & h_1 & h_5 \end{pmatrix}$$

After Shapley and Scarf's paper, a series of papers proved that the *core* of a housing market has really nice properties when it is used as a direct mechanism:

Theorem 4 (Roth and Postlewaite 1977) *The core of a housing market has exactly one matching which is also the unique matching that can be sustained at a competitive equilibrium.*

The above result together with Shapley and Scarf's result implies that the core can be used as a mechanism, and Gale's TTC can be used to find it. By definition, the core is Pareto-efficient and individually rational. The following theorem shows that this mechanism also has good incentive properties:

Theorem 5 (Roth 1982a) *The core mechanism is strategy-proof.*

Moreover, there is no other mechanism with these properties:

Theorem 6 (Ma 1994) *The core mechanism is the only mechanism that is individually rational, Pareto-efficient, and strategy-proof for a housing market.*

Thus, from theoretical, practical, and economic points of view, the core is the best solution concept for housing markets. It is the decentralized solution concept and can be implemented in a centralized manner. In economics, there are very few problem domains with such a property. For example, in exchange economies with divisible goods, the competitive equilibrium allocation is a subset of the core, but both the competitive equilibrium and any other core selection are manipulable as a direct mechanism.^{3,4}

In the next subsection, we focus on a market design problem that has the features of both housing markets and house allocation problems.

³Positive results of this section no longer hold in an economy in which one agent can consume multiple houses or multiple types of houses. Even the core may be empty (Konishi, Quint, and Wako 2001). Also see Pápai (2003), Wako (2005), and Klaus (2008) on the subject under different preference assumptions.

On the other hand, if there are no initial property rights, serial dictatorships can still be used for strategy-proof and Pareto-efficient allocation (see Klaus and Miyagawa 2002). Also see Pápai (2003) and Ehlers and Klaus (2003) for other characterizations under different preference assumptions.

⁴See Quinzii (1984) for the existence results of core allocations and competitive equilibria in a generalized model with both discrete and divisible goods. See Bevia, Quinzii, and Silva (1999) for a generalization of this model when an agent can consume multiple indivisible goods.

2.3 House Allocation with Existing Tenants

In some US universities, a probabilistic version of the serial dictatorship is used for allocating dormitory rooms to students. By a (usually equally weighted) lottery, a priority ordering is determined and students reveal a preference ordering over possible dormitory rooms. Then the induced serial dictatorship is used to allocate these rooms to students. This is known as the *housing lottery* at campuses.

Motivated by real-life **on-campus housing** practices, Abdulkadiroğlu and Sönmez (1999) introduced a **house allocation problem with existing tenants**: A set of houses shall be allocated to a set of agents by a centralized clearing house. Some of the agents are **existing tenants**, each of whom already occupies a house, referred to as an **occupied house**, and the rest of the agents are **newcomers**. Each agent has strict preferences over houses. In addition to occupied houses, there are **vacant houses**. Existing tenants are entitled not only to keep their current houses but also to apply for other houses.

Here, existing tenants can be likened to the current college students who occupy on-campus houses (or dormitory rooms, condos, etc.) from the previous year. The newcomers can be likened to the freshman class and any other current student who does not already occupy a house. Vacant houses are the houses vacated by the graduating class and the students who no longer need on-campus housing.

The mechanism known as the **random serial-dictatorship (RSD) with squatting rights** is used in most real-life applications of these problems. Some examples include undergraduate housing at Carnegie Mellon, Duke, Michigan, Northwestern, and Pennsylvania. This mechanism works as follows:

Algorithm 3 *The RSD with squatting rights:*

1. *Each existing tenant decides whether she will enter the housing lottery or keep her current house (or dormitory room). Those who prefer keeping their houses are assigned their houses. All other houses (vacant houses and houses of existing agents who enter the lottery) become available for allocation.*
2. *An ordering of agents in the lottery is randomly chosen from a given distribution of orderings. This distribution may be uniform or it may favor some groups.*
3. *Once the agents are ordered, available houses are allocated using the induced serial dictatorship: The first agent receives her top choice, the next agent receives her top choice among the remaining houses, and so on.*

Since it does not guarantee each existing tenant a house that is as good as what she already has, some existing tenants may choose to keep their houses even though they wish to move, and this may result in a loss of potentially large gains from trade.

In contrast, Abdulkadiroğlu and Sönmez propose a mechanism that has the features of both the core in housing markets and serial dictatorships in house allocation problems. We refer to this mechanism as the "**You request my house - I get your turn**" (**YRMH-IGYT**) mechanism. Let $f \in \mathcal{F}$ be a priority ordering of agents in A . Each f defines a YRMH-IGYT mechanism. The corresponding YRMH-IGYT algorithm clears as follows:

Algorithm 4 *The YRMH-IGYT algorithm induced by f :*

- *Assign the first agent her top choice, the second agent her top choice among the remaining houses, and so on, until someone requests the house of an existing tenant.*
- *If at that point the existing tenant whose house is requested is already assigned another house, then do not disturb the procedure. Otherwise modify the remainder of the ordering by inserting the existing tenant before the requestor at the priority order and proceed with the first step of procedure through this existing tenant.*
- *Similarly, insert any existing tenant who is not already served just before the requestor in the priority order once her house is requested by an agent.*
- *If at any point a cycle forms, it is formed by exclusively existing tenants and each of them requests the house of the tenant who is next in the cycle. (A cycle is an ordered list $(h_{a_1}, a_1, \dots, h_{a_k}, a_k)$ of occupied houses and existing tenants where agent a_1 demands the house of agent a_2 , h_{a_2} , agent a_2 demands the house of agent a_3 , h_{a_3}, \dots , agent a_k demands the house of agent a_1, h_{a_1} .) In such cases, remove all agents in the cycle by assigning them the houses they demand and proceed similarly.*

Below, we present an example showing how the algorithm clears:

Example 2 *The execution of the YRMH-IGYT algorithm*

$A_E = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ is the set of existing tenants,

$A_N = \{a_{10}, a_{11}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}\}$ is the set of newcomers, and

$H_V = \{h_{10}, h_{11}, h_{12}, h_{13}, h_{14}, h_{15}, h_{16}\}$ is the set of vacant houses.

Suppose that each existing tenant a_k occupies h_k for each $k \in \{1, \dots, 9\}$. Let the preference profile \succ

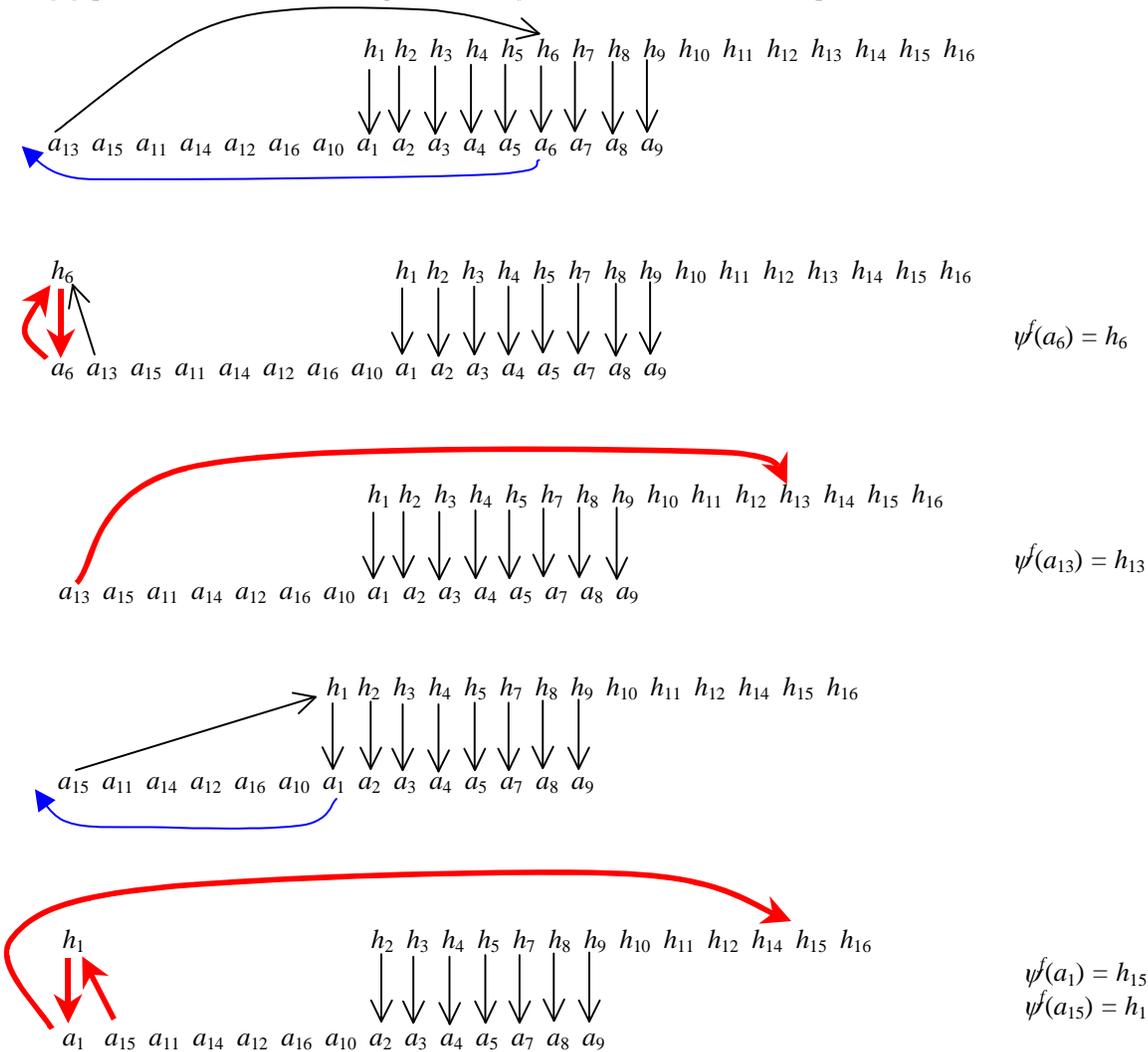
be given as:

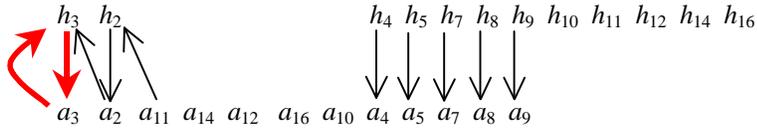
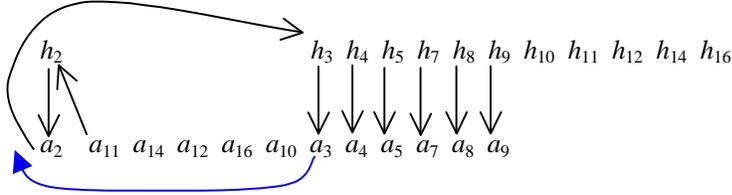
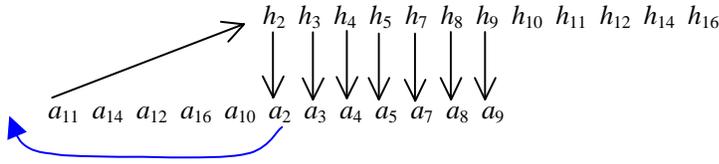
A_E									A_N						
a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
h_{15}	h_3	h_1	h_2	h_9	h_6	h_7	h_6	h_{11}	h_7	h_2	h_4	h_6	h_8	h_1	h_5
\vdots	h_4	h_3	\vdots	\vdots	\vdots	\vdots	h_{12}	\vdots	h_3	h_4	h_{14}	h_{13}	\vdots	\vdots	\vdots
	\vdots	\vdots					\vdots		h_{12}	h_{16}	\vdots	\vdots			
									h_{10}	\vdots					
									\vdots						

Let

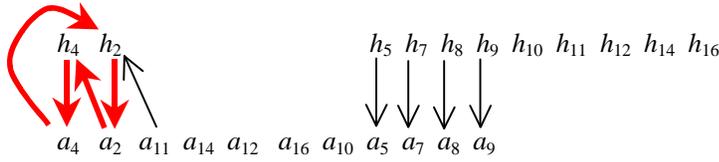
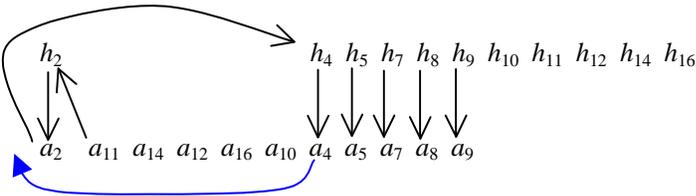
$$f = (a_{13}, a_{15}, a_{11}, a_{14}, a_{12}, a_{16}, a_{10}, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9)$$

be the ordering of the agents. We will denote the outcome of the mechanism by $\psi^f [\succ]$. The following series of figures illustrates the dynamics of the YRMH-IGYT algorithm:



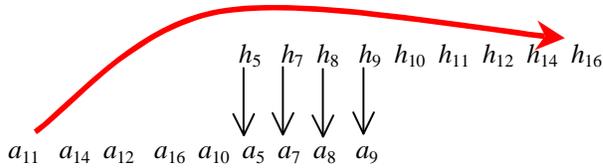


$$\psi^f(a_3) = h_3$$

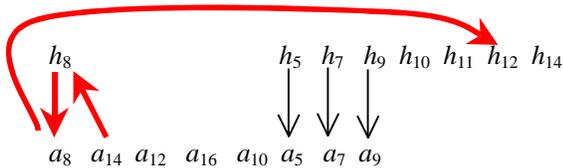
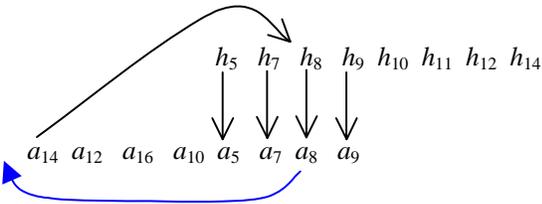


$$\psi^f(a_4) = h_2$$

$$\psi^f(a_2) = h_4$$

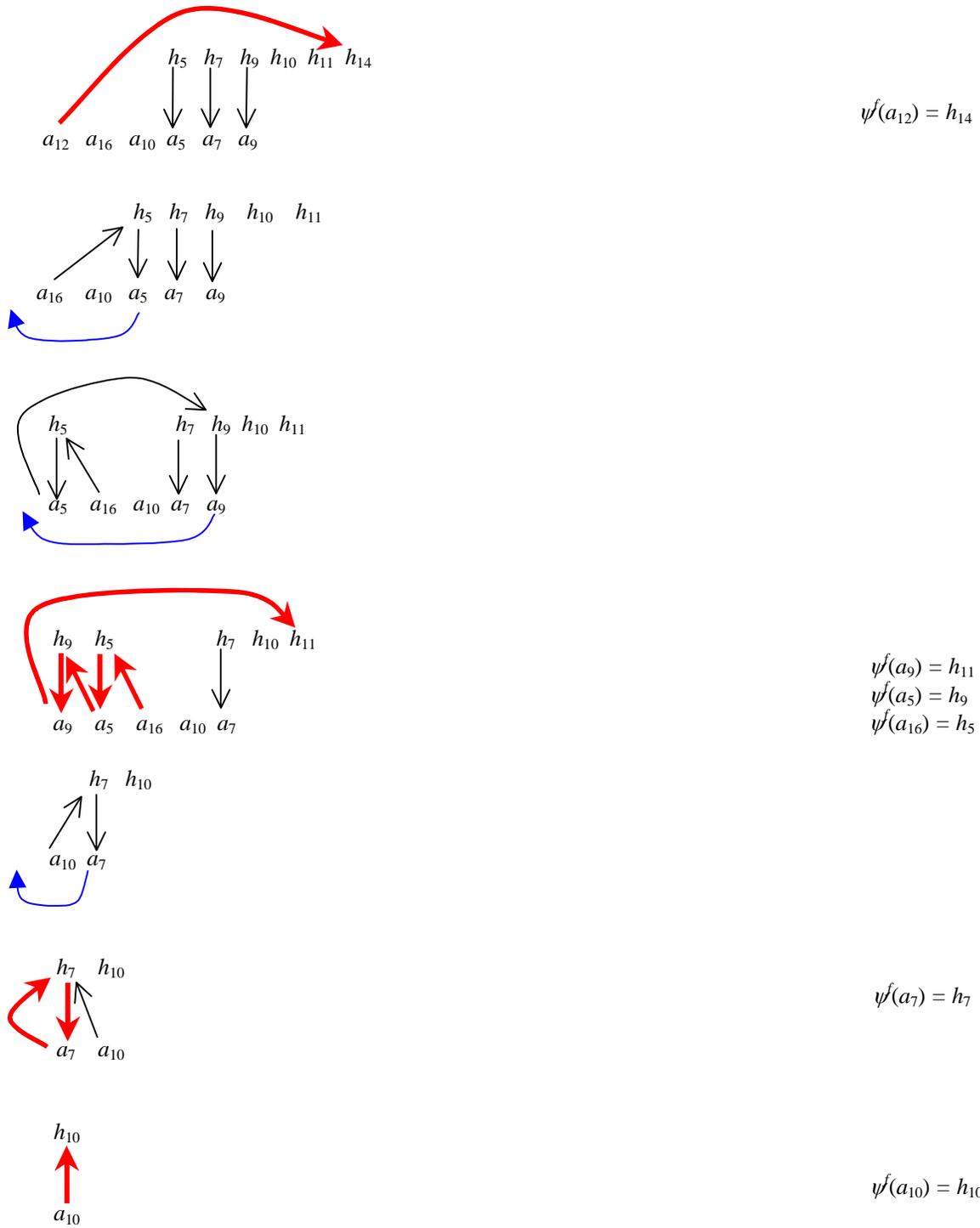


$$\psi^f(a_{11}) = h_{16}$$



$$\psi^f(a_8) = h_{12}$$

$$\psi^f(a_{14}) = h_8$$



The outcome of the algorithm is

$$\mu = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ h_{15} & h_4 & h_3 & h_2 & h_9 & h_6 & h_7 & h_{12} & h_{11} & h_{10} & h_{16} & h_{14} & h_{13} & h_8 & h_1 & h_5 \end{pmatrix}$$

The following theorem shows that this mechanism has desirable properties:

Theorem 7 (*Abdulkadiroğlu and Sönmez 1999*) *Any YRMH-IGYT mechanism is individually rational, Pareto-efficient, and strategy-proof.*

Thus, the YRMH-IGYT mechanisms have nice features. The priority ordering can be determined through a lottery. Chen and Sönmez (2002) showed through a laboratory experiment that this mechanism is practically better than the RSD mechanism with squatting rights. The treatments of the YRMH-IGYT mechanism were more efficient than the RSD with squatting rights mechanism, and manipulation did not occur to a significant degree.

Moreover, it is the unique mechanism that satisfies certain desirable properties:

A mechanism is **coalitionally strategy-proof** if for any problem there is no coalition of agents who can jointly misreport their preferences and all weakly benefit while at least one in the coalition strictly benefits.

A mechanism is **consistent** if, we remove the agents and their assigned houses by the mechanism from the problem together with some unassigned houses, provided that in the remaining problem if an existing tenant remains her occupied house also remains, then rerunning the mechanism for this subproblem does not change the assignments of agents in the subproblem.

A mechanism is **weakly neutral** if, when the vacant houses are relabeled and the mechanism is rerun, then every agent who was assigned a vacant house in the original problem is assigned the relabeled version of the vacant house, and every agent who was assigned an occupied house in the original problem is assigned the same occupied house.

The characterization theorem is as follows:

Theorem 8 (*Sönmez and Ünver 2006*) *A mechanism is coalitionally strategy-proof, individually rational, Pareto-efficient, weakly neutral, and consistent if and only if it is equivalent to a YRMH-IGYT mechanism.*

We conclude by stating some other characterizations regarding restricted domains. In the restricted domains, the mechanisms characterized are equivalent to YRMH-IGYT mechanisms.

Theorem 9 (*Svensson 1999*) *In the house allocation problem, a mechanism is coalitionally strategy-proof, and (weakly) neutral if and only if it is equivalent to a serial dictatorship.*

Theorem 10 (*Ergin 2000*) *In the house allocation problem, a mechanism is Pareto-efficient, (weakly) neutral, and consistent if and only if it is equivalent to a serial dictatorship.*

On the other hand, when there are no newcomers, as in the housing market domain, Theorems 5 by Roth (1982a) and 6 by Ma (1994) imply that the core mechanism is the only desirable mechanism: A mechanism is individually rational, strategy-proof, and Pareto-efficient if and only if it is equivalent to the core mechanism. Observe that these three theorems do not follow from Theorem 8, since smaller sets of axioms are needed in characterization in the more restricted domains.

Chapter 3

Kidney Exchange

In the recent years, the design of kidney exchange (in the medical literature also known as kidney paired donation) mechanisms has been one of the important market design applications of the house allocation and exchange models. A new theory has been developed to accommodate the institutional restrictions imposed by the nature of the problem. This chapter surveys three articles on this design problem (Roth, Sönmez, and Ünver, 2004, 2005a, 2007).

Transplantation is the preferred treatment for the end-stage kidney disease. There are more than 70000 patients waiting for a kidney transplant in the US. In 2005, only 16500 transplants were conducted, 9800 from **deceased donors** and 6570 from **living donors**, while 29160 new patients joined the deceased donor waiting list and 4200 patients died while waiting for a kidney.¹ Buying and selling a body part is illegal in many countries in the world including the US. **Donation** is the only source of kidneys in many countries. There are two sources of donation:

1. **Deceased donors:** In the US and Europe a centralized priority mechanism is used for the allocation of deceased donor kidneys. The patients are ordered in a waiting list, and the first available donor kidney is given to the patient who best satisfies a metric based on the quality of the match, waiting time in the queue, age of the patient, and other medical and fairness criteria.
2. **Living donors:** Generally friends or relatives of a patient (due to the "no buying and selling" constraint) would like to donate one of their kidneys to a designated patient.² Live donations have been an increasing source of donations in the last decade. The design problem determines in the most efficient manner of allocating the kidneys of these donors.

¹According to SRTR/OPTN national data retrieved at <http://www.optn.org> on 2/27/2007.

²Although the number of "non-directed" good Samaritan altruistic donors has steadily been increasing, it is still small relative to the number of "directed" live donors.

3.1 Directed Live Donations and Donor Exchanges

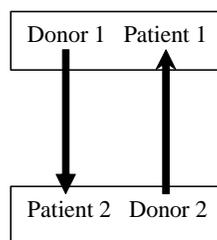
After a patient identifies a willing donor, the transplant is carried out if the donor kidney is compatible with the patient. There are two tests that a donor should pass before she is deemed compatible with the patient:

1. **Blood compatibility test:** There are 4 blood types, "O," "A," "B," and "AB." "O" type kidneys are blood-type compatible with all patients; "A" type kidneys are blood-type compatible with "A" and "AB" type patients; "B" type kidneys are blood-type compatible with "B" and "AB" type patients; and "AB" type kidneys are only blood-type compatible with "AB" type patients.
2. **Tissue compatibility test (or crossmatch test):** 6 HLA (short for human leukocyte antigen) proteins (3 inherited from the mother and 3 inherited from the father) located on patient and donor DNA helices respectively play two roles in determining tissue compatibility. If antibodies exist in the patient blood against the donor HLA, then the donor kidney cannot be transplanted to the patient and it is deemed tissue-type incompatible. It is reported that, on average, there is only 11% chance of tissue-type incompatibility for a random donor and patient (Zenios, Woodle, and Ross, 2001).

Exact HLA match is not required for tissue compatibility; however, there is a debate in the medical literature about how important the closeness of HLA proteins of the patient and donor are for the long-run survival rate of a transplanted kidney.

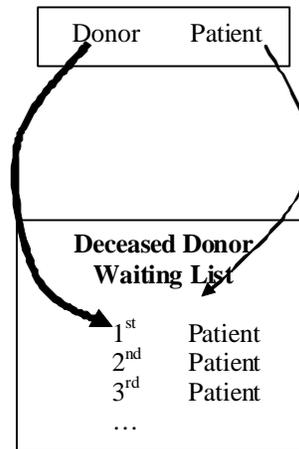
Traditionally, if either test fails, the patient remains on the deceased donor waiting list and the donor goes home unutilized. However, the medical community came up with two ways of utilizing these "unused" donors.

An (**paired**) **exchange** involves two incompatible patient-donor pairs such that the patient in each pair feasibly receives a transplant from the donor in the other pair. This pair of patients exchange donated kidneys. For example, see the figure below:



Of course the number of pairs in a paired exchange can be larger than two.

A **list exchange** involves an exchange between one incompatible patient-donor pair and the deceased donor waiting list. The patient in the pair becomes the first priority person on the deceased donor waiting list in return for the donation of her donor's kidney to someone on the waiting list.



List exchanges can potentially harm O blood-type patients waiting on the deceased donor waiting list. Since the O blood type is the most common blood type, a patient with an incompatible donor is most likely to have O blood herself and a non-O blood-type incompatible donor. Thus, after the list exchange, the blood type of the donor sent to the deceased donor waiting list has generally non-O blood, while the patient placed at the top of the list has O blood. Thus, list exchanges are deemed ethically controversial. Only the New England region in the US adopted list exchange. A list exchange can also involve more pairs than one. Doctors also use non-directed live altruistic donors instead of deceased donors. There is no uniform national policy regarding the handling of non-directed live donors. Many regions conduct exchanges induced by non-directed live donors. Since live donor kidneys are better quality than deceased donor kidneys, such exchanges create better participation incentives for patients and their live paired donors.

3.2 The Designs

Two live donor exchange programs have already been established in the US through collaboration between economists and medical doctors, one in New England and one in Ohio. A national exchange program is being developed.

The surveyed designs illustrate how the exchange system may be organized from the point of view of **efficiency**, providing consistent **incentives** to patients-donors-doctors. Although medical compatibilities are important for matching, the incentives to patients and doctors are also quite important. Patients (doctors) hold *private* information about their (their patients') preferences over several dimensions such as the geographic distance of the match or the number of willing donors they

have. Under some designs, they may not want to reveal this information truthfully, since they (their patients) may benefit from manipulation of information revelation. The initial two designs we will discuss in this survey extract the private information truthfully from patients (doctors) under any circumstance (*strategy-proofness*). We impose several other important *economic* or *normative* criteria on our designs besides incentive compatibility, such as Pareto efficiency and fairness. For fairness, we consider two different approaches: (1) giving *priorities* to patients based on their exogenous characteristics or (2) making every patient as equally well off as the medical constraints permit (also known as *egalitarianism*). Finally, the last design we will discuss refines Pareto efficiency and focuses on aggregate efficiency concerns.

3.3 The Model

A **kidney exchange problem** consists of

- a set of **donor (kidney)-(transplant) patient pairs** $\{(k_1, t_1), \dots, (k_n, t_n)\}$,
- a set of **compatible kidneys** $K_i \subset K = \{k_1, \dots, k_n\}$ for each patient t_i , and
- a **strict preference relation** \succ_i over $K_i \cup \{k_i, w\}$ where w refers to the priority in the waiting list in exchange for kidney k_i .

An outcome of a problem is a **matching** of kidneys/waiting list option to patients such that multiple patients can be matched with the w option (and lotteries over matchings are possible). A kidney exchange **mechanism** is a systematic procedure to select a matching for each kidney exchange problem (and lottery mechanisms are possible).³

A matching is **Pareto-efficient** if there is no other matching that makes everybody weakly better off and at least one patient strictly better off. A mechanism is **Pareto-efficient** if it always chooses Pareto-efficient matchings.

A matching is **individually rational** if each patient is matched with an option that is weakly better than her own paired-donor. A mechanism is **individually rational** if it always selects an individually rational matching.⁴

A mechanism is **strategy-proof** if it is always the best strategy for each patient to

1. reveal her preferences over other available kidneys truthfully, and

³For the time being, we exclude the possibility of non-directed altruistic donors. However, such donors can be incorporated into the problem easily as w option. But there is one difference: an altruistic donor cannot be matched to more than one patient.

⁴We will assume that an incompatible own paired-donor is the opt-out option of a patient.

2. declare the whole set of her donors (in case she has multiple donors) to the system without hiding any (the model treats each patient as having a single donor, but the extension to multiple donors is straightforward).

3.4 Multi-Way Kidney Exchanges with Strict Preferences

The first design and the set of results are due to Roth, Sönmez, and Ünver (2004). Unless otherwise noted, all stated results are from this paper. In this design the underlying assumptions are as follows:

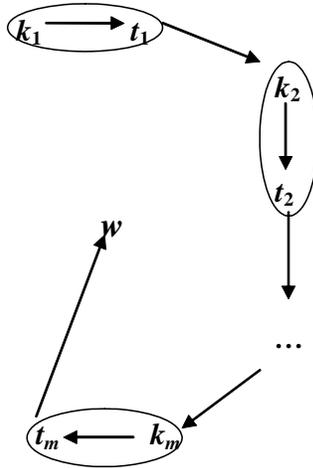
- Any number of patient-donor pairs can participate in an exchange, i.e., exchanges are possibly multi-way.
- Patients have heterogeneous preferences over compatible kidneys; in particular, no two kidneys have the same quality, i.e., the preferences of a patient are strict and they linearly order compatible kidneys, the waiting list option, and her own paired-donor. Opelz (1997) shows in his data set that among compatible donors, the increase in the number of HLA protein mismatches decreases the likelihood of kidney survival. Body size, age of donor etc. also affect kidney survival.
- List exchanges are allowed.

Under these assumptions, this model is very similar to the house allocation model with existing tenants. We will consider a class of mechanisms that clear through an iterative algorithm.

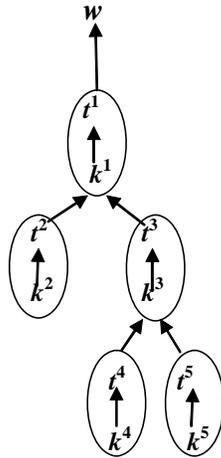
Since the mechanism relies on an algorithm consisting of several rounds, let's first focus on some of the graph-theoretical objects encountered by the algorithm. In each step

- each patient t_i points either toward a kidney in $K_i \cup \{k_i\}$ or toward w , and
- each kidney k_i points to its paired patient t_i .

In such a directed graph, we are interested in two types of subgraphs: One is a **cycle** (as defined in housing markets, where agents refer to patients and houses refer to kidneys). Each cycle is of even size and no two cycles can intersect. The other is a new concept. A **w -chain** is an ordered list of kidneys and patients $(k_1, t_1, k_2, t_2, \dots, k_m, t_m)$ such that k_i points to t_i for each patient, t_i points to k_{i+1} for each $i \neq m$, and t_m points to w .



We refer to the last pair (k_m, t_m) as the **head** and the first pair (k_1, t_1) as the **tail** in such a w -chain. A w -**chain** is also of even size but, unlike in a cycle, a kidney or a patient can be part of several w -chains (see the figure below):



In this figure, there are 5 w -chains initiated by each of the 5 pairs: (k_1, t_1) , (k_2, t_2, k_1, t_1) , (k_3, t_3, k_1, t_1) , $(k_4, t_4, k_3, t_3, k_1, t_1)$, and $(k_5, t_5, k_3, t_3, k_1, t_1)$

One practical possibility is choosing among w -chains with a well-defined chain selection rule. The choice of chain selection rule has implications for efficiency and incentive-compatibility.

We can now state our first result of this section:

Lemma 2 *Consider a graph in which both the patient and the kidney of each pair are distinct nodes as is the waiting list option w . Suppose each patient points either toward a kidney or w , and each kidney points to its paired patient. Then either there exists a cycle or each pair initiates a w -chain. Moreover, when cycles exist, no two cycles intersect.*

Based on this lemma, we can formulate the following exchange procedure that is referred to as the **top trading cycles and chains algorithm (TTCC) algorithm**. Fix a chain selection rule. At a given time and for a given kidney exchange problem, the TTCC mechanism determines the exchanges as follows:

Algorithm 5 *The TTCC algorithm with a chain selection rule:*

1. *Initially all kidneys are available and all agents are active. At each stage of the procedure*
 - *each remaining active patient t_i points to the best remaining unassigned kidney or to the waiting list option w , whichever is more preferred,*
 - *each remaining passive patient continues to point to her assignment, and*
 - *each remaining kidney k_i points to its paired patient t_i .*
2. *By Lemma 2, there is either a cycle, or a w -chain, or both.*
 - (a) *Proceed to Step 3 if there are no cycles. Otherwise locate each cycle and carry out the corresponding exchange. Remove all patients in a cycle together with their assignments.*
 - (b) *Each remaining patient points to her top choice among remaining choices and each kidney points to its paired patient. Proceed to Step 3 if there are no cycles. Otherwise locate all cycles, carry out the corresponding exchanges, and remove them.*
 - (c) *Repeat Step 2b until no cycle exists.*
3. *If there are no pairs left, we are done. Otherwise by Lemma 2, each remaining pair initiates a w -chain. Select only one of the chains with the chain selection rule. The assignment is final for the patients in the selected w -chain. In addition to selecting a w -chain, the chain selection rule also determines*
 - (a) *whether the selected w -chain is removed, or*
 - (b) *the selected w -chain remains in the procedure although each patient in it is henceforth passive.*

If the w -chain is removed, then the tail kidney is assigned to a patient in the deceased donor waiting list. Otherwise, the tail kidney remains available in the problem for the remaining steps.
4. *Each time a w -chain is selected, a new series of cycles may form. Repeat Steps 2 and 3 with the remaining active patients and unassigned kidneys until no patient is left. If there exist some tail kidneys of w -chains remaining at this point, remove all such kidneys and assign them to the patients in the deceased-donor waiting list.*

Below we list a number of plausible chain selection rules:

- a. Choose minimal w -chains and remove them.
- b. Choose the longest w -chain and remove it.
- c. Choose the longest w -chain and keep it.
- d. Prioritize patient-donor pairs in a single list. Choose the w -chain starting with the highest priority pair and remove it.
- e. Prioritize patient-donor pairs in a single list. Choose the w -chain starting with the highest priority pair and keep it.

Each w -chain selection rule induces a TTCC mechanism. The removal and non-removal of w -chains has implications for efficiency.

Theorem 11 (*Roth, Sönmez, and Ünver 2004*) *Consider a chain selection rule where any w -chain selected at a non-terminal step remains in the procedure and thus the kidney at its tail remains available for the next step. The TTCC mechanism induced by any such chain selection rule is Pareto-efficient.*

In the absence of list exchanges, the kidney exchange problem is a direct application of housing markets, and therefore, Theorem 5 implies that TTCC is strategy-proof. What happens when list exchanges are allowed?

Theorem 12 (*Roth, Sönmez, and Ünver 2004*) *The TTCC mechanism induced by chain selection rules (a), (d), or (e) is strategy-proof. On the other hand, the TTCC mechanism induced by chain selection rules (b) or (c) is not strategy-proof.*

We mentioned that the current model is very similar to the house allocation model with existing tenants. There is also a close relationship between the TTCC algorithm and YRMH-IGYT algorithm, when we introduce to the house allocation problem with existing tenants a house similar to the w option of the kidney exchange problem.

Proposition 1 (*Krishna and Wang 2007*) *The TTCC algorithm induced by chain selection rule (e) is equivalent to the YRMH-IGYT algorithm.*

3.5 Two-Way Kidney Exchanges with 0-1 Preferences

Although the previous model is a variation of the house allocation and exchange model, there are intricate restrictions of the kidney exchange problem that this model cannot handle.

Since kidney donation is considered a gift, a donor cannot be forced to sign a contract regarding the donation. Thus, all transplants in an exchange should be conducted simultaneously, since otherwise a donor in the exchange could potentially back out after her paired-patient receives a kidney. This is an important restriction and almost always respected in real life. Since there should be a separate transplant team of doctors present for each donation and consequent transplant, this constraint puts a physical limit on the number of pairs that can participate simultaneously in one exchange. Because of this restriction, most of the real-life exchanges have been two-way exchanges including two pairs in one exchange. Roth, Sönmez, and Ünver (2005a) considered a model of kidney exchange using this restriction.

Another controversial issue in the market design for kidneys concerns the preferences of patients over kidneys. In the previous model, the assumption was that these preferences are heterogeneous. Although this is certainly the correct modeling approach from a theoretical point of view, small differences in quality may be only of secondary importance. Indeed, in the medical empirical literature several authors make this claim. In this second model, we will assume that all compatible kidneys have the same likelihood of survival, following Delmonico (2004) and Gjertson and Cecka (2000) who statistically show this in their data set. The medical doctors also point out that if the paired-donor of a patient is compatible with her, she will directly receive a kidney from her paired-donor and will not participate in the exchange.

The following model and the results are due to Roth, Sönmez, and Ünver (2005a), unless otherwise noted.

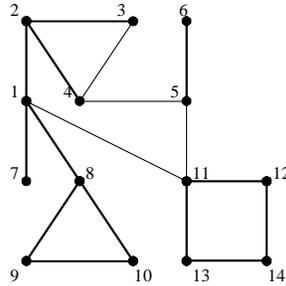
Let N be the set of pairs of all and only incompatible donors and their patients. Preferences are restricted further such that, for each pair $i \in N$, and $k, k' \in K_i$, $k \sim_i k'$, i.e. a patient is indifferent among all compatible kidneys. Moreover we restrict our attention to individually rational and two-way exchanges in this subsection. That is, for any $\mu \in \mathcal{M}$ and pair i , if $\mu(t_i) = k_j$ for some pair j then $\mu(t_j) = k_i$, and $k_j \in K_i$, $k_i \in K_j$. By a slight abuse of notation, we treat both the patient and the donor as one entity, and rewrite $\mu(i) = j$, meaning that patient t_i is matched with donor k_j , instead of $\mu(t_i) = k_j$.⁵ Since we focus on two-way exchanges, we need to define the following concept: Pairs i, j are **mutually compatible** if j has a compatible donor for the patient of i and i has a compatible donor for the patient of j , that is, $k_j \in K_i$ and $k_i \in K_j$. We can focus on a *mutual compatibility matrix* that summarizes the feasible exchanges and preferences. A **mutual**

⁵Moreover, throughout this section, whenever it is appropriate, we will use the term “patient” instead of “pair.”

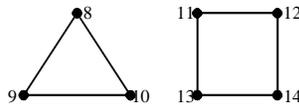
compatibility matrix $R = [r_{i,j}]_{i \in N, j \in N}$ is defined as for any $i, j \in N$,

$$r_{i,j} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are mutually compatible} \\ 0 & \text{otherwise} \end{cases} .$$

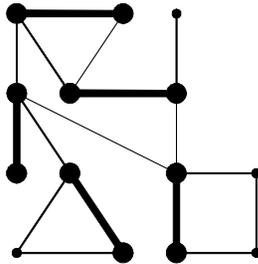
A **two-way kidney exchange problem** is denoted by (N, R) . The following figure depicts an undirected graph representation of a kidney exchange problem with $N = \{1, 2, \dots, 14\}$, Problem (N, R) is given as



where the edges are the set of feasible two-way exchanges and the vertices are the incompatible pairs. A **subproblem** of (N, R) is denoted as (I, R_I) where $I \subseteq N$ and R_I is the restriction of R to the pairs in I . For example, the following graph depicts subproblem (I, R_I) of the above problem with $I = \{8, 9, 10, 11, 12, 13, 14\}$:



Below, we depict with boldface edges a matching for the above problem (R, N) .



A Pareto-inefficient matching

A problem is **connected** if the corresponding graph of the problem is connected, i.e., one can traverse between any two nodes of the graph using the edges of the graph. A **component** is a largest connected subproblem. In the above problem (R, N) , there is only one component, the problem itself.

On the other hand, in the above subproblem (I, R_I) , there are two components, the first consisting of pairs 8, 9, and 10 and the second consisting of pairs 11, 12, 13, and 14. We refer to a component as **odd** if it has an odd number of pairs, and as **even** if it has an even number of pairs. In the above example, the first component is odd and the second component is even.

Besides deterministic outcomes, we will also define stochastic outcomes. A stochastic outcome is a **lottery** $\lambda = (\lambda_\mu)_{\mu \in \mathcal{M}}$ that is a probability distribution on all matchings. Although in many matching problems, there is no natural definition of von Neumann - Morgenstern utility functions, there is one for this problem: It takes value 1 if the patient is matched and 0 otherwise. We can define the (**expected**) **utility** of a patient t_i under a lottery λ as the probability of the patient getting a transplant and we denote it by $u_i(\lambda)$. The **utility profile** of lottery λ is denoted by $u(\lambda) = (u_i(\lambda))_{i \in N}$.

A matching is **Pareto-efficient** if there is no other matching that makes every patient weakly better off and some patient strictly better off. A lottery is **ex-post efficient** if it gives positive weight to only Pareto-efficient matchings. A lottery is **ex-ante efficient** if there is no other lottery that makes every patient weakly better off and some patient strictly better off. Although in many matching domains ex-ante and ex-post efficiency are not equivalent (for example, see Bogomolnaia and Moulin, 2001), because of the following lemma, they are equivalent for two-way kidney exchanges with 0-1 preferences.

Lemma 3 (Roth, Sönmez, and Ünver 2005a) *The same number of patients are matched at each Pareto-efficient matching, which is the maximum number of pairs that can be matched.*

Thus, finding a Pareto-efficient matching is equivalent to finding a matching that matches the maximum number of pairs. In graph theory, such a problem is known as a *cardinality matching problem* (see e.g. Korte and Vygen 2002, for an excellent survey of this and other optimization problems regarding graphs), and various intuitive polynomial time algorithms are known to find one Pareto-efficient matching starting with Edmonds' (1965) algorithm.

This lemma would not hold if exchange were possible among three or more patients. Moreover, we can state the following lemma regarding efficient lotteries:

Lemma 4 (Roth, Sönmez, and Ünver 2005a) *A lottery is ex-ante efficient if and only if it is ex-post efficient.*

There are many Pareto-efficient matchings, and finding all of them is not computationally feasible (i.e. NP-complete). Therefore, we will focus on two selections of Pareto-efficient matchings and lotteries that have nice fairness features.

3.5.1 Priority Mechanism

In many situations a natural priority ordering may arise that naturally orders patients. For example, the sensitivity of a patient to the tissue types of others, known as PRA, is a good criterion accepted also by medical doctors. Some patients may be sensitive to almost all tissue types other than their own and have a PRA=99%, meaning that they will reject 99% of donors from a random sample based solely on tissue incompatibility. So, one can order the patients from high to low with respect to their PRA's and use the following *priority mechanism*:

Algorithm 6 *The two-way priority (kidney exchange) mechanism:*

*Given a priority ordering of patients, a **priority mechanism** matches **Priority 1** patient if she is mutually compatible with a patient, and skips her otherwise.*
 \vdots
*matches **Priority k** patient in addition to all the previously matched patients if possible, and skips her otherwise.*

Thus, the mechanism determines which patients are to be matched first, and then one can select a Pareto-efficient matching that matches those patients. Thus, the mechanism is only unique-valued for the utility profile induced. Any matching inducing this utility profile can be the final outcome. The following result makes a priority mechanism very appealing:

Theorem 13 *A two-way priority mechanism is Pareto-efficient and strategy-proof.*

Although the above model did not consider multiple paired-donors, the extension of the model to multiple paired-donors is straightforward.

One can find additional structure about Pareto-efficient matchings (even though finding all such matchings is exhaustive) thanks to the results of Gallai (1963, 1964) and Edmonds (1965) in graph theory and combinatorial optimization. We can partition the patients (as a matter of fact, the incompatible pairs) into three sets as N^U, N^O, N^P . The members of these sets are defined as follows:

An **underdemanded patient** is one for whom there exists a Pareto-efficient matching that leaves her unmatched. Set N^U is formed by underdemanded patients, and we will refer to this set as the set of underdemanded patients. An **overdemanded patient** is one who is not underdemanded, yet is mutually compatible with an underdemanded patient. Set N^O is formed by overdemanded patients. A **perfectly matched patient** is one that is neither underdemanded nor mutually compatible with any underdemanded patient. Set N^P is formed by perfectly matched patients.

3.5.2 The Structure of Pareto-Efficient Matchings

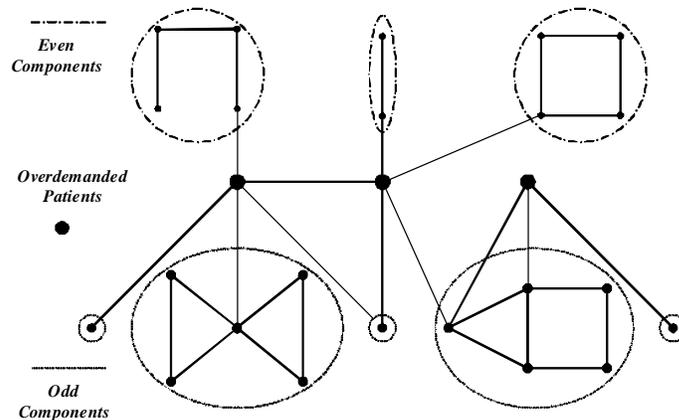
The following result, due to Gallai and Edmonds, is the key to understand the structure of Pareto-efficient matchings:

Lemma 5 Gallai (1963,1964)-Edmonds (1965) Decomposition (GED): Let μ be any Pareto-efficient matching for the original problem (N, R) and (I, R_I) be the **subproblem** for $I = N \setminus N^O$. Then we have:

1. Any overdemanded patient is matched with an underdemanded patient under μ .
2. $J \subseteq N^P$ for any **even component** J of the subproblem (I, R_I) and all patients in J are matched with each other under μ .
3. $J \subseteq N^U$ for any **odd component** J of the subproblem (I, R_I) and for any patient $i \in J$, it is possible to match all remaining patients with each other under μ . Moreover, under μ
 - either one patient in J is matched with an overdemanded patient and all others are matched with each other,
 - or
 - one patient in J remains unmatched while the others are matched with each other.

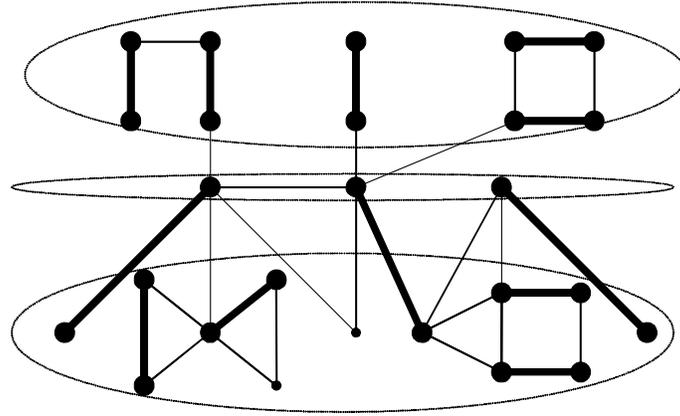
One can interpret this lemma as follows: There exists a competition among odd components of the subproblem (I, R_I) for overdemanded patients. Let $\mathcal{D} = \{D_1, \dots, D_p\}$ be the set of odd components remaining in the problem when overdemanded patients are removed. By the GED Lemma, all patients in each odd-component are matched but at most one, and all of the other patients are matched under each Pareto-efficient matching. Thus, such a matching leaves unmatched $|\mathcal{D}| - |N^O|$ patients each of whom is in a distinct odd component.

A depiction of the GED Lemma for a problem is given in the following figure:



First suppose that we determine the set of overdemanded patients, N^O . After removing those from the problem, we mark the patients in odd components as *underdemanded*, and patients in even components as *perfectly matched*. Moreover, we can think of each odd component as a single entity, which is competing to get one overdemanded patient for its patients under a Pareto-efficient

matching. An example of a Pareto-efficient matching is given in the following figure for the above problem:



It turns out that the sets N^U , N^O , N^P and the GED decomposition can also be found in polynomial time thanks to Edmonds' algorithm.

Below, we introduce another mechanism that takes into consideration another notion of fairness. This mechanism is also due to Roth, Sönmez, and Ünver (2005a).

3.5.3 Egalitarian Mechanism

Recall that the utility of a patient under a lottery is the probability of receiving a transplant. Equalizing utilities as much as possible may be considered very plausible from an equity perspective, which is also in line with the Rawlsian notion of fairness (Rawls 1971). We define a central notion in Rawlsian egalitarianism:

A feasible utility profile is **Lorenz-dominant** if

- the least fortunate patient receives the highest utility among all feasible utility profiles, and
- \vdots
- the sum of utilities of the k least fortunate patients is the highest among all feasible utility profiles.⁶

Is there a feasible Lorenz-dominant utility profile? Roth, Sönmez, and Ünver answer this question affirmatively. It is constructed with the help of the GED of the problem. Let

- $\mathcal{J} \subseteq \mathcal{D}$ be an arbitrary set of odd components of the subproblem obtained by removing the overdemanded patients,

⁶By *k least fortunate patients* under a utility profile, we refer to the k patients whose utilities are lowest in this utility profile.

- $I \subseteq N^O$ be an arbitrary set of overdemanded patients, and
- $C(\mathcal{J}, I)$ denote the **neighbors** of \mathcal{J} among I , that is, each overdemanded patient in $C(\mathcal{J}, I)$ is in I and is mutually compatible with a patient in an odd component of the collection \mathcal{J} .

Suppose only overdemanded patients in I are available to be matched with underdemanded patients in $\bigcup_{J \in \mathcal{J}} J$. Then, what is the upper bound of the utility that can be received by the *least fortunate* patient in $\bigcup_{J \in \mathcal{J}} J$? The answer is

$$f(\mathcal{J}, I) = \frac{|\bigcup_{J \in \mathcal{J}} J| - (|\mathcal{J}| - |C(\mathcal{J}, I)|)}{|\bigcup_{J \in \mathcal{J}} J|}$$

and it can be received only if

1. all underdemanded patients in $\bigcup_{J \in \mathcal{J}} J$ receive the same utility, and
2. all overdemanded patients in $C(\mathcal{J}, I)$ are committed for patients in $\bigcup_{J \in \mathcal{J}} J$.

The function f is the key in constructing an egalitarian utility profile. The following procedure can be used to construct it:

Algorithm 7 The construction of the egalitarian utility profile u^E :

Partition \mathcal{D} as $\mathcal{D}_1, \mathcal{D}_2, \dots$ and N^O as N_1^O, N_2^O, \dots as follows:

Step 1.

$$\mathcal{D}_1 = \arg \min_{\mathcal{J} \subseteq \mathcal{D}} f(\mathcal{J}, N^O) \text{ and}$$

$$N_1^O = C(\mathcal{D}_1, N^O)$$

\vdots

Step k .

$$\mathcal{D}_k = \arg \min_{\mathcal{J} \subseteq \mathcal{D} \setminus \bigcup_{\ell=1}^{k-1} \mathcal{D}_\ell} f\left(\mathcal{J}, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right) \text{ and}$$

$$N_k^O = C\left(\mathcal{D}_k, N^O \setminus \bigcup_{\ell=1}^{k-1} N_\ell^O\right)$$

Construct the vector $u^E = (u_i^E)_{i \in N}$ as follows:

1. For any overdemanded patient and perfectly matched patient $i \in N \setminus N^U$,

$$u_i^E = 1.$$

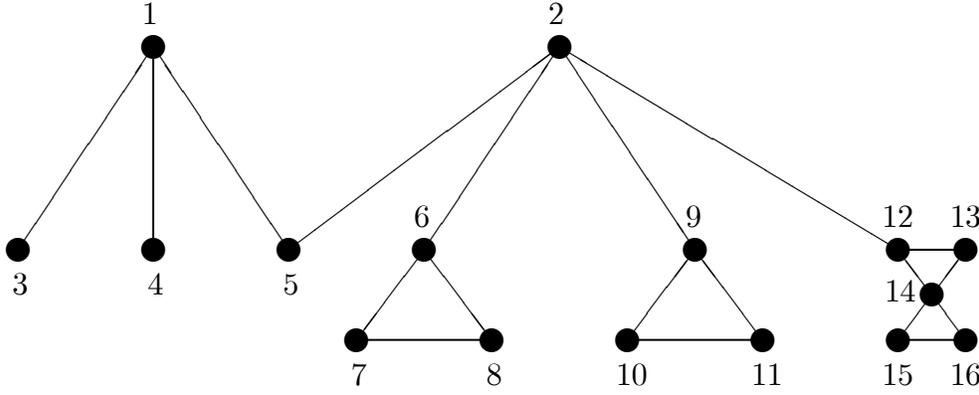


Figure 3.1: Graphical Representation for Example 2.

2. For any underdemanded patient i whose odd component left the above procedure at Step $k(i)$,

$$u_i^E = f(\mathcal{D}_{k(i)}, N_{k(i)}^O).$$

We provide an example explaining this construction:

Example 3 Let $N = \{1, \dots, 16\}$ be the set of patients and let the reduced problem be given by the graph in Figure 3.1. Each patient except 1 and 2 can be left unmatched at some Pareto-efficient matching and hence $N^U = \{3, \dots, 16\}$ is the set of underdemanded patients. Since both patients 1 and 2 have links with patients in N^U , $N^O = \{1, 2\}$ is the set of overdemanded patients.

$$\mathcal{D} = \{D_1, \dots, D_6\}$$

where

$$\begin{aligned} D_1 &= \{3\}, D_2 = \{4\}, D_3 = \{5\}, D_4 = \{6, 7, 8\} \\ D_5 &= \{9, 10, 11\}, D_6 = \{12, 13, 14, 15, 16\} \end{aligned}$$

Consider $\mathcal{J}_1 = \{D_1, D_2\} = \{\{3\}, \{4\}\}$. Note that by the GED Lemma, an odd component that has k patients guarantees $\frac{k-1}{k}$ utility for each of its patients. Since $f(\mathcal{J}_1, N^O) = \frac{1}{2} < \frac{2}{3} < \frac{4}{5}$, none of the multi-patient odd components is an element of \mathcal{D}_1 . Moreover, patient 5 has two overdemanded neighbors and $f(\mathcal{J}, N^O) > f(\mathcal{J}_1, N^O)$ for any $\mathcal{J} \subseteq \{\{3\}, \{4\}, \{5\}\}$ with $\{5\} \in \mathcal{J}$. Therefore

$$\begin{aligned} \mathcal{D}_1 &= \mathcal{J}_1 = \{\{3\}, \{4\}\}, \quad N_1^O = \{1\}, \\ u_3^E &= u_4^E = \frac{1}{2}. \end{aligned}$$

Next consider $\mathcal{J}_2 = \{D_3, D_4, D_5\} = \{\{5\}, \{6, 7, 8\}, \{9, 10, 11\}\}$. Note that $f(\mathcal{J}_2, N^O \setminus N_1^O) = \frac{7-(3-1)}{7} = \frac{5}{7}$. Since $f(\mathcal{J}_2, N^O \setminus N_1^O) = \frac{5}{7} < \frac{4}{5}$, the 5-patient odd component D_6 is not an element

of \mathcal{D}_2 . Moreover,

$$\begin{aligned} f(\{D_3\}, N^O \setminus N_1^O) &= f(\{D_4\}, N^O \setminus N_1^O) \\ &= f(\{D_5\}, N^O \setminus N_1^O) = 1, \\ f(\{D_3, D_4\}, N^O \setminus N_1^O) &= f(\{D_3, D_5\}, N^O \setminus N_1^O) = \frac{3}{4}, \\ f(\{D_4, D_5\}, N^O \setminus N_1^O) &= \frac{5}{6}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathcal{D}_2 &= \mathcal{J}_2 = \{\{5\}, \{6, 7, 8\}, \{9, 10, 11\}\}, \\ N_2^O &= \{2\}, \\ \text{and } u_5^E &= \dots = u_{11}^E = \frac{5}{7}. \end{aligned}$$

Finally since $N^O \setminus (N_1^O \cup N_2^O) = \emptyset$,

$$\begin{aligned} \mathcal{D}_3 &= \{\{12, 13, 14, 15, 16\}\}, \\ N_3^O &= \emptyset, \\ \text{and } u_{12}^E &= \dots = u_{16}^E = \frac{4}{5}. \end{aligned}$$

Hence the egalitarian utility profile is

$$u^E = (1, 1, \frac{1}{2}, \frac{1}{2}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{5}{7}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}).$$

Roth, Sönmez, and Ünver (2005a) proved the following results:

Theorem 14 (Roth, Sönmez, and Ünver 2005a) *The vector u^E is a feasible utility profile.*

In particular, the proof of Theorem 14 shows how a lottery that implements u^E can be constructed.

Theorem 15 (Roth, Sönmez, and Ünver 2005a) *The utility profile u^E Lorenz-dominates any other feasible utility profile (efficient or not).*

The egalitarian mechanism is a lottery mechanism that selects a lottery whose utility profile is u^E . It is only unique-valued for the utility profile induced. As a mechanism, the egalitarian approach has also appealing properties:

Theorem 16 (Roth, Sönmez, and Ünver 2005a) *The egalitarian mechanism is Pareto-efficient and strategy-proof.*

The egalitarian mechanism can be used for cases in which there is no exogenous way to distinguish among patients. The related literature for this subsection include two other papers, one by Bogomolnaia and Moulin (2004), who inspected a two-sided matching problem with the same setup as the model above, and one by Dutta and Ray (1989), who introduced the egalitarian approach for convex TU-cooperative games.

3.6 Multi-Way Kidney Exchanges with 0-1 Preferences

Roth, Sönmez, and Ünver (2007) inspected what is lost when the central authority conducts only two-way kidney exchanges rather than multi-way exchanges. More specifically, they inspected the upper bound of marginal gains from conducting 2&3-way exchanges instead of only two-way exchanges, 2&3&4-way exchanges instead of only 2&3-way exchanges, and unrestricted multi-way exchanges instead of only 2&3&4-way exchanges. The setup is very similar to the previous subsection with only one difference: a matching does not necessarily consist of two-way exchanges. All results in this subsection are due to Roth, Sönmez, and Ünver (2007) unless otherwise noted.

An example helps illustrate why the possibility of a 3-way exchange is important:

Example 4 *Consider a sample of 14 incompatible patient-donor pairs. A pair is denoted as type x - y if the patient and donor are ABO blood-types x and y respectively. There are nine pairs, who are blood-type incompatible, of types A - AB , B - AB , O - A , O - A , O - B , A - B , A - B , A - B , and B - A ; and five pairs, who are incompatible because of tissue rejection, of types A - A , A - A , A - A , B - O , and AB - O . For simplicity in this example there is no tissue rejection between patients and other patients' donors.*

- *If only two-way exchanges are possible:*

$(A$ - B , B - A); $(A$ - A , A - A); $(B$ - O , O - B); $(AB$ - O , A - AB) is a possible Pareto-efficient matching.

- *If three-way exchanges are also feasible:*

$(A$ - B , B - A); $(A$ - A , A - A , A - A); $(B$ - O , O - A , A - B); $(AB$ - O , O - A , A - AB) is a possible maximal Pareto-efficient matching.

The three-way exchanges allow

1. *an odd number of A - A pairs to be transplanted (instead of only an even number with two-way exchanges), and*
2. *a pair with a donor who has a blood type more desirable than her patient's to facilitate three transplants rather than only two. Here, the AB - O type pair helps two pairs with patients having less desirable blood type than their donors (O - A and A - AB), while the B - O type pair helps one pair with a patient having a less desirable blood type than her donor (O - A) and a pair of type A - B . Here, note that another A - B type pair is already matched with a B - A type, and this second A - B type pair is in excess.*

First we introduce two upper-bound assumptions and find the size of Pareto-efficient exchanges with only two-way exchanges:

Assumption 1 (Upper Bound Assumption) *No patient is tissue-type incompatible with another patient's donor.*

Assumption 2 (Large Population of Incompatible Patient-Donor Pairs) *Regardless of the maximum number of pairs allowed in each exchange, pairs of types O-A, O-B, O-AB, A-AB, and B-AB are on the "long side" of the exchange in the sense that at least one pair of each type remains unmatched in each feasible set of exchanges.*

The first result is about the greatest lower bound of the size of two-way Pareto-efficient matchings:

Proposition 2 (Roth, Sönmez, and Ünver 2007) The Maximal Size of Two-Way Matchings: *For any patient population obeying Assumptions 1 and 2, the maximum number of patients who can be matched with only two-way exchanges is:*

$$\begin{aligned} & 2(\#(A-O) + \#(B-O) + \#(AB-O) + \#(AB-A) + \#(AB-B)) \\ & + (\#(A-B) + \#(B-A) - |\#(A-B) - \#(B-A)|) \\ & + 2 \left(\left\lfloor \frac{\#(A-A)}{2} \right\rfloor + \left\lfloor \frac{\#(B-B)}{2} \right\rfloor + \left\lfloor \frac{\#(O-O)}{2} \right\rfloor + \left\lfloor \frac{\#(AB-AB)}{2} \right\rfloor \right) \end{aligned}$$

where $\lfloor a \rfloor$ refers to the largest integer smaller than or equal to a and $\#(x-y)$ refers to the number of $x-y$ type pairs.

We can generalize the above example in a proposition for three-way exchanges. We introduce an additional assumption for ease of notation. The symmetric case implies replacing types "A" with "B" and "B" with "A" in all of the following results.

Assumption 3 $\#(A-B) > \#(B-A)$.

The following is a simplifying assumption.

Assumption 4 *There is either no type A-A pair or there are at least two of them. The same is also true for each of the types B-B, AB-AB, and O-O.*

When three-way exchanges are also feasible, as we noted earlier, Lemma 3 no longer holds. Thus, we consider the largest of the Pareto-efficient matchings under 2&3-way matching technology.

Proposition 3 (Roth, Sönmez, and Ünver 2007) The Maximal Size of 2&3-Way Matchings: *For any patient population for which Assumptions 1-4 hold, the maximum number of patients who can be matched with two-way and three-way exchanges is:*

$$\begin{aligned} & 2(\#(A-O) + \#(B-O) + \#(AB-O) + \#(AB-A) + \#(AB-B)) \\ & + (\#(A-B) + \#(B-A) - |\#(A-B) - \#(B-A)|) \\ & + (\#(A-A) + \#(B-B) + \#(O-O) + \#(AB-AB)) \\ & + \#(AB-O) \\ & + \min\{(\#(A-B) - \#(B-A)), (\#(B-O) + \#(AB-A))\} \end{aligned}$$

And to summarize, the marginal effect of availability of 2&3-way kidney exchanges over two-way exchanges is:

$$\begin{aligned} & \#(A-A) + \#(B-B) + \#(O-O) + \#(AB-AB) \\ & - 2 \left(\left\lfloor \frac{\#(A-A)}{2} \right\rfloor + \left\lfloor \frac{\#(B-B)}{2} \right\rfloor + \left\lfloor \frac{\#(O-O)}{2} \right\rfloor + \left\lfloor \frac{\#(AB-AB)}{2} \right\rfloor \right) \\ & + \#(AB-O) \\ & + \min\{(\#(A-B) - \#(B-A)), (\#(B-O) + \#(AB-A))\} \end{aligned}$$

What about the marginal effect of 2&3&4-way exchanges over 2&3-way exchanges? It turns out that there is only a slight improvement in the maximal matching size with the possibility of four-way exchanges. We illustrate this using the above example:

Example 5 (Example 4 Continued) *If four-way exchanges are also feasible, instead of the exchange (AB-O, O-A, A-AB) we can now conduct a four-way exchange (AB-O, O-A, A-B, B-AB). Here, the valuable AB-O type pair helps an additional A-B type pair in excess in addition to two pairs with less desirable blood-type donors than their patients.*

Proposition 4 (Roth, Sönmez, and Ünver 2007) The Maximal Size of 2&3&4-Way Matchings: *For any patient population in which Assumptions 1-4 hold, the maximum number of patients who can be matched with two-way, three-way, and four-way exchanges is:*

$$\begin{aligned} & 2(\#(A-O) + \#(B-O) + \#(AB-O) + \#(AB-A) + \#(AB-B)) \\ & + (\#(A-B) + \#(B-A) - |\#(A-B) - \#(B-A)|) \\ & + (\#(A-A) + \#(B-B) + \#(O-O) + \#(AB-AB)) \\ & + \#(AB-O) \\ & + \min\{(\#(A-B) - \#(B-A)), \\ & \quad (\#(B-O) + \#(AB-A) + \#(AB-O))\} \end{aligned}$$

Therefore, in the absence of tissue-type incompatibilities between patients and other patients' donors, the marginal effect of four-way kidney exchanges is bounded from above by the rate of the very rare AB-O type.

It turns out that under the assumptions above, larger exchanges do not help to match more patients. This is stated as follows:

Theorem 17 (Roth, Sönmez, and Ünver 2007) Availability of Four-Way Exchange Suffices: *Consider a patient population for which Assumptions 1, 2, 4 hold and let μ be any maximal matching (when there is no restriction on the size of the exchanges). Then there exists a maximal matching ν that consists only of two-way, three-way, and four-way exchanges, under which the same set of patients benefits from exchange as in matching μ .*

In fact, Roth, Sönmez, and Ünver proved a more general theorem, which states that as long as there are n object types (e.g. for kidneys, 4 blood-types) and compatibility is determined by a partial order (i.e. a transitive, reflexive, anti-symmetric binary relation, e.g. blood-type compatibility is a partial order with "O" at the highest level, "A" and "B" incomparable with each other at the next level, and "AB" at the bottom level of compatibility), if Assumptions 2 and 4 hold, and μ is any maximal matching, then there exists a maximal matching ν which consists only of 2&3&...& n -way exchanges, in which the same agents are matched as in μ .

The strategic properties of multi-way kidney exchange mechanisms are inspected by Hatfield (2005) in the 0-1 preference domain. This result is a generalization of Theorem 13.

A deterministic kidney exchange mechanism is **consistent*** if whenever it only selects a multi-way matching in set $\mathcal{X} \subseteq \mathcal{M}$ as its outcome, where all matchings in \mathcal{X} generate the same utility profile when the set of feasible individually rational matchings is \mathcal{M} , then for any other problem for the same set of pairs such that the set of feasible individually rational matchings is $\mathcal{N} \subset \mathcal{M}$ with $\mathcal{X} \cap \mathcal{N} \neq \emptyset$, it selects a multi-way matching in set $\mathcal{X} \cap \mathcal{N}$.^{7,8} The last result of this section is as follows:

Theorem 18 (Hatfield 2005): *If a deterministic mechanism is non-bossy and strategy-proof then it is consistent*. Moreover, a consistent* mechanism is strategy-proof.*⁹

Thus, it is trivial to create strategy-proof mechanisms using *maximal-priority* or *priority* multi-way exchange rules. By maximal-priority mechanisms, we mean mechanisms that maximize the number of patients matched (under an exchange restriction such as 2, 3, 4, etc., or no exchange size restriction) and then use a priority criterion to select among such matchings.

3.7 Recent Developments and Related Literature

In closing of this section, we would like to note that New England Program for Kidney Exchange (NEPKE)¹⁰ is using a priority-based mechanism that incorporates 2&3&4-way paired exchanges, list exchanges, and non-directed altruistic donor exchanges (similar to the list exchanges, instead of the pair initiating a list exchange, an altruistic donor is used, e.g. see Sönmez and Ünver, 2006 and Roth, Sönmez, Ünver, Delmonico, and Saidman, 2006; also see Roth, Sönmez, and Ünver, 2005b). The Alliance for Paired Donation (APD)¹¹ is another kidney exchange program that has been established

⁷Recall that a kidney exchange mechanism may select many matchings that are utilitywise equivalent in the 0-1 preference domain. A two-way priority mechanism is an example.

⁸We use the * superscript to distinguish this new property from the consistency property we introduced in the house allocation problem.

⁹When there are possible indifferences in preferences, non-bossiness and strategy-proofness together are not necessarily equivalent to coalitional strategy-proofness.

¹⁰See <http://www.nepke.org> retrieved on 10/16/2008.

¹¹See <http://www.paireddonation.org> retrieved on 10/16/2008.

with the help of economists. This program is larger than its New England counterpart in number of transplant centers participating. In 2007, remarkably most of the kidney exchanges conducted in NEPKE and APD were chain exchanges initiated by a non-directed altruistic donor.

At the time of the preparation of this survey, the United Network for Organ Sharing (UNOS), the contractor for the federal Organ Procurement and Transplant Network (OPTN) that is in charge of the allocation of deceased donor kidneys in the US, has been designing the national kidney exchange program in collaboration with medical doctors, economists, and computer scientists.

Finding maximal multi-way matchings with a size limit is an NP-complete problem unlike its counterpart for two-way exchanges. Especially in large patient pools this may create a computational handicap. In the computer science literature, Abraham, Blum, and Sandholm (2007) introduced an integer programming algorithm that can compute the maximal multi-way exchanges with size-limit in a fast fashion exploiting the special structure of the multi-way kidney exchange problem. They use the Roth, Sönmez, and Ünver (2007) formulation of the multi-way exchange problem in their algorithm. Ünver (2007) considered a dynamic exchange problem where pairs arrive at the pool under a stochastic Poisson process. He finds optimal dynamic matching in this framework and shows that it may always not be optimal to conduct the largest exchange currently possible. Yilmaz (2008a) found an egalitarian mechanism that allows multi-way list and paired exchanges under compatibility-based preferences.

Zenios (2002) studied the optimal control of a paired and list exchange program. In addition to the simulations reported in Roth, Sönmez, and Ünver (2004, 2005b, and 2007), in the medical literature starting with Segev et al. (2005), who simulated possible gains in the US population using Edmonds' (1965) algorithm from weight-maximal two-way exchanges, several papers reported Monte-Carlo simulations estimating possible gains from various ideas in kidney exchange.

In the algorithmic design literature, there are theoretically related studies to the kidney exchange problem such as Abraham et al. (2005), Cechlárová, Fleiner, and Manlove (2005), Biró and Cechlárová (2007), Irving (2007), and Biró and McDermid (2008). These studies study computational complexity of different proposed solutions to the house allocation and kidney exchange problems.

Chapter 4

School Admissions

4.1 College Admissions

In Gale and Shapley's (1962) seminal model, there exist two sides of agents referred to as colleges and students. Each student would like to attend a college and has preferences over colleges and the option of remaining unmatched. Each college would like to recruit a maximum number of students determined by their exogenously given capacity. They have preferences over individual students, which translate into preferences over groups of students under a responsiveness (Roth 1985) assumption. More specifically, a **college admissions problem** consists of

- a finite set of **students** I ,
- a finite set of **schools** S ,
- a **quota** vector $q = (q_s)_{s \in S}$ such that $q_s \in \mathbb{Z}_{++}$ is the quota of school s ,
- a **preference profile for students** $\succ_I = (\succ_i)_{i \in I}$ such that \succ_i is a strict preference relation over schools and remaining unmatched, denoting the strict preference relation of student i , and
- a **preference profile for schools over individual students** $\succ_S = (\succ_s)_{s \in S}$ such that \succ_s is a strict preference relation over students and remaining unmatched, such that when such a relation is extended over groups of students it satisfies the following two restrictions known as **responsiveness** (Roth 1985):¹
 - whenever $i, j \in I$ and $J \subseteq I \setminus \{i, j\}$, $i \cup J \succ_s j \cup J$ if and only if $i \succ_s j$,
 - whenever $i \in I$ and $J \subseteq I \setminus i$, $i \cup J \succ_s J$ if and only if $i \succ_s \emptyset$, which denotes the remaining unmatched option for a school (and for a student).

¹By an abuse of notation, we will denote a singleton without $\{\}$.

A **matching** is the outcome of a problem, and is defined by a function $\mu : I \cup S \rightarrow 2^S \cup 2^I$ such that for each student $i \in I$, $\mu(i) \in 2^S$ with $|\mu(i)| \leq 1$, for each school s , $\mu(s) \in 2^I$ with $|\mu(s)| \leq q_s$, and $\mu(i) = s$ if and only if $i \in \mu(s)$. A (**deterministic direct**) **mechanism** selects a matching for each problem.

The central solution concept in the literature is *stability* (Gale and Shapley 1962). A matching μ is **stable** if

- each match is **individually rational**, i.e., there is no **blocking agent** x and a partner $y \in \mu(x)$ such that $\mu(x) \setminus y \succ_x \mu(x)$, that is, no agent would rather not be matched with one of her mates under μ (if x is a student, then she prefers remaining unmatched to her mate), and
- there is no **blocking pair** $(i, s) \in I \times S$ such that
 - $s \succ_i \mu(i)$ and
 - $i \cup (\mu(s) \setminus x) \succ_i \mu(s)$ for some $x \in \mu(s)$ or $|\mu(s)| < q_s$ and $\mu(s) \cup i \succ_s \mu(s)$,

that is, there exists no student-school pair who would prefer to be matched with each other rather than at most one of their current mates under μ .

Gale and Shapley prove that for each market there exists a stable matching that can be found through the **school-proposing** or **student-proposing** versions of the **deferred acceptance (DA) algorithm**. We state these algorithms below:

Algorithm 8 *The school-proposing DA algorithm:*

Step 1: *Each school s proposes to its top choice q_s students (if it has fewer individually rational choices than q_s , then it proposes to all its individually rational students). Each student rejects any individually irrational proposals and, if more than one individually rational proposal is received, "holds" the most preferred.*

⋮

Step k : *Any school s that was rejected in the previous step by ℓ students makes a new proposal to its most preferred ℓ students who haven't yet rejected it (if there are fewer than ℓ individually rational students, it proposes to all of them). Each student "holds" her most preferred individually rational offer to date and rejects the rest.*

The algorithm terminates after a step where no rejections are made by matching each student to the school (if any) whose proposal she is "holding."

Algorithm 9 *The student-proposing DA algorithm:*

Step 1: *Each student proposes to her top-choice individually rational school (if she has one). Each school s rejects any individually irrational proposals and, if more than q_s individually rational proposals are received, "holds" the most preferred q_s of them and rejects the rest.*

⋮

Step k : Any student who was rejected in the previous step makes a new proposal to her most preferred individually rational school that hasn't yet rejected her (if there is one). Each school s "holds" at most q_s best student proposals to date, and rejects the rest.

The algorithm terminates after a step where no rejections are made by matching each school to the students (if any) whose proposals it is "holding."

These algorithms have desirable properties:

Theorem 19 (Gale and Shapley 1962) *The student- and school-proposing DA algorithm each converge to a stable matching in a finite number of steps.*

Moreover, these algorithms can be used to determine the outcomes of important stable mechanisms:

Theorem 20 (Gale and Shapley 1962) *The outcome of the student-proposing DA algorithm is at least as good as any other stable matching for all students. The outcome of the school-proposing DA algorithm is at least as good as any other stable matching for all schools.*

We will refer to the mechanism whose outcome is reached by the student-proposing DA algorithm as the **student-optimal stable mechanism** and the mechanism whose outcome is reached by the school-proposing DA algorithm as the **school-optimal stable mechanism**.²

Stability implies Pareto efficiency. However, it imposes many restrictions on mechanisms:

Theorem 21 (Roth 1982b) *There is no stable and strategy-proof college admissions mechanism.*

Yet, a partially positive result exists:

Theorem 22 (Dubins and Freedman 1981, Roth 1982b) *It is a weakly dominant strategy for students to tell the truth under the student-optimal stable mechanism.*

However, we have a negative result for schools' incentives under stable mechanisms:

Theorem 23 (Roth 1985) *There exists no stable mechanism that makes it a dominant strategy for each school to state its preferences over the students truthfully.*

²See Roth and Sotomayor (1990) for other properties of stable matchings, such as the lattice property, conflict of interest, and parallels between the model in which a school can also be matched with a single student (also known as the *one-to-one matching market* or *marriage market*) and the college admissions model.

While these results are true in the college admissions setting, the hospital-intern entry-level labor markets in the US can be modeled using the same framework. In the US, the National Residency Matching Program (NRMP) oversees this matching procedure. Roth (1984) showed that the previous NRMP mechanism that was in use from 1950s to 1997 was equivalent to the school-optimal stable mechanism. Roth (1991) observed that several matching mechanisms that have been used in Britain for hospital-intern matching were unstable and as a result were abandoned, while stable mechanisms survived. This key observation helped to pin down stability as a key property of matching mechanisms in the college admissions framework. Roth and Peranson (1999) introduced a new design for the NRMP matching mechanism based on the student-optimal stable mechanism. Interestingly, the replacement of the older stable mechanism with the newer mechanism was partially attributed to the positive and negative results in Theorems 22 and 23, respectively.

4.1.1 Differences between College Admissions, Student Placement, and School Choice Problems

Although Gale and Shapley named their model as the college admissions problem, not all college admission procedures can be studied within this framework. For example, US college admissions are usually decentralized. However, there are countries, such as Turkey, Greece, and China, where the process of college admissions is centralized. In such countries, colleges are not strategic agents unlike in the college admissions model, while students potentially are. School seats are objects to be consumed, and there are priority orderings for each school over students based on their exam scores. We will refer to such a problem as a *student placement problem* (Balinski and Sönmez 1999). In the US, K-12 public school admissions are centralized in many states. Moreover, there is relative freedom of school choice freedom, i.e., students do not have to attend the neighborhood school, but have the chance to attend a different school. In such a problem, schools seats are objects to be consumed, and students are potential strategic agents. Priorities that order students for each school are exogenously determined by geography and demographics. We will refer to such a problem as a *school choice problem* (Abdulkadiroğlu and Sönmez 2003a). We explore these models and real-life mechanisms below.

4.2 Student Placement

A **student placement problem** consists of

- a finite set of **students** I ,
- a finite set of **schools** S ,
- a **quota** vector $q = (q_s)_{s \in S}$ such that $q_s \in \mathbb{Z}_{++}$ is the quota of school s ,

- a **preference profile for students** $\succ_I = (\succ_i)_{i \in I}$ such that \succ_i is a strict preference relation over schools and remaining unmatched option, denoting the strict preference relation of student i ,
- a finite set of **categories for schools** C ,
- an **exam score profile** for students $e = (e^i)_{i \in I}$ such that for any $i \in I$ and $e^i = (e_c^i)_{c \in C}$ where for each category $c \in C$, $e_c^i \in \mathbb{R}_+$ is the exam score of student i in this category and there are no other students $j \in I \setminus \{i\}$ such that $e_c^i = e_c^j$, and
- a **type function** mapping each school to a category type, $t : S \rightarrow C$.

Throughout this subsection we fix I, S, C , and t . Thus a placement problem is denoted through a triple (\succ_I, q, e) .

Each school s admits students according to the exam scores of students in category $t(s)$.

For each student placement problem, we can construct an **associated college admissions problem** by assigning each school s a preference relation \succ_s based on the ranking in its category $t(s)$.

We will define a matching and mechanism in this domain together with a new concept.

A **matching** is a function $\mu : I \rightarrow S \cup \{\emptyset\}$ such that no school is assigned to more students than its capacity. When $\mu(i) = \emptyset$, we say that student i is *unmatched* or *matched to no school option*.

A **tentative student placement** is a correspondence $\mu : I \rightrightarrows S \cup \{\emptyset\}$ such that no school is assigned to more students than its capacity. Observe that a *tentative student placement* allows a student to be assigned to more than one school.

A **mechanism** is a function that assigns a *matching* for each student placement problem.

Next, we will define desirable properties of student placement mechanisms.

A matching μ **eliminates justified envy** if, whenever a student i prefers another student j 's assignment $\mu(j)$ to her own, she ranks worse than j in the category of school $\mu(j)$.

A matching μ is **non-wasteful** if, whenever a student i prefers a school s to her own, there is no empty slot at school s under μ .

We introduced these new concepts to relate *elimination of justified envy*, *non-wastefulness*, and *individual rationality* to *stability* in the college admissions model as follows:

Proposition 5 (Balinski and Sönmez 1999) *A school placement matching eliminates justified envy and is non-wasteful and individually rational if and only if the matching is stable in the associated college admissions problem. That is, there is an isomorphism with stable college admissions.*

Elimination of justified envy is a critical property in the context of Turkish college admissions. In Turkey, colleges have schools in different areas such as medicine, engineering, humanities, social sciences, and management. The score categories for these schools are typically different from each

other. Medical schools usually admit based on a science-weighted score, engineering schools use a math-weighted score, management schools use an equal-language-math-weighted score, and many social sciences and humanities use a social-science-weighted score. Elimination of justified envy is used as the basic notion of fairness in Turkish placement system.

A mechanism **eliminates justified envy** (or is **non-wasteful**) if it always selects a matching that eliminates justified envy (is non-wasteful).

4.2.1 Simple Case: One Skill Category

If there is a single category, then the following proposition follows:

Proposition 6 (*Balinski and Sönmez 1999*) *If there is only one category (and hence only one ranking) then there is only one mechanism that is Pareto-efficient and eliminates justified envy: The simple serial dictatorship induced by this ranking.*

It is also useful to observe that there is a unique stable matching in the associated college admissions model that coincides with the outcome of the above serial dictatorship.

An example from Turkey is again useful in this context. There exist merit-based Turkish high schools that admit their students using the results of a centralized exam. This exam has a single score and category. It turns out that the mechanism used in Turkey, developed independently by computer programmers, is the induced serial-dictatorship.

4.2.2 Current Mechanism in Turkish College Student Placement: Multi-Category Serial Dictatorship

Currently, the Turkish centralized mechanism uses the following iterative algorithm:

Algorithm 10 *The multi-category serial dictatorship:*

Step 1:

- *For each category c : Consider the ranking induced by the exam scores in this category and assign the school seats in this category to students with the induced simple serial dictatorship.*
- *Assign the "no school" option to all students who are not assigned a school.*
- *This in general leads to a tentative student placement.*
- *For each student i construct \succ_i^1 from \succ_i as follows:*

– *If the student is not assigned more than one school then $\succ_i^1 = \succ_i$.*

- If the student is assigned more than one school then obtain \succ_i^1 by moving the "no school" option \emptyset right after the best of these assignments, otherwise keeping the ranking of the schools the same.

Let $\succ^1 = (\succ_i^1)_{i \in I}$ be the list of adjusted preferences.

⋮

Step k: Construct \succ^k from \succ^{k-1} as it is described in Step 1.

The procedure terminates at the step in which no student is assigned more than one school. The **multi-category serial dictatorship** selects this matching.

We give an example to show how this algorithm works.

Example 6 $I = \{i_1, i_2, i_3, i_4, i_5\}$, $S = \{s_1, s_2, s_3\}$, $q = (q_{s_1}, q_{s_2}, q_{s_3}) = (2, 1, 1)$, $C = \{c_1, c_2\}$, $t(s_1) = c_1$, $t(s_2) = t(s_3) = c_2$, with preference profile \succ and exam score profile e given as:

$$\begin{array}{ll} i_1 : s_2 - s_1 - \emptyset & e^{i_1} = (9, 9) \\ i_2 : s_1 - s_2 - s_3 - \emptyset & e^{i_2} = (8, 6) \\ i_3 : s_1 - s_3 - s_2 - \emptyset & e^{i_3} = (7, 7) \\ i_4 : s_1 - s_2 - \emptyset & e^{i_4} = (6, 8) \\ i_5 : s_2 - s_3 - s_1 - \emptyset & e^{i_5} = (5, 5) \end{array}$$

Note that these scores induce the following rankings in categories c_1 and c_2 :

$$\begin{array}{l} c_1 : i_1 \ i_2 \ i_3 \ i_4 \ i_5 \\ c_2 : i_1 \ i_4 \ i_3 \ i_2 \ i_5 \end{array}$$

Step 1: In Step 1 we first find the serial dictatorship outcomes for \succ :

$$\begin{array}{ll} c_1 : & i_1 \ i_2 \\ & s_1 \ s_1 \\ c_2 : & i_1 \ i_4 \ i_3 \\ & s_2 \ \emptyset \ s_3 \end{array}$$

Step 1 yields the following tentative student placement:

$$\nu^1 = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_1, s_2 & s_1 & s_3 & \emptyset & \emptyset \end{pmatrix}$$

Since student i_1 is assigned two schools her preferences are truncated:

$$i_1 : s_2 - \emptyset$$

For other students: $\succ_{i_2}^1 = \succ_{i_2}$, $\succ_{i_3}^1 = \succ_{i_3}$, $\succ_{i_4}^1 = \succ_{i_4}$, and $\succ_{i_5}^1 = \succ_{i_5}$.

Step 2: In Step 2 we first find the serial dictatorship outcomes for \succ^1 :

$$c_1 : \begin{array}{ccc} i_1 & i_2 & i_3 \\ \emptyset & s_1 & s_1 \end{array} \quad c_2 : \begin{array}{ccc} i_1 & i_4 & i_3 \\ s_2 & \emptyset & s_3 \end{array}$$

Step 2 yields the following tentative student placement:

$$\nu^2 = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_2 & s_1 & s_1, s_3 & \emptyset & \emptyset \end{pmatrix}.$$

Since student i_3 is assigned two schools her preferences are truncated:

$$i_3 : s_1 - \emptyset$$

For other students: $\succ_{i_1}^2 = \succ_{i_1}^1$, $\succ_{i_2}^2 = \succ_{i_2}^1$, $\succ_{i_4}^2 = \succ_{i_4}^1$, and $\succ_{i_5}^2 = \succ_{i_5}^1$.

Step 3: In Step 3 we first find the serial dictatorship outcomes for \succ^2 :

$$c_1 : \begin{array}{ccc} i_1 & i_2 & i_3 \\ \emptyset & s_1 & s_1 \end{array} \quad c_2 : \begin{array}{cccc} i_1 & i_4 & i_3 & i_2 \\ s_2 & \emptyset & \emptyset & s_3 \end{array}$$

Step 3 yields the following tentative student placement:

$$\nu^3 = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_2 & s_1, s_3 & s_1 & \emptyset & \emptyset \end{pmatrix}$$

Since student i_2 is assigned two schools her preferences are truncated:

$$i_2 : s_1 - \emptyset$$

For other students: $\succ_{i_1}^3 = \succ_{i_1}^2$, $\succ_{i_3}^3 = \succ_{i_3}^2$, $\succ_{i_4}^3 = \succ_{i_4}^2$, and $\succ_{i_5}^3 = \succ_{i_5}^2$.

Step 4: In Step 4 we first find the serial dictatorship outcomes for \succ^3 .

$$c_1 : \begin{array}{ccc} i_1 & i_2 & i_3 \\ \emptyset & s_1 & s_1 \end{array} \quad c_2 : \begin{array}{ccccc} i_1 & i_4 & i_3 & i_2 & i_5 \\ s_2 & \emptyset & \emptyset & \emptyset & s_3 \end{array}$$

Step 4 yields the following tentative student placement (which is also a matching):

$$\nu^4 = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_2 & s_1 & s_1 & \emptyset & s_3 \end{pmatrix}$$

Since no student is assigned more than one school in ν^4 the algorithm terminates and $\varphi^{msd}(\succ_I, e, q) = \nu^4$.

4.2.3 Mechanisms via the Associated College Admissions Problem

We can introduce two desirable mechanisms using the isomorphism between the student placement and school admissions models:

- The **Gale-Shapley school-optimal stable mechanism**: The mechanism that selects the school-optimal stable matching of the associated college admissions problem for each student placement problem.
- The **Gale-Shapley student-optimal stable mechanism**: The mechanism that selects the student-optimal stable matching of the associated college admissions problem for each student placement problem.

The following theorem proves the relationship between the Gale-Shapley mechanisms and the multi-category serial dictatorship.

Theorem 24 (*Balinski and Sönmez 1999*) *The multi-category serial dictatorship is equivalent to the Gale-Shapley school-optimal stable mechanism.*

Next, we comment on the properties of this mechanism:

4.2.4 Pareto Efficiency and Elimination of Justified Envy

Although all stable mechanisms (including Gale and Shapley's) are Pareto-efficient in the college admissions model, in the student placement model, this is no longer true. The reason can be summarized as follows: Since schools are no longer agents in the latter model, we are no longer interested in their welfare. Moreover, unstable matchings can raise the welfare of students over the student-optimal stable matching in the college admissions model. These two results together imply that the outcome of any stable mechanism can be Pareto-inefficient in the student placement model:

Example 7 *There are three students i_1, i_2, i_3 and three schools s_1, s_2, s_3 , each of which has only one seat and admit according to the following two categories c_1 and c_2 as $t(s_1) = c_1, t(s_2) = c_2$, and $t(s_3) = c_3$. The preferences and exam scores are as follows:*

$$\begin{array}{ll} i_1 : s_2 - s_1 - s_3 - \emptyset & e^{i_1} = (10, 7) \\ i_2 : s_1 - s_2 - s_3 - \emptyset & e^{i_2} = (8, 8) \\ i_3 : s_1 - s_2 - s_3 - \emptyset & e^{i_3} = (9, 3) \end{array}$$

These exam scores induce the following ranking for categories:

$$\begin{array}{l} c_1 : i_1 - i_3 - i_2 \\ c_2 : i_2 - i_1 - i_3 \end{array}.$$

Only μ eliminates justified envy but it is Pareto-dominated by ν :

$$\mu = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_1 & s_2 & s_3 \end{pmatrix} \quad \nu = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}$$

However, the multi-category serial dictatorship mechanism is not even Pareto-efficient within the set of mechanisms that eliminate justified envy.

Example 8 Let $I = \{i_1, i_2\}$ $S = \{s_1, s_2\}$ $q = (1, 1)$ $C = \{c_1, c_2\}$, $t(s_1) = c_1, t(s_2) = c_2$. The preferences of students are given as follows:

$$\begin{aligned} i_1 : s_1 - s_2 - \emptyset & \quad e^{i_1} = (6, 8) \\ i_2 : s_2 - s_1 - \emptyset & \quad e^{i_2} = (8, 6) \end{aligned}$$

The algorithm terminates in one step resulting in the following Pareto-inefficient matching:

$$\varphi^{msd}[\succ_I, e, q] = \begin{pmatrix} i_1 & i_2 \\ s_2 & s_1 \end{pmatrix}$$

It is Pareto-dominated by the following matching that eliminates justified envy:

$$\mu = \begin{pmatrix} i_1 & i_2 \\ s_1 & s_2 \end{pmatrix}.$$

On the other hand, we can adopt Theorem 20 (due to Gale and Shapley) in the school placement domain for the Gale-Shapley student-optimal stable mechanism as follows:

Theorem 25 (Gale and Shapley 1962) The Gale-Shapley student-optimal stable mechanism Pareto-dominates any other mechanism that eliminates justified envy.

4.2.5 Strategy-Proofness and Elimination of Justified Envy

On the other hand, strategy-proofness is no longer at odds with the elimination of justified envy, yet the multi-category serial dictatorship is not strategy-proof:

Example 9 (Example 8 continued) Recall that

$$\varphi^{msd}[\succ_I, e, q] = \begin{pmatrix} i_1 & i_2 \\ s_2 & s_1 \end{pmatrix}$$

Now suppose i_1 announces a fake preference relation \succ'_{i_1} where only s_1 is individually rational.

In this case

$$\varphi^{msd}[\succ'_{i_1}, \succ_{i_2}, e, q] = \begin{pmatrix} i_1 & i_2 \\ s_1 & s_2 \end{pmatrix}$$

and hence, student i_1 successfully manipulates the multi-category serial dictatorship.

A mechanism is **strategy-proof** if truth telling is a weakly dominant strategy for each student in its associated preference revelation game. We can adopt Theorem 22 for the student placement model:

Theorem 26 (Dubins and Freedman 1981, Roth 1982b): *The Gale-Shapley student-optimal stable mechanism is strategy-proof.*

The following theorem shows that there is no other desirable mechanism:

Theorem 27 (Alcalde and Barberà 1994): *The Gale-Shapley student-optimal stable mechanism is the only mechanism that eliminates justified envy, and is individually rational, non-wasteful, and strategy-proof.*

4.2.6 Respecting Improvements

Example 10 Example 8 continued Recall that

$$\varphi^{msd} [\succ_I, e, q] = \begin{pmatrix} i_1 & i_2 \\ s_2 & s_1 \end{pmatrix}.$$

Now suppose student i_1 scores worse in both tests and her new exam scores are $e^{i_1} = (5, 5)$. In this case

$$\varphi^{msd} [\succ_I, e^{i_1}, e^{i_2}, q] = \begin{pmatrix} i_1 & i_2 \\ s_1 & s_2 \end{pmatrix}.$$

and student i_1 is rewarded by getting her top choice as a result of worse performance!

Note the example is about rewarding worse performance, not respecting better performance. We define this as a property: A mechanism **respects improvements** if a student never receives a worse assignment as a result of an increase in one or more of her exam scores. The following theorems give another characterization of the Gale-Shapley student-optimal stable mechanism:

Theorem 28 (Balinski and Sönmez 1999): *The Gale-Shapley student-optimal stable mechanism respects improvements.*

Theorem 29 (Balinski and Sönmez 1999): *The Gale-Shapley student-optimal stable mechanism is the only mechanism that is individually rational and non-wasteful, and that eliminates justified envy and respects improvements.*

Thus, the Gale-Shapley student-optimal stable mechanism is the clear winner for student placement, while the Turkish student placement system uses a mechanism that is equivalent to the Gale-Shapley school-optimal stable mechanism.³

³See Ehlers and Klaus (2006) and Kojima and Manea (2007b) for two other characterizations regarding the Gale-Shapley student-optimal stable mechanism in resource allocation problems.

4.3 School Choice

Next, we discuss the third model in this section: A *school choice problem* (Abdulkadiroğlu and Sönmez 2003a) models the school choice in public schools in many school districts in the US, such as Boston, St. Petersburg (Florida), Minneapolis, etc. It consists of a number of students, each of whom should be assigned a seat at one of a number of schools. Each school has a maximum capacity but there is no shortage of the total seats. Each student has preferences over all schools and each school has a priority ordering of all students. The priorities are exogenous.

Formally, a **school choice problem** consists of

- a finite set of **students** I ,
- a finite set of **schools** S ,
- a **quota** vector $q = (q_s)_{s \in S}$ such that $q_s \in \mathbb{Z}_{++}$ is the quota of school s ,
- a **preference profile for students** $\succ_I = (\succ_i)_{i \in I}$ such that \succ_i is a strict preference relation over schools and remaining unmatched, denoting the strict preference relation of student i ,
- a **priority profile for schools** $\succ_S = (\succ_s)_{s \in S}$ such that for each school $s \in S$, \succ_s is a binary relation over the set of students that is complete, reflexive, and transitive. That is, $i \succ_s j$ means that student i has at least as high priority as student j at school s . Two distinct students i and j can have the same priority at school s , which is denoted as $i \sim_s j$ (i.e., \sim_s is the cyclic part of \succ_s). If i has higher priority than j at s , we denote it as $i \succ_s j$ (i.e., \succ_s is the antisymmetric part of \succ_s).

This problem has a number of differences from the college admissions problem and the student placement problem:

- Differences from college admissions:
 - Students are (possibly strategic) agents; school seats are objects to be consumed.
 - Elimination of justified envy is *plausible* but not a *must*. If imposed, then the school choice problem is *isomorphic* to stable college admissions.
- Differences from student placement:
 - Priorities are exogenous, and
 - Elimination of justified envy is *plausible* but *not a must*.

4.3.1 The Boston School Choice Mechanism

The most commonly used school choice mechanism is that used by the Boston Public Schools (BPS) until 2005:

Algorithm 11 *The Boston (school choice) mechanism:*

1. *For each school a priority ordering is exogenously determined. (In case of Boston, priorities depend on home address, whether the student has a sibling already attending a school, and a lottery number to break ties.)*
2. *Each student submits a preference ranking of the schools.*
3. *The final phase is the student assignment based on preferences and priorities:*

Step 1: *In Step 1 only the top choices of the students are considered. For each school, consider the students who have listed it as their top choice and assign seats of the school to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her top choice.*

⋮

Step k : *Consider the remaining students. In Step k only the k^{th} choices of these students are considered. For each school still with available seats, consider the students who have listed it as their k^{th} choice and assign the remaining seats to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her k^{th} choice.*

4.3.2 Incentives, Pareto Efficiency, and Justified-Envy-Freeness with Strict and Weak Priorities

The major difficulty with the Boston mechanism is that it is not strategy-proof. Moreover, it is almost straightforward to manipulate it. Even if a student has a very high priority at school s , unless she lists it as her top choice she loses her priority to students who have top ranked school s . Hence the Boston mechanism gives parents strong incentives to overrank schools where they have high priority.

There is also some evidence in the popular media regarding the ease of manipulation of this mechanism. Consider the following quotation from the St. Petersburg Times (09/14/2003):

"Make a realistic, informed selection on the school you list as your first choice. It's the cleanest shot you will get at a school, but if you aim too high you might miss. Here's why: If the random computer selection rejects your first choice, your chances of getting

your second choice school are greatly diminished. That's because you then fall in line behind everyone who wanted your second choice school as their first choice. You can fall even farther back in line as you get bumped down to your third, fourth and fifth choices."

Further evidence comes from the 2004-2005 BPS School Guide:

"For a better choice of your 'first choice' school ... consider choosing less popular schools."

The Boston mechanism does not eliminate justified envy, either. Priorities are lost unless the school is ranked as the top choice. In the previous section we argued that if *elimination of justified envy* is plausible, then the Gale-Shapley student-optimal stable mechanism is the big winner! However, unlike in the student placement problem, in which ties in student exam scores are rare, there are possibly many students who have the same priority in the school choice problem. For example, in Boston, all students who live in the walking zone of a school and have no siblings attending the school have the same priority. Thus, the student-proposing DA algorithm can be used after breaking the tie among equal priority students through a single even lottery. This lottery preserves the strategy-proofness and justified-envy-freeness of the Gale-Shapley mechanism.

The following theorem is about the Nash equilibria of the Boston Mechanism revelation game:

Theorem 30 (Ergin and Sönmez 2006): *When priorities are strict, the set of Nash equilibrium outcomes of the preference revelation game induced by the Boston mechanism is equal to the set of stable matchings of the associated college admissions game under true preferences.*

Thus, we can state the following corollary regarding the Boston mechanism and the Gale-Shapley student-optimal stable mechanism:

Corollary 1 *When priorities are strict, the dominant-strategy equilibrium outcome of the Gale-Shapley student-optimal stable mechanism either Pareto-dominates or is equal to the Nash equilibrium outcomes of the Boston mechanism.*

The preference revelation game induced by the Boston mechanism is a "coordination game" among large numbers of parents in which there is incomplete information. So it is unrealistic to expect to reach a Nash equilibrium in practice.

On the other hand, if there is a limit to the number of schools that a student can reveal to the centralized match (as in Boston and New York City), then Corollary 1 no longer holds, while Theorem 30 still holds:

Theorem 31 (Haeringer and Klijn 2007) *When priorities are strict and students can reveal only a limited number of schools in their preference lists, the Gale-Shapley student-optimal stable mechanism may have Nash equilibria in undominated strategies that induce justified envy.*

Haeringer and Klijn (2007) also found the sufficient conditions when equilibria of the above game eliminate justified envy.

On the other hand, the following nice property of the Gale-Shapley mechanism relates its efficiency properties to any other strategy-proof and Pareto-efficient mechanism:

Theorem 32 (Kesten 2008a) *When priorities are strict, the Gale-Shapley student-optimal stable mechanism is not Pareto-dominated by any other Pareto-efficient mechanism that is strategy-proof.*

When a school has the same priority for two or more students, some results under strict priorities extend, while some don't.

Under weak priorities, there can be many **student-optimal justified-envy-free** matchings, matchings that are not Pareto-dominated by any other justified-envy-free matching and Pareto-dominate any justified-envy-free matching that is not student optimal. Recall that when priorities are strict, there is a unique such matching (see Theorem 25). The above mechanism also has desirable properties for recovering such matchings:

Theorem 33 (Ehlers 2006, Erdil and Ergin 2008) *When priorities are weak, all student-optimal justified-envy-free matchings can be found by different tie-breaking rules among equal priority students using the student-proposing DA algorithm.*

This above result is a generalization of an earlier result of Abdulkadiroğlu and Sönmez (1998) who showed that when all students have the same priority, all Pareto-efficient matchings can be achieved through different serial dictatorships.

The following is a stronger generalization of the earlier result of Kesten (2008a) (Theorem 32) for weak priorities:

Theorem 34 (Abdulkadiroğlu, Pathak, and Roth 2008) *When priorities are weak, the Gale-Shapley student-optimal stable mechanism with any tie breaking rule is not Pareto-dominated by any other mechanism that is strategy-proof.*

On the other hand, the Gale-Shapley student-optimal stable mechanism is not Pareto-efficient. As we discussed in the previous section, there is an efficiency cost to the elimination of justified envy. We restate a version of Example 7 below. Observe that this result does need strict priorities among at least three students to hold:

Example 11 *There are three students i_1, i_2, i_3 and three schools s_1, s_2, s_3 , each of which has only one seat. Priorities and preferences are as follows:*

$$\begin{array}{ll}
 s_1 : i_1 - i_3 - i_2 & i_1 : s_2 - s_1 - s_3 \\
 s_2 : i_2 - i_1 - i_3 & i_2 : s_1 - s_2 - s_3 \\
 s_3 : i_2 - i_1 - i_3 & i_3 : s_1 - s_2 - s_3
 \end{array}$$

Only μ eliminates justified envy but it is Pareto-dominated by ν :

$$\mu = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_1 & s_2 & s_3 \end{pmatrix} \quad \nu = \begin{pmatrix} i_1 & i_2 & i_3 \\ s_2 & s_1 & s_3 \end{pmatrix}$$

Actually, the efficiency cost of justified envy is much more severe with weak priorities. The following result can be contrasted with Theorems 25 and 26, which show that the Gale-Shapley student-optimal stable mechanism is strategy-proof and Pareto-dominant among mechanisms that eliminate justified envy when priorities are strict:

Theorem 35 (Erdil and Ergin 2008) *When priorities are weak, there is no mechanism that is constrained Pareto-efficient (within the justified-envy-free class) among (lottery) mechanisms that eliminate justified envy and are (weakly) strategy-proof.*^{4,5}

To summarize, with weak priorities, the above results show the tension between strategy-proofness and constrained efficiency for justified-envy-free mechanisms. The Gale-Shapley student-optimal stable mechanism (with a tie-breaking rule that makes it strategy-proof, such as a single tie-breaking lottery) is strategy-proof and Pareto-undominated by other strategy-proof mechanisms. Yet, there exist justified-envy-free and non-strategy-proof mechanisms that Pareto-dominate this mechanism. Two examples of constrained efficient and justified-envy-free mechanisms are given by Erdil and Ergin (2008) and Kesten (2008a). These mechanisms are non-strategy-proof.

4.3.3 The School Choice TTC Mechanism

Given these negative results, one can argue that Pareto efficiency is a more important property than elimination of justified envy. School boards can interpret priorities as trading rights to a particular school. In this case a version of the TTC mechanism becomes very plausible. Abdulkadiroğlu and Sönmez (2003a) introduced a mechanism whose outcome can be determined by the following algorithm:

Algorithm 12 *The school choice TTC algorithm:*

- *Break the ties among equal priority students of each school through a single even lottery.*
- *Assign a counter for each school that keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools.*

⁴"Weak" strategy-proofness is defined for lottery mechanisms, and requires existence of at least one von Neumann-Morgenstern utility function compatible with preferences, under which truth telling is a dominant strategy.

⁵Yilmaz (2008b) obtained a similar impossibility result for the house allocation with existing tenants domain.

- *Each student "points to" her favorite school. Each school points to the student who has the highest priority.*
- *There is at least one cycle (by Lemma 2). Every student in a cycle is assigned a seat at the school she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. Counters of all other schools stay put.*
- *Repeat above steps for the remaining "economy."*

TTC simply trades priorities of students among themselves starting with the students with highest priorities. TTC inherits the plausible properties of Gale's TTC:

Theorem 36 (*Abdulkadiroğlu and Sönmez 2003a*) *The school choice TTC mechanism is Pareto-efficient and strategy-proof.*

Chen and Sönmez (2006) conducted an experimental study and found that the Gale-Shapley mechanism outperforms TTC and the Boston mechanism in terms of truthful revelation of preferences and overall efficiency. They related this result to the fact that TTC has a tedious algorithmic description with respect to the Gale-Shapley mechanism; thus students understood the second algorithm, better than the first one, under which they tried to manipulate their preferences. On the other hand, Pais and Pintér (2007) showed that when the same games are played in an incomplete information setting then TTC resulted with more efficiency than the Gale-Shapley mechanism and the Boston mechanism.

4.4 Recent Developments and Related Literature

In New York City (Abdulkadiroğlu, Pathak, and Roth 2005), the Gale-Shapley student-optimal stable mechanism was adopted in Fall 2003. The New York City school choice problem is a hybrid between college admissions and school choice, since there are some strategic schools. In Boston (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2005, 2006), though TTC had a head start, the Gale-Shapley student-optimal stable mechanism was selected to replace the Boston mechanism.

Ergin (2002) showed that under an acyclicity condition of priorities, the Gale-Shapley mechanism finds Pareto-efficient outcomes in the school admissions domain. Moreover, the Gale-Shapley mechanism is coalitionally strategy-proof in this case.

Since in the adopted mechanisms we discussed above, ties among equal priority students are broken randomly, we may observe some unnecessary inefficiency under the Gale-Shapley student-optimal stable mechanism.

Kesten (2008a) introduced a hybrid approach for the school choice domain that compensates for the inefficiency of the Gale-Shapley student-optimal stable mechanism through a compromise

mechanism that introduces minimal instability while creating more efficient outcomes. Moreover, the instability is created with the consent of participating students: a blocking student will never be worse off if she gives consent for such stability violations.

Erdil and Ergin (2008) recognized that the artificial tie-breaking of priorities induces inefficiencies under the Gale-Shapley student-optimal stable mechanism. Therefore, after the algorithm converges they proposed a second stage. This is also an iterative procedure. They proposed a random trading stage so that each student can trade her seat as long as other students agree. However, not all trades are acceptable. Trades involving students with the highest priority are deemed feasible. After a "stable" trading cycle is randomly found, the trades are realized. Thus this process does not induce further inefficiencies. One can conduct feasible trades again and repeat the above procedure until no stable trades are left. Although the Erdil-Ergin mechanism is ex-post Pareto efficient, it is not strategy-proof, and yet truth-telling is an ordinal Bayesian-Nash equilibrium in a low and symmetric information setting. Using data from NYC schools, Abdulkadiroğlu, Pathak, and Roth (2008) showed that over 1,500 student applicants among 8th graders could have improved their assignment in the Erdil-Ergin mechanism among 90,000 students, if the same student preferences would have been revealed.

Pathak and Sönmez (2008) inspected the Boston mechanism's revelation game when not all students are sophisticated. Sincere players are restricted to report their true preferences, while strategic players play a best response. Although there are multiple equilibrium outcomes, a sincere student receives the same assignment in all equilibria. Finally, the assignment of any strategic student under the Pareto-dominant Nash equilibrium of the Boston mechanism is weakly preferred to her assignment under the student-optimal stable mechanism.

Abdulkadiroğlu and Ehlers (2007) inspected the school choice problem, when there are minimum quotas for students from different backgrounds at schools. These minimum quotas in general lead to non-existence of justified-envy-free matchings. Thus, they introduced a new definition of justified-envy-freeness. Under this new definition, they showed that a justified-envy-free matching always exists in a "controlled" school choice problem.

There is also an emerging literature regarding the lottery mechanisms in school choice. We cite some of the recent papers below.

Abdulkadiroğlu, Che, and Yasuda (2008) introduced a new tie-breaking rule: each student has the option to designate a target school besides revealing her preferences. Whenever tie-breaking is needed among multiple students for a school, students who designate this school as target get priority in tie-breaking. Then the Gale-Shapley student-optimal stable mechanism is applied on the modified priority structure. The authors found plausible properties of this mechanism over the Gale-Shapley version.

Pathak (2006) inspected lottery design in the school choice domain. He proved an equivalence result between RSD and random school-choice TTC mechanism, when all priority orders of schools

are independently and uniformly randomly drawn. This corresponds to two versions of tie-breaking among equal priority students: tie-breaking for all schools using a single lottery or tie-breaking independently for each school. However, such an equivalence does not exist for random multiple tie-breaking version of the Gale-Shapley student-optimal stable mechanism and RSD (which is equivalent to Gale-Shapley mechanism with random single tie-breaking).

Kesten and Ünver (2008) introduced two lottery mechanisms which result in lotteries over student-optimal justified-envy-free matchings according to two new definitions of justified-envy-freeness. This is the first study that employed an "ex-ante" lottery design approach in school choice, while the previous approaches were "ex-post." All previous approaches that use "ex-ante" random tie-breaking followed by an "ex-post" implementation of a deterministic mechanism, including Erdil and Ergin's, resulted with efficiency loss that could be avoided using an alternative "ex-ante" approach.

Featherstone and Niederle (2008) observed that Boston mechanism resulted with better efficiency than the Gale-Shapley student-optimal mechanism in laboratory experiments, when ties are broken randomly, and preferences are private information. Thus, Boston mechanism is effectively manipulated by the students in these experiments. They also prove this result in a symmetric environment in theory.

Chapter 5

Axiomatic Mechanisms and Market Design

5.1 House Allocation and Hierarchical Exchange

In the house allocation domain, Pápai (2000) introduced a wide class of mechanisms called *hierarchical exchange mechanisms* that are inspired by Gale's TTC algorithm and serial dictatorships such that they uniquely characterize the class of Pareto-efficient, reallocation-proof, and coalitionally strategy-proof mechanisms.

A mechanism ϕ is **reallocation-proof** if for any problem \succ , there is no pair of agents a and b and two preference relations \succ'_a and \succ'_b such that $\phi[\succ'_a, \succ_{-a}](a) = \phi[\succ](a)$ and $\phi[\succ'_b, \succ_{-b}](b) = \phi[\succ](b)$ and yet $\phi[\succ'_a, \succ'_b, \succ_{-\{a,b\}}](b) \succ_a \phi[\succ](a)$ and $\phi[\succ'_a, \succ'_b, \succ_{-\{a,b\}}](a) \succ_b \phi[\succ](b)$.

The idea behind hierarchical exchange mechanisms is as follows:

Suppose that we assign houses to the agents initially according to an inheritance rule that is described by the mechanism. As the agents who have the property rights of the houses leave the market while the houses remain unmatched, their property rights are passed to other agents according to the inheritance rule.

A **submatching** is the matching of a subset of agents $B \subseteq A$ to houses $G \subseteq H$, i.e., a one-to-one and onto function $\sigma : B \rightarrow G$. Let $A_\sigma = B$ and $H_\sigma = G$. Let \mathcal{S} be the set of submatchings. For each house h , let \mathcal{S}_{-h} be the set of submatchings that do not assign house h .

Note that a **matching** is a submatching σ with $A_\sigma = A$. Let $\mathcal{M} \subset \mathcal{S}$ be the set of matchings, as before.

Formally, a hierarchical exchange mechanism is described through an inheritance function $f = (f_h)_{h \in H}$ such that each $f_h : \mathcal{S}_{-h} \setminus \mathcal{M} \rightarrow A$ determines who has the property rights of house h , once a submatching is already fixed. That is, for any $\sigma \in \mathcal{S}_{-h} \setminus \mathcal{M}$, $f_h(\sigma) \in A \setminus A_\sigma$, such that $f_h(\sigma)$ is the

agent who has the property right of house h when the submatching σ is already fixed.¹

We have the following restriction on f_h : For all $\sigma \subseteq \sigma'$ with $f_h(\sigma) \notin A_{\sigma'}$, we have $f_h(\sigma') = f_h(\sigma)$. That is, if an agent has the right of a house, when more matches are determined, and this agent is not matched, she does not lose her right for this house. Let \mathcal{F} be the set of such f functions. Each $f \in \mathcal{F}$ induces a **hierarchical exchange mechanism**, let ϕ^f be this mechanism.

An iterative algorithm is used to find the allocation under a hierarchical exchange mechanism:

Algorithm 13 *The hierarchical exchange induced by f :*

Step k : *Suppose σ^k is a submatching already determined at the end of the previous step (we start with $\sigma^1 = \emptyset$ initially at $k = 1$). If σ^k is a matching then we terminate the algorithm, and σ^k is the outcome of the algorithm. Otherwise, each remaining house h points to its inheritance right holder $f_h(\sigma^k)$, each remaining agent points to her top choice house among the remaining houses, and we obtain a directed graph. There exists at least one cycle (by Lemma 2). We clear each cycle by assigning each agent in the cycle the house she is pointing to. Let σ^{k+1} be the submatching that is determined by clearing these cycles, and the matches already determined under σ^k . We continue with Step $k+1$.*

Below, we give examples about the relationship of hierarchical exchange and other mechanisms we introduced in the previous chapters of this survey:

Example 12 *Suppose that μ is a matching, and for each agent $a \in A$, $f_{\mu(a)}(\emptyset) = a$. Then this inheritance rule gives a house to each agent initially. The rest of the inheritance rule is defined arbitrarily.*

*The induced hierarchical exchange algorithm is equivalent to **Gale's top trading cycles** algorithm and finds the core of the housing market induced by initial endowment μ .*

Example 13 *Let $p = (a_1, \dots, a_n)$ be an ordering of agents in A . Suppose that for all h and all σ , $f_h(\sigma) = a_k$ where k is the lowest index such that a_k not matched under σ .*

*This inheritance rule gives the control rights of all houses to the same agent as long as that agent is available. That is, the induced hierarchical exchange mechanism is the **serial dictatorship** induced by p .*

Example 14 *Suppose that there are two types of agents and houses, A_E , A_N and H_O , H_V , respectively. For each $a \in A_E$, $h_a \in H_O$, we set $f_{h_a}(\emptyset) = a$. Moreover, suppose there is an ordering of agents $p = (a_1, \dots, a_n)$ such that for all $h \in H_V$, $f_h(\sigma) = a_k$ where a_k is the agent with lowest k such that a_k is not matched under σ . For all $h_a \in H_O$, whenever a is matched under σ but h_a is not, then $f_{h_a}(\sigma) = a_k$ where a_k is the agent with lowest k such that a_k is not matched under σ .*

*The induced hierarchical exchange mechanism is the **YRMH-IGYT mechanism** induced by priority order p .*

¹This simplified definition is due to Pycia and Ünver (2007).

Example 15 Suppose the property rights of the houses are given according to the following inheritance table for houses $H = \{h_1, h_2, h_3, h_4\}$ over $A = \{1, 2, 3, 4\}$.

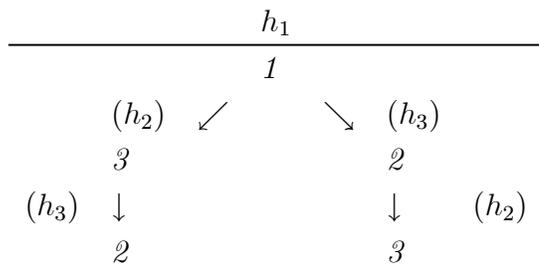
h_1	h_2	h_3	h_4
1	1	2	4
2	2	3	3
3	3	1	2
4	4	4	1

An inheritance table refers to a specific inheritance rule profile such that regardless of the assigned house of the owner of a remaining house, this remaining house is inherited by the same new owner. The induced inheritance profile f by the above table is as follows: $f_{h_1}(\emptyset) = 1$, $f_{h_1}(\{(1, x)\}) = 2$ for any $x \in \{h_2, h_3, h_4\}$ (that is, when 1 is matched, the right goes to 2), $f_{h_1}(\{(1, x), (2, y)\}) = 3$ for all $\{x, y\} \subseteq \{h_2, h_3, h_4\}$. $f_{h_1}(\{(1, x), (2, y), (3, z)\}) = 4$ for all $\{x, y, z\} \subseteq \{h_2, h_3, h_4\}$. The rights for houses h_2, h_3 , and h_4 are similarly defined.

One interpretation of the above table is that the inheritance table gives the **priority profile** of houses over the students (for example, houses are school seats and the agents in A are students, and the priority profile is induced by f). Then the induced **school choice top trading cycles mechanism** (Abdulkadiroğlu and Sönmez 2003a) is a hierarchical exchange mechanism.

Hierarchical exchange mechanisms constitute a proper superset of the mechanisms we introduced earlier. We illustrate this with an example, in which the hierarchical exchange mechanism introduced is neither a serial dictatorship, the core mechanism, a YRMH-IGYT mechanism, nor a school choice TTC mechanism:

Example 16 Let $A = \{1, 2, 3\}$, $H = \{h_1, h_2, h_3\}$. Suppose the inheritance rule profile f induces a tree for house h_1 :



This means, $f_{h_1}(\emptyset) = 1$, $f_{h_1}(\{(1, h_2)\}) = 3$, $f_{h_1}(\{(1, h_3)\}) = 2$, $f_{h_1}(\{(1, h_2), (3, h_3)\}) = 2$, $f_{h_1}(\{(1, h_3), (2, h_2)\}) = 3$. Suppose for houses h_2 and h_3 we have the following inheritance table for f_{h_2} and f_{h_3} :

h_2	h_3
1	2
2	3
3	1

Let the preferences of the agents be given as:

1	2	3
h_2	h_2	h_1
h_3	h_1	h_2
h_1	h_3	h_3

The induced hierarchical exchange outcome is found as follows through the directed graphs formed:

Step 1: $1 \rightarrow h_2 \rightarrow 1$, $2 \rightarrow h_2 \rightarrow 1$, $3 \rightarrow h_1 \rightarrow 1$.

There is only one cycle: $1 \rightarrow h_2 \rightarrow 1$, agent 1 is assigned h_2 .

Now according to h_1 's inheritance tree the right of house h_1 goes to agent 3.

Step 2: $2 \rightarrow h_1 \rightarrow 3$, $3 \rightarrow h_1 \rightarrow 3$.

There is one cycle: $3 \rightarrow h_1 \rightarrow 3$, agent 3 is assigned house h_1 .

Step 3: $2 \rightarrow h_3 \rightarrow 2$, there is one cycle: $2 \rightarrow h_3 \rightarrow 2$.

No agent is left, thus the algorithm terminates. The outcome of the hierarchical exchange mechanism is given as

$$\mu = \begin{pmatrix} 1 & 2 & 3 \\ h_2 & h_3 & h_1 \end{pmatrix}.$$

Our result of this chapter is as follows:

Theorem 37 (Pápai 2000) *A mechanism is reallocation-proof, Pareto-efficient, and coalitionally strategy-proof if and only if it is a hierarchical exchange mechanism.*

5.2 Trading Cycles with Brokers and Owners

In this section, we introduce a new algorithm called *trading cycles with brokers and owners* (Pycia and Ünver, 2007), which is more general than hierarchical exchange. This will remove the reallocation-proofness axiom from the above characterization result.

The algorithm works as follows: In each round, it assigns the control rights of each unremoved house to some unremoved agent. This agent controls this house as an "owner" or as a "broker." The hierarchical exchange only designated control rights holders as "owners". Thus "brokers" are innovation of this new algorithm. In either case, this house cannot be matched in this round unless its control rights holder is matched. The algorithm is based on the top-trading cycles idea, yet it is substantially different.

The assignment produced by this algorithm depends on the structure of control rights. Let us define this new concept first. A **structure of control rights** (a^c, h^b) consists of a profile of **control functions** $a^c = (a_h^c : \mathcal{S}_{-h} \rightarrow A)_{h \in H}$ such that for all h and all $\sigma \in \mathcal{S}_{-h}$, $a_h^c(\sigma) \in A - A_\sigma$; and a **brokered house function** $h^b : \mathcal{S} - \mathcal{M} \rightarrow H \cup \{\emptyset\}$ such that for all $\sigma \in \mathcal{S} - \mathcal{M}$, if $|A_\sigma| = |A| - 1$, then $h^b(\sigma) = \emptyset$.

For all control rights structures, the assignment of houses to agents is determined by an iterative algorithm that we refer to as the **trading-cycles-with-brokers-and-owners algorithm (TCBO algorithm)** for short).

Algorithm 14 *The trading cycles with brokers and owners (TCBO) induced by (*

Since we assign at least one agent a house in every round, and since there are finitely many agents, the algorithm stops after finitely many rounds.

The terminology of owners and brokers is motivated by the trading analogy. In each round of the algorithm, an owner can either trade a house he controls for another house (in a cycle of several exchanges), or can leave in this round matched with a house he owns. A broker can trade the house he owns for another house (in a cycle of several exchanges), but cannot leave in this round matched with the house he brokers. One interpretation of this is that the owner can consume his house, but the broker cannot.

Example 17 (Execution of the TCBO algorithm) *Let $A = \{1, 2, 3\}$ and $H = \{h_1, h_2, h_3\}$. Suppose the control rights structure is such that*

- h_1 is owned by 1 as long as 1 and h_1 are unmatched, is owned by 2 when 2 and h_1 are unmatched and 1 is matched, and is owned by 3 when 3 and h_1 are unmatched and 1 and 2 are matched,
- h_2 is owned by 2 as long as 2 and h_2 are unmatched, is owned by 1 when 1 and 2 are unmatched and 2 is matched, and is owned by 3 when 3 and h_2 are unmatched, and 1 and 2 are matched,
- h_3 is controlled by 3; he has the brokerage right as long as either 1 and 2 are unmatched and the ownership right when 1 and 2 are matched (notice that we do not need to specify who inherits h_3 when 3 is matched, because 3 may be matched only in a cycle that also contains h_3).

The above structure of control rights may be represented as follows:

$a_{h_1}^c$	$a_{h_2}^c$	$a_{h_3}^c$
1	2	3^b
2	1	
3	3	

The ^b sign, above, next to 3 in h_3 's control right column, shows that h_3 is a brokered-house (when some agents other than 3 who controls h_3 are unmatched). The preferences of the agents are given as follows:

$$\begin{aligned}
 \text{agent 1: } & h_3 \succ_1 h_2 \succ_1 h_1 \\
 \text{agent 2: } & h_3 \succ_2 h_1 \succ_2 h_2 \\
 \text{agent 3: } & h_3 \succ_3 h_1 \succ_3 h_2
 \end{aligned}$$

We run the algorithm as follows:

Step 1. Owned-house h_1 points to $a_{h_1}^c(\emptyset) = 1$, owned-house h_2 points to $a_{h_2}^c(\emptyset) = 2$, brokered-house $h^b(\emptyset) = h_3$ points to $a_{h^b(\emptyset)}^c(\emptyset) = 3$. Agents 1 and 2 point to h_3 and broker 3 points to his first choice owned-house, that is h_1 . There exists one cycle

$$h_1 \rightarrow 1 \rightarrow h_3 \rightarrow 3 \rightarrow h_1,$$

and by removing it, we obtain

$$\sigma^1 = \{(1, h_3), (3, h_1)\}$$

Step 2. O-house h_2 points to $a_{h_2}^c(\sigma^1) = 2$ and agent 2 points to h_2 . There exists one cycle

$$h_2 \rightarrow 2 \rightarrow h_2,$$

and by removing it, we obtain

$$\sigma^2 = \{(1, h_3), (2, h_2), (3, h_1)\}.$$

This is a matching, since no agents are left.

We terminate the algorithm, and the outcome of the mechanism is σ^2 .

Observe that this outcome cannot be reproduced by a hierarchical exchange mechanism. Consider a modified problem obtained by changing preferences of agent 3 so that h_2 is preferred to h_1 :

$$\begin{aligned} \text{agent 1: } & h_3 \succ_1 h_1 \succ_1 h_2 \\ \text{agent 2: } & h_3 \succ_2 h_2 \succ_2 h_1 \\ \text{agent 3: } & h_3 \succ'_3 h_2 \succ'_3 h_1 \end{aligned}$$

In this case, the TCBO outcome is

$$\sigma' = \{(1, h_1), (2, h_3), (3, h_2)\}.$$

However, any hierarchical exchange mechanism that assigns h_3 to 1 in the first problem should continue to do so in the second problem. Thus, no hierarchical exchange mechanism can reproduce this TCBO's outcome.

We are ready to formally define TCBO mechanism class (Pycia and Ünver 2007). A control rights structure (a^c, h^b) is **compatible** if for all submatchings $\sigma \in \mathcal{S} - \mathcal{M}$,

C1. Persistence of ownership: If agent a owns house h at σ , and a and h are unmatched at $\sigma' \supset \sigma$, then a owns h at σ' .

C2. No ownership for brokers: If agent b is a broker at σ , then h^b does not own any house at σ .

C3. Limited persistence of brokerage: If agent h^b brokers house f at σ , agent $a' \neq b$ and house $g \neq f$ are unmatched at σ , and b does not broker f at submatching $\sigma \cup \{(a', g)\}$, then either

- **Broker-to-heir transition:** (i) there is exactly one agent a who owns a house both at σ and $\sigma \cup \{(a', g)\}$, (ii) agent a owns house f at $\sigma \cup \{(a', g)\}$, and (iii) at submatching $\sigma \cup \{(a', g), (a, f)\}$, agent b owns all houses that a owns at σ ,
or
- **Direct exit from brokerage:** there is no agent who owns a house at both σ and $\sigma \cup \{(a', g)\}$.

Each *compatible* pair (a^c, h^b) induces a **trading-cycles-with-brokers-and-owners mechanism** (**TCBO mechanism** for short). Its outcome is found through the TCBO algorithm that was introduced earlier. The control rights structure introduced in the previous example is compatible, thus the mechanism implemented is TCBO.

The main theorem regarding this larger class is proven by Pycia and Ünver (2007) and removes reallocation-proofness property of Pápai from the axiomatic characterization. We further assume that $|H| \geq |I|$:

Theorem 38 (Pycia and Ünver 2007) *A mechanism is coalitionally strategy-proof and Pareto-efficient if and only if it is a TCBO mechanism.*

The characterization does not need Pareto-efficiency, if the mechanisms have *full range*, i.e., mechanism ϕ has **full range** if for every matching $\mu \in \mathcal{M}$, there exists some preference profile \succ such that $\phi[\succ] = \mu$.

Corollary 2 (Pycia and Ünver 2007) *A full-range mechanism is coalitionally strategy-proof if and only if it is a TCBO mechanism.*

As an example of a mechanism design problem in which brokerage rights are useful, consider a manager who assigns n tasks t_1, \dots, t_n to n employees w_1, \dots, w_n with strict preferences over the tasks. The manager wants the allocation to be Pareto-efficient with regard to the employees' preferences. Within this constraint, she would like to avoid assigning task t_1 to employee w_1 . She wants to use a coalitionally strategy-proof direct mechanism, because she does not know employees' preferences. The only way to do it using the previously known mechanisms is to endow employees w_2, \dots, w_n with the tasks, let them find the Pareto-efficient allocation through a top-trading cycles procedure, such as Pápai's (2000) hierarchical exchange, and then allocate the remaining task to employee w_1 . Ex ante each such procedure is unfair to the employee w_1 . Using a trading-cycles-with-brokers-and-owners mechanism, the manager can achieve her objective without the extreme discrimination of the employee w_1 . She makes w_1 the broker of t_1 , allocates the remaining tasks among w_2, \dots, w_n (for instance she may make w_i the owner of t_i , $i = 2, \dots, n$), and runs trading cycles with brokers and owners.

5.3 Related Literature

Unlike the core mechanism for housing markets (see Theorem 6), there are many desirable mechanisms in the house allocation (with existing tenants) domain. We already stated some axiomatic characterization results in Theorems 8, 9, and 10. Also in the school admissions domain, we stated two characterization results (see Theorems 27 and 29, see also Ehlers and Klaus 2006, and Kojima and Manea 2007b, for other characterizations in the same domain).

We will cite several other papers below:

On the other hand, if we do not insist on strict preferences, coalitional strategy-proofness and Pareto efficiency are incompatible in general. Ehlers (2002) found the largest possible preference domain under which these two properties are not at odds, and characterized the set of coalitionally strategy-proof and Pareto-efficient mechanisms. Similarly, Bogomolnaia, Deb, and Ehlers (2005) characterized two classes of strategy-proof mechanisms in the same preference domain.

There are several other axiomatic studies that focus on more specialized properties of mechanisms in different domains, such as Ehlers, Klaus, and Pápai (2002), Miyagawa (2002), Ehlers and Klaus (2007), Pápai (2007), Velez (2008), and Kesten (2008b).

Chapter 6

Concluding Remarks

We would like to conclude by commenting on the literature that we left out of this survey. Our attention to axiomatic mechanism design was brief. Similarly, we did not explore lottery mechanisms in depth. Such explorations deserve their own survey papers. We give a brief summary of the literature on lottery mechanisms below, since the literature may have important implications for market design.

6.1 Lottery Mechanisms in Matching

In the house allocation domain, a study by Chambers (2004) showed that a probabilistic consistency property is difficult to achieve if fairness is also imposed. He showed that a uniform lottery allocation of houses is the unique stochastically consistent mechanism that is also fair in the sense of equal treatment of equals. Clearly, such an allocation is not Pareto-efficient.

On the positive side, Bogomolnaia and Moulin (2001) introduced an algorithm class, which we can refer to as *eating algorithms*, that implement different lottery mechanisms. Randomization is used to sustain fairness among the agents, since as we have seen, desirable deterministic mechanisms impose an artificial hierarchical structure that can favor some agents over others. A central mechanism in the class, which gives "equal eating speeds" to all agents, is known as the *probabilistic serial (PS) mechanism*.

One shortcoming of the PS mechanism is that it is not strategy-proof. Yet, all mechanisms induced by eating algorithms including PS are ordinally efficient, in the sense that the probability distribution of houses assigned is not first-order stochastically dominated by any other (lottery) mechanism. In fact, a mechanism is ordinally efficient if and only if its outcome can be found through an eating algorithm.¹

¹Crés and Moulin (2001) and Bogomolnaia and Moulin (2002) introduced a strategy-proof and ordinally efficient lottery mechanism in a preference domain where relative preferences of the agents are identical for the houses, but opting-out can be ranked differently for each different agent.

On the other hand, another central mechanism, obtained by randomly drawing a priority ordering of agents and implementing the resulting serial dictatorship, is not ordinally efficient. This is a surprising result, since serial dictatorships are Pareto-efficient mechanisms. On the other hand, this lottery mechanism, known as the *random serial dictatorship (RSD)* is strategy-proof.

PS and RSD mechanisms are both fair (in the sense of equal treatment of equals). Yet, it turns out that ordinal efficiency, equal treatment of equals, and strategy-proofness are incompatible properties. Thus, PS favors ordinal efficiency, while RSD favors strategy-proofness. RSD is only ex-post efficient and PS is only weakly strategy-proof.

Kojima and Manea (2007a) showed that manipulability of the PS mechanism may not be a big problem. If there are sufficiently many copies of the houses (e.g., when "houses" represent "slots at schools" in the school choice domain), then PS will be a strategy-proof mechanism. In such cases, one can claim that PS is a superior mechanism to RSD.²

Abdulkadiroğlu and Sönmez (1998) gave a theoretical intuition in support of the use of RSD. One can imagine another fair mechanism as follows: randomly assign houses to agents and find the core of the resulting housing market (core from random endowments). It turns out that this mechanism is equivalent to RSD through their result. On the other hand, Sönmez and Ünver (2005) showed that in the house allocation with existing tenants domain, randomly endowing newcomers with vacant houses, and finding the core of the resulting housing market in which existing tenants initially own their occupied houses, is equivalent to randomly drawing a priority order of agents in which existing tenants are always ordered after the newcomers and implementing the induced YRMH-IGYT mechanism. Thus, the core idea favors newcomers by giving all rights to vacant houses to newcomers.

Abdulkadiroğlu and Sönmez (2003b) explored why serial dictatorships, Pareto-efficient mechanisms, could result in an ordinally inefficient probability distribution over assigned houses when they are used following a uniformly random priority order drawing (i.e., RSD). They discovered that the probability distribution induced by RSD can also be generated by equivalent lotteries over *inefficient quasi-matchings*. Moreover, they also found a full characterization of ordinally efficient matchings through this intuition.

Kesten (2008c) explored the origins of ordinal inefficiencies under RSD (equivalently core from random endowments) from a different point of view. He discovered that these inefficiencies are not the results of the allocation or trading procedures used, but the deterministic problem definition. That is, if we can allocate or endow agents fractions of houses (equivalent to probabilities) through the algorithms we introduced, then RSD, PS, and Gale's TTC are essentially equivalent.

Katta and Sethuraman (2006) generalized the PS mechanism when indifferences are allowed in preferences. Yilmaz (2008b,c) included individual rationality constraints as in the house allocation

²See Manea (2006) and Che and Kojima (2008) about results on asymptotic ordinal *inefficiency* and efficiency of RSD in different large economies, respectively.

with existing agents domain and introduced a natural generalization of the PS mechanism with and without indifferences in preferences. Athanassoglou and Sethuraman (2007) allowed fractional house endowments in the house allocation domain (i.e., the existing tenants initially own a probability distribution over houses) and found a generalization of Yilmaz's mechanisms.

6.2 Other Literature

We end with a series of citations pointing out new and emerging areas in discrete resources allocation and exchange problems.

First of all, there is an emerging literature on generalizations of the matching problem to different domains which simultaneously include hedonic games, housing markets, two-sided matching problems, etc (see for example Sönmez 1996, 1999, and Pápai 2007).

Additionally, Ben-Shoham, Serrano, and Volij (2004) looked into the evolutionary dynamics that drive decentralized robust exchange in a housing market (for a generalization of this process to multiple house consumption see Bochet, Klaus, and Walzl 2007). Kandori, Serrano, and Volij (2008) inspected a similar decentralized process for housing markets with transfers when there are random and persistent shocks to the preferences of agents.

Recently, Bade (2008) studied rationalizable and non-rationalizable behavior of agents in housing problems and markets.

Market design has recently been the driving force in the advance of theory in discrete resource allocation and exchange problem. Market design applications are not limited to the ones discussed throughout this survey. Guillen and Kesten (2008) discovered that the mechanism used to assign students to rooms in an MIT dormitory is essentially equivalent to a version of the Gale-Shapley student-optimal stable mechanism that takes into consideration individual rationality constraints, and compared YRMH-IGYT and the MIT dormitory allocation mechanisms experimentally. In another market design study, Kesten and Yazici (2008) introduced an ex-post fair "discrete resource" allocation mechanism for possible applications in large corporations and organizations such as the navy or a university. However, in general such an allocation is not efficient. When multiple objects, such as courses, are being distributed to agents, such as students at a university, competitive equilibrium from equal (artificial) budgets is a natural candidate for sustaining ex-post fairness and efficiency together. Since a competitive equilibrium may not exist in general, Sönmez and Ünver (2003) introduced a "course" allocation mechanism based on bidding under equal budgets, which can replace the most popular course bidding mechanism used in many business schools. This bidding mechanism was intended to create competitive equilibrium under equal budgets, but it fails by the impossibility result. Even under a modified definition of competitive equilibrium, this mechanism is not a competitive mechanism, while the Sönmez and Ünver proposal is. Krishna and Ünver (2008) showed that the Sönmez and Ünver (2003) proposal is superior to the current bidding mechanisms in

a designed experimental environment and in a field experiment at University of Michigan Business School. Harvard Business School course bidding mechanism tries to achieve ex-post fairness using a series of serial dictatorships with reversal of priority orders in each round of course allocation. Budish and Cantillon (2009) tested the Harvard Business School course allocation scheme in a field experiment and showed that it is manipulable and causes significant welfare losses. Budish (2009) endogenized competitive prices and bidding using a direct mechanism. He proposed an approximate competitive equilibrium concept and a mechanism which finds such equilibria. The proposed direct mechanism calculates an approximate competitive equilibrium by finding approximately market clearing prices from approximately equal (artificial bid) budgets for students. This equilibrium is also approximately strategy-proof and ex-post envy-free.

There are other experimental studies on matching market design that we did not mention earlier. Calsamiglia, Haeringer, and Klijn (2007) supported the Haeringer and Klijn (2007) theoretical study on constrained school choice with laboratory experiments and complemented the Chen and Sönmez (2006) experimental study on unconstrained school choice. In the marketing literature, Wang and Krishna (2006) made an experimental study of the TTCC mechanism of Roth, Sönmez, and Ünver (2004), which was employed for time-share summer housing exchange.

Dynamic models of house allocation and exchange have been attracting attention recently: In addition to Ünver (2007), Bloch and Cantala (2008), and Kurino (2008) considered intertemporal house allocation when some agents leave and new agents join the agent population over time. Abdulkadiroğlu and Loertscher (2007) considered dynamic house allocation when the preferences of agents are uncertain.

Bibliography

- [1] Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2008) "Expanding 'Choice' in School Choice." Working paper.
- [2] Abdulkadiroğlu, Atila and Lars Ehlers (2007) "Controlled School Choice." Working paper.
- [3] Abdulkadiroğlu, Atila and Simon Loertscher (2007) "Dynamic House Allocation." Working paper.
- [4] Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2005) "The New York City High School Match." *American Economic Review, Papers and Proceedings*, 95(2), 364-367.
- [5] Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2008) "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review*, forthcoming.
- [6] Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005) "The Boston Public School Match." *American Economic Review, Papers and Proceedings*, 95(2), 368-371.
- [7] Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2006) "Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access." Working paper.
- [8] Abdulkadiroğlu, Atila and Tayfun Sönmez (1998) "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems." *Econometrica* 66, 689-701.
- [9] Abdulkadiroğlu, Atila and Tayfun Sönmez (1999) "House Allocation with Existing Tenants." *Journal of Economic Theory*, 88, 233-260.
- [10] Abdulkadiroğlu, Atila and Tayfun Sönmez (2003a) "School Choice: A Mechanism Design Approach." *American Economic Review*, 93, 729-747.
- [11] Abdulkadiroğlu, Atila and Tayfun Sönmez (2003b) "Ordinal Efficiency and Dominated Sets of Assignments." *Journal of Economic Theory*, 112, 157-172.

- [12] Abraham, David J., Katarína Ceclárová, David F. Manlove, and Kurt Mehlhorn (2005) "Pareto Optimality in House Allocation Problems." in *Lecture Notes in Computer Science*, Volume 3827, Springer, pp. 1163-1175.
- [13] Abraham, David J., Avrim Blum, and Tuomas Sandholm (2007) "Clearing Algorithms for Barter Exchange Markets: Enabling Nationwide Kidney Exchanges." in *Proceedings of ACM-EC 2007: the Eighth ACM Conference on Electronic Commerce*.
- [14] Alcalde, José and Salvador Barberà (1994) "Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems." *Economic Theory*, 4, 417-435.
- [15] Athanassoglou, Stergios and Jay Sethuraman (2007) "House Allocation with Fractional Endowments." Working paper.
- [16] Bade, Sophie (2008) "Housing Problems with Non-Rationalizable Behavior." Working paper.
- [17] Balinski, Michel and Tayfun Sönmez (1999) "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory*, 84, 73-94.
- [18] Ben-Shoham, A., Roberto Serrano, and Oscar Volij (2004) "The Evolution of Exchange." *Journal of Economic Theory*, 114, 310-328.
- [19] Bevia, Carmen, Martine Quinzii, and Jose A. Silva (1999) "Buying Several Indivisible Goods." *Mathematical Social Sciences*, 37, 1:25.
- [20] Biró, Peter, and Katarína Ceclárová (2007) "Inapproximability of the kidney exchange problem." *Information Processing Letters*, 101, 199-202.
- [21] Biró, Peter and Eric McDermid (2008) "Three-sided Stable Matchings with Cyclic Preferences and the Kidney Exchange." Ulle Endriss and Paul W. Goldberg (eds.) *COMSOC-2008: Proceedings of the 2nd International Workshop on Computational Social Choice*, pp. 97-108.
- [22] Bloch, Francis, and David Cantal (2008) "Markovian Assignment Rules." Working paper.
- [23] Bochet, Olivier, Bettina Klaus, and Markus Walzl (2007) "Dynamic Recontracting Processes with Multiple Indivisible Goods." Working paper.
- [24] Bogomolnaia, Anna, Rajat Deb, and Lars Ehlers (2005) "Strategy-Proof Assignment on the Full Preference Domain." *Journal of Economic Theory*, 123, 161-186.
- [25] Bogomolnaia, Anna and Hervé Moulin (2001) "A New Solution to the Random Assignment Problem." *Journal of Economic Theory*, 100, 295-328.

- [26] Bogomolnaia, Anna and Hervé Moulin (2002) "A Simple Random Assignment Problem with a Unique Solution." *Economic Theory*, 19, 623-635.
- [27] Bogomolnaia, Anna and Hervé Moulin (2004) "Random Matching under Dichotomous Preferences." *Econometrica*, 72, 257-279.
- [28] Budish, Eric (2009) "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes." Working Paper.
- [29] Budish, Eric and Estelle Cantillon (2009) "Strategic Behavior in Multi-Unit Assignment Problems: Lessons for Market Design." Working Paper.
- [30] Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn (2007) "On the Robustness of Recombinant Estimation: Efficiency in School Choice." Working Paper.
- [31] Cechlárová, Katarína, Tamás Fleiner, and David F. Manlove (2005) "The kidney exchange game." Proceedings of SOR'05: the 8th International Symposium on Operations Research in Slovenia, pp. 77–83.
- [32] Chambers, Christopher P. (2004) "Consistency in the Probabilistic Assignment Model." *Journal of Mathematical Economics*, 40, 953-962.
- [33] Che, Yeon-Koo and Fuhito Kojima (2008) "Asymptotic Equivalence of Random Priority and Probabilistic Serial Mechanisms." *Econometrica*, forthcoming.
- [34] Chen, Yan and Tayfun Sönmez (2002) "Improving Efficiency of On-Campus Housing: An Experimental Study." *American Economic Review*, 92, 1669-1686.
- [35] Chen, Yan and Tayfun Sönmez (2006) "School Choice: An Experimental Study" *Journal of Economic Theory*, 127, 2002-2031.
- [36] Crès, Hervé and Hervé Moulin (2001) "Scheduling with Opting Out: Improving upon Random Priority." *Operations Research*, 49, 565 - 577.
- [37] Delmonico, Francis L. (2004) "Exchanging Kidneys—Advances in Living-Donor Transplantation." *New England Journal of Medicine*, 350, 1812–1814.
- [38] Dubins, Lester E. and D. A. Freedman (1981) "Machiavelli and the Gale-Shapley Algorithm." *American Mathematical Monthly*, 88, 485-494.
- [39] Dutta, Bhaskar and Debraj Ray (1989) "A Concept of Egalitarianism under Participation Constraints." *Econometrica*, 57, 615-635.

- [40] Edmonds, Jack (1965) "Paths, Trees, and Flowers." *Canadian Journal of Mathematics*, 17, 449–467.
- [41] Ehlers, Lars (2002) "Coalitional Strategy-Proof House Allocation." *Journal of Economic Theory*, 105, 298–317.
- [42] Ehlers, Lars (2006) "Respecting Priorities when Assigning Students to Schools." Working Paper.
- [43] Ehlers, Lars and Bettina Klaus (2006) "Efficient Priority Rules." *Games and Economic Behavior*, 55, 372–384.
- [44] Ehlers, Lars and Bettina Klaus (2007) "Consistent House Allocation." *Economic Theory*, 30, 561–574.
- [45] Ehlers, Lars, Bettina Klaus, Szilvia Pápai (2002) "Strategy-Proofness and Population-Monotonicity for House Allocation Problems." *Journal of Mathematical Economics*, 38, 329–339.
- [46] Erdil, Aytok and Haluk Ergin (2008) "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice." *American Economic Review*, 98, 669–689.
- [47] Ergin, Haluk (2000) "Consistency in House Allocation Problems." *Journal of Mathematical Economics*, 34, 77–97.
- [48] Ergin, Haluk (2002) "Efficient Resource Allocation on the Basis of Priorities." *Econometrica*, 70, 2489–2497.
- [49] Ergin, Haluk and Tayfun Sönmez (2006) "Games of School Choice under the Boston Mechanism." *Journal of Public Economics*, 90, 215–237.
- [50] Featherstone, Clayton and Muriel Niederle (2008) "Manipulation in School Choice Mechanisms." Working paper.
- [51] Gale, David and Lloyd Shapley (1962) "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69, 9–15.
- [52] Gallai, Tibor (1963) "Kritische Graphen II." *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 8, 373–395.
- [53] Gallai, Tibor (1964) "Maximale Systeme unabhängiger kanten." *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 9, 401–413.
- [54] Gjertson, David W. and J. Michael Cecka (2000) "Living Unrelated Donor Kidney Transplantation." *Kidney International*, 58, 491–499.

- [55] Guillen, Pablo and Onur Kesten (2008) "On-Campus Housing: Theory and Experiment." Working paper.
- [56] Gusfield, Dan and Robert W. Irving (1989) *The Stable Marriage Problem: Structure and Algorithms*. MIT Press.
- [57] Haeringer, Guillaume and Flip Klijn (2007) "Constrained School Choice." Working paper.
- [58] Hatfield, John W. (2005) "Pairwise Kidney Exchange: Comment." *Journal of Economic Theory*, 125, 189-193.
- [59] Hatfield, John W. and Paul R. Milgrom (2005) "Matching with Contracts." *American Economic Review*, 95, 913-935.
- [60] Hylland, Aanund, and Richard Zeckhauser (1979) "The Efficient Allocation of Individuals to Positions." *Journal of Political Economy*, 87, 293-314.
- [61] Irving, Robert W. (2007) "The cycle roommates problem: a hard case of kidney exchange." *Information Processing Letters*, 103, 1-4.
- [62] Kandori, Michihiro, Roberto Serrano, and Oscar Volij (2008) "Decentralized trade, random utility and the evolution of social welfare." *Journal of Economic Theory*, 140, 328-338.
- [63] Katta, Akshay-Kumar and Jay Sethuraman (2006) "A Solution to the Random Assignment Problem on the Full Preference Domain." *Journal of Economic Theory*, 131, 231-250.
- [64] Kelso, Alexander S. and Vincent P. Crawford (1982) "Job Matchings, Coalition Formation, and Gross Substitutes." *Econometrica*, 50, 1483-1504.
- [65] Kesten, Onur (2008a) "An Alternative Mechanism Design Approach to School Choice in the United States." Working paper.
- [66] Kesten, Onur (2008b) "Coalitional Strategy-Proofness and Resource Monotonicity for House Allocation Problems." *International Journal of Game Theory*, forthcoming.
- [67] Kesten, Onur (2008c) "Why Do Popular Mechanisms Lack Efficiency in Random Environments." Working paper.
- [68] Kesten, Onur and Ayse Yazici (2008) "A Mechanism Design Approach to a Common Distributional Problem of Centrally Administered Organizations." Working paper.
- [69] Kesten, Onur and M. Utku Unver (2008) "A Theory of School Choice Lotteries: Why Ties Should Not Be Broken Randomly." Working paper.

- [70] Klaus, Bettina (2008) "The Coordinate-Wise Core for Multiple-Type Housing Markets is Second-Best Incentive Compatible." *Journal of Mathematical Economics*, 44, 919-924.
- [71] Klemperer, Paul (2004) *Auctions: Theory and Practice*. Princeton University Press.
- [72] Kojima, Fuhito and Mihai Manea (2007a) "Strategy-Proofness of the Probabilistic Serial Mechanism in Large Random Assignment Problems." Working paper.
- [73] Kojima, Fuhito and Mihai Manea (2007b) "Axioms for Deferred Acceptance." *Econometrica*, forthcoming.
- [74] Konishi, Hideo, Thomas Quint, and Jun Wako (2001) "On the Shapley-Scarf Economy: The Case of Multiple Types of Indivisible Goods." *Journal of Mathematical Economics*, 35, 1-15.
- [75] Korte, Bernhard and Jens Vygen (2002) *Combinatorial Optimization: Theory and Algorithms*, 2nd ed., Springer.
- [76] Krishna, Aradhna and M. Utku Ünver (2008) "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies." *Marketing Science*, 27, 262-282.
- [77] Krishna, Aradhna and Yu Wang (2007) "The Relationship between Top Trading Cycles Mechanism and Top Trading Cycles and Chains Mechanism." *Journal of Economic Theory*, 132, 539-547.
- [78] Kurino, Morimitsu (2008) "House Allocation with Overlapping Agents: A Dynamic Mechanism Design Approach." Working paper.
- [79] Ma, Jinpeng (1994) "Strategy-Proofness and the Strict Core in a Market with Indivisibilities." *International Journal of Game Theory*, 23, 75-83.
- [80] Manea, Mihai (2006) "Asymptotic Ordinal Inefficiency of Random Serial Dictatorship." Working paper.
- [81] Milgrom, Paul (2000) "Putting Auction Theory to Work: The Simultaneous Ascending Auction." *Journal of Political Economy*, 108, 245-272.
- [82] Milgrom, Paul R. (2004) *Putting Auction Theory to Work*. Cambridge University Press.
- [83] Milgrom, Paul R. (2007) "Package Auctions and Package Exchanges." *Econometrica*, 75: 935-966.
- [84] Miyagawa, Eiichi (2002) "Strategy-Proofness and the Core in House Allocation Problems." *Games and Economic Behavior*, 38, 347-361.

- [85] Opelz, Gerhard (1997) "Impact of HLA Compatibility on Survival of Kidney Transplants from Unrelated Live Donors." *Transplantation*, 64, 1473–1475.
- [86] Pais, Joana and Ágnes Pintér (2007) "School Choice and Information: An Experimental Study on Matching Mechanisms." *Games and Economic Behavior*, forthcoming.
- [87] Pápai, Szilvia (2000) "Strategyproof Assignment by Hierarchical Exchange." *Econometrica*, 68, 1403-1433.
- [88] Pápai, Szilvia (2001) "Strategyproof and Nonbossy Multiple Assignments." *Journal of Public Economic Theory*, 3, 257-271.
- [89] Pápai, Szilvia (2003) "Strategyproof Exchange of Indivisible Goods." *Journal of Mathematical Economics*, 39: 931-959.
- [90] Pápai, Szilvia (2007) "Exchange in a General Market with Indivisible Goods." *Journal of Economic Theory*, 132, 208-235.
- [91] Pathak, Parag A. (2006) "Lotteries in Student Assignment." Working paper.
- [92] Pathak, Parag A. and Tayfun Sönmez (2008) "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*, 98, 1636-1652.
- [93] Pycia, Marek and M. Utku Ünver (2007) "A Theory of House Allocation and Exchange Mechanisms." Working paper.
- [94] Quinzii, Martine (1984) "Core and Competitive Equilibria with Indivisibilities." *International Journal of Game Theory*, 13, 41:60.
- [95] Rawls, John (1971) *A Theory of Justice*, Harvard University Press, Cambridge.
- [96] Roth, Alvin E. (1982a) "Incentive Compatibility in a Market with Indivisibilities." *Economics Letters*, 9, 127-132.
- [97] Roth, Alvin E. (1982b) "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research* 7, 617-628.
- [98] Roth, Alvin E. (1984) "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92, 991-1016.
- [99] Roth, Alvin E. (1985) "The College Admissions Problem is not Equivalent to the Marriage Problem." *Journal of Economic Theory*, 36, 277-288.

- [100] Roth, Alvin E. (1991) "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom." *American Economic Review*, 81: 415-40.
- [101] Roth, Alvin E. (2002) "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica*, 70, 1341-1378.
- [102] Roth, Alvin E. (2008a) "Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions." *International Journal of Game Theory*, 36, 537-569.
- [103] Roth, Alvin E. (2008b) "What have we learned from market design?" *Economic Journal*, 118: 285-310.
- [104] Roth, Alvin E. and Elliott Peranson (1999) "The Redesign of the Matching Markets for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89, 748-780.
- [105] Roth, Alvin E. and Andrew Postlewaite (1977) "Weak versus Strong Domination in a Market with Indivisible Goods." *Journal of Mathematical Economics*, 4, 131-137.
- [106] Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2004) "Kidney Exchange." *Quarterly Journal of Economics*, 119, 457-488.
- [107] Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005a) "Pairwise Kidney Exchange." *Journal of Economic Theory*, 125, 151-188.
- [108] Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2005b) "A Kidney Exchange Clearinghouse in New England." *American Economic Review Papers and Proceedings*, 95(2): 376-380.
- [109] Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2007) "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences." *American Economic Review*, 97(3), 828-851.
- [110] Roth, Alvin E., Tayfun Sönmez, M. Utku Ünver, Francis L. Delmonico, and Susan L. Saidman (2006) "Utilizing List Exchange and Nondirected Donation through 'Chain' Paired Kidney Donations." *American Journal of Transportation*, 6, 2694-2705.
- [111] Roth, Alvin E. and Marilda Sotomayor (1990) *Two-Sided Matching: A Study on Game-Theoretic Modelling*. Cambridge University Press.
- [112] Satterthwaite, Mark A. and Hugo Sonnenschein (1981) "Strategy-Proof Allocation Mechanisms at Differentiable Points." *Review of Economic Studies*, 48, 587-597.

- [113] Segev, Dorry, Sommer Gentry, Daniel S. Warren, Brigitte Reeb, and Robert A. Montgomery (2005) "Kidney paired donation: Optimizing the use of live donor organs." *Journal of the American Medical Association*, 293: 1883-1890.
- [114] Shapley, Lloyd and Herbert Scarf (1974) "On Cores and Indivisibility." *Journal of Mathematical Economics*. 1, 23-28.
- [115] Shapley, Lloyd and Martin Shubik (1972) "The Assignment Game I: The Core." *International Journal of Game Theory*, 1, 111-130.
- [116] Sönmez, Tayfun (1996) "Implementation in Generalized Matching Problems." *Journal of Mathematical Economics*, 26, 429-439.
- [117] Sönmez, Tayfun (1999) "Strategy-Proofness and Essentially Single-Valued Cores." *Econometrica*, 67, 677-690.
- [118] Sönmez, Tayfun and M. Utku Ünver (2003) "Course Bidding at Business Schools." *International Economic Review*, forthcoming.
- [119] Sönmez, Tayfun and M. Utku Ünver (2005) "House Allocation with Existing Tenants: An Equivalence." *Games and Economic Behavior*, 52, 153-185.
- [120] Sönmez, Tayfun and M. Utku Ünver (2006) "Kidney Exchange with Good Samaritan Donors: A Characterization." Working paper. Revised as (2008) "House Allocation with Existing Tenants: A Characterization." Working paper.
- [121] Svensson, Lars-Gunnar (1994) "Queue allocation of indivisible goods." *Social Choice and Welfare*, 11, 323-330.
- [122] Svensson, Lars-Gunnar (1999) "Strategyproof Allocation of Indivisible Goods." *Social Choice and Welfare*, 16, 557-567.
- [123] Ünver, M. Utku (2007) "Dynamic Exchange Mechanisms." *Review of Economic Studies*, forthcoming.
- [124] Velez, Rodrigo (2008) "Revisiting Consistency in House Allocation Problems." Working paper.
- [125] Vickrey, William (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders." *Journal of Finance*, 16, 8-37.
- [126] Wako, Jun (2005) "Coalition-Proof Nash Allocation in a Barter Game with Multiple Indivisible Goods." *Mathematical Social Sciences*, 49, 179-199.

- [127] Wang, Yu and Aradhna Krishna (2006) "Timeshare Exchange Mechanisms." *Management Science*, 52(8), 1223-1237.
- [128] Wilson, Robert (2002) "Architecture of Power Markets." *Econometrica*, 70, 1299-1340.
- [129] Yilmaz, Özgür (2008a) "Kidney Exchange: An Egalitarian Mechanism." Working paper.
- [130] Yilmaz, Özgür (2008b) "House Allocation with Existing Tenants: A New Solution." Working paper.
- [131] Yilmaz, Özgür (2008c) "Random Assignment under Weak Preferences." *Games and Economic Behavior*, forthcoming.
- [132] Zenios, Stefanos A. (2002) "Optimal Control of a Paired-Kidney Exchange Program." *Management Science*, 48, 328-342.
- [133] Zenios, Stefanos A., E. Steve Woodle, Lainie Friedman Ross (2001) "Primum non nocere: avoiding increased waiting times for individual racial and blood-type subsets of kidney wait list candidates in a living donor/cadaveric donor exchange program." *Transplantation*, 72, 648-654.