Market Design

Fuhito Kojima, Fanqi Shi, and Akhil Vohra

March 31, 2017

1 Introduction

Matching theory in economics began with the seminal contribution by Gale and Shapley (1962). Ever since, the theory has advanced considerably and has been applied to an increasing number of economic problems. Notably, it has proved useful in helping designs of mechanisms in a variety of markets. Examples include medical match (Roth, 1984; Roth and Peranson, 1999) and other entry-level labor markets (Roth, 1991), school choice (Abdulkadiroğlu and Sönmez, 2003), course allocation in education (Sönmez and Ünver, 2010; Budish and Cantillon, 2012), and organ donation (Roth, Sönmez and Ünver, 2004, 2005, 2007). Application of matching theory to these and other practical problems is known as "market design." Although market design is often used to refer to other types of research as well, in this article, we focus on market design as application of matching theory.

This paper describes matching theory and its applications. We begin by describing standard models in two-sided and one-sided (object allocation) models in some detail, and then describe economic applications. By now, there are many surveys of this literature, most notably the celebrated work by Roth and Sotomayor (1990), and more recently by Abdulkadiroğlu and Sönmez (2013), Roth (2008a, 2008b), Sönmez and Ünver (2009), Pathak (2015), Kojima and Troyan (2011), Kojima (2015), and many others. Given the rich set of existing surveys, in this article, we try to balance between the basic models and recent applications. We also try to differentiate by choosing several specific topics which we regard as promising for further investigation.

The rest of this paper goes as follows. Section 2 presents the standard models of two-sided matching. In Section 3, we describe the models of one-sided matching. Section 4 discusses various applications. Section 5 concludes by discussing several future research directions.

2 Two-sided Matching

In two-sided matching, there are two groups of participants. Each participant may be matched to a participant on the other side of the market (or remain unmatched), and she has preferences over these options. Two canonical examples are college admissions and firm-worker assignment. In the first example, students try to get into their "ideal" colleges, and colleges seek to admit their most preferred students. In the second example, workers look for their "dream" jobs and firms attempt to fill the openings with their desired talents. In the presence of mutual interests and potential conflicts, of course, it is generally not possible to completely fulfill all participants' desires. In the face of this constraint, it would be desirable if a procedure can help match the students (workers) with colleges (firms) in a fair and efficient manner. That is precisely what we will take up in the current section.

On the other hand, despite their close resemblance, there is at least one notable difference between the college admissions problem and the firm-worker assignment: the terms governing the match of a student and a college are almost always identical, so the preference of a student/college depends only on the identity of her/its partners (if we ignore the differences in fellowship and other flexible terms). By comparison, labor contracts may vary much, with wage differences as a prominent example.[1] As such, a worker/firm not only cares about who he/it is matched with, but also the contracting details as well. Because of this additional complication, the firm-worker assignment is naturally more involved than the college admissions problem.

We will begin by presenting the basic Gale-Shapley (1962) two-sided matching model in Subsection 2.1, with the college admissions problem as the leading example. In Subsection 2.2, we discuss the matching with contracts model (Hatfield and Milgrom, 2005), with the firm-worker assignment in mind.

2.1 Basic Two-Sided Matching Model

Adopting the language of Gale and Shapley (1962) and Roth (1985), we describe the basic model in terms of colleges and students. However, we note that it can be applied to any matching model where both sides of the markets have preferences and the contracting details are standardized (e.g. medical residency matching).

There is a finite set I of students and a finite set C of colleges. A student can be matched to at most one college, while each college c has capacity q_c, i.e., the college can be matched to at most q_c students. Each student i has a strict preference $\succ_i$ over $C \cup \{\emptyset\}$, where $\emptyset$ denotes the outcome in which the student is unmatched, and each college has a strict preference $\succ_c$ over sets of students 2^I. For student i, we write $c_1 \succeq_i c_2$ if and only if $c_1 \succ_i c_2$ or $c_1 = c_2$. Similarly, for college c, we write $J_1 \succeq_c J_2$ if and only if $J_1 \succ_c J_2$ or

[1] Note, however, terms may not vary substantially in some labor markets, especially in standardized entry-level markets. In such a case, the matching model in Subsection 2.1 may be appropriate.

$J_1 = J_2$. Note we implicitly assume a student/college only cares about his/its own match and that indifferences do not occur.

Throughout the current subsection, we assume the preference of each college c is **responsive**, or the relative desirability of sets of students does not depend on the composition of the current assignment of college c. More formally, $\succ_c$ is **responsive** if:

1. For any $J \subset I$ with $|J| < q_c$ and any $i \in I \backslash J$, $(J \cup i) \succ_c J \Leftrightarrow i \succ_c \emptyset$ and
2. For any $J \subset I$ with $|J| < q_c$ and any $i, j \in I \backslash J$, $(J \cup i) \succ_c (J \cup j) \Leftrightarrow i \succ_c j$.[2]

We write the set of all strict responsive preference profiles as:

$$\mathcal{R} = \{(\succ_l)_{l \in I \cup C} | \succ_c \text{ is responsive}, \forall c \in C\}.$$

A **matching** is a function $\mu : I \to C \cup \{\emptyset\}$. For each $c \in C$, we define $\mu(c) = \{i \in I | \mu(i) = c\}$. We say that a matching μ is **feasible** if $|\mu(c)| \leq q_c$ for all $c \in C$. Simply put, in a feasible matching, each college is matched with a set of students not exceeding its capacity. For the rest of the discussion, we only look at feasible matchings and will simply refer to them as matchings. Let $\mathcal{M}$ be the set of all (feasible) matchings.

As mentioned at the beginning of the section, our goal is to find a systematic procedure that can help match students with colleges in a fair and efficient manner. Before proceeding, we first make precise what we mean by "a systematic procedure" and "fair and efficient manner." Formally, a **(direct) mechanism** is a function that produces a (random) matching (outcome) for each preference profile, or $\varphi : \mathcal{R} \to \Delta\mathcal{M}$. For fairness and efficiency, one possible criterion is **stability**. Formally, a matching μ is **individually rational** if $\mu(l) \succeq_l \emptyset$, $\forall l \in I \cup C$. A matching μ is **blocked by a pair** $(i, c) \in I \times C$ if:

1. $c \succ_i \mu(i)$ and
2. $|\mu(c)| < q_c$ and $i \succ_c \emptyset$ *or* $|\mu(c)| = q_c$ and $i \succ_c j$ for some $j \in \mu(c)$.

We say a matching is **stable** if it is individually rational and not blocked by any pair; a mechanism is **stable** if it produces a stable matching for any preference profile in any realization. Intuitively, a matching is individually rational if the assignment of each student and college is acceptable (at least as good as being unmatched). A matching is not blocked by any pair if whenever a student prefers a college to his current assignment, either the student is not acceptable to the college or the college has reached its capacity and it prefers every current student to the new student (note that we have implicitly made use of the assumption that preferences are responsive). A matching is stable if both conditions are satisfied. Stability is a desirable criterion for at least two reasons: to begin with, in an unstable mechanism, a student/college or a pair may want to deviate from the proposed

[2]To simplify notation, we denote a singleton set $\{s\}$ as s whenever there is no confusion.

outcome, which is problematic if we require voluntary participation. In addition, stability implies standard concepts of efficiency and fairness.

To see this, recall a matching μ is **(strongly) Pareto efficient** if there is no other matching ν such that $\nu(l) \succeq_l \mu(l)$ for all $l \in I \cup C$ and $\nu(l) \succ_l \mu(l)$ for some $l \in I \cup C$. Given a matching μ, a student i has **justified envy** toward student $j \in \mu(c)$, if $c \succ_i \mu(i)$ and $i \succ_c j$. We say μ is **justified envy-free** if it is individually rational and no student has justified envy. Roughly speaking, (strong) Pareto efficiency says the assignment of a student can only be improved at the expense of another student or college. Justified envy-freeness says if a student prefers the college of another student to his current matching, then it must be the case the college ranks the other student higher. If we take (strong) Pareto-efficiency and envy-freeness as the natural efficiency and fairness requirement, then the following proposition shows they are both implied by stability.

Proposition 1. *If a matching μ is stable, then it is both (strongly) Pareto efficient and envy-free.*

Having established a desirable property of a matching (and thus a mechanism), one may wonder whether a stable matching can be achieved with every preference profile. Gale and Shapley (1962) gave a positive answer with the construction of the following mechanism. Since then, the mechanism and its variations have played important roles in real-life matching markets.

(Student-proposing) Deferred Acceptance Algorithm:

- Step 1: Each student applies to his most preferred college. Each college tentatively keeps its acceptable students up to the capacity (on hold), and reject all others.

In general, for any $t = 1, 2, \ldots$

- Step t: Each student who is not currently on hold applies to his next preferred acceptable college (he does not make an application if there is none). Each college considers all new applicants, together with the students on hold, and tentatively keeps its acceptable students up to the capacity, while rejecting all others.

The algorithm terminates when there is no new application. (Clearly, it terminates in a finite number of steps because the number of students and colleges are both finite.)

Theorem 1 (Theorem 1 in Gale and Shapley 1962)**.** *For any (strict, responsive) preference profile, the (student-proposing) deferred acceptance algorithm gives a stable matching. In other words, it is a stable mechanism.*

Given the student-proposing deferred acceptance algorithm, one may naturally imagine a corresponding version of the college-proposing deferred acceptance algorithm. Such a

mechanism does exist and is also stable. In fact, the set of stable matchings is not necessarily a singleton set, and the student-proposing and college-proposing deferred acceptance algorithms can give rise to different stable matchings. Given the (potential) multiplicity of stable matchings, one may wonder which one to implement in practice. The following proposition suggests the answer depends on our evaluation of the relative welfare of the two sides of the markets.

Proposition 2 (Theorem 2* in Roth 1985)**.**

1. *There exists a student-optimal stable matching, i.e. a stable matching that every student likes at least as well as any other stable matching. Moreover, the student-proposing deferred acceptance algorithm always yields the student-optimal stable matching.*

2. *There exists a college-optimal stable matching, i.e. a stable matching that every college likes at least as well as any other stable matching. Moreover, the college-proposing deferred acceptance algorithm always yields the college-optimal stable matching.*

3. *The student-optimal stable matching is the least preferred stable matching for each college. Likewise, the college-optimal stable matching is the least preferred stable matching for each student.*

Roughly speaking, Proposition 2 says that despite the competition among students/colleges, there is considerable coincidence of interests on either side of the market if we restrict to stable matchings. By comparison, the interests of the two sides of the markets are not always aligned. In fact, they are almost opposite if we restrict our attention to stable matchings.

Another important concern for mechanisms is incentives. Apart from ease of implementation, an incentive compatible mechanism induces students and colleges to reveal their preferences truthfully. Such a property is necessary for our fairness and efficiency criteria to be well-grounded.[3] We say a mechanism is **strategy-proof** if it is a weakly dominant strategy for every player to (always) report his/her true preferences. With this definition, we have the following result for incentive properties of the student-proposing deferred acceptance algorithm (and any stable mechanism).

Theorem 2 (Theorem 5* in Roth 1985)**.** *The student-proposing deferred acceptance algorithm is strategy-proof for students. However, (when colleges have responsive preferences), no stable mechanism is strategy-proof for colleges.*

Simply put, Theorem 2 suggests that the student-proposing deferred acceptance algorithm is "safe" to play for students: it is always optimal for students to simply report their

[3]Chen and Sonmez (2006) show evidence in lab experiments that strategy-proof school choice mechanisms indeed induce true preferences more often than those without truthful revelation. Nevertheless, there is some evidence in real-life matching markets that some agents still misreport even in strategy-proof mechanisms. See for instance Rees-Jones (2017) and Hassidim et al. (2017).

true preferences. The rough intuition is that because of "deferred" acceptance, a student stands nothing to lose by applying to his most preferred (remaining) college in each step. Furthermore, even though it is not always in the colleges' interests to report truthfully, the problem is not confined to the particular mechanism, but stable mechanisms in general. In other words, incentive compatibility (on the college side) is not compatible with stability.

To see how colleges may gain by misreporting under the student-proposing deferred acceptance algorithm, consider the following example:

Example 1: There are two students i_1, i_2 and two colleges c_1, c_2. Each college has capacity 1 ($q_{c_1} = q_{c_2} = 1$) and the preferences are as follows:[4]

$$\begin{array}{ll} \succ_{i_1}\colon c_1, c_2 & \succ_{i_2}\colon c_2, c_1 \\ \succ_{c_1}\colon i_2, i_1 & \succ_{c_2}\colon i_1, i_2 \end{array}$$

It is easy to see the outcome of the student-proposing deferred acceptance algorithm under the true preferences is:

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

Now suppose college 1 misreports by stating that only student 2 is acceptable, while all other agents continue to report their true preferences. Then, the outcome of the student-proposing deferred acceptance algorithm under this preference profile is:

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

Thus, college 1 strictly benefits from the misreport.

2.2 Matching with Contracts

This subsection introduces the model of "matching with contracts" due to Hatfield and Milgrom (2005). It is a generalization of the basic model described in Section 2.1. It also incorporates, as special cases, some other matching/auction models in the existing literature. The primary example we have in mind is labor market matching, and so we describe the model in terms of workers and firms. However, it can be applied in any setting where contract terms play an important role.

There is a finite set I of workers, a finite set F of firms, and a finite set of contracts X. Each contract $x \in X$ is bilateral, so it is associated with one worker $x_I \in I$ and one firm $x_F \in F$. For instance, in the labor market matching model, a contract specifies a firm, a worker, and a wage. So we have $X = I \times F \times W$, where W is a (finite) set of possible wages.

Each worker i can sign at most one contract, and his preferences over possible contracts (plus the outcome in which she signs no contract, i.e., the empty set $\emptyset$, which we sometimes

[4]When denoting an agent's preference, we list her acceptable choices in order of her preference. Similar notation is used throughout the paper.

refer to as the null contract) are described by the strict total order $\succ_i$. We say a contract x is **acceptable** for worker i if $x \succ_i \emptyset$. With workers' preferences well-defined, we know a worker's choice when faced with a set of contracts. Formally, given a set of contracts $X' \subset X$, define worker i's chosen set $C_i(X')$ of contracts as follows:

$$C_i(X') = \begin{cases} \emptyset & \text{if } \{x \in X' | x_I = i, x \succ_i \emptyset\} = \emptyset \\ \max_{\succ_i}\{x \in X' | x_I = i\} & \text{otherwise} \end{cases}$$

In other words, a worker's chosen set is simply the most preferred, *acceptable* contract from those that are available. If none are acceptable, then his choice is the null contract (note that a worker's chosen set is either a singleton or an empty set).

On the other hand, each firm f can sign multiple contracts, so its preferences are over sets of contracts. Let firm f's preference be described by $\succ_f$, a strict total order over 2^X. Given a set of contracts $X' \subset X$, we can similarly define firm f's chosen set $C_f(X')$. Notice that a firm can sign at most one contract with any given worker, so for all $f \in F, X' \subset X$, and $x, x' \in C_f(X')$, if $x \neq x'$, then $x_I \neq x'_I$.

Given $X' \subset X$ and the chosen set of each worker, we can define the contracts chosen by the worker side as $C_I(X') = \cup_{i \in I} C_i(X')$. The remaining offers in X' are the rejected set by the worker side: $R_I(X') = X' - C_I(X')$. Similarly, the chosen and rejected sets by the firm side are denoted by $C_F(X') = \cup_{f \in F} C_f(X')$ and $R_F(X') = X' - C_F(X')$.

If we imagine some sort of stable matchings similar to the one in the basic model, the chosen and rejected sets will prove useful both in the definition and in actually finding them. Intuitively, given a starting set of contracts, the rejected sets (by either side) cannot be in any stable allocation, so we know, at the very least, a stable allocation is a fixed point of the "chosen set (by either side)" operator. Moreover, we will need to require that there is no coalition of workers and firms who prefer to sign contracts among themselves rather than follow the prescribed allocation.

To make this precise, we present the following definition of a stable allocation: a set of contracts $X' \in X$ is a **stable allocation** if:

1. $C_I(X') = C_F(X') = X'$ and
2. There exists no firm f and a set of contracts $X'' \neq C_f(X')$ such that $X'' = C_f(X' \cup X'') \subset C_I(X' \cup X'')$.

Intuitively, the first requirement corresponds to "individual rationality". The second requirement says there cannot be an alternative set of contracts such that a particular firm and its matched workers all prefer, which is similar to the "no blocking" condition we discussed in the basic model.

In order to guarantee the existence of a stable allocation, we need an additional restriction on the preferences of firms: substitutability. Elements of a set of contracts X

are **substitutes** for firm f if $\forall X' \subset X'' \subset X$, we have $R_f(X') \subset R_f(X'')$. In words, the restriction says if a particular firm is faced with a larger choice set, it has to reject (weakly) more contracts.

Now we are ready to present the existence result, based on an iterative algorithm. As it turns out, the iteration we shall apply is on the product set $X \times X$ instead of the set of all contracts X. Similar to most iteration procedures, we hope to obtain monotonicity. For monotonicity to be well-defined, a partial order on $X \times X$ is needed. We define it as follows: given $X_I, X_I', X_F, X_F' \subset X$, $((X_I, X_F) \geq (X_I', X_F')) \Leftrightarrow (X_I' \subset X_I$ and $X_F \subset X_F')$.

With the order $\geq$, and given a starting set (X_I, X_F), we can define the **generalized deferred acceptance algorithm** as the iterated applications of the function $F : X \times X \to X \times X$, defined by:

$$F_1(X') = X - R_F(X')$$
$$F_2(X') = X - R_I(X')$$
$$F(X_I, X_F) = (F_1(X_F), F_2(F_1(X_I))).$$

With the generalized deferred acceptance algorithm in hand, the following theorem and proposition tell us how they can direct us to find stable allocations and shed light on the tradeoff in welfare between the workers and firms.

Theorem 3 (Theorem 3 in Hatfield and Milgrom 2005)**.** *Suppose elements of X are substitutes for all the firms. Then,*

1. *The set of fixed points of F on $X \times X$ includes a smallest element $(\underline{X}_I, \underline{X}_F)$ and a largest element $(\bar{X}_I, \bar{X}_F)$;*
2. *Starting at $(X_I, X_F) = (X, \emptyset)$, the generalized deferred acceptance algorithm converges monotonically to the largest fixed point $(\bar{X}_I, \bar{X}_F) = \sup\{(X', X'') | F(X', X'') \geq (X', X'')\}$; and*
3. *Starting at $(X_I, X_F) = (\emptyset, X)$, the generalized deferred acceptance algorithm converges monotonically to the smallest fixed point $(\underline{X}_I, \underline{X}_F) = \inf\{(X', X'') | F(X', X'') \leq (X', X'')\}$.*

Proposition 3 (Theorem 4 in Hatfield and Milgrom 2005)**.** *Suppose elements of X are substitutes for all the firms, then $\bar{X}_I \cap \bar{X}_F$ and $\underline{X}_I \cap \underline{X}_F$ are both stable allocations. Moreover, every worker likes $\bar{X}_I \cap \bar{X}_F$ at least as well as any other stable allocation, and every firm likes any other stable allocation at least as well as $\bar{X}_I \cap \bar{X}_F$. Similarly, every firm likes $\underline{X}_I \cap \underline{X}_F$ at least as well as any other stable allocation and every worker likes any other stable allocation at least as well as $\underline{X}_I \cap \underline{X}_F$.*

To understand Theorem 3 and Proposition 3, consider first the generalized deferred acceptance algorithm starting with the contract tuple $(X, \emptyset)$. Here, workers first choose

from the set of all available contracts, which is very similar to the student-proposing deferred acceptance algorithm (where students first propose to their most preferred colleges). Accordingly, we land at the worker-optimal stable allocation. Similarly, the generalized deferred acceptance algorithm starting with the contract tuple $(\emptyset, X)$ gives us the firm-optimal stable allocation. For Proposition 3, the first part of why $\bar{X}_I \cap \bar{X}_F$ and $\underline{X}_I \cap \underline{X}_F$ are stable needs some additional reasoning (which we shall not give here), but the second part incorporates the insight from the basic model: if we focus on stable allocations, there is little conflict of interests among agents on the same side of the market, but there is substantial conflict of interests between the worker side and the firm side.

Similar to the basic model, incentives are also an important concern in the matching with contracts model, but we shall omit it here due to space constraint.

3 One-sided Matching

Similar to two-sided matching, in a one-sided matching problem, there are still two groups. Nevertheless, one side has no (intrinsic) preference for the other side and are simply "objects to be consumed". Two typical examples are house allocation and kidney exchange. In each problem, each individual tries to obtain her most preferred object, and if she has an initial endowment, may obtain her desired object through exchange. Our goal is to find a systematic procedure to allocate the houses (kidneys) to tenants (patients) in a fair and efficient way.

Even if we ignore the contracting details such as side payments (as in two-sided matching), complications arise depending on the structure of initial endowments. The first case we consider is pure exchange: each individual brings an item to the market and tries to achieve an efficient outcome that everyone is willing to accept. The next case is pure allocation: nobody has anything a priori, and there are a number of items to be allocated to the individuals. A more involved case is allocation with existing owners: some individuals come to the market with an item while others do not, and there are new items available for allocation. In this situation, we want to find an allocation that appeals to both the existing owners and the newcomers. We will present the three different models in Subsections 3.1, 3.2 and 3.3. To highlight the similarities and differences, we describe all three models in the language of housing markets.

3.1 House Exchange

As the name suggests, in a house exchange problem, each individual is initially endowed with a house and there is no new house available. Thus, the only possible re-allocation is through exchange between the agents. The model was first described by Shapley and Scarf

(1974).[5]

There is a finite set H of houses and a finite set I of house-owners. Initially, individual i is the owner of house h_i, with $|H| = |I|$. Therefore, everyone is endowed with exactly one house to begin with, and there is no leftover house. Each agent i demands exactly one house, and has a strict preference $\succ_i$ over H (we assume all houses are acceptable, though it is enough to assume agent i's initial endowment h_i is acceptable). Similar to two-sided matching, we write $h_1 \succeq_i h_2$ if and only if $h_1 \succ_i h_2$ or $h_1 = h_2$. We also implicitly assume a house-owner only cares about her own assignment and indifferences between houses do not occur. We write the set of all strict preference profiles as $\mathcal{R} = \{(\succ_i)_{i\in I}\}$. A matching is a one-to-one correspondence $\mu : I \to H$ (or equivalently represented as a permutation on $\{1, 2, \ldots, |I|\}$). Note that there is no feasibility concern once we restrict to one-to-one correspondences. Let $\mathcal{M}$ be the set of all matchings. A (direct) mechanism in the one-sided matching model is a function $\varphi : \mathcal{R} \to \Delta\mathcal{M}$.

Given one-sided preference and no priority, one possible criterion in the current setup is the **(strong) core** standard in cooperative games. Given $(\succ_i)_{i\in I}$, recall a matching μ is a **(strong) core** allocation if there is no other matching $\mu' \in \mathcal{M}$ and a subset $S \subset I$ such that:

1. $\mu'(i) \in \{h_j\}_{j\in S}, \forall i \in S$;
2. $\mu'(i) \succeq_i \mu(i), \forall i \in S$ and
3. $\mu'(j) \succ_j \mu(j)$ for some $j \in S$.

In words, a matching is a strong core allocation if no sub-group can come up with an allocation using only their endowments that every group member weakly prefers to the original allocation, and some group member strictly prefers. Similar to stability, (strong) core reflects the idea of voluntary participation. Moreover, if we take S to be the singleton set i and the whole set I, we see the definition implies individual rationality and strong Pareto efficiency.[6]

Proposition 4. *If a matching μ is a (strong) core allocation, then it is both individually rational and (strongly) Pareto efficient.*

Notice that with one-sided preference and no priority list, fairness is not too much of a concern once we respect individual rationality. Given the desirable criterion, our goal is to find a strong core allocation (if any) for any preference profile. Fortunately, David Gale (described in Shapley and Scarf (1974)) gave a satisfactory answer with the construction of the top trading cycle algorithm.

[5] To better conform with the two-sided matching model, our notation will be different from theirs. However, the model and the main results follow theirs.

[6] Here the outside option of being unmatched is modified to be the initial endowment in the definition of individual rationality.

Top Trading Cycle Algorithm:

- Step 1: Each agent points to her most preferred house and each house points to its owner. Remove all agents and houses in a cycle (at least one cycle exists). For any agent removed, assign her the house she points to.

In general, for any $t = 1, 2, \ldots$

- Step t: Each remaining agent points to her most preferred house left and each house points to its owner (who must still be in the mechanism). Remove all agents and houses in a cycle (at least one cycle exists). For any agent removed, assign her the house she points to.

The algorithm terminates when there is no agent or house left. (An equal number (greater than or equal to one) of agents and houses are removed in each step, so the mechanism must terminate in a finite number of steps.)

Theorem 4 (Theorem in Shapley and Scarf 1974 and Theorem 2 in Roth and Postlewaite 1977)**.** *For any strict preference profile, the top trading cycle algorithm gives the unique strong core allocation.*

Intuitively, the top trading cycle gives a strong core allocation roughly because (given the cycles removed earlier), the group of agents removed in each step receive their best available houses.

As in two-sided markets, another important concern in house exchange is incentives. After all, it is only with truthful revelation that the strong core requirement has important welfare implications. Fortunately, the following theorem says the top trading cycle algorithm has good incentive properties.

Theorem 5 (Theorem in Roth 1982)**.** *The top trading cycle mechanism is strategy-proof.*

To obtain intuition for Theorem 5, recall truthful revelation is a weakly dominant strategy for students in the student-proposing deferred acceptance algorithm. The insight here is somewhat similar: a house will not leave the market unless its (initial) owner gets her most preferred (remaining) house. Therefore, an agent will not "lose" a house unless she cannot get it anyway. Hence, she may as well report her most preferred (remaining) house in each step.

Given the relative complexity of the top trading cycle algorithm, we end the subsection with an example:

Example 2: There are four house-owners i_1, i_2, i_3, i_4, with their respective (initial) houses h_1, h_2, h_3, h_4. The preferences of the house-owners are as follows:

$$\succ_{i_1}\colon h_2, h_3, h_4, h_1 \quad \succ_{i_2}\colon h_2, h_1, h_3, h_4$$
$$\succ_{i_3}\colon h_2, h_1, h_3, h_4 \quad \succ_{i_4}\colon h_2, h_1, h_3, h_4$$

Step 1: All agents point to h_2 and the houses point to their respective owners. There is a one-cycle: i_2 is assigned h_2.

Step 2: i_1 points to h_3, i_3 and i_4 point to h_1 and the three remaining houses point to their respective owners. There is a two-cycle: i_1 is assigned h_3 and i_3 is assigned h_1.

Step 3: i_4 points to h_4, which points backwards. There is a one-cycle: i_4 is assigned h_4.

It follows that the outcome of the top trading cycle algorithm is:

$$\begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ h_3 & h_2 & h_1 & h_4 \end{pmatrix}$$

3.2 House Allocation with No Existing Owner

In a pure house allocation problem, no agent has a house to begin with, and there are a number of houses to be allocated. A similar problem was first studied by Hylland and Zeckhauser (1979). The model we present here is a simplified version of theirs.

There is finite set H of empty houses to be allocated among a finite set I of agents. Each agent i demands exactly one house, and has a strict preference $\succ_i$ over H. (We assume all houses are acceptable.) It may be the case that $|H| > |I|$, $|H| < |I|$ or $|H| = |I|$, so the houses may be in over-supply, under-supply or just balances with the number of agents. Similar to the house exchange problem, we write $h_1 \succeq_i h_2$ if and only if $h_1 \succ_i h_2$ or $h_1 = h_2$. Implicit in the assumption is that an agent only cares about her own assignment and that indifferences between houses do not occur. We write the set of all strict preference profiles as $\mathcal{R} = \{(\succ_i)_{i \in I}\}$. A matching is a one-to-one function $\mu : I \to H \cup \emptyset$.[7] Let $\mathcal{M}$ be the set of all matchings. A (direct) mechanism is a function $\varphi : \mathcal{R} \to \Delta\mathcal{M}$.

Given the lack of existing owners, the primary criterion here is (strong) Pareto efficiency. Nevertheless, randomization may be particularly useful in the current setup if there are fairness concerns. Once we introduce randomization, at least two versions of (strong) Pareto efficiency arises. A mechanism is **ex-ante Pareto efficient** if its assignment of lotteries is Pareto efficient relative to agents' preferences over lotteries. By comparison, a mechanism is **ex-post Pareto efficient** if its final allocation is Pareto efficient given any strict preference profile. It can be readily shown that ex-ante Pareto efficiency implies ex-post Pareto efficiency but not vice versa.[8]

Before introducing a desirable algorithm, one more definition is needed: a **(rank) ordering** is a permutation of I, or a one-to-one correspondence $\sigma : \{1, 2, \ldots, |I|\} \to I$. The following mechanism and its variations are widely used in real-life house allocation problems:

[7] As the number of houses and agents need not balance, in general, a matching here is not bijective and cannot be equivalently represented as a permutation on $\{1, 2, \ldots, |I|\}$.

[8] Ex-ante Pareto efficiency implies ex-post Pareto efficiency because if any final allocation resulting from a lottery is not ex-post Pareto efficient, then the lottery can be improved by replacing the particular allocation with a more efficient one, implying that the lottery is not ex-ante Pareto efficient.

Serial Dictatorship Algorithm:

- Step 0: Fix a rank ordering σ.
- Step 1: Assign $\sigma(1)$ her most preferred house.

In general, for any $t = 1, 2, \ldots$

- Step t: Assign $\sigma(t)$ her most preferred remaining house.

The algorithm terminates when there is no agent or house left. If there are still agents left, then they are not assigned a house. (Given each step reduces the number of agents and houses both by 1, the algorithm must terminate in a finite number of steps.)

Intuitively, the mechanism works as if an agent is the dictator when it is her turn to choose. At that time, she picks her most preferred house out of those available (note that the agent does not care about the allocation of any other agent). As mentioned earlier, randomization is often introduced when implementing the mechanism in practice. This can be done by modifying Step 0 as follows:

- Step 0: Pick a rank ordering σ uniformly at random from the set of all rank orderings.

The resulting mechanism is called **random serial dictatorship**. The following theorem gives the desirable properties of random serial dictatorship.

Theorem 6 (Variation of Lemma 1 in Abdulkadiroğlu and Sönmez 1998)**.** *The random serial dictatorship algorithm is ex-post Pareto efficient. Moreover, two agents with the same preferences receive the same random allocation to each other.*

In other words, the random serial dictatorship mechanism has decent fairness and efficiency properties. Intuitively, ex-post efficiency is achieved because an agent is made as well off as possible given the allocation of the agents in earlier steps. (This is true in every realization, so "ex-post" with randomization.) The second part of this theorem describes a fairness property of this mechanism, and it follows immediately from uniform randomization over rank orderings used in Step 0' of the algorithm.

The following theorem says random serial dictatorship also has good incentive properties.

Theorem 7. *The random serial dictatorship mechanism is strategy-proof.*

The main insight of Theorem 7 is that each agent is essentially the dictator when it comes to her turn, so she cannot gain by misreporting.

Unfortunately, even random serial dictatorship is not without its own problems. For one, the mechanism is not ex-ante Pareto efficient. Some research has been done to tackle the

problem. Nevertheless, it is found that ex-ante Pareto efficiency and fairness (as defined in Theorem 6) are incompatible with strategy-proofness. (See Bogomolnaia and Moulin (2001).) Partly because of this, random serial dictatorship is still probably among the most popular mechanisms when it comes to real-life object allocation.

3.3 House Allocation with Existing Owners

Given the discussions of Subsections 3.1 and 3.2, one may imagine a situation where existing house-owners and new entrants coexist. A few more specific real-life examples are college dorm allocations and office assignment. The problem was first studied by Abdulkadiroğlu and Sönmez (1999).

There are a finite set of houses H and a finite set of agents I. Of all the houses in H, a subset H_O are currently occupied, each belonging to a distinct member of the existing house-owners $I_E \subset I$ (so $|H_O| = |I_E|$). The remaining houses $H_V = H - H_O$ are currently vacant and can be freely allocated. The remaining agents $I_N = I - I_E$ are new entrants and do not have a house. Each agent $i \in I$ demands exactly one house and has a strict preference $\succ_i$ over H. (We assume for simplicity that all houses are acceptable for all the agents.) We write $h_1 \succeq_i h_2$ if and only if $h_1 \succ_i h_2$ or $h_1 = h_2$. Implicit in the assumption is that an agent only cares about her own assignment and that indifferences between houses do not occur. We write the set of all strict preference profiles as $\mathcal{R} = \{(\succ_i)_{i \in I}\}$. A matching is a one-to-one function $\mu : I \to H \cup \emptyset$.[9] Let $\mathcal{M}$ be the set of all matchings. A (direct) mechanism is a function $\varphi : \mathcal{R} \to \Delta\mathcal{M}$.

Given the presence of both existing house-owners and new entrants, one possible desirable criterion of a matching is Pareto efficiency. In light of the top trading cycles and serial dictatorship algorithms of previous subsections, we have the following two generalizations as natural candidates. Indeed, Abdulkadiroğlu and Sönmez (1999) show for any preference profile, outcomes from these two mechanisms coincide and satisfy Pareto efficiency.

(Generalized) Top Trading Cycles Algorithm:

- Step 0: Fix a rank ordering σ.
- Step 1: Define the set of available houses to be the vacant houses (H_V). Each agent points to her most preferred house. Each occupied house points to its owner, and each available house points to $\sigma(1)$. Remove all agents and houses in a cycle (at least one cycle exists). For any agent removed, assign her the house she points to.

In general, for any $t = 1, 2, \ldots$

- Step t: Update the set of available houses to be the current vacant houses. Each remaining agent points to her most preferred house left. Each remaining occupied

[9] Similar to the pure house allocation problem, a matching here need not be bijective.

house points to its owner, and each available house points to the remaining agent with the highest priority ($\sigma(j)$, where j is the smallest among the remaining agents). Remove all agents and houses in a cycle (at least one cycle exists). For any agent removed, assign her the house she points to.

The algorithm terminates when there is no agent or house left. If there are still agents left, then they are not assigned a house. Since each step reduces the number of agents and houses both by at least 1, the algorithm must terminate in a finite number of steps.

You Request My House-I Get Your Turn (YRMH-IGYT) Algorithm:

- Step 0: Fix a rank ordering σ.
- Step 1: Agent $\sigma(1)$ points to her most preferred house. If the house she points to is vacant (in H_V) or her own house, she is assigned the house she points to. Otherwise, modify σ so that the owner is at the top of the list (the other relative orderings unchanged) and proceed to the next step.

In general, for any $t = 1, 2, \ldots$

- Step t: The remaining agent with the highest priority ($\sigma(j)$, where j is the smallest among the remaining agents) points to her most preferred house. If the house she points to is *currently* vacant (which may or may not be in H_V) or her own house, she is assigned the house she points to. If the house she points to is occupied by another remaining agent, modify σ so that the owner is at the top of the list (the other relative orderings unchanged). At this point, if a loop forms (no house is a assigned in the process where the rank ordering is back to an earlier one), every agent is assigned the house she points to. Otherwise, proceed to the next step.

The algorithm terminates when there is no agent or house left. If there are still agents left, then they are not assigned a house. (It can be shown the algorithm terminates in a finite number of steps.)

Intuitively, the (generalized) top trading cycles algorithm is a direct generalization of top trading cycles in Section 3.1, with all remaining vacant houses pointing to the remaining agent with the highest priority. On the other hand, YRMH-IGYT is a direct generalization of serial dictatorship in Section 3.2, with the added twist that the owner is granted the opportunity to choose before her house is gone.

Theorem 8 (Theorem 3 in Abdulkadiroğlu and Sönmez 1999)**.** *Given a rank ordering σ and for any (strict) preference profile, the YRMH-IGYT algorithm yields the same matching as the generalized top trading cycles algorithm.*

Theorem 9 (Propositions 1 and 2 in Abdulkadiroğlu and Sönmez 1999)**.** *Given a rank ordering σ and for any (strict) preference profile, the matching given by YRMH-IGYT and generalized top trading cycles algorithms is strongly Pareto efficient.*

Moreover, the following theorem reveals that incentives do not pose a problem either.

Theorem 10 (Theorem 1 in Abdulkadiroğlu and Sönmez 1999)**.** *For any rank ordering σ, both the YRMH-IGYT and the generalized top trading cycles algorithms are strategy-proof.*

Given the close relationships, the intuitions of Theorems 9 and 10 are very similar to the counterparts of top trading cycles and serial dictatorship algorithms. (Theorems 4 through 7)

To illustrate the two algorithms and the main insights of Theorem 8, we conclude the subsection with an example.

Example 3: There are four agents and three houses. Agents i_1 and i_2 are current house-owners, with their respective houses h_1 and h_2. Agents i_3 and i_4 are new entrants. House h_3 is currently available. There preferences of the agents are as follows:

$$\succ_{i_1}\colon h_3, h_2, h_1 \quad \succ_{i_2}, \succ_{i_3}, \succ_{i_4}\colon h_1, h_3, h_2$$

Fix the rank ordering $\sigma = (3, 4, 1, 2)$.

Generalized Top Trading Cycles:

Step 1: i_1 points h_3, and the remaining agents point to h_1. h_1 points to i_1, h_2 points to i_2 and h_3 points to i_3. There is a two-cycle: i_1 is assigned h_3 and i_3 is assigned h_1.

Step 2: i_2 and i_4 both point to h_2 and h_2 points to i_2. There is a one-cycle: i_2 is assigned h_2.

It follows that the outcome of the generalized top trading cycle algorithm is:

$$\begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ h_3 & h_2 & h_1 & \emptyset \end{pmatrix}$$

YRMH-IGYT:

Step 1: i_3 points to h_1, which is currently occupied. The ranking ordering σ is modified to $(1, 3, 4, 2)$.

Step 2: i_1 points to h_3, which is currently vacant and i_1 is assigned h_3.

Step 3: i_3 points to h_1, which is currently vacant and i_3 is assigned h_1.

Step 4: i_4 points to h_2, which is currently occupied. The ranking ordering σ is modified to $(1, 3, 2, 4)$.

Step 5: i_2 points to h_2, which is her own house and i_2 is assigned h_2.

It follows that the outcome of the YRMH-IGYT algorithm is also:

$$\begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ h_3 & h_2 & h_1 & \emptyset \end{pmatrix}$$

4 Applications

The theories described in the previous sections have found applications in a wide variety of areas. While we are not able to survey all of them extensively, we have selected some of the most prominent examples in order to highlight how the theory can be utilized. We begin by discussing the following topics.

1. Medical Residency Matching (most closely related to the model in Subsection 2.1)
2. Kidney Exchange (most closely related to the model in Subsection 3.3)
3. School Choice (most closely related to the models in Subsections 2.1 and 3.3)

Then, building on an understanding of the problems encountered when applying the tools to the situations above, we transition to a relatively new area of matching theory called "matching with constraints".

4.1 Medical Residency Matching

Among the most common applications of two-sided matching algorithms is the medical residency programs. In 2016, roughly 43,000 medical school graduates registered for the National Resident Match Program (NRMP), where students are matched to teaching hospitals through a variant of a deferred acceptance algorithm.[10] This service has been in operation since 1952 and its longevity is ascribed to the fact that the matchings produced are stable (Roth, 1984; Roth and Sotomayor, 1990).

What makes NRMP's matching problem complex, though, is the existence of "couples". While some students apply independently and rank their preferences accordingly, individuals who have a significant other in the residency match program are allowed to apply together as couples so they can work in areas close to one another. In this context, stability requires there be no coalition of students and hospitals who prefer to match among themselves than follow the prescribed matching.[11] The presence of couples who submit joint preference lists complicates the problem significantly as stability is not guaranteed. Take the following example (from Roth, 1984):

Example 4: There are four medical school graduates i_1, i_2, i_3, i_4 and four hospitals h_1, h_2, h_3, h_4 each with capacity 1 ($q_{h_1} = q_{h_2} = q_{h_3} = q_{h_4} = 1$). (i_1, i_2) and (i_3, i_4) are couples with preferences over *ordered pairs* of hospitals. The exact preferences of the couples and the hospitals are as follows:

[10] For more detailed statistics, one can visit http://www.nrmp.org/match-data/main-residency-match-data/.

[11] The main difference of this definition from the one in the basic model of Subsection 2.1 is that we consider a coalition composed of a couple of doctors and two hospitals each of which seeks to match with a member of the couple. See Roth (1984) for detail.

$\succ_{(i_1,i_2)}$: $(h_1,h_2),(h_4,h_1),(h_4,h_3),(h_4,h_2),(h_1,h_4),(h_1,h_3),(h_3,h_4),(h_3,h_1),(h_3,h_2),(h_2,h_3),(h_2,h_4),(h_2,h_1)$
$\succ_{(i_3,i_4)}$: $(h_4,h_2),(h_4,h_3),(h_4,h_1),(h_3,h_1),(h_3,h_1),(h_3,h_2),(h_3,h_4),(h_2,h_1),(h_2,h_3),(h_1,h_2),(h_1,h_4),(h_1,h_3)$
$\succ_{h_1}$: i_4, i_2, i_1, i_3
$\succ_{h_2}$: i_4, i_3, i_2, i_1
$\succ_{h_3}$: i_2, i_3, i_1, i_4
$\succ_{h_4}$: i_2, i_4, i_1, i_3

It is straightforward, if tedious, to check that no stable matching exists in this example. Given that an increasing number of medical students marry other medical students, it would seem then that finding a stable matching for NRMP would be impossible. Even determining, in a given instance, whether a stable matching exists is a computationally hard problem.[12] As a result, to replace the original method, Roth and Peranson (1999) proposed a heuristic modification of the deferred acceptance algorithm in place to accommodate couples' preferences. Although this algorithm is not guaranteed to always produce a match that is stable with respect to the reported preferences, it has done so in almost all instances.

Why does the algorithm in NRMP find a stable matching despite the theoretical possibility of nonexistence? Kojima et al. (2013) show that in a setting where applicant preferences are drawn independently from a distribution, as the size of the market increases and the proportion of couples approaches 0, the Roth and Peranson algorithm terminates in a stable matching with high probability. Thus, one of the reasons the NRMP algorithm finds stable matchings in most cases may be because the size of NRMP is large while the proportion of couples in the market is small, roughly between 5% and 10%. By contrast, Biro and Kljin (2013) and Ashlagi, Braverman and Hassidim (2014) have shown, in separate settings, that as the proportion of couples increases, this algorithm frequently fails to terminate in a stable matching. This may be important given that residency matching is not the only environment with a "couples" issue. In other such settings, couples could make up much more of the market (Biro and Kljin (2013) provide the example of assigning high school teachers in Hungary to majors, where almost all teachers need to be assigned to two majors; in this setting, the percentage of "couples" is nearly 100%).

Given these difficulties in the "couples" problem, Nguyen and Vohra (2017) propose an alternative approach. They allow for perturbations of hospital capacities to find a "nearby" instance of the matching problem that is guaranteed to have a stable matching. They find that the necessary perturbations are small, especially when hospital/school/firm capacities are large. Specifically, given capacities q_h for each hospital h, there is a redistribution of the slots, q'_h, satisfying $|q_h - q'_h| \leq 2$ for all hospitals h *and* $\sum q_h \leq \sum q'_h \leq \sum q_h + 4$. Thus, the perturbations change the capacity of each individual hospital by at most 2, and increase

[12] More precisely, this problem is in the class of "NP-hard" problems. NP-hardness is a notion in computational complexity theory describing the complexity of computation, which we will not describe in detail here.

the total number of positions in hospitals by no more than 4 while never decreasing it.[13]

The complication surrounding matching with couples turns out to be a specific instance of a more general issue economists have sought to understand: matching with complementarities. In two-sided matching markets, substitutability of agent preferences (see Section 2.2), i.e., the lack of complementarity, is "necessary" for guaranteeing the existence of a stable matching.[14] The existence of couples leads to a violation of substitutability because a pair of positions close to each other works as complements for the couple. Recent research by Che, Kim, and Kojima (2017) and Azevedo and Hatfield (2017) examine matching with complementarities in large markets settings with a continuum of agents. They have found positive results describing sufficient conditions for the existence of stable matchings.

4.2 Kidney Exchange

The application of matching theory to kidney exchange has been discussed often and is quite thorough, so we will be relatively brief in our exposition. For an extensive survey, we refer the reader to Sönmez and Ünver (2011). In the kidney "market" (using the term loosely), the National Organ Transplant Act of 1984 made it illegal to buy or sell a kidney in the US. Similar legal prohibitions are nearly universal around the globe. Thus, donation is the only viable option for kidney transplantation for most patients.

The initial foundational contribution to kidney exchange came with Roth, Sönmez, and Ünver (2004). They used a variation of the Shapley-Scarf house exchange model (Section 3.1) to represent the kidney-exchange market. In their model, agents enter in pairs composed of a patient and his potential donor. Applying the top trading cycles (TTC) mechanism where potential donors substitute "houses" of the original Shapley-Scarf model, one can produce a matching between donors and patients in a Pareto-efficient and strategy-proof way.

The way economists model kidney exchange has progressed as we now know many ways in which the assumptions in the original 2004 paper do not seem to be the best representation of the real kidney market. As economists have advanced into the area of matching under general constraints and dynamic matching, they have attempted to employ other mechanisms different from TTC. For instance, because all transplantations in any kidney exchange need to be carried out simultaneously, long cycles that could be conducted using the TTC mechanism might not be feasible in practice. Roth, Sönmez, and Ünver (2005) provided strategy-proof, constrained-efficient mechanisms of kidney exchange where only pairwise exchanges are permitted. They showed that finding a constrained-efficient match-

[13] How the authors proceed from the setup is notable as they approach the problem from a linear programming perspective. Formulating the matching problem as a linear program and applying the celebrated Scarf's lemma, they find a random matching that satisfies a notion of stability. They then use an iterative rounding method to find an actual matching (corresponding to a $0-1$ solution) such that the resulting matching satisfies stabilitiy. Such rounding corresponds to the perturbation of the capacities.

[14] See Hatfield and Kojima (2008) and Sönmez and Ünver (2010) for formal statements.

ing in their model relates to the cardinality matching problem discussed in the graph theory literature.[15] In a 2007 paper, these same authors showed that under certain conditions on kidney supply and demand levels that could normally be expected, full efficiency can be extracted by using exchanges that involve no more than four pairs.

In the papers discussed above, agents and the market itself are static. What if the exchange pool changes over time? Should we conduct exchanges immediately, or if there is no urgency, is it more efficient to wait? These issues are not addressed formally in the aforementioned papers. Ünver (2010) tackles the question of how to conduct barter exchanges in a centralized mechanism when the agent pool evolves over time: he characterizes the efficient two-way and multi-way exchange mechanisms that maximize total exchange surplus. The study of dynamic matching environments has attracted the interest of not only economists but computer scientists and operation research specialists as well. Notable contributions include Anderson et al. (2015) and Akbarpour et al. (2016). There are still many questions left to be addressed, which makes the kidney exchange market one of the great interests amongst researchers and practitioners today.

4.3 School Choice

The third prominent area matching theory is applied to is that of school choice and student assignment policy. School choice has become one of the most important and contentious debates in modern education policy. School Choice is a policy that allows parents the opportunity to choose the school their child will attend. Traditionally, children are assigned to public schools according to where they live. Wealthy parents already have school choice, because they can enroll their children in private schools or have the ability to move to a different district entirely. Supporters have argued that school choice helps lower income families by providing them the freedom to send their children to different schools within and across districts. In addition, the increased competition schools face under school choice should incentivize them to increase their quality.[16] Since it is not possible to assign each student to her top choice school, a central issue in school choice is the design of a student assignment mechanism. One of the first to rigorously and formally tackle this issue with matching theory is Abdulkadiroğlu and Sönmez (2003). The model they propose, which has been regarded as the canonical model, consists of a set of students and schools where,

1. Each student i has a preference relation $\succ_i$ over the schools and
2. Each school c has capacity q_c and priority ordering $\succ_c$ over the students.

One of the reasons we refer to the school's ordering as a priority ordering is because in some school choice programs, orderings are given exogenously (mandated by law for

[15] In addition, the 2005 paper assumed that each patient is indifferent among all kidneys that are compatible to her, based on certain medical evidence.

[16] For a more extensive survey on empirical and theoretical literature around school choice, see Pathak (2011).

example). Pathak (2011) describes a variety of orderings in different districts as follows: "In Boston's school choice plan, for instance, elementary school applicants obtain walk-zone priority if they reside within 1 mile of the school. In other districts, schools construct an ordering of students, as in two-sided problems. In Chicago, for instance, students applying for admissions to selective high schools take an admissions test."[17] When evaluating a matching in this setting, two notions are of primary interest: Pareto efficiency and stability. Stability is defined in the standard manner as in Subsection 2.1, while Pareto efficiency only considers students' allocations and does not take into account the schools' priority ordering.[18] Abdulkadiroğlu and Sönmez (2003) compare three mechanisms: the student-proposing deferred acceptance algorithm, an adaptation of the top trading cycles mechanism (referred to as TTC in this subsection), and the Boston mechanism.

As the deferred acceptance mechanism is familiar to the reader by now, we describe the other two mechanisms. We start with the Boston mechanism, which, as its name suggests, was in use in school choice programs in the city of Boston (before being replaced by the deferred acceptance algorithm):

- Step 0: Each school orders students by priority block.[19] Within each block, students are ordered via a lottery system.
- Step 1: In this step, 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 is no seat left or there is no student left who has listed it as his first choice.

In general, for any $t = 1, 2, \ldots$

- Step t: In this step, only the t^{th} choices of the students are considered. For each school, consider the students who have listed it as their t^{th} choice and assign seats of the school to these students one at a time following their priority order until either there is no seat left or there is no student left who has listed it as his t^{th} choice.

In Boston, the Boston mechanism was originally implemented in July, 1999 but was abandoned in 2005. One of the central reasons it was abandoned is that it is not strategy-proof for students, i.e., families have an incentive to strategically misreport their preferences. Variations of this mechanism, however, are common in many other school districts.

[17] Later, Boston's school choice plan implemented a reform which eliminated the use of walk-zone priority.

[18] As the literature has grown and evolved, generalizations of the notion of stability have been discussed, which we will examine in Section 4.4.

[19] In Boston, first priority consisted of students who lived in a proximal neighborhood and had a sibling that attended the school. The second tier consisted of students with a sibling at the school. Third priority is of the students who live in the "relevant" area. Finally, the remaining students are grouped within the last priority block.

The next procedure presented by Abdulkadiroğlu and Sönmez (2003) is the TTC mechanism, which is implemented in the following manner:[20]

- Step 1: Assign a counter for each school which keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools. Each student points to her favorite school under her announced preferences. Each school points to the student who has the highest priority for the school. Since the number of students and schools are finite, there is at least one cycle. Moreover, each school can be part of at most one cycle. Similarly, each student can be part of at most one cycle. Every student in a cycle is assigned a seat at the school she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. The counters of the schools not in a cycle remain the same.

In general, for any $t = 1, 2, \ldots$

- Step t: Each remaining student points to her favorite school among the remaining schools and each remaining school points to the student with highest priority among the remaining students. There is at least one-cycle. Every student in a cycle is assigned a seat at the school that she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero the school is also removed.

This algorithm is very similar to the top trading cycles mechanisms described in Subsections 3.1 and 3.3, except that agents are not initially endowed with any good. In this adaptation, students are essentially swapping priority orderings with each other. Note that if every school has the same priority ordering, this mechanism reduces to serial dictatorship where the rank ordering is determined by the priority ranking.

Some of the main properties of TTC are different from those of the deferred acceptance mechanism although both mechanisms are strategy-proof. The student-proposing deferred acceptance mechanism is stable, but the resulting outcome is not necessarily Pareto efficient for students, while the top trading cycles mechanism is not stable but produces a Pareto efficient outcome for students. Whether efficiency or stability is more important is a question that may be an important determinant for the choice of the mechanism. Also, it is then natural to ask whether one can construct an efficient, strategy-proof mechanism which also produces a stable outcome whenever it exists. Kesten (2010) shows that this is impossible.

Much work within school choice literature has expounded on the results of Abdulkadiroğlu and Sönmez (2003). It is important, though, to address criticisms and weakness of the model as well as some difficulties in application of matching theory to the analysis of the policy.

[20]The description of the mechanism is taken directly from Abdulkadiroğlu and Sönmez (2003).

One of the central assumptions of the above model is that students have an exogenous preference over schools that is independent of the other students who are assigned to the same school. This is rather problematic if the quality of a school is effected by the composition of the student body (this is referred to as *peer effect*). The second issue is that the effect of a school choice mechanism on school quality is exogenously given and fixed in the canonical model, although the issue of improving schools takes a center stage of school choice debate in practice.[21] Another major difficulty is that the information submitted by students is ordinal and does not necessarily convey information on preference intensities. Abdulkadiroğlu, Che and Yasuda (2011, 2015) and Carroll (2017) analyze this issue theoretically. Agarwal and Somaini's (2016) empirical analysis on strategic reporting in school choice mechanisms highlighted the importance of further study on mechanisms that use the intensity of student preferences. Given these issues (and others we are not discussing here), it is still not completely clear whether the current notions of stability and Pareto efficiency are the most relevant measures by which to evaluate school choice mechanisms.

4.4 Matching with Constraints

We now proceed to a discussion of a relatively new area of research within matching theory and market design application, matching with constraints. This field seeks to study allocations and matching when characteristics and constraints other than the common individual capacity limits are regarded as desirable or required for feasibility. Schools, hospitals or firms (to use the language of our previous models) may be not only worried about the obvious limit on total individuals they can accept but also about the quantity of types of individuals that are admitted. With the prevalence of affirmative action and the goal of creating a diverse student/employee body, understanding the implementation and impact of such policies is crucial. The desire for diversity ranges beyond just race and gender: in universities, for instance, having students all interested in one or two academic areas is often considered disadvantageous because it may stymie the intellectual growth of its student population.

Abdulkadiroğlu and Sönmez (2003) model a simple affirmative action policy of type-specific quotas and propose mechanisms that satisfy the affirmative action constraints. Under the same type of affirmative action policy, Abdulkadiroğlu (2005) shows that a stable matching can be found using a strategy-proof, student-proposing deferred acceptance algorithm. These papers pushed affirmative action into mainstream matching literature, whereas traditional papers on affirmative action were based on "classical" mechanism design theory.[22] Kojima (2012) demonstrated various impossibility results that can arise when attempting to implement affirmative action policies in a matching environment. There are

[21] See Hatfield, Kojima and Narita (2016) for an analysis of this topic.

[22] The study of employment discrimination began in the second half of the 20th century. The two main theories of discrimination are a theory based on tastes, pioneered by Becker (1957), and a statistical theory, pushed forth by Phelps (1972) and Arrow (1973). Economists such as Glenn Loury and Roland Fryer have further developed the literature around race-based affirmative action.

situations where affirmative action policies inevitably hurt every minority student under *any stable matching mechanism*. Furthermore, similar impossibility results hold when using TTC. Hafalir, Yenmez and Yildirim (2013) further expound on these phenomena and show that the use of a "quota" vs "reserve" affirmative action system can have significant consequences on the resulting allocation. With minority reserves, schools give higher priority to minority students up to the point that the minorities fill the reserves. They show that the deferred acceptance algorithm with minority reserves is Pareto superior for students to the one with majority quotas.

Kamada and Kojima (2015) advance the idea by looking at matching environments with more general distributional constraints. One example is the Japan Residency Matching Program which imposes regional caps on the numbers of prospective residents so as to limit the concentration of residents in urban areas such as Tokyo. They point out that the mechanisms used in that market and others with constraints suffer from instability and inefficiency. To remedy this problem, they create a modified version of the deferred acceptance algorithm which is strategy-proof for students, constrained efficient, and stable in an appropriate sense. Kamada and Kojima (2016, 2017) and Goto et al. (2016) further explore various stability concepts and characterize environments in which stability and other desirable properties such as stratgy-proofness can be guaranteed.

There are still many issues and problems in the area that are unresolved and worth pursuing. How to address more general types of constraints, especially lower-bound constraints, is still a difficult problem and being actively studied (see Fragiadakis and Troyan (2016) for instance). New mathematical tools from discrete convex analysis have been applied to matching with constraints (Kojima, Tamura, and Yokoo 2016), but the use of such mathematical tools may warrant further investigation.

5 Conclusion

As indicated throughout this article, matching theory has expanded vastly since the seminal work by Gale and Shapley (1962). Although the theory has advanced considerably, there are many new questions and issues waiting to be explored further.

To begin with, almost all research in the existing literature defines stability under the assumption of complete information, but this is at best a rough approximation of reality. Liu, Mailath, Postlewaite, and Samuelson (2014) investigate stability under incomplete information in two-sided matching markets with transfer, while Bikhchandani (2014) studies a similar concept in the no-transfer setting.

Once incomplete information is taken seriously, it is natural to consider "informational externality", i.e., interdependence in valuations. Chakraborty, Citanna, and Ostrovsky (2010, 2015) study two-sided matching with interdependent values, while Che, Kim, and Kojima (2015) study one-sided matching with interdependent values. In both cases, the

possibility of extending desirable matching mechanisms from the standard private values models proved to be severely limited. Designing satisfactory mechanisms under interdependent values is a promising, if challenging, avenue for future research.[23]

Another important limitation of the existing literature is that the models tend to be static. Although some matching markets could be approximated well by a model of a static market (e.g., yearly medical residency matching or school choice), others may be better modeled as a dynamic market (e.g., daycare slots assignment with arrival and departure of children and the ongoing kidney exchange program). In addition to papers on dynamic kidney exchange already discussed, there is a burgeoning literature on dynamic two-sided matching markets. Kurino (2009), Du and Livne (2016), Doval (2017), and Kadam and Kotowski (2017) propose concepts of dynamic stability and analyze existence under various assumptions on commitment technologies and preferences. This literature is so young that several alternative stability concepts are being studied, but a consensus on the appropriate definition has not been reached yet. In the future, a consensus on the appropriate stability definition may emerge, but it is also possible that different stability concepts are appropriate in different types of dynamic markets. Reaching conclusions on this and other questions awaits further research.

References

[1] Abdulkadiroğlu, A. (2005). College admissions with affirmative action. *International Journal of Game Theory*, 33, 535–549.

[2] Abdulkadiroğlu, A. & Sönmez, T. (2013). Matching Markets: Theory and Practice. In D. Acemoglu et al. (Eds.), *Advances in Economics and Econometrics*. Cambridge University Press.

[3] Abdulkadiroğlu, Che. Y.K. & Yasuda, Y. (2011). Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered. *American Economic Review*, 101(1), 399–410.

[4] Abdulkadiroğlu, Che. Y.K. & Yasuda, Y. (2015). Expanding "choice" in school choice. *American Economic Journal: Microeconomics*, 7(1), 1–42.

[5] Abdulkadiroğlu, A. & Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3).

[6] Abdulkadiroğlu, A. & Sönmez, T. (1999). House allocation with existing tenants. *Journal of Economic Theory*, 88(2), 233-260.

[7] Abdulkadiroğlu, A. & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3), 729–747.

[23] See Hashimoto (2015) and Pakzad-Hurson (2016) for notable advances.

[8] Agarwal, N. & Somaini. (2016). Demand analysis using strategic reports: An application to a school choice mechanism. Working paper.

[9] Akbarpour, M., et al. (2016). Thickness and information in dynamic matching markets. Working paper.

[10] Anderson, R., Ashlagi, I., Gamarnik, D., & Kanoria, Y. (2015). Efficient dynamic barter exchange. *Operations Research*, forthcoming.

[11] Arrow, K. (1973). The theory of discrimination. In A.H. Pascal (Eds.), *Racial discrimination in economic life*. D.C. Heath.

[12] Ashlagi, I., Braverman, M. & Hassidim, A. (2014). Stability in large matching markets with complementarities. *Operations Research*, 62(4), 713–732.

[13] Azevedo, E.M. & Hatfiled, J.W (2017). Existence of equilibrium in large matching markets with complementarities. Working paper.

[14] Becker, G.S. (1957). *The economics of discrimination*. Chicago: University of Chicago Press.

[15] Bikhchandani, S. (2014). Two-sided matching with incomplete information. Working paper.

[16] Biro, P. & Klijn, F. (2013). Matching with couples: A multidisciplinary Survey. *International Game Theory Review* 15(2), 1–18.

[17] Bogomolnaia, A. & Moulin, H. (2001). A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2), 295–328.

[18] Budish, E. & Cantillon, E. (2012). The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *American Economic Review*, 102(5), 2237–2271.

[19] Carroll, G. (2017). On mechanisms eliciting ordinal preferences. Working paper.

[20] Chakraborty, A., Citanna, A., & Ostrovsky, M. (2010). Two-sided matching with interdependent values. *Journal of Economic Theory*, 145(1), 85–105.

[21] Chakraborty, A., Citanna, A., & Ostrovsky, M. (2015). Group stability in matching with interdependent values. *Review of Economic Design*, 19(1), 3–24.

[22] Che, Y. K., Kim, J., & Kojima, F. (2015). Efficient assignment with interdependent values. *Journal of Economic Theory*, 158, 54–86.

[23] Che. Y.K., Kim, J. & Kojima, F. (2017). Stable matching in large markets. Working paper.

[24] Chen, Y. & Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory*, 127(1), 202–231.

[25] Doval, L. (2017). A theory of stability in dynamic matching markets. Working paper.

[26] Du, S., & Livne Y. (2016). Rigidity of transfers and unraveling in matching markets. Working paper.

[27] Fragiadakis, D. & Troyan, P. (2016). Improving matching under hard distributional constraints. *Theoretical Economics*, forthcoming.

[28] Gale, D. & Shapley, L.S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 68(1), 9–15.

[29] Goto, M. et al. (2017). Designing matching mechanisms under general distributional constraints. *American Economic Journal: Microeconomics*, forthcoming.

[30] Hafalir, I.E., Yenmez, M.B., & Yildirim, M.A. (2013). Effective affirmative action in school choice. *Theoretical Economics*, 8(2), 325–363.

[31] Hashimoto, T. (2016). The generalized random priority mechanism with budgets. Working paper.

[32] Hassidim, A. et al. (2017). The mechanism is truthful, why aren't you? *American Economic Review Papers and Proceedings*, forthcoming.

[33] Hatfield, J.W., & Kojima, F. (2008). Matching with contracts: Comment. *American Economic Review*, 98, 1189-1194.

[34] Hatfield, J.W., Kojima, F. & Narita, Y. (2016). Improving schools through school choice: A mechanism design approach. *Journal of Economic Theory*, 166, 186-211.

[35] Hatfield, J.W., & Milgrom, P.R. (2005). Matching with contracts. *The American Economic Review*, 95(4), 913–935.

[36] Hylland, A. & Zeckhauser, R. (1979). The efficient allocation of individuals to positions. *Journal of Political Economy*, 87(2), 293–314.

[37] Kadam, S. & Kotowski, M. (2017) Multi-period matching. Working paper.

[38] Kamada, Y. & Kojima, F.(2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1), 67–99.

[39] Kamada, Y. & Kojima, F. (2016). Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition. Working paper.

[40] Kamada, Y. & Kojima, F.(2017). Stability concepts in matching under distributional constraints. *Journal of Economic Theory*, 168, 107–142.

[41] Kesten, O. (2010). School choice with consent. *The Quarterly Journal of Economics*, 125(3), 1297–1348.

[42] Kojima, F. (2015). Recent developments in matching theory and its practical applications. *Advances in Economics and Econometrics*. Cambridge University Press.

[43] Kojima, F. et al. (2013). Matching with couples: Stability and incentives in large markets. *The Quarterly Journal of Economics*, 128(4), 1585–1632.

[44] Kojima, F., Tamura, A., & Yokoo, M. (2016). Designing matching mechanisms under constraints: An approach from discrete convex analysis. Working paper.

[45] Kojima, F. & Troyan, P. (2011). Matching and market design: an introduction to selected topics. *Japanese Economic Review*, 62(1), 82–98.

[46] Kurino, M. (2009). Credibility, efficiency, and stability: A theory of dynamic matching markets. Working paper.

[47] Liu, Q., Mailath, G. J., Postlewaite, A., & Samuelson, L. (2014). Stable matching with incomplete information. *Econometrica*, 82(2), 541–587.

[48] Nguyen, T. & Vohra, R. (2017). Near feasible stable matchings with couples. Working paper.

[49] Pakzad-Hurson, B. (2016). Crowdsourcing and Optimal Market Design. Working Paper.

[50] Pathak, P.A. (2011). The mechanism design approach to student assignment. *Annual Review of Economics*, 3(1), 513–536.

[51] Pathak, P.A. (2015). What really matters in designing school choice mechanisms. *Advances in Economics and Econometrics*. Cambridge University Press.

[52] Phelps, E.S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(4), 659–661.

[53] Rees-Jones, A. (2017). Mistaken play in the deferred acceptance algorithm: Implications for positive assortative matching. *American Economic Review Papers and Proceedings*, forthcoming.

[54] Roth, A.E. (1982). Incentive compatibility in a market with indivisible goods. *Economics Letters*, 9(2), 127–132.

[55] 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(6), 991–1016.

[56] Roth, A.E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36(2), 277–288.

[57] Roth, A.E. (1991). A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *American Economic Review*, 81(3), 415–440.

[58] Roth, A.E. (2008a). Deferred acceptance algorithms: History, theory, practice, and open questions. *international Journal of game Theory*, 36, 537–569.

[59] Roth, A.E. (2008b). What we have learned from market design. *Economic Journal*, 118(527), 285–310.

[60] Roth, A.E & Peranson, E. (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4), 748–780.

[61] Roth, A.E. & Postlewaite, A. (1977). Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4(2), 131–137.

[62] Roth, A.E, Sönmez, T. & Ünver, M.U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, 119(2), 457–488.

[63] Roth, A.E, Sönmez, T. & Ünver, M.U. (2005). Pairwise kidney exchange. *Journal of Economic Theory*, 125(2), 151–188.

[64] Roth, A.E, Sönmez, T. & Ünver, M.U. (2007). Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review*, 97(3), 828–851.

[65] Roth, A.E., & Sotomayer, M.A.O. (1990). *Two-sided matching*. Cambridge: Cambridge University Press.

[66] Shapley, L. & Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1(1), 23–37.

[67] Sönmez, T. & Ünver, M.U. (2009). Matching, allocation, and the exchange of discrete resources. In J. Benhabib et al. (Eds.), *The Handbook of Social Economics*. Elsevier.

[68] Sönmez, T. & Ünver, M.U. (2010). Course bidding at business schools. *International Economic Review*, 51(1), 99–123.

[69] Sönmez, T. & Ünver, M.U. (2011). Market design for kidney exchange. In Z. Neeman et al. (Eds.), *The handbook of market design*. Oxford University Press.

[70] Ünver, M.U. (2010). Dynamic kidney exchange. *Review of Economic Studies*, 77(1), 372–414.