Matching Markets: Theory and Practice*

Atila Abdulkadiroğlu[†] Tayfun Sönmez[‡]
Duke University *Boston College*

1 Introduction

It has been almost half a century since David Gale and Lloyd Shapley published their pathbreaking paper "College admissions and the stability of marriage" in *American Mathematical Monthly*. It is hard to know whether Gale and Shapley expected the literature they initiated to be used to improve lives of masses of people all around the world. We are very fortunate to see that this is happening today.

The model Gale and Shapley presented is very simple. A number of boys and girls have preferences for each other and would like to be matched. The question Gale and Shapley were especially interested in was whether there is a "stable" way to match each boy with a girl so that no unmatched pair can later find out that they can both do better by matching each other. They found that there indeed is such a stable matching, and they presented a deferred acceptance algorithm that achieves this objective. Versions of this algorithm are used today to match hospitals with residents and students with public schools in New York City and Boston.

In 1974, Lloyd Shapley and Herbert Scarf published a somewhat related paper, "On Cores and Indivisibility," in the first issue of the *Journal of Mathematical Economics*. Their model was arguably the simplest exchange economy one could think of. Each agent comes to the market with one indivisible good and seeks to trade it for possibly more preferred ones that might be brought by other agents. In their simple model agents are restricted to consume only one good. They were interested in whether a core allocation exists in their model. They showed that there indeed

*We thank Yeon-Koo Che and Eddie Dekel for their extensive comments that improved this survey.
[†]Department of Economics, Duke University, 213 Social Sciences Bldg., Durham NC, 27708.
[‡]Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467.

exists an allocation in the core, and they presented a Top Trading Cycles algorithm, which they attribute to David Gale, that achieves this objective. The basic ideas of Gale's Top Trading Cycles algorithm and its extensions resulted in organized kidney exchange in various parts of the world almost thirty years later, saving thousands of lives.

There are a number of well-written surveys on matching markets. The best-known of these by Roth and Sotomayor (1990), covers the literature on two-sided matching markets until 1990. More recently, Roth (2008) focuses on the history of the deferred acceptance algorithm. Both these surveys are essentially of two-sided matching markets. The content of our survey is closest to that of Sönmez and Ünver (2010), which also focuses on one-sided matching as well as two-sided matching. In this survey we pay particular attention to the formal relations and links between these two original contributions and the recent matching literature that has had a policy impact on various areas.

Here is a brief overview of our survey. In Section 2 we introduce and briefly review some of the key results of the two-sided matching model by Gale and Shapley (1962). In Section 3 we introduce the "housing market" model by Shapley and Scarf (1974) as well as a number of more recent one-sided matching models, some of which are closely related to two-sided matching models. In Section 4 we present the recent developments in School Choice, and in Section 5 we present the recent developments in Kidney Exchange. We conclude in Section 6.

2 Two-Sided Matching

One of the key observations relating Gale and Shapley's seminal contribution to real-life applications was made by Alvin Roth in 1984. Roth (1984) showed that the algorithm that was used to match medical residents to hospitals since 1950s by the National Resident Matching Program is equivalent to a version of the celebrated deferred acceptance algorithm. Since then, similar equivalences have been made by several authors. In this section we give a brief summary of the two-sided matching literature. We focus on discrete two-sided matching models without money. There is an important literature studying versions of these models with money. Shapley and Shubik (1972) wrote the first paper to consider "continuous" matching markets widely knows as "assignment games." This model was later studied by Crawford and Knoer (1981) and Kelso and Crawford (1982), showing many parallels with the discrete model. More recently, Hatfield and Milgrom (2005) provide a unified framework for the discrete and continuous models.

2.1 One-to-One Matching: Marriage Problems

A **marriage problem** (Gale and Shapley 1962) is a triple $\langle M, W, \succsim \rangle$ where M is a finite set of men, W is a finite set of women, and $\succsim = (\succsim_i)_{i \in M \cup W}$ is a list of preferences. Here $\succsim_m$ denotes the preference relation of man m over $W \cup \{m\}$, $\succsim_w$ denotes the preference relation of woman w over $M \cup \{w\}$, and $\succ_i$ denotes the strict preferences derived from $\succsim_i$ for agent $i \in M \cup W$.

Consider man m:

- $w \succ_m w'$ means that man m prefers woman w to woman w'
- $w \succ_m m$ means that man m prefers woman w to remaining single, and
- $m \succ_m w$ means that woman w is unacceptable to man m

We use similar notation for women.

Assumption 2.1 *Unless otherwise mentioned all preferences are* strict.

The outcome of a marriage problem is a **matching**. Formally, a matching is a function $\mu : M \cup W \to M \cup W$ such that:

1. $\mu(m) \notin W \Rightarrow \mu(m) = m \quad$ for all $m \in M$,
2. $\mu(w) \notin M \Rightarrow \mu(w) = w \quad$ for all $w \in W$, and
3. $\mu(m) = w \Leftrightarrow \mu(w) = m \quad$ for all $m \in M$, $w \in W$.

Here $\mu(i) = i$ means that agent i remains single under matching μ.

Assumption 2.2 *There are no consumption externalities: An individual i prefers a matching μ to a matching ν if and only if he/she prefers $\mu(i)$ to $\nu(i)$.*

A matching μ is **Pareto efficient** if there is no other matching ν such that $\nu(i) \succsim_i \mu(i)$ for all $i \in M \cup W$ and $\nu(i) \succ_i \mu(i)$ for some $i \in M \cup W$.

A matching μ is **blocked by an individual** $i \in M \cup W$ if $i \succ_i \mu(i)$. A matching is **individually rational** if it is not blocked by any individual. A matching μ is **blocked by a pair** $(m, w) \in M \times W$ if they both prefer each other to their partners under μ, i.e.

$$w \succ_m \mu(m) \text{ and } m \succ_w \mu(w).$$

A matching is **stable** if it is not blocked by any individual or a pair. The next result follows immediately by definition.

Proposition 2.1 Stability implies Pareto efficiency.

The following algorithm and its versions played a central role for almost 50 years in not only matching theory but also its applications in real-life matching markets.

Men-Proposing Deferred Acceptance Algorithm

Step 1. Each man m proposes to his first choice (if he has any acceptable choices). Each woman rejects any offer except the best acceptable proposal and "holds" the most-preferred acceptable proposal (if any).

In general, at

Step k. Any man who was rejected at step $k-1$ makes a new proposal to his most-preferred acceptable potential mate who has not yet rejected him. (If no acceptable choices remain, he makes no proposal.) Each woman "holds" her most-preferred acceptable proposal to date, and rejects the rest.

The algorithm terminates when there are no more rejections. Each woman is matched with the man she has been holding in the last step. Any woman who has not been holding an offer or any man who was rejected by all acceptable woman remains single.

Theorem 2.1 (Theorems 1,2 in Gale and Shapley 1962): The men-proposing deferred acceptance algorithm gives a stable matching for each marriage problem. Moreover, every man weakly prefers this matching to any other stable matching.

Hence we refer to the outcome of the men-proposing deferred acceptance algorithm as the **man-optimal stable matching** and denote its outcome by μ^M. The algorithm where the roles of men and women are reversed is known as the **women-proposing deferred acceptance algorithm** and we refer to its outcome μ^W as the **woman-optimal stable matching**.

Theorem 2.2 (McVitie and Wilson 1970): The set of agents who are matched is the same for all stable matchings.

Let μ, μ' be two stable matchings. The function $\mu \vee^M \mu' : M \cup W \to M \cup W$ (**join** of μ and μ') assigns each man the more preferred of his two assignments under μ and μ' and each woman the less preferred of her two assignments under μ and μ'. That is, for any man m and woman w:

$$\mu \vee^M \mu'(m) = \begin{cases} \mu(m) & \text{if } \mu(m) \succsim_m \mu'(m) \\ \mu'(m) & \text{if } \mu'(m) \succsim_m \mu(m) \end{cases}$$

$$\mu \vee^M \mu'(w) = \begin{cases} \mu(w) & \text{if } \mu'(w) \succsim_w \mu(w) \\ \mu'(w) & \text{if } \mu(w) \succsim_w \mu'(w) \end{cases}$$

Define the function $\mu \wedge^M \mu' : M \cup W \to M \cup W$ (**meet** of μ and μ') similarly, by reversing the preferences.

Given a pair of arbitrary matchings, neither the join nor the meet needs to be a matching. However, for a pair of stable matchings, not only are meet and join both matchings, they are also stable. The following result in Knuth (1976) is attributed to John Conway.

Theorem 2.3 (Conway): If μ and μ' are stable matchings, then not only are the functions $\mu \vee^M \mu'$ and $\mu \wedge^M \mu'$ both matchings, they are also both stable.

Of particular interest, is the following corollary.

Corollary 2.1 Every man weakly prefers any stable matching to woman-optimal stable matching.

If we can match a man with a woman who finds him unacceptable, then there may be a matching where all man receive better mates than under the man-optimal stable matching. If, however, we are seeking an individually rational matching while some man can receive better mates without hurting any man, it is not possible to match all man with strictly more-preferred mates.

Theorem 2.4 (Theorem 6 in Roth 1982): There is no individually rational matching ν where $\nu(m) \succ_m \mu^M(m)$ for all $m \in M$.

The next example by Roth (1982) shows that some of the men can receive more-preferred mates than under the man-optimal stable matching.

Example 2.1 There are 3 men and 3 women with the following preferences:

$$\begin{array}{llcll}
\succ_{m_1}: & w_1\ w_2\ w_3\ m_1 & \qquad & \succ_{w_1}: & m_2\ m_1\ m_3\ w_1 \\
\succ_{m_2}: & w_2\ w_1\ w_3\ m_2 & & \succ_{w_2}: & m_1\ m_3\ m_2\ w_2 \\
\succ_{m_3}: & w_1\ w_2\ w_3\ m_3 & & \succ_{w_3}: & m_1\ m_2\ m_3\ w_3
\end{array}$$

Here both m_1 and m_2 prefer matching ν to man-optimal stable matching μ^M where

$$\mu^M = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \end{pmatrix} \quad \textit{and} \quad \nu = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

A matching μ is in the **core** if there exists no matching ν and coalition $T \subseteq M \cup W$ such that $\nu(i) \succ_i \mu(i)$ and $\nu(i) \in T$ for any $i \in T$. The next result follows directly from definitions.

Proposition 2.2 The set of stable matchings is equal to the core.

2.2 One-to-One Matching: Incentives

Throughout this subsection we fix M, W so that each preference profile $\succsim$ defines a marriage problem.

Let $\mathcal{R}_i$ denote the set of all preference relations for agent i , $\mathcal{R} = \mathcal{R}_{m_1} \times \cdots \times \mathcal{R}_{m_p} \times \mathcal{R}_{w_1} \times \cdots \times \mathcal{R}_{m_q}$ denote the set of all preference profiles, and $\mathcal{R}_{-i}$ denote the set of all preference profiles for all agents except agent i. Let $\mathcal{M}$ denote the set of all matchings.

A **(direct) mechanism** is a systematic procedure that determines a matching for each marriage problem. Formally, it is a function $\varphi : \mathcal{R} \to \mathcal{M}$.

A mechanism φ is **stable** if $\varphi(\succsim)$ is stable for any $\succsim \in \mathcal{R}$. Similarly a mechanism is **Pareto efficient** if it always selects a Pareto efficient matching, and it is **individually rational** if it always selects an individually rational matching. Clearly any stable mechanism is both Pareto efficient and also individually rational.

Let ϕ^M be the mechanism that selects the man-optimal stable matching for each problem and ϕ^W be the mechanism that selects the woman-optimal stable matching for each problem.

Each mechanism φ induces a preference revelation game for each problem where the set of players is $M \cup W$, the strategy space for player i is the set of his/her preferences $\mathcal{R}_i$, and the outcome is determined by the mechanism φ. A mechanism is **strategy-proof** if truthful preference revelation (or simply truth-telling) is a weakly dominant strategy equilibrium of the induced preference revelation game.

Formally, a mechanism φ is strategy-proof if

$$\forall i \in M \cup W,\ \forall \succsim_i, \succsim_i' \in \mathcal{R}_i,\ \forall \succsim_{-i} \in \mathcal{R}_{-i} \qquad \varphi[\succsim_{-i}, \succsim_i](i) \ \succsim_i \ \varphi[\succsim_{-i}, \succsim_i'](i)$$

While strategy-proofness is a very plausible requirement, it is not compatible with stability.

Theorem 2.5 (Theorem 3 in Roth 1982): There exists no mechanism that is both stable and strategy-proof.

The following simple example is enough to prove this impossibility result.

Example 2.2 Consider the following 2-men, 2-women problem with the following preferences.

$$\begin{array}{llll} \succsim_{m_1}: & w_1\ w_2\ m_1 & \qquad \succsim_{w_1}: & m_2\ m_1\ w_1 \\ \succsim_{m_2}: & w_2\ w_1\ m_2 & \qquad \succsim_{w_2}: & m_1\ m_2\ w_2 \end{array}$$

In this problem there are only two stable matchings:

$$\mu^M = \begin{pmatrix} m_1 & m_2 \\ w_1 & w_2 \end{pmatrix} \quad \text{and} \quad \mu^W = \begin{pmatrix} m_1 & m_2 \\ w_2 & w_1 \end{pmatrix}$$

Let φ be any stable mechanism. Then $\varphi[\succsim] = \mu^M$ or $\varphi[\succsim] = \mu^W$.

If $\varphi[\succsim] = \mu^M$ then woman w_1 can report a fake preference $\succsim'_{w_1}$ where only her top choice m_2 is acceptable and force her favorite stable matching μ^W to be selected by φ since it is the only stable matching for the manipulated economy $(\succsim_{-w_1}, \succsim'_{w_1})$.

If, on the other hand, $\varphi[\succsim] = \mu^W$, then man m_1 can report a fake preference $\succsim'_{m_1}$ where only his top choice w_1 is acceptable and force his favorite stable matching μ^M to be selected by φ since it is the only stable matching for the manipulated economy $(\succsim_{-m_1}, \succsim'_{m_1})$.

Indeed, strategy-proofness is not only incompatible with stability but also with Pareto efficiency and individual rationality.

Theorem 2.6 (Proposition 1 in Alcalde and Barbera 1994): There exists no mechanism that is Pareto efficient, individually rational, and strategy-proof.

On the positive side, stability is compatible with truth-telling in marriage problems for only one side of the market.

Theorem 2.7 (Theorem 9 in Dubins and Freedman 1981, Theorem 5 in Roth 1982): Truth-telling is a weakly dominant strategy for any man under the man-optimal stable mechanism. Similarly truth-telling is a weakly dominant strategy for any woman under the woman-optimal stable mechanism.

For any man, any strategy that agrees with truth-telling for the set of acceptable women as well as their relative ranking is also a weakly dominant strategy. We consider any such strategy also as truth-telling (since the relative ranking of unacceptable women is irrelevant under any individually rational mechanism). Any other strategy is a weakly dominated strategy. Clearly an agent can play a weakly dominated strategy at a Nash equilibrium of a game. A Nash equilibrium where no agent plays a weakly dominated strategy is called a **Nash equilibrium in undominated strategies.**

Theorem 2.8 (Theorem 1 in Roth 1984, Theorem 2 in Gale and Sotomayor 1985): Fix a marriage problem $\succsim$ and consider the preference revelation game induced by the man-optimal stable mechanism ϕ^M. A matching is stable under $\succsim$ if and only if it is a Nash equilibrium outcome of ϕ^M in undominated strategies.

2.3 Many-to-One Matching: College Admissions

A **college admissions problem** (Gale and Shapley 1962) is a four-tuple $\langle C, I, q, \succsim\rangle$ where C is a finite set of colleges, I is finite a set of students, $q = (q_c)_{c \in C}$ is a vector of college capacities, and $\succsim = (\succsim_\ell)_{\ell \in C \cup I}$ is a list of preferences. Here $\succsim_i$ denotes the preferences of student i over $C \cup \{\emptyset\}$, $\succsim_c$ denotes the preferences of college c over 2^I, and $\succ_c, \succ_i$ denote strict preferences derived from $\succsim_c, \succsim_i$.

Throughout this section we assume that whether a student is acceptable for a college or not does not depend on other students in her class. Similarly, we assume that the relative desirability of students does not depend on the composition of the class. This latter property is known as **responsiveness** (Roth 1985).

Formally, college preferences $\succsim_c$ are responsive iff

1. for any $J \subset I$ with $|J| < q_c$ and any $i \in I \setminus J$,
$$(J \cup \{i\}) \succ_c J \quad \Leftrightarrow \quad \{i\} \succ_c \emptyset,$$
2. for any $J \subset I$ with $|J| < q_c$ and any $i, j \in I \setminus J$,
$$(J \cup \{i\}) \succ_c (J \cup \{j\}\}) \quad \Leftrightarrow \quad \{i\} \succ_c \{j\}.$$

Notions of a matching, individual rationality, and stability naturally extend to college admissions. A matching for college admissions is a correspondence $\mu : C \cup I \Longrightarrow 2^{C \cup I}$ such that:

1. $\mu(c) \subseteq I$ such that $|\mu(c)| \leq q_c \quad$ for all $c \in C$,
2. $\mu(i) \subseteq C$ such that $|\mu(i)| \leq 1 \quad$ for all $i \in I$, and
3. $i \in \mu(c)$ if and only if $\mu(i) = \{c\} \quad$ for all $c \in C$ and $i \in I$.

A matching μ is **blocked by a college** $c \in C$ if there exists $i \in \mu(c)$ such that $\emptyset \succ_c i$. A matching μ is **blocked by a student** $i \in I$ if $\emptyset \succ_i \mu(i)$. A matching is **individually rational** if it is not blocked by any college or student. A matching μ is **blocked by a pair** $(c, i) \in C \times I$ if

1. $c \succ_i \mu(i)$, and
2. (a) either there exists $j \in \mu(c)$ such that $\{i\} \succ_c \{j\}$, or
   (b) $|\mu(c)| < q_c$ and $\{i\} \succ_c \emptyset$.

Observe that this version of blocking by a pair is plausible only under responsiveness. A matching is **stable** if it is not blocked by any agent or pair.

The deferred acceptance algorithm naturally extends to college admissions.

College-Proposing Deferred Acceptance Algorithm

Step 1. Each college c proposes to its top q_c acceptable students (and if it has less acceptable choices than q_c, then it proposes to all its acceptable students). Each student rejects any unacceptable proposals and, if more than one acceptable proposal is received, she "holds" the most-preferred and rejects the rest.

In general, at

Step k. Any college c who was rejected at step $k-1$ by any student proposes to its most-preferred q_c acceptable students who have not yet rejected it (and if among the remaining students there are fewer than q_c acceptable students, then it proposes to all). Each students "holds" her most-preferred acceptable offer to date and rejects the rest.

The algorithm terminates when there are no more rejections. Each student is matched with the college she has been holding in the last step.

Theorem 2.1 extends to college admissions as well.

Theorem 2.9 (Theorem 1 in Gale and Shapley 1962): The college-proposing deferred acceptance algorithm gives a stable matching for each college admissions problem.

Indeed, many results for marriage problems extend to college admissions problems. The following "trick" is very useful to extend some of these. Given a college admissions problem $\langle C, I, q, \succsim \rangle$, construct a **related marriage problem** as follows:

- "Divide" each college c_ℓ into q_{c_ℓ} seperate pieces $c_\ell^1, \ldots, c_\ell^{q_{c_\ell}}$ where each piece has a capacity of one; and let each piece have the same preferences over I as college c has. (Since college preferences are responsive, $\succsim_c$ is consistent with a unique ranking of students.)

  C^* : The resulting set of college "pieces" (or seats).

- For any student i, extend her preferences to C^* by replacing each college c_ℓ in her original preferences $\succsim_i$ with the block $c_\ell^1, \ldots, c_\ell^{q_{c_\ell}}$ in that order.

So, in the related marriage problem, each seat of a college c is an individual unit that has the same preferences with college c, and students rank seats at different colleges as they rank the colleges whereas they rank seats at the same college based on the index of the seat.

Given a matching for a college admissions problem, it is straightforward to define a **corresponding matching** for its related marriage problem: Given any college c, assign the students who were assigned to c in the original problem one at a time to pieces of c starting with lower index pieces.

Lemma 2.1 (Lemma 1 in Roth and Sotomayor 1989): A matching of a college admissions problem is stable if and only if the corresponding matching of its related marriage problem is stable.

Lemma 2.1 can be used to extend the following results for marriage problems to college admissions:

1. There exists a **student-optimal stable matching** μ^I that every student likes at least as well as any other stable matching. Furthermore, the outcome of the student-proposing deferred acceptance algorithm yields the student-optimal stable matching.

2. There exists a **college-optimal stable matching** μ^C that every college likes at least as well as any other stable matching. Furthermore, the outcome of the college-proposing deferred acceptance algorithm yields the college-optimal stable matching.

3. The student-optimal stable matching is the worst stable matching for each college. Similarly, the college-optimal stable matching is the worst stable matching for each student.

4. The set of students filled and the set of positions filled is the same at each stable matching.

5. The join as well as the meet of two stable matchings is each a stable matching.

6. There is no individually rational matching ν where $\nu(i) \succ_i \mu^I(i)$ for all $i \in I$.

There are also some new results for college admissions.

Theorem 2.10 (Theorem 1 in Roth 1986): Any college that does not fill all its positions at some stable matching is assigned precisely the same set of students at every stable matching.

Theorem 2.11 (Theorem 4 in Roth and Sotomayor 1989): Let μ and ν be two stable matchings. For any college c,

- *either* $\{i\} \succ_c \{j\}$ *for all* $i \in \mu(c) \setminus \nu(c)$ *and* $j \in \nu(c) \setminus \mu(c)$, *or*
- $\{j\} \succ_c \{i\}$ *for all* $i \in \mu(c) \setminus \nu(c)$ *and* $j \in \nu(c) \setminus \mu(c)$.

In contrast to the marriage problem, there can be an individually rational matching in college admissions where each college receives a strictly better assignment than under the college-optimal stable matching. Loosely speaking, there is an implicit competition between different seats of a college in the sense that a college may lose a student in one of its seats in order to get a more-preferred student for another seat.

Example 2.3 There are 2 colleges c_1, c_2 with $q_{c_1} = 2$, $q_{c_2} = 1$, and 2 students i_1, i_2. The preferences are as follows:

$$\begin{array}{llll} \succsim_{i_1}: & \{c_1\}\ \{c_2\}\ \emptyset & \qquad \succsim_{c_1}: & \{i_1, i_2\}\ \{i_2\}\ \{i_1\}\ \emptyset \\ \succsim_{i_2}: & \{c_2\}\ \{c_1\}\ \emptyset & \qquad \succsim_{c_2}: & \{i_1\}\ \{i_2\}\ \emptyset \end{array}$$

Here both colleges strictly prefer ν to μ^C where:

$$\mu^C = \begin{pmatrix} c_1 & c_2 \\ i_1 & i_2 \end{pmatrix} \quad \text{and} \quad \nu = \begin{pmatrix} c_1 & c_2 \\ i_2 & i_1 \end{pmatrix}.$$

2.3.1 College Admissions: Preference Manipulation

Any impossibility result obtained for a smaller class of problems immediately extends to a larger class. Therefore, the following two results are immediate.

Theorem 2.12 (Theorem 3 in Roth 1982b): There exists no mechanism that is stable and strategy-proof.

Theorem 2.13 (Proposition 1 in Alcalde and Barbera 1994): There exists no mechanism that is Pareto efficient, individually rational, and strategy-proof.

The following positive result is a direct implication of the Stability Lemma and the corresponding positive result for marriage problems:

Theorem 2.14 (Theorem 5 in Roth 1986): Truth-telling is a weakly dominant strategy for all students under the student-optimal stable mechanism.

For colleges, however, the situation is different. The following example makes this point.

Example 2.4 There are 2 colleges c_1, c_2 with $q_{c_1} = 2$, $q_{c_2} = 1$, and 2 students i_1, i_2. The preferences are as follows:

$$\begin{array}{llll} \succsim_{i_1}: & \{c_1\}\ \{c_2\}\ \emptyset & \qquad \succsim_{c_1}: & \{1_1, i_2\}\ \{i_2\}\ \{i_1\}\ \emptyset \\ \succsim_{i_2}: & \{c_2\}\ \{c_1\}\ \emptyset & \qquad \succsim_{c_2}: & \{i_1\}\ \{i_2\}\ \emptyset \end{array}$$

The only stable matching for this problem is:

$$\mu^C = \begin{pmatrix} c_1 & c_2 \\ i_1 & i_2 \end{pmatrix}$$

Now suppose college c_1 submits the manipulated preferences $\succsim'_{c_1}$ where only student i_2 is acceptable. For problem $(\succsim_{-c_1}, \succsim'_{c_1})$ the only stable matching is:

$$\begin{pmatrix} c_1 & c_2 \\ i_2 & i_1 \end{pmatrix}.$$

Hence college c_1 benefits by manipulating its preferences under any stable mechanism (including the college-optimal stable mechanism).

Theorem 2.15 (Theorem 4 in Roth 1985): There exists no stable mechanism where truth-telling is a weakly dominant strategy for all colleges.

2.3.2 College Admissions: Capacity Manipulation

For this section we fix I, C and indicate each college admissions problem with a preference profile-capacity vector pair.

A college c **manipulates mechanism $\boldsymbol{\varphi}$ via capacities** at problem $(\succsim, q)$ if

$$\varphi[\succsim, q_{-c}, q'_c](c) \succ_c \varphi[\succsim, q](c) \qquad\qquad \text{for some } q'_c < q_c.$$

A mechanism is **immune to manipulation via capacities** if it can never be manipulated via capacities.

The following simple example shows that the college-optimal stable mechanism is manipulable via capacities.

Example 2.5 There are 2 colleges c_1, c_2 with $q_{c_1} = 2$, $q_{c_2} = 1$, and 2 students i_1, i_2. The preferences are as follows:

$$\begin{array}{llll} \succsim_{i_1}: & \{c_1\}\ \{c_2\}\ \emptyset & \qquad \succsim_{c_1}: & \{i_1, i_2\}\ \{i_2\}\ \{i_1\}\ \emptyset \\ \succsim_{i_2}: & \{c_2\}\ \{c_1\}\ \emptyset & \qquad \succsim_{c_2}: & \{i_1\}\ \{i_2\}\ \emptyset \end{array}$$

Let $q'_{c_1} = 1$ be a potential capacity manipulation by college c_1. Then:

$$\phi^C[\succsim, q] = \begin{pmatrix} c_1 & c_2 \\ i_1 & i_2 \end{pmatrix}, \qquad \phi^C[\succsim, q'_{c_1}, q_{c_2}] = \begin{pmatrix} c_1 & c_2 \\ i_2 & i_1 \end{pmatrix}.$$

Hence college c_1 benefits by reducing the number of its positions under ϕ^C.

Indeed, there is no stable mechanism that escapes capacity manipulation.

Theorem 2.16 (Theorem 1 in Sönmez 1997): Suppose there are at least 2 colleges and 3 students. Then there exists no stable mechanism that is immune to manipulation via capacities.

If, however, colleges prefer larger groups of students to smaller groups of students, then a positive result is obtained. College preferences are **strongly monotonic** if

$$\forall c \in C, \ \ \forall\, J, J' \subset I \qquad\qquad |J'| < |J| \leq |q_c| \Rightarrow J \ \succ_c \ J'.$$

Theorem 2.17 (Theorem 5 in Konishi and Ünver 2006): Suppose college preferences are strongly monotonoic. Then the student-optimal stable mechanism is immune to manipulation via capacities.

Remarkably, Example 2.5 shows that the college-optimal stable mechanism is manipulable via capacities even under strongly monotonic preferences.

3 One-Sided Matching

3.1 House Allocation and Housing Markets

A **house allocation problem** (Hylland and Zeckhauser 1979) is a triple $\langle I, H, \succ \rangle$ where I is a set of agents, H is a set of indivisible objects (henceforth houses), and $\succ = (\succ_i)_{i \in I}$ is a list of preferences over houses. Throughout this section we assume $|H| = |I|$ and the preferences are strict.

The outcome of a house allocation problem is simply an assignment of houses to agents such that each agent receives a distinct house. Formally, a **(house) matching** $\mu : I \rightarrow H$ is a one-to-one and onto function from I to H.

A matching μ **Pareto dominates** another matching ν if $\mu(i) \succeq_i \nu(i)$ for all $i \in I$ and $\mu(i) \succ_i \nu(i)$ for some $i \in I$. A matching is **Pareto efficient** if it is not Pareto dominated by any other matching.

A house allocation problem is simply a collective ownership economy where a number of houses shall be assigned to a number of agents. It is an economy where the grand coalition I owns the set of all houses H, but no strict subset of I has any say over a house or a set of houses. In contrast, the following economy is a private ownership economy, where each agent holds the property rights of a specific house.

A **housing market** (Shapley and Scarf 1974) is a four-tuple $\langle I, H, \succ, \mu \rangle$ where I is a set of agents, H is a set of houses with $|H| = |I|$, $\succ = (\succ_i)_{i \in I}$ is a list of strict preferences over houses, and μ is an initial endowment matching. Formally, a housing market is a house allocation problem along with a matching that is interpreted as the initial endowment. Let $h_i = \mu(i)$ denote the initial endowment of agent $i \in I$.

3.1.1 Core of a Housing Market

A matching η is **individually rational** if $\eta(i) \succeq_i h_i$ for all $i \in I$. A matching η is in the **core** of the housing market $(I, H, \succ, \mu)$ if there is no coalition $T \subseteq I$ and matching ν such that

1. $\nu(i) \in \{h_j\}_{j \in T}$ for all $i \in T$,
2. $\nu(i) \succeq_i \eta(i)$ for all $i \in T$,
3. $\nu(i) \succ_i \eta(i)$ for some $i \in T$.

The following algorithm, along with the deferred acceptance algorithms, plays a key role in the matching literature.

Gale's Top Trading Cycles (TTC) Algorithm

Step 1. Each agent "points to" the owner of his favorite house. Since there is a finite number of agents, there is at least one **cycle** of agents pointing to one another. Each agent in a cycle is assigned the house of the agent he points to and removed from the market with his assignment. If there is at least one remaining agent, proceed with the next step.

In general, at

Step k. Each remaining agent points to the owner of his favorite house among the remaining houses. Every agent in a cycle is assigned the house of the agent he points to and removed from the market with his assignment. If there is at least one remaining agent, proceed with the next step.

This influential algorithm is first described by Shapley and Scarf (1974) and attributed to David Gale. Shapley and Scarf used TTC to show that the core of each housing market is non-empty. Roth and Postlewaite later showed that this algorithm has some additional remarkable properties when preferences are strict.

Theorem 3.1 (Theorem 2 in Roth and Postlewaite 1977): The outcome of Gale's TTC algorithm is the unique matching in the core of each housing market. Moreover, this matching is the unique competitive allocation.

Indeed, the mechanism that chooses the unique core matching for each housing market has some remarkable incentives properties.

Theorem 3.2 (Theorem 1 in Roth 1982): The core (as a direct mechanism) is strategy-proof.

Hence, unlike one-sided matching problems, there exists a plausible mechanism for housing markets where truth-telling is a dominant strategy for all agents. Indeed core is the only "plausible" mechanism that is strategy-proof.

Theorem 3.3 (Theorem 1 in Ma 1994): Core is the only mechanism that is Pareto efficient, individually rational, and strategy-proof.

Hence the core as a mechanism is the unambiguous winner for housing markets.

3.1.2 House Allocation Mechanisms

Unlike housing markets, a "perfect solution" to house allocation is not possible. There are, however, several quite plausible mechanisms.

An **ordering** $f : \{1, \ldots, n\} \to I$ is a one-to-one and onto function. Each ordering induces the following simple mechanism, which is especially plausible if there is a natural hierarchy of agents.

Simple serial dictatorship induced by f: The agent who is ordered first (by the ordering f) gets her top choice; the agent ordered second gets his top choice among those remaining; and so on.

Hence, if there are n agents, there are $n!$ simple serial dictatorships. Here an ordering is exogenously defined.

Another class of mechanisms, although initially it sounds unnatural, is based on the simple observation that a housing market is simply a house allocation problem along with a matching (which represents the initial endowment). Hence, one can simply assign a fixed matching to be interpreted as an initial endowment and find the core of the resulting housing market as a solution to house allocation.

Core from assigned endowments μ: For any house allocation problem $\langle I, H, \succ \rangle$, select the core of the housing market $\langle I, H, \succ, \mu \rangle$.

Not only are both of the above mechanisms Pareto efficient, but also any Pareto efficient matching can be obtained through them (very much in the spirit of first and second welfare theorems).

Theorem 3.4 (Lemma 1 in Abdulkadiroğlu and Sönmez 1998): For any ordering f and any matching μ, the simple serial dictatorship induced by f and the core from assigned endowments μ both yield *Pareto efficient* matchings. Moreover, for any *Pareto efficient* matching η, there is a *simple serial dictatorship* and a *core from assigned endowments* that yields it.

Prioritizing agents is quite common in many applications of these problems and in that sense simple serial dictatorships are quite natural. Interpreting core-based mechanisms is a lot harder unless there is a natural candidate to be interpreted as the initial endowment. One natural possibility is to extend the set of outcomes to allow for stochastic matchings and then choosing the core of the resulting housing market when the initial endowment is randomly determined with uniform distribution. Loosely speaking, this is in line with interpreting a house allocation problem to be a housing market where the initial endowment is stochastic, where each agent holds the ownership of $1/n$ of each house.

A **lottery** is a probability distribution over the set of matchings. A **(direct) lottery mechanism** is a systematic procedure to select a lottery for each problem.

Two natural lottery mechanisms are:

Random serial dictatorship (RSD): Randomly select an ordering with uniform distribution and then use the induced simple serial dictatorship.

Core from random endowments: Randomly select a matching (to be interpreted as the initial endowment) and select the core of the induced housing market.

Surprisingly, the above two mechanisms are identical. That is, they both choose the same lottery.

Theorem 3.5 (Theorem 2 in Abdulkadiroğlu and Sönmez 1998): The random serial dictatorship is equivalent to the core from random endowments.

This equivalence, along with its strategy-proofness, provides strong support for RSD, which is widely used in real-life applications. Unfortunately RSD has a major deficiency as was shown by Bogomolnaia and Moulin (2001). We need some additional notation to elaborate on this point.

A **random consumption** is a probability distribution over all houses H. Let $\triangle H$ denote the set of all random consumptions.

We denote a lottery by $\mathcal{L} = \sum \alpha_\mu.\mu$ where $\alpha_\mu \in [0,1]$ is the probability weight of matching μ. Let $\mathcal{L}(i)$ denote the random consumption given under $\mathcal{L}$ to agent i.

Each lottery induces a **random assignment** $P = [p_{ih}]_{i\in I, h\in H}$ where $p_{ih} \in [0,1]$ is the probability of agent i receiving house h. Let $\pi(\mu)$ be the permutation matrix that represents the matching μ. Then the random assignment that is induced by lottery $\sum \alpha_\mu.\mu$ is $P = \sum \alpha_\mu.\pi(\mu)$. Given any random assignment P, by the Birkhoff-von Neumann theorem there exists at least one lottery $\mathcal{L}$ that induces P. Let $U(\succ_i, h) \equiv \{h' \in H : h' \succeq_i h\}$ denote the upper contour set of h at $\succ_i$.

Given a pair of distinct random consumptions P_i, Q_i and a strict preference relation $\succ_i$, P_i **stochastically dominates** Q_i under $\succ_i$ if and only if

$$\forall h \in H, \qquad \sum_{h' \in U(\succ_i, h)} p_{ih'} \geq \sum_{h' \in U(\succ_i, h)} q_{ih'}.$$

That is, P_i stochastically dominates Q_i if and only if

- the probability of receiving the first choice is at least as much in P_i as in Q_i, and in general,
- for any k, the probability of receiving one of top k choices is at least as much in P_i as in Q_i.

Given a pair of distinct random assignments P, Q and a preference list $\succ$, P **stochastically dominates** Q under $\succ$ if and only if for any agent i either $P_i = Q_i$ or P_i stochastically dominates Q_i under $\succ_i$. A random assignment P is **ordinally efficient** (Bogomolnaia and Moulin 2001) at $\succ$ if and only if it is not stochastically dominated by any other random assignment under $\succ$.

The following example by (Bogomolnaia and Moulin 2001) shows that RSD is not ordinally efficient.

Example 3.1 $I = \{1,2,3,4\}$, $H = \{a,b,c,d\}$. The preferences are as follows:

$$\begin{array}{llll} \succ_1: & a\ b\ c\ d & \succ_3: & b\ a\ d\ c \\ \succ_2: & a\ b\ c\ d & \succ_4: & b\ a\ d\ c \end{array}$$

RSD induces the following random assignment:

$$P = \begin{array}{c|c|c|c|c|} & a & b & c & d \\ \hline 1 & 5/12 & 1/12 & 5/12 & 1/12 \\ \hline 2 & 5/12 & 1/12 & 5/12 & 1/12 \\ \hline 3 & 1/12 & 5/12 & 1/12 & 5/12 \\ \hline 4 & 1/12 & 5/12 & 1/12 & 5/12 \\ \hline \end{array}$$

Next consider the lottery

$$\mathcal{L} = 0.5 \begin{pmatrix} 1 & 2 & 3 & 4 \\ a & c & b & d \end{pmatrix} + 0.5 \begin{pmatrix} 1 & 2 & 3 & 4 \\ c & a & d & b \end{pmatrix}$$

which induces the random assignment

$$Q = \begin{array}{c|c|c|c|c|} & a & b & c & d \\ \hline 1 & 1/2 & 0 & 1/2 & 0 \\ \hline 2 & 1/2 & 0 & 1/2 & 0 \\ \hline 3 & 0 & 1/2 & 0 & 1/2 \\ \hline 4 & 0 & 1/2 & 0 & 1/2 \\ \hline \end{array}$$

Note that

* the random assignment P assigns everyone their 1st choices with 5/12 probability, 2nd choices with 1/12 probability, 3rd choices with 5/12 probability, and 4th choices with 1/12 probability,

* whereas Q assigns everyone their 1st choices with 1/2 probability and 3rd choices with 1/2 probability.

Hence Q stochastically dominates P, and therefore RSD is not ordinally efficient.

It turns out that, there is no "fair" mechanism that is strategy-proof and ordinally efficient.

Theorem 3.6 (Theorem 2 in Bogomolnaia and Moulin 2001): Suppose there are at least four agents. Then there is no mechanism that satisfies ordinal efficiency, strategy-proofness, and equal treatment of equals (i.e. agents with same preferences should receive the same random consumption).

An ordinally efficient and fair mechanism, the **Probabilistic Serial** (PS) mechanism, is introduced by Bogomolnaia and Moulin (2001) as an alternative to RSD. PS is based on the following **simultaneous eating algorithm**:

- Interpret each house as 1 unit of a divisible commodity and let all agents start "eating" from their top choices (accumulating probabilities) at the same speed.
- Once any person's top choice is completely consumed, she starts eating from her best choice among the remaining houses that still have available "pieces."
- The process ends when each agent consumes a total of 1 unit.

The following result summarizes the trade-off between PS and RSD.

Theorem 3.7 (Theorem 1, Proposition 1 in Bogomolnaia and Moulin 2001): PS satisfies both ordinal efficiency and equal treatment of equals, but it is not strategy-proof.

3.2 A Hybrid Model: House Allocation with Existing Tenants

Many real-life applications of one-sided matching use a hybrid between housing markets and house allocation. A typical scenario is: A set of houses should 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, and the rest of the agents are newcomers. In addition to occupied houses, there are vacant houses. Existing tenants are not only entitled to keep their current houses but can also apply for other houses. This motivates the following model:

A **house allocation problem with existing tenants** (Abdulkadiroğlu and Sönmez 1999) is a five-tuple $\langle I_E, I_N, H_O, H_V, \succ\rangle$ where I_E is a finite set of existing tenants, I_N is a finite set of newcomers, $H_O = \{h_i\}_{i\in I_E}$ is a finite set of occupied houses, H_V is a finite set of vacant houses where $h_0 \in H_V$ denotes the null house, and $\succ= (\succ_i)_{i\in I_E\cup I_N}$ is a list of strict preference relations. The null house h_0 (i.e., receiving nothing) has multiple copies, and for the purposes of this section it is the last choice for each agent.

A matching in this context is an assignment of houses to agents such that every agent is assigned one house, and only the null house h_0 can be assigned to more than one agent.

A matching is Pareto efficient if there is no other matching that makes all agents weakly better off and at least one agent strictly better off. A matching is individually rational if no existing tenant strictly prefers his endowment to his assignment. A lottery is ex-post Pareto efficient if it only gives positive weight to Pareto efficient matchings, and it is ex-post individually rational if it only gives positive weight to individually rational matchings.

A popular real-life mechanism is a simple modification of the RSD mechanism.

Random Serial Dictatorship with Squatting Rights

- Each existing tenant decides whether she will enter the housing lottery or keep her current house. Those who prefer keeping their houses are assigned their houses. All other houses become available for allocation.

- 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.

- Once the agents are ordered, available houses are allocated using the induced *simple serial dictatorship*: The first agent receives her top choice, the next agent receives her top choice among the remaining houses, and so on so forth.

Although popular in real-life applications, its outcome is not necessarily ex-post individually rational or ex-post Pareto efficient. Fortunately, a generalization of Gale's TTC mechanism by Abdulkadiroğlu and Sönmez (1999) overcomes this difficulty.

Top Trading Cycles (TTC) Mechanism

Fix an ordering f of agents. Interpret this as a **priority ordering**. The outcome of the TTC mechanism is obtained in several steps. Define the set of *available houses* for Step 1 to be the set of vacant houses. In general, at

Step t: The set of **available houses** for Step t is defined at the end of Step (t-1).

- Each remaining agent points to his favorite house among the remaining houses;

- each remaining occupied house points to its occupant; and

- each available house points to the agent with the highest priority among the remaining agents.

There is at least one cycle. Every agent in a cycle is assigned the house that he points to and removed from the market with his assignment. Whenever there is an available house in a cycle, the agent with the highest priority is also in the same cycle. If the most senior (remaining) agent is an existent tenant and his house is vacated, then it is added to the set of available houses for the next step. All available houses that are not removed remain available. If there is at least one remaining agent and one remaining house then we proceed with the next step.

TTC reduces to Gale's TTC for housing markets and RSD for house allocation problems. Moreover:

Theorem 3.8 (Propositions 1,2 and Theorem 1 in Abdulkadiroğlu and Sönmez 1999): For any ordering f, the induced TTC mechanism is individually rational, Pareto efficient, and strategy-proof.

We can find the outcome of the TTC mechanism using the following **you request my house-I get your turn (YRMH-IGYT) algorithm**:

1. For any given ordering f, assign the first agent his top choice, the second agent his top choice among the remaining houses, and so on, until someone demands the house of an existing tenant.

2. If at that point the existing tenant whose house is demanded is already assigned a house, then do not disturb the procedure.

   Otherwise modify the remainder of the ordering by inserting him at the top and proceed with the procedure.

3. Similarly, insert any existing tenant who is not already served at the top of the line once his house is demanded.

4. If at any point a loop forms, it is formed by exclusively existing tenants and each of them demands the house of the tenant next in the loop. (A **loop** is an ordered list of agents $(i_1, i_2, \ldots, i_k)$ where agent i_1 demands the house of agent i_2, agent i_2 demands the house of agent i_3,..., agent i_k demands the house of agent i_1.) In such cases remove all agents in the loop by assigning them the houses they demand and proceed with the procedure.

Theorem 3.9 (Theorem 3 in Abdulkadiroğlu and Sönmez 1999): For a given ordering f , YRMH-IGYT algorithm yields the same outcome as the TTC algorithm.

One might be tempted to consider the following simpler mechanism when we have $|H_V| = |I_N|$ (so that there are the same number of agents and houses).

1. Construct an initial allocation by assigning each existing tenant her own house and randomly assigning the vacant houses to newcomers with uniform distribution, and

2. choose the core of the induced housing market to determine the final outcome.

Clearly this mechanism is also Pareto efficient, individually rational, and strategy-proof. There is, however, a hidden bias in this mechanism that favors the newcomers. Loosely speaking, the existing tenants forfeit their portion of property rights on vacant houses when these houses are randomly assigned to newcomers to be treated as the initial endowment.

Theorem 3.10 (Theorem 1 in Sönmez and Ünver 2005): The above mechanism is equivalent to an extreme case of TTC where newcomers are randomly ordered first and existing tenants are randomly ordered next.

3.3 Priority-Based Indivisible Goods Allocation

Some of the real-life applications of one-sided matching problems share many common elements with the two-sided matching theory of Gale and Shapley (1962). While one side of the market consists of objects under one-sided matching, often there are priority orders over these objects. These priorities may be same for all objects or they may differ. And technically, these priority orders are mathematical objects similar to preferences. One possible "trick" is to treat these priorities as "object preferences" and effectively transform the one-sided matching problem into a two-sided matching problem. One important difficulty in applying this trick is that a key axiom "stability" is imposed in the entire two-sided matching literature. In the absence of a meaningful or plausible counterpart for stability, this trick is without much merit. However, it turns out that stability in two-sided matching markets is closely related to a very basic fairness axiom, *elimination of justified envy*, in the context of priority-based indivisible goods allocation. The *Student Placement* model of Balinski and Sönmez (1999) was the first model to utilize this isomorphism. The motivation for this model was centralized school admissions in Turkey, where priorities over school seats are determined by a centralized exam. A closely related model is the School Choice model of Abdulkadiroğlu and Sönmez (2003), which models assignment of public school seats to K-12 children. One key difference between these two models is that while elimination of justified envy is an absolute necessity in student placement, this is not so in school choice. This is mostly due to the fact that while priorities are exam based in student placement and need to be strictly respected, in school choice they depend on more flexible factors such as home address, lottery numbers, etc.

In this section we briefly review the student placement model. We devote an entire section to the school choice model since it initiated a literature of its own within matching and resulted in the reform of student admissions to public schools in many U.S. school districts, including Boston and NYC.

A **student placement problem** consists of a finite set of students I, a finite set of colleges S, a list of student preferences $\succsim = (\succsim_i)_{i \in I}$, a vector of college capacities $q = (q_s)_{s \in S}$, a set of skill categories $T = \{t_1, \ldots, t_k\}$, a list of test scores $f = (f^i)_{i \in I}$, and a function $t : S \longrightarrow T$ that maps each college to a skill category. Here q_s is the capacity of college s, $\succsim_{i_\ell}$ is the preference of student i_ℓ over colleges and the no-college option, and $f^{i_\ell} = (f_t^{i_\ell})_{t \in T}$ is a vector that gives the test score of student i_ℓ in each category. Each college s only cares for the test scores in category $t(s)$.

For each student placement problem, we can construct an **associated college admissions problem** by assigning each college s a preference relation $\succ_s$ based on the ranking in its category $t(s)$. A matching is defined the same way as it is defined in college admissions. The following notion is closely related to stability in college admissions. The exact relation will be given shortly.

A matching μ **eliminates justified envy** if whenever a student i prefers another student j's assignment $\mu(j)$ to his own, he ranks worse than j in the category of college $\mu(j)$. That is, no student i shall find himself in a situation where another student j receives a seat at a school s where not only i prefers s to his assignment but also student i has better test scores than student j in the relevant test for school s. A mechanism **eliminates justified envy** if it always selects a matching that eliminates justified envy.

A simple case to analyze is when there is only one skill category (i.e. all schools rank students based on the same exam). Given a priority ranking, the induced **simple serial dictatorship** assigns the first student his top choice, the next student his top choice among remaining seats, etc.

Proposition 3.1 (Corollary to Lemma 2 in Balinski and Sönmez 1999): Suppose there is only one category and no two students have the save exam score. Then there is only one Pareto efficient mechanism that eliminates justified envy: the simple serial dictatorship induced by this ranking.

A matching is **individually rational** if no student prefers the no-college option to his assignment and it is **non-wasteful** if no student prefers a college with one or more empty slots to his assignment. The following lemma relates elimination of justified envy and stability.

Lemma 3.1 (Lemma 2 in Balinski and Sönmez 1999): A matching is individually rational, non-wasteful, and eliminates justified envy if and only if it is stable for its associated college admissions problem.

Hence, elimination of justified envy is a slightly weaker requirement than stability.

The first paper to relate formal matching theory to real-life applications was by Roth (1984), who showed that the mechanism that assigned medical residents to hospitals in U.S. at the time was equivalent to a hospital-optimal stable mechanism. Balinski and Sönmez (1999) make a similar observation. The mechanism that assigns Turkish high school graduates to colleges is equivalent to college-optimal stable mechanism. There is, however, an important difference between the two applications. Using a hospital-optimal stable mechanism is plausible in a model where hospitals are agents with their welfare having some weight in the social welfare. Using college-optimal stable mechanism makes a lot less sense in a model where college seats are merely objects to be consumed and exams are conducted merely to ration these scarce resources. The student-optimal stable mechanism is a much more plausible alternative in the context of student placement. The following series of results formalizes this point.

For any student placement problem, let the college-optimal stable mechanism (COSM) be the mechanism which selects the college-optimal stable matching of the associated college admissions problem and let the student-optimal stable mechanism (SOSM) be the mechanism that selects the student-optimal stable matching of the associated college admissions problem. The following result is a slight generalization of Theorem 2.1 of Gale and Shapley.

Theorem 3.11 (Theorem 2 in Balinski and Sönmez 1999): The Gale-Shapley student-optimal stable mechanism Pareto dominates any other mechanism that eliminates justified envy.

Theorem 2.7 and Proposition 1 by Alcalde and Barberà (1994) in the two-sided matching literature can be reinterpreted in the context of student placement.

Theorem 2.7 (Theorem 9 in Dubins and Freedman 1981, Theorem 5 in Roth 1982b): The student-optimal stable mechanism is strategy-proof.

Theorem 3.12 (Proposition 1 in Alcalde and Barberà 1994): The student-optimal stable mechanism is the only mechanism that eliminates justified envy and is individually rational, non-wasteful, and strategy-proof.

Another deficiency of COSM is that a student may receive an inferior outcome as a result of an improvement in his exam scores. A mechanism **respects improvements** if a student never receives a worse assignment as a result of an increase in one or more of his test scores.

Theorem 3.13 (Theorem 5 in Balinski and Sönmez 1999): The student-optimal stable mechanism respects improvements. Moreover, it is the only mechanism that is individually rational, non-wasteful, eliminates justified envy, and respects improvements.

Based on these results, SOSM is the unambiguous winner, provided that priorities be strictly respected in the sense of elimination of justified envy. This plausible requirement is hard to give up when priorities are obtained via standardized exams. Elimination of justified envy, however, is not without cost. Unfortunately, it may conflict with Pareto efficiency. A reinterpretation of Example 2.1 (Roth 82) makes this simple point.

Example 3.2 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:

$$\begin{array}{lcl} \succsim_{s_1}: i_1\ i_3\ i_2 & \qquad\qquad & \succsim_{i_1}: s_2\ s_1\ s_3 \\ \succsim_{s_2}: i_2\ i_1\ i_3 & & \succsim_{i_2}: s_1\ s_2\ s_3 \\ \succsim_{s_3}: i_2\ i_1\ i_3 & & \succsim_{i_3}: s_1\ s_2\ s_3 \end{array}$$

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

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

Hence, while SOSM Pareto dominates any other mechanism that eliminates justified envy, it may not itself be Pareto efficient. Of course, this can only happen if there are at least two skill categories with different student rankings.

In the next section we turn our attention to the closely related school choice model by Abdulkadiroğlu and Sönmez (2003). One major difference in school choice is that, while elimination of justified envy is still plausible, most school districts have historically relied on mechanisms that do not satisfy it. Given its efficiency cost, it is important to study other plausible mechanisms that may not eliminate justified envy but are Pareto efficient.

4 School Choice

Traditionally, in most parts of the world students are assigned a neighborhood school. Since in 1980s, more and more school districts in the U.S. have been offering parents a choice over public schools in order to extend families' access to schools beyond their residence area. Student assignment is an integral part of such school choice programs. Abdulkadiroğlu and Sönmez (2003) study the problem of student assignment in public school choice from a mechanism-design perspective in a matching model, where, in addition to student's having preferences over schools, schools rank students in priority order.

As in the case of the student placement model of Balinski and Sönmez (1999), schools are not strategic agents with preferences over students; instead, school priorities are determined by exogenous factors such as a student's home address, which schools her siblings are attending, whether her current school is a feeder for another school, etc. This modeling choice is quite realistic for a vast majority of practical applications, although there are exceptions. For example in New York City High School Match, some schools (potentially) strategically rank students in preference order as part of the admissions process. In a school choice problem, schools are simply objects to be allocated and only student welfare matters. Consequently, stability, a key notion in two-sided matching, does not imply efficiency. On a positive side, stability is no longer incompatible with strategy-proofness. Many school choice programs have been relying on mechanisms that are not stable, and since stability comes with a potential efficiency cost, this may indeed be the way to go. We briefly explore these issues and recent developments in school choice below. We refer the reader to Abdulkadiroğlu (2011) and Pathak (2011) for comprehensive surveys on school choice.

4.1 Model

A school choice problem is another application of priority-based indivisible goods allocation with more flexibility in school priorities. Formally a **school choice problem** (Abdulkadiroğlu and Sönmez 2003) is a five-tuple $\langle I, S, q, \succsim_S, \succ_I \rangle$ where I is a finite set of students, S is a finite set of schools, $q = (q_s)_{s \in S}$ is a capacity profile for schools where q_s is the number of available seats at school $s \in S$, $\succsim_S = (\succsim_s)_{s \in S}$ is a profile of weak priority relations for schools where $\succsim_s$ is a complete, reflexive and transitive binary relations over $I \cup \varnothing$ for school $s \in S$, and $\succ_I = (\succ_i)_{i \in I}$ is a profile of strict preference relations for students where $\succ_i$ is a complete, irreflexive, and transitive binary relation over $S \cup \{\varnothing\}$ for student $i \in I$. Here $\varnothing$ represents remaining unmatched.

For $i \in I$, let $\succsim_i$ be the symmetric extension of $\succ_i$. That is, for all $s, j \in S \cup \{\varnothing\}$, if $s \succsim_i j$ then $s \succ_i j$ or $s = j$. Let the indifference relation $\sim_i$ denote the symmetric part of $\succsim_i$.

A **matching** of students to schools is a function $\mu : I \cup S \longrightarrow 2^{I \cup S}$ such that

1. $\mu(i) \subset S$ with $|\mu(i)| \leq 1$ for all $i \in I$; and

2. $\mu(s) \subset I$ with $|\mu(s)| \leq q_s$ for all $s \in S$; and

3. $s \in \mu(i)$ if and only if $i \in \mu(s)$ for all $i \in I$ and $s \in S$.

In the context of school choice, a matching μ is **feasible** if $i \succsim_s \varnothing$ for all $i \in \mu(s)$ and $s \in S$. As in the student placement model, a matching is **individually rational** if no student prefers being unmatched to his assignment, and it is **non-wasteful** if no student prefers a school with one or more empty seats to his assignment. A matching μ **eliminates justified envy** if no student i prefers the assignment of another student j while at the same time having higher priority at school $\mu(j)$. A matching μ is **stable** if it is individually rational, non-wasteful, and eliminates justified envy. Define a stable matching to be **student-optimal** if it is not Pareto dominated by another stable matching.

The main axioms we have been focusing on, namely efficiency, stability, and strategy-proofness, all have policy appeal in the context of school choice. An efficient matching optimizes student welfare in the Pareto sense. Stability eliminates justified envy and avoids wastefulness in the sense of Balinski and Sönmez (1999). A student-optimal stable matching optimizes student welfare subject to a stability constraint. Strategy-proofness simplifies the decision making for parents regardless of how sophisticated these parents are, leveling the playing field.

The following so-called **Boston mechanism**, which was brought to light by Abdulkadiroğlu and Sönmez (2003) and was in use in Boston until 2005, has been adopted widely by school districts in the U.S.:

The Boston Mechanism

1. For each school, a priority ordering is exogenously determined. (In the case of Boston, priorities depend on home address, whether 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:

   Round 1: In Round 1 only the first choices of the students are considered. For each school, consider the students who have listed it as their first 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 first choice.

   In general, at

   Round k: Consider the remaining students. In Round k only the k$^{\text{th}}$ choices of these students are considered. For each school with available seats, consider the students who have listed it as their k$^{\text{th}}$ 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$^{\text{th}}$ choice.

At each round, every assignment is final and the algorithm terminates when no more students are assigned.

The Boston mechanism assigns as many students as possible to their first choices based on their submitted preferences; next, as many students as possible to their second choices; and so on. The major drawback of this widely used mechanism is its lack of strategy-proofness:

Example 4.1 There are three students $\{i_1, i_2, i_3\}$ and three schools $\{s_1, s_2, s_3\}$ each with one seat. Student preferences and school priorities are given as follows:

$$\begin{array}{ll} \succ_{i_1}: & s_2\ s_1\ s_3 \\ \succ_{i_2}: & s_1\ s_2\ s_3 \\ \succ_{i_3}: & s_1\ s_2\ s_3 \end{array} \quad \textit{and} \quad \begin{array}{ll} \succsim_{s_1}: & i_1\ i_3\ i_2 \\ \succsim_{s_2}: & i_2\ i_1\ i_3 \\ \succsim_{s_3}: & i_2\ i_1\ i_3 \end{array}$$

When the students report their true preferences, the Boston mechanism produces the following matching:

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

If i_2 reports her preferences as $s_2 \succ'_{i_2} s_1 \succ'_{i_2} s_3$ instead, the Boston mechanism produces the following matching

$$\mu' = \begin{pmatrix} 1 & 2 & 3 \\ s_3 & s_2 & s_1 \end{pmatrix}$$

and student i_2 benefits from submitting a false preference list.

As seen in this example, a student who ranks a school as her second choice loses her priority to students who rank it as their first choice, so that it is risky for the student to use her first choice at a highly sought-after school if she has relatively low priority there. So the Boston mechanism gives students incentive to misrepresent their preferences by improving the ranking of schools in their choice lists for which they have high priority. Indeed, The BPS school guide [2004, p3] explicitly advises parents to strategize when submitting their preferences (quotes in original):

> For a better chance of your "first choice" school . . . consider choosing less popular schools. Ask Family Resource Center staff for information on "underchosen" schools.

The feature that one may gain from manipulating her choice list in the Boston mechanism is also recognized by parents in Boston and elsewhere. Indeed, the West Zone Parent Group (WZPG), a parent group in Boston, recommends strategies to take advantage of the mechanism (introductory meeting minutes on 10/27/03):

> One school choice strategy is to find a school you like that is undersubscribed and put it as a top choice, OR, find a school that you like that is popular and put it as a first choice and find a school that is less popular for a "safe" second choice.

Abdulkadiroğlu and Sönmez (2003) offer two strategy-proof solutions to the problem, one based on Gale and Shapley's student-proposing deferred acceptance algorithm, which produces the student-optimal stable matching, and another one inspired by Gale's top trading cycles algorithm, which produces an efficient matching. We refer to the former as Gale-Shapley student-optimal stable mechanism (SOSM) and the latter as Top Trading Cycles mechanism (TTC). In the rest of this section, we will assume strict school priorities, that is, for all $i, j \in I$, $i \neq j$, and $s \in S$, either $i \succ_s j$ or $j \succ_s i$.

The following results on SOSM follow from earlier discussions:

Theorem 2.1 (Theorems 1,2 in Gale and Shapley 1962): Given $(\succ_I, \succ_S)$, SOSM produces a matching that is stable at $(\succ_I, \succ_S)$, which is also at least as good for every student as any other stable matching at $(\succ_I, \succ_S)$.

Furthermore,

Theorem 2.7 (Theorem 9 in Dubins and Freedman 1981, Theorem 5 Roth 1982b): Given fixed priorities $\succ_S$, SOSM is strategy-proof.

Yet, as we have shown in Example 3.2, the outcome of SOSM is not necessarily efficient in the context of school choice. Ergin (2002) shows that the outcome of SOSM is Pareto efficient if and only if school priorities satisfy a certain acyclicity condition. Ehlers and Erdil (2010) generalize that result when school priorities are coarse. This can be interpreted as a negative result for the efficiency of SOSM, since school priorities are not likely to satisfy the acyclicity conditions of Ergin (2002) and Ehlers and Erdil (2010) in applications. In the context of school choice, Pareto efficiency may well be the primary objective. Abdulkadiroğlu and Sönmez (2003) proposed the following Pareto efficient mechanism.

Top Trading Cycles (TTC)

Students submit an ordered-choice list of schools. Given $(\succ_I, \succ_S)$, in every round every student points to her most-preferred school that still has available seats; every school s with available seats points to the unmatched students that are ranked highest in $\succ_s$. At each round there exists at least one cycle where a cycle of students and schools $(i_1, s_1, ..., i_K, s_K)$ is such that each element of the sequence points to the next whereas the last element s_K points to the first element i_1. Then student i_k is matched with school s_k and the capacity of school s_k is decreased by 1 for every $k = 1, ..., K$. Every matching is final. The algorithm terminates when no more students are assigned.

TTC is inspired by Gale's top trading cycles mechanism for housing markets (Shapley and Scarf 1974,) and, loosely speaking, it allows students to trade their priorities at schools to be matched with more preferred choices. In addition,

Theorem 4.1 (Propositions 3,4 in Abdulkadiroğlu and Sönmez 2003): TTC is strategy-proof and Pareto efficient.

Although TTC is Pareto efficient and SOSM is not, the two are not Pareto ranked in general, as can be seen in the next example.

Example 4.2 There are three students $\{i_1, i_2, i_3\}$ and three schools $\{s_1, s_2, s_3\}$ each with one seat. Student preferences and school priorities are given as follows:

$$\begin{array}{llcll} \succ_{i_1}: & s_2\ s_1\ s_3 & & \succsim_{s_1}: & i_1\ i_3\ i_2 \\ \succ_{i_2}: & s_1\ s_2\ s_3 & \text{and} & \succsim_{s_2}: & i_2\ i_1\ i_3 \\ \succ_{i_3}: & s_1\ s_2\ s_3 & & \succsim_{s_3}: & i_2\ i_1\ i_3 \end{array}$$

The outcomes of SOSM and TTC are

$$\mu^{SOSM} = \begin{pmatrix} 1 & 2 & 3 \\ s_2 & s_3 & s_1 \end{pmatrix} \quad and \quad \mu^{TTC} = \begin{pmatrix} 1 & 2 & 3 \\ s_2 & s_1 & s_3 \end{pmatrix}$$

where neither matching Pareto dominates the other one.

This example raises the following natural question: Is there a strategy-proof and Pareto efficient mechanism that Pareto dominates SOSM? Kesten (2010) answers this question negatively.

Theorem 4.2 (Proposition 4 in Kesten 2010): Given strict school priorities, no Pareto efficient and strategy-proof mechanism Pareto dominates SOSM.

Kesten (2010) also proposes a new algorithm that eliminates the efficiency loss associated with SOSM by allowing students to give up certain priorities whenever it does not hurt them to do so. Despite the lack of a Pareto ranking between SOSM and TTC, there exists a clear-cut comparison between dominant strategy equilibria of SOSM and Nash of the Boston mechanism.

Theorem 4.3 (Theorem 1 in Ergin and Sönmez 2006): Given strict school priorities, the set of Nash equilibrium outcomes of the Boston mechanism is equal to the set of stable matchings under true preferences. Therefore, the dominant strategy equilibrium of SOSM weakly Pareto dominates every Nash equilibrium outcome of the Boston mechanism.

SOSM and TTC are both strategy-proof mechanisms and hence their analysis is robust to information structure. Analysis of Boston mechanism is much harder and its equilibria relies on information structure. Ergin and Sönmez (2006) show that Theorem 4.3 do not extend to a Bayesian framework.

Formulation of the school choice as a matching problem has fostered the design of student-assignment mechanisms in several cities. In Boston, while initially TTC was recommended by the Family Task Force to replace the Boston mechanism, eventually SOSM was adopted in 2005 (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2005, 2006). In New York City, SOSM was adopted in 2003 (Abdulkadiroğlu, Pathak, and Roth 2005). These market-design exercises introduce new theoretical problems, which we discuss below.

4.2 Extensions

As in any market-design exercise, choice programs offered by school districts may involve distinctive features that are not captured by the basic model. Next we discuss some of these features brought to light by applications and the developments in the literature.

4.2.1 Ties in School Priorities

Much of the theory of two-sided matching concentrates on the case that all parties have strict preferences, mainly because of the theoretical limitations with indifferences. In contrast, a primary feature of school choice is that there are indifferences– ties– in how students are ordered by at least some schools, which brings in new trade-offs among efficiency, stability, and strategy-proofness (Erdil and Ergin 2008; Abdulkadiroğlu, Pathak and Roth 2009).

Since both SOSM and TTC take strict school priorities as input, one needs to break ties in school priorities when either mechanism is adopted. The choice over how to break ties with either mechanism is far from obvious. For example, in the course of designing the New York City high school match, policy makers from the Department of Education were concerned with the fairness of tie breaking; they believed that each student should receive a different random number at each program they applied to and this number should be used to construct strict preferences of schools for students (Abdulkadiroğlu, Pathak and Roth 2009, Pathak and Sethuraman 2011).

TTC remains Pareto efficient and strategy-proof with single and multiple tie-breakers. Furthermore, when there are no priorities at schools, i.e. all students tie in priority at every school, TTC produces the same probability distribution over matchings when a single or a multiple tie-breaker is drawn uniformly randomly (Pathak and Sethuraman 2011).

The use of a single tie-breaker under SOSM was first advocated by Abdulkadiroğlu and Sönmez (2003). With weak school priorities, there may be multiple student-optimal stable matchings that are not Pareto ranked with each other. However, each one can be obtained by a SOSM with some tie-breaker (Ehlers 2006) or even with a SOSM with a single tie-breaker (Abdulkadiroğlu, Pathak and Roth 2009; Erdil and Ergin 2006). The latter result implies that if there is a matching that can be produced by SOSM with a multiple tie-breaker but not with any single tie-breaker, that matching cannot be student-optimal. Indeed, Abdulkadiroğlu, Pathak and Roth (2009) show via simulations with NYC High School Match data that significantly more students get their first choices when ties are broken by a single lottery.

However, a single tie-breaker is not sufficient for student-optimality (Erdil and Ergin 2008). Motivated by this observation, Erdil and Ergin (2008) introduce a novel stable improvement cycles (SIC) algorithm to find a student-optimal stable matching. More interestingly, they show that ties in school priorities introduce a trade-off between efficiency and strategy-proofness; namely no student-optimal stable mechanism is strategy-proof when school priorities are weak (Example 2 in Erdil and Ergin 2008). In contrast, Abdulkadiroğlu, Pathak and Roth (2009) show that, given a tie-breaker, the associated SOSM is not Pareto-dominated by any strategy-proof mechanism. In other words, removal of any inefficiency produced by SOSM under single tie-breaker is at the expense of students' incentives. This observation generalizes the following two earlier results: SIC is not strategy-proof (Erdil and Ergin 2008) and no Pareto efficient and strategy-proof mechanism Pareto-dominates SOSM (Kesten 2010).

The existing literature, in particular all the results stated so far, rely on a notion of efficiency from an ex-post point of view, that is after the resolution of all potential uncertainties. Abdulkadiroğlu, Che and Yasuda (2008, 2011) recognize that ties in school priorities introduce scope for efficiency from an ex-ante point of view. When schools are indifferent among all students and students have the same ordinal ranking of schools but their cardinal utilities are private information that are drawn from a commonly known distribution, they show that each student's expected utility in every symmetric Bayesian equilibrium of the Boston mechanism in mixed strategies is weakly greater than her expected utility in the dominant-strategy equilibrium of SOSM. This finding contrasts but does not contradict with Ergin and Sönmez (2006), who analyze a complete information setup with strict school priorities and heterogenous ordinal preferences for students. Motivated by their observation for the Boston mechanism, Abdulkadiroğlu, Che and Yasuda (2008) propose an SOSM with "preferential" tie breaking with superior ex-ante efficiency in large economies.

4.2.2 Strategy-Proofness as a Notion of Fairness

The Boston experience presents the first case in which strategy-proofness has been adopted as a public policy concern related to transparency, fairness, and equal access to public facilities (Abdulkadiroğlu, Pathak, Roth, and Sönmez 2006). In July 2005, the Boston School Committee voted to adopt SOSM, which removes the incentives to "game the system" that handicapped the Boston mechanism. In his memo to the School Committee on May 25, 2005, Superintendent Payzant wrote:

> The most compelling argument for moving to a new algorithm is to enable families to list their true choices of schools without jeopardizing their chances of being assigned to any school by doing so... A strategy-proof algorithm levels the playing field by diminishing the harm done to parents who do not strategize or do not strategize well.

Pathak and Sönmez (2008) investigate this issue by studying a model with both sincere students who always submit their true preferences, and sophisticated students who respond strategically. They analyze Nash equilibria of the Boston mechanism and compare the equilibrium outcomes with the dominant-strategy outcome of SOSM. Their results confirm the intuitive idea that replacing the Boston mechanism with the strategy-proof SOSM "levels the playing field." They show that the set of Nash equilibrium outcomes of the Boston mechanism is equal to the set of stable matchings of a modified economy where sincere students lose their priorities to sophisticated students at all but their first choice schools; furthermore, every sophisticated student weakly prefers her assignment under the Pareto-dominant Nash equilibrium outcome of the Boston mechanism to the dominant-strategy outcome of SOSM.

Boston Public Schools is not the only school district that has decided to abandon the Boston mechanism based on the high-stakes gamble it imposes on families. Pathak and Sönmez (2011) focus on recent school choice reforms in Chicago and England, where the reaction to Boston mechanism was made very clear. In fall 2009, officials from Chicago Public Schools gave up a version of the Boston mechanism for coveted spots at selective college preparatory high schools in the middle of the assignment process. After asking about 14,000 applicants to submit their preferences for schools under the Boston mechanism, the district asked them resubmit their preferences under a new mechanism. Officials were concerned that "high-scoring kids were being rejected simply because of the order in which they listed their college prep preferences." Perhaps more striking is the ban of the Boston mechanism and similar "first preference first" mechanisms in England with the 2007 Admissions Code. The rationale stated by Englands Department for Education and Skills is that the "first preference first criterion made the system unnecessarily complex to parents" (School Code 2007, Foreword, p. 7). A story in the *Guardian*, a British newspaper, emphasizes that "the new School Admissions Code will end the practice called 'first preference first' which forces many parents to play an 'admissions game' with their childrens future, and unnecessarily complicates the admissions system" (Smith 2007).

It is worth emphasizing that the main objection of policy makers to Boston mechanism is its vulnerability to strategic manipulation and not its efficiency performance. In a Bayesian framework there are situations where certain disadvantaged students may receive favorable outcomes under the Boston mechanism. Neighborhood priorities are a common feature of many school choice programs. For instance, Boston Public Schools gives priority to students who live within 1 mile from an elementary school in attending those schools. Abdulkadiroğlu, Che and Yasuda (2011) observe that the extent to which the neighborhood priority inhibits the access to good schools by students in failing schools districts may differ across mechanisms. They provide examples in which the Boston mechanism provides greater access to good schools for students without neighborhood priority at those schools.

4.2.3 Controlled Choice

Controlled school choice in the U.S. attempts to provide parents with choice over public schools while maintaining racial, ethnic, and socioeconomic balance at schools. Boston's renowned controlled choice program emerged out of concerns for economically and racially segregated neighborhoods that were caused by traditional neighborhood-based assignment to public schools. Today, many school districts adopt desegregation guidelines either voluntarily or because of a court order.

Diversity control can be incorporated to mechanisms via quotas for various types of students. TTC with quotas is constrained efficient and strategy-proof (Abdulkadiroğlu and Sönmez 2003). SOSM with quotas is strategy-proof (Abdulkadiroğlu 2005, Hatfield and Milgrom 2005) and it produces a stable matching that respects quotas and is weakly preferred by every student to any other stable matching that respects quotas (Roth 1984).

Kojima (2010) observes that such quotas can make majority students and every minority student worse-off under both SOSM and TTC. Hafalir, Yenmez and Yildirim (2011) offer an alternative policy that gives preferential treatment to minorities for a number of reserved seats at schools. They also provide a group strategy-proof mechanism, which gives priority to minority students for reserved seats at schools. Their mechanism Pareto dominates SOSM with quotas.

4.2.4 Short preference lists

Some school districts impose a limit on the number schools that can be listed in an application. For instance students could list at most five schools in Boston before 2005; and students can list at most twelve schools in NYC High School Admissions.

Haeringer and Klijn (2009) study the preference revelation game induced by different mechanisms when students can only list up to a fixed number of schools. They focus on the stability and efficiency of the Nash equilibrium outcomes. They show that, when school priorities are strict and students can list a limited number of schools, (i) SOSM may have a Nash equilibrium in undominated strategies that produce a matching that is not stable under true preferences; (ii) TTC may have a Nash equilibrium in undominated strategies that produce a matching that is not Pareto efficient under true preferences.

5 Kidney Exchange

Transplantation is the preferred treatment for the most serious forms of kidney disease. As of July 2010, there are more than 85,000 patients waiting for a kidney transplant in the United States. In 2009, only 16,829 transplants were conducted, of which 10,442 were from deceased donors and 6,387 were from living donors. In the same year, 35,123 new patients joined the deceased donor waiting list, and 4,789 patients died while waiting for a kidney. Although there is a substantial organ shortage, buying and selling an organ is illegal in many countries and therefore donation is almost the only source for kidney transplants. While living donor donations make up a very substantial portion of all donations, many potential living donors cannot be utilized because of blood-type incompatibility or tissue rejection.

Rapaport (1986) was the first to propose paired kidney exchange between two such incompatible pairs in case each donor can feasibly donate a kidney to the patient of the other pair. Ross et. al (1997) reinforced this idea, and in 2000, the transplantation community issued a consensus statement declaring kidney exchange to be ethically acceptable (Abecassis et. al 2000). Another possibility is an indirect exchange (or list exchange) involving an exchange between one incompatible patient-donor couple, and the cadaver queue (Ross and Woodle 2000). In this kind of exchange, the patient in the couple receives high priority on the cadaver queue, in return for the donation of his donor's kidney to someone on the queue. In the period 2000-2004, feasible exchanges were sought in an unorganized way in parts of the U.S., and a relatively small number of them have been carried out.

In 2003, Alvin Roth, Tayfun Sönmez, and Utku Ünver observed that there are some important parallels between the above two types of kidney exchanges and the one-sided matching literature. Paired exchanges are the most basic cases of "cycles" in TTC algorithm (Shapley and Scarf 1974), whereas indirect exchanges are the most basic cases of "chains" in the YRMH-IGYT algorithm (Abdulkadiroğlu and Sönmez 1999). This observation, along with the lack of organized exchanges in transplantation community, motivated Roth, Sönmez, and Ünver (2004) to introduce the first formal mechanism, Top Trading Cycles and Chains (TTCC), for kidney exchanges. Shortly after the publication of this first paper in kidney exchange, the authors contacted Doctors Frank Delmonico of the New England Organ Bank and Susan Saidman of Massachusetts General Hospital and with their partnership they formed the New England Program for Kidney Exchange (NEPKE), the first regional kidney exchange program in the United States. Several transplant centers across the world have launched kidney exchange programs since then, and in fall 2010 the Organ Procurement and Transplantation Network has started a pilot national kidney exchange program in the U.S. In this section we briefly review the model and results of Roth, Sönmez, and Ünver (2004) since that paper is directly linked to one-sided matching literature. (See Sönmez and Ünver 2010 for a comprehensive survey on Kidney Exchange.)

Three assumptions are important in modeling kidney exchange:

1. Patient preferences over compatible kidneys.

    (a) The "European" view: The graft survival rate increases as the tissue type mismatch decreases (Opelz *Transplantation* 1997).

    (b) The "American" view: The graft survival rate is the same for all compatible kidneys (Gjertson and Cecka *Kidney International* 2000, Delmonico *NEJM* 2004).

2. The number of simultaneous transplants.

3. Feasibility of indirect exchanges.

Roth, Sönmez, and Ünver (2004) follow the "European" view, assuming strict preferences over compatible kidneys; it allows for any size exchange and allows for indirect exchanges.

A **kidney exchange problem**, or simply a problem, consists of a set of donor-transplant patient pairs $\{(k_1, t_1), \ldots, (k_n, t_n)\}$, a set of compatible kidneys $K_i \subseteq K = \{k_1, \ldots, k_n\}$ for each patient t_i, and a strict preference relation $\succsim_i$ for each patient t_i over $K_i \bigcup \{k_i, w\}$ where w represents the exchange of his donor with a priority increase at cadaveric wait list. For simplicity of presentation we often refer to a donor as a kidney in this section. His attached donor k_i may or may not be compatible with patient t_i but even if the kidney is not compatible he has preferences for it for its option value.

The outcome of a problem is a matching of kidneys/waitlist option to patients such that

1. each patient is either assigned a compatible kidney, or her donor's kidney, or the waitlist option, and
2. no kidney can be assigned to more than one patient, although the waitlist option w can be assigned to several patients.

A kidney exchange mechanism selects a matching for each kidney exchange problem. We are almost ready to introduce the Top Trading Cycles and Chains (TTCC) mechanism. First, we give a few definitions and observations to facilitate the description of the mechanism.

The mechanism relies on an algorithm consisting of several rounds. In each round each patient t_i points either to a kidney in $K_i \bigcup \{k_i\}$ or to w, and each kidney k_i points to its paired recipient t_i. A **cycle** is an ordered list of kidneys and patients $(k'_1, t'_1, \ldots, k'_m, t'_m)$ such that each element of the list points to the next element where kidney k'_1 is considered the next element of the patient t'_m. Cycles larger than a single pair are associated with direct exchanges, very much like the paired-kidney-exchange programs, but may involve more than two pairs. Note that each kidney or patient can be part of at most one cycle, and thus no two cycles intersect. A **w-chain** is an ordered list of kidneys and patients $(k'_1, t'_1, \ldots, k'_m, t'_m)$ such that each element of the list points to the next element, except the very last element patient t'_m who points to w. We refer to the pair (k'_m, t'_m) whose patient receives a cadaver kidney in a w-chain as the head, and the pair (k'_1, t'_1) whose donor donates to someone on the cadaver queue as the tail of the w-chain. W-chains are associated with indirect exchanges, but unlike in a cycle, a kidney or a patient can be part of several w-chains. One practical possibility is choosing among w-chains with a well-defined chain selection rule, very much like the rules that establish priorities on the cadaveric waiting list. The pilot indirect exchange program used in New England prior to the establishment of NEPKE was simply choosing the minimal w-chains, consisting of a single donor-recipient pair, but this in general is not efficient. Selection of longer w-chains will benefit other patients as well, and therefore the choice of a chain selection rule has efficiency implications. Chain selection rules may also be used for specific policy objectives such as increasing the inflow of type O living donor kidneys to the cadaveric waiting list.

The following lemma is key for the mechanics of the TTCC mechanism.

Lemma 5.1 (Lemma 1 in Roth, Sönmez, and Ünver 2004): Consider a graph in which both the patient and the kidney of each pair are distinct nodes as is the waitlist option w. Suppose each patient points either to a kidney or w, and each kidney points to its paired recipient. Then either there exists a cycle or each pair is the tail of a w-chain.

Top Trading Cycles and Chains (TTCC) Mechanism

Fix a chain selection rule. The TTCC mechanism determines the exchanges as follows:

1. Initially all kidneys are available and all agents are active. At each stage
    - each remaining *active* patient t_i points to the best remaining unassigned kidney or to the waitlist option w, whichever is more preferred,
    - each remaining passive patient continues to point to his assignment, and
    - each remaining kidney k_i points to its paired recipient t_i.
2. *By Lemma 5.1, 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 his top choice among remaining choices and each kidney points to its paired recipient. Proceed to Step 3 if there are no cycles. Otherwise locate all cycles, carry out the corresponding exchanges, and remove them. Repeat this step until no cycle exists.
3. If there are no pairs left, then we are done. Otherwise by Lemma 1, *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 passive henceforth.
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.

Here are some plausible chain-selection rules:

a. Choose the longest w-chain and remove it.

b. Choose the longest w-chain and keep it.

c. Prioritize patient-donor pairs in a single list. Choose the w-chain starting with the highest-priority pair and remove it.

d. Prioritize patient-donor pairs in a single list. Choose the w-chain starting with the highest-priority pair and keep it.

The choice of the chain selection rule is important and it has efficiency and incentives implications:

Theorem 5.1 (Theorem 1 in Roth, Sönmez, and Ünver 2004): Consider a chain-selection rule where any w-chain selected at a non-terminal round remains in the procedure and thus the kidney at its tail remains available for the next round. The TTCC mechanism, implemented with any such chain-selection rule, is Pareto efficient.

The chain-selection rules (b) and (d) described above result in a Pareto efficient execution of the TTCC mechanism.

Theorem 5.2 (Theorem 2 in Roth, Sönmez and Ünver 2004): Consider the priority-based chain-selection rules (c) and (d). The TTCC mechanism implemented with either of these chain-selection rules is strategy-proof.

Hence the TTCC mechanism, implemented with chain-selection rule (d), is Pareto-efficient and strategy-proof.

We have already indicated that TTCC is motivated by the YRMH-IGYT algorithm of Abdulkadiroğlu and Sönmez (1999). Krishna and Wang (2007) show that the two mechanisms are indeed equivalent if chain-selection rule (d) is used and wait-list option w is reinterpreted as a number of vacant houses in Abdulkadiroğlu and Sönmez (1999).

Theorem 5.3 (Proposition 1 in Krishna and Wang 2007): The TTCC algorithm executed with chain-selection rule (d) is equivalent to the YRMH-IGYT algorithm.

The interaction of Roth, Sönmez and Ünver with doctors in the transplantation community resulted in a followup paper by Roth, Sönmez and Ünver (2005), which followed the "American" view assuming patients are indifferent over compatible kidneys and only considered two-way exchanges to accommodate logistical constraints. Although initially there was a tendency in the transplantation community to utilize versions of the priority mechanism proposed in this paper, the large efficiency gains shown by Roth, Sönmez and Ünver (2007) resulted in the inclusion of 3-way exchanges in major kidney exchange programs. More recently chains that are initiated by good samaritan donors have begun to play an important role in kidney exchange. Roth, Sönmez, Ünver, Delmonico, and Saidman (2006) argued that transplantations in such chains do not need to be conducted simultaneously. This idea has been recently embraced by several kidney exchange programs and, most notably in 2010, more than half of the kidneys transplanted via kidney exchanges by the Alliance for Paired Donation were part of such chains.

6 Conclusion

The evolution of matching theory has gone through several phases over the last fifty years. Starting with the seminal paper of Gale and Shapley (1962), much of the earlier literature was devoted to understanding the theoretical structure of these problems. The 1980s were especially fruitful for matching theory. Following the developments in the mechanism design literature, economists started analyzing the incentives properties of various matching mechanisms. Another milestone came with Roth (1984), who showed that the National Resident Matching Program had been using the college-optimal stable mechanism since the 1950s to match medical residents and hospitals in the U.S. That paper was the first to relate real-life markets to matching theory. While the theory of two-sided matching markets was well-developed by 1990s, only a handful of papers by then were devoted to one-sided matching models originally introduced by Shapley and Scarf (1974). The 1990s and early 2000s have been fruitful not only for the theory of one-sided matching models, but also for models that have aspects of both one-sided and two-sided matching. This has prepared the foundation for new applications such as school choice and kidney exchange. In the past ten years the attention has shifted to mechanism/market-design aspects of real-life matching markets, and not only have these applications been influencing the direction of the theory but matching theory also has been making an important impact on policy. A few examples of this recent trend include the reform of student assignment mechanisms in Boston Public Schools and New

York City high schools, as well as the development of a number of kidney exchange programs in the U.S. such as the New England Program for Kidney Exchange and the Alliance for Paired Donation.

References

[1] Abdulkadiroğlu, A. (2005), College Admissions with Affirmative Action, *International Journal of Game Theory*, 33: 535-549.

[2] Abdulkadiroğlu, A. (2011), School Choice, Z. Neeman, M. Niederle, N. Vulkan (eds.) *Oxford Handbook of Market Design*, forthcoming.

[3] Abdulkadiroğlu, A. Y-K. Che and Y. Yasuda (2008), Expanding Choice in School Choice, mimeo.

[4] Abdulkadiroglu, A. Y-K. Che and Y. Yasuda (2009), Resolving Con- icting Preferences in School Choice: The "Boston" Mechanism Reconsidered, *American Economic Review*, 101(1), 399-410.

[5] Abdulkadiroğlu, A. P. A. Pathak and A. E. Roth (2005), The New York City High School Match, *American Economic Review, Papers and Proceedings*, 95, 364-367.

[6] Abdulkadiroğlu, A. P. A. Pathak and A. E. Roth (2009), Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match, *American Economic Review*, 99(5): 1954-78.

[7] Abdulkadiroğlu, A. P. A. Pathak, A. E. Roth, and T. Sönmez (2005), The Boston Public School Match, *American Economic Review, Papers and Proceedings*, 95, 368-371.

[8] Abdulkadiroğlu, A. P. A. Pathak, A. E. Roth, and T. Sönmez (2006), Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access, mimeo.

[9] Abdulkadiroğlu, A. and T. Sönmez (1998), Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problem, *Econometrica*, 66, 689-701.

[10] Abdulkadiroğlu, A and T. Sönmez (1999), House Allocation with Existing Tenants, *Journal of Economic Theory*, 88, 233-260.

[11] Abdulkadiroğlu, A. and T. Sönmez (2003), School Choice: A Mechanism Design Approach, *American Economic Review*, 93, 729-747.

[12] Abecassis, M. , M. Adams, P. Adams, R. M. Arnold, C. R. Atkins, M. L. Barr, W. M. Bennett, M. Bia, D. M. Briscoe, J. Burdick, R. J. Corry, J. Davis, F. L. Delmonico, R. S. Gaston, W. Harmon, C. L. Jacobs, J. Kahn, A. Leichtman, C. Miller, D. Moss, J. M. Newmann, L. S. Rosen, L. Siminoff, A Spital, V. A. Starnes, C. Thomas, L. S. Tyler, L. Williams, F. H. Wright, and S. Youngner (2000) "Consensus statement on the live organ donor." *Journal of the American Medical Association*, 284, 2919-2926.

[13] Alcalde J. and S. Barberá (1994), Top Dominance and the Possibility of Strategy-Proof Stable Solutions to Matching Problems *Economic Theory*, 4(3), 417-435.

[14] Balinski, M. and T. Sönmez (1999), A Tale of Two Mechanisms: Student Placement, *Journal of Economic Theory*, 84, 73-94.

[15] Bogomolnaia, A. and H. Moulin (2001), A New Solution to the Random Assignment Problem, *Journal of Economic Theory*, 100, 295-328.

[16] Crawford, V. P. and E. M. Knoer (1981), Job Matching with Heterogeneous Firms and Workers, *Econometrica,* 49, 437-450.

[17] Delmonico, F. L. (2004), Exchanging Kidneys–Advances in Living-Donor Transplantation, *New England Journal of Medicine*, 350, 1812-1814.

[18] Dubins, L. E. and D. A. Freedman (1981), Machiavelli and the Gale-Shapley Algorithm, *American Mathematical Monthly*, 88, 485-494.

[19] Ehlers, L. (2006), Respecting Priorities when Assigning Students to Schools, mimeo.

[20] Ehlers, L. and A. Erdil (2010), Efficient Assignment Respecting Priorities, *Journal of Economic Theory*, 145 (3), 1269-1282.

[21] Erdil, A. and H. Ergin (2006), Two-Sided Matching with Indi?erences, mimeo.

[22] Erdil, A. and H. Ergin (2008), What's the Matter with Tie Breaking? Improving Efficiency in School Choice, *American Economic Review*, 98, 669-689.

[23] Ergin, H. (2002), Efficient Resource Allocation on the Basis of Priorities, *Econometrica*, 70, 2489-2497.

[24] Ergin, H. and T. Sönmez (2006), Games of School Choice under the Boston Mechanism, *Journal of Public Economics*, 90, 215-237.

[25] Gale, D. and L. Shapley (1962), College Admissions and the Stability of Marriage, *American Mathematical Monthly*, 69, 9-15.

[26] Gale, D. and M. Sotomayor (1985), Ms Machiavelli and the stable matching problem, *American Mathematical Monthly*, 92, 261-268.

[27] Gjertson, D. W. and J. M. Cecka (2000), Living Unrelated Donor Kidney Transplantation, *Kidney International*, 58, 491-499.

[28] Haeringer, G. and F. Klijn (2009), Constrained School Choice, *Journal of Economic Theory,* 144, 1921-1947.

[29] Hafalir I, M. B. Yenmez, and M. A. Yildirim (2011), Effective Affirmative Action in School Choice, mimeo.

[30] Hatfield, J. W. and P. R. Milgrom (2005), Matching with Contracts, *American Economic Review*, 95, 913-935.

[31] Hylland, A. and R. Zeckhauser (1979), The Efficient Allocation of Individuals to Positions, *Journal of Political Economy*, 87, 293-314.

[32] Kelso, A. S. and V. P. Crawford (1982), Job Matchings, Coalition Formation, and Gross Substitutes, *Econometrica*, 50, 1483-1504.

[33] Kesten, O. (2010), School Choice with Consent, *Quarterly Journal of Economics*, 125(3), 1297-1348.

[34] Kojima, F. (2010), School Choice: Impossibilities for Affirmative Action, Stanford University working paper.

[35] Knuth, D. E. (1976), Mariages Stables. Les Presses de LUniversite de Montreal.

[36] Konishi, F. and U. Ünver (2006), Games of Capacity Manipulation in Hospital-InternMarkets, *Social Choice and Welfare*, 27: 3-24.

[37] Krishna, A. and Y. Wang (2007), The Relationship between Top Trading Cycles Mechanism and Top Trading Cycles and Chains Mechanism, *Journal of Economic Theory*, 132, 539-547.

[38] Ma, J. (1994), Strategy-Proofness and the Strict Core in a Market with Indivisibilities, *International Journal of Game Theory*, 23, 75-83.

[39] McVitie D.G. and L.B. Wilson (1970), Stable marriage assignment for unequal sets, *BIT* 10, 295309.

[40] Opelz, G. (1997), Impact of HLA Compatibility on Survival of Kidney Transplants from Unrelated Live Donors, *Transplantation*, 64, 1473-1475.

[41] Pathak, P. A. (2011), The Mechanism Design Approach to Student Assignment *Annual Reviews of Economics*, volume 3, in press.

[42] Pathak, P. A. and J. Sethuraman (2011), Lotteries in Student Assignment: An Equivalence Result, *Theoretical Economics*, 6, 117.

[43] Pathak, P. A. and T. Sönmez (2008), Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism, *American Economic Review*, 98, 1636-1652.

[44] Pathak, P. A. and T. Sönmez (2011), School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation, NBER Working Paper 16783.

[45] Rapaport, F. T. (1986), The case for a living emotionally related international kidney donor exchange registry, *Transplantation Proceedings*, 18, 5-9.

[46] Ross, L. F., D. T. Rubin, M. Siegler, M. A. Josephson, J. R. Thistlethwaite, Jr., and E S. Woodle (1997), Ethics of a paired-kidney-exchange program, *The New England Journal of Medicine*, 336, 1752-1755.

[47] Ross, Laine Friedman, and E. Steve Woodle (2000), Ethical issues in increasing living kidney donations by expanding kidney paired exchange programs, *Transplantation*, 69, 1539-1543.

[48] Roth, A. E. (1982a), Incentive Compatibility in a Market with Indivisibilities, *Economics Letters*, 9, 127-132.

[49] Roth, A. E. (1982b), The Economics of Matching: Stability and Incentives, *Mathematics of Operations Research*, 7, 617-628.

[50] Roth, A. 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.

[51] Roth, A. E. (1984), Misrepresentation and stability in the marriage problem, *Journal of Economic Theory*, 34, 383-387.

[52] Roth, A. E. (1986), On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets, *Econometrica*, 54, 425-427.

[53] Roth, A. E. (2008), Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions, *International Journal of Game Theory*, 36, 537-569.

[54] Roth, A. E. and A. Postlewaite (1977), Weak versus Strong Domination in a Market with Indivisible Goods, *Journal of Mathematical Economics*, 4, 131-137.

[55] Roth, A. E. and M. Sotomayor (1989), The College Admissions Problem Revisited, *Econometrica*, 57, 559-570.

[56] Roth, A. E. and M. Sotomayor (1990), Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis, Econometric Society Monograph Series, Cambridge University Press, 1990.

[57] Roth, A. E., T. Sönmez, and M. U. Ünver (2004), Kidney Exchange, *Quarterly Journal of Economics*, 119, 457-488.

[58] Roth, A. E., T. Sönmez, and M. U. Ünver (2005), Pairwise Kidney Exchange, *Journal of Economic Thory*, 125, 151-188.

[59] Roth, A. E., T. Sönmez, and M. U. Ünver (2007), Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences, *American Economic Review*, 97-3: 828-851.

[60] Roth, A. E. T. Sönmez, M. U. Ünver, F. Delmonico, and S. Saidman (2006), Utilizing List Exchange and Non-directed Donation through "Chain" Paired Kidney Donations, *American Journal of Transplantation*, 6(11): 2694-2705.

[61] Shapley, L. and H. Scarf (1974), On Cores and Indivisibility, *Journal of Mathematical Economics*, 1, 23-28.

[62] Shapley, L. and M. Shubik (1972), The Assignment Game I: The Core, *International Journal of Game Theory*, 1, 111-130.

[63] Smith, A. (2007), Schools admissions code to end covert selection, *Guardian*, Education section, January 9.

[64] Sönmez, T. (1997), Manipulation via Capacities in Two-Sided Matching Markets, *Journal of Economic Theory*, 77, 1, November, 197-204.

[65] Sönmez, T. and M. U. Ünver (2005), House Allocation with Existing Tenants: An Equivalence, *Games and Economic Behavior*, 52, 153-185.

[66] Sönmez, T. and M. U. Ünver (2010), Matching, Allocation, and Exchange of Discrete Resources, J. Benhabib, A. Bisin, and M. Jackson (eds.), *Handbook of Social Economics*, Elsevier.

[67] Sönmez, T. and M. U. Ünver (2010), Market Design for Kidney Exchange, Z. Neeman, M. Niederle, N. Vulkan (eds.) *Oxford Handbook of Market Design*, forthcoming.