PART I

NONSTANDARD MARKETS

Cambridge University Press
978-1-107-01604-0 - Advances in Economics and Econometrics: TenthWorld Congress, Volume I: Economic Theory
Edited by Daron Acemoglu, Manuel Arellano and Eddie Dekel
Excerpt
More information

ONE

Matching Markets: Theory and Practice

Atila Abdulkadiroğlu and Tayfun Sönmez

1.0 Introduction

It has been almost a half-century since David Gale and Lloyd Shapley (1962) published their pathbreaking paper, "College Admissions and the Stability of Marriage," in *American Mathematical Monthly*. It is difficult to know whether Gale and Shapley expected the literature they initiated to be used to improve the lives of masses of people all around the world. We are fortunate to see that this is happening today.

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

In 1974, Lloyd Shapley and Herbert Scarf published a related paper, "On Cores and Indivisibility," in the first issue of the *Journal of Mathematical Economics*. Arguably, their model was the simplest exchange economy we could imagine. Each agent comes to the market with one indivisible good and seeks to trade it for more preferred goods that might be brought by other agents. In their simple model, agents are restricted to consuming

We thank Yeon-Koo Che and Eddie Dekel for their extensive comments that improved this survey.

only one good. Shapley and Scarf were interested in whether a core allocation exists in their model; they showed that indeed it does, and they presented a *Top Trading Cycles (TTC) algorithm* – which they attributed to David Gale – that achieves this objective. The basic ideas of Gale's TTC algorithm and its extensions resulted almost 30 years later in an organized kidney exchange in various parts of the world, which saves thousands of lives.

There are several 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) focused on the history of the deferred-acceptance algorithm. Essentially, both surveys are of two-sided matching markets. The content of our survey is closest to that of Sönmez and Ünver (2010), which focuses on one-sided as well as two-sided matching. In this survey, we give particular attention to the formal relations and links between these two original contributions and the recent matching literature that has impacted policy in various areas.

Following is an overview of our survey. In Section 2.0, we introduce and review key results of the two-sided matching model by Gale and Shapley (1962). In Section 3.0, we discuss the housing-market model by Shapley and Scarf (1974) and several more recent one-sided matching models, some of which are closely related to two-sided matching models. In Section 4.0, we present the recent developments in school choice. In Section 5.0, we discuss recent developments in kidney exchange. We present our conclusions in Section 6.0.

2.0 Two-Sided Matching

One key observation relating to Gale and Shapley's seminal contribution to real-life applications was made by Alvin Roth in 1984. He showed that the algorithm used to match medical residents to hospitals since the 1950s by the National Resident Matching Program is equivalent to a version of the celebrated deferred-acceptance algorithm (Roth 1984a). Since then, similar equivalences have been demonstrated by several authors. In this section, we summarize the two-sided matching literature, with a focus on discrete two-sided matching models without money. There is an important literature that studies versions of these models with money. Shapley and Shubik (1972) wrote the first paper to consider "continuous" matching markets, widely known as "assignment games." This model was studied later by Crawford and Knoer (1981) and Kelso and Crawford (1982), who showed many parallels with the discrete model. More recently, Hatfield and Milgrom (2005) provided 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
- $m \succ_m w$ means that woman w is unacceptable to man m

We use a 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. $\mu(m) \notin W \Rightarrow \mu(m) = m$ for all $m \in M$.
2. $\mu(w) \notin M \Rightarrow \mu(w) = w$ for all $w \in W$.
3. $\mu(m) = w \Leftrightarrow \mu(w) = m$ 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 or 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 one another to their partner under μ; that is:

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

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 not only in matching theory but also in 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 women who has not been holding an offer or any man who was rejected by all acceptable women remains single.

Theorem 2.1 (Theorems 1 and 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 in which 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 $\mu \vee^M \mu' : M \cup W \rightarrow M \cup W$ (*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:

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

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

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

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

Theorem 2.3: 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 also are 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 in which all men receive better mates than under the man-optimal stable matching. If, however, we are seeking an individually rational matching in which some man can receive better mates without hurting any man, it is not possible to match all men with strictly more-preferred mates.

Theorem 2.4 (Theorem 6 in Roth 1982b): 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 (1982b) shows that some of the men can receive more-preferred mates than under the man-optimal stable matching.

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

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

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

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

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 these 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 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 in which the set of players is $M \cup W$, the strategy space for player i is the set of his or 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:

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

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

Theorem 2.5 (Theorem 3 in Roth 1982b): 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 two-man, two-woman problem with the following preferences:

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

In this problem, there are only two stable matchings:

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

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, thereby forcing her favorite stable matching μ^W to be selected by φ because it is the only stable matching for the manipulated economy $(\succsim_{-w_1}, \succsim'_{w_1})$.

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

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

Theorem 2.6 (Proposition 1 in Alcalde and Barberá 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 1982b): 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 also is a weakly dominant strategy. We consider any such strategy also as truth-telling because 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 in which no agent plays a weakly dominated strategy is called a *Nash equilibrium in undominated strategies.*

Theorem 2.8 (Theorem 1 in Roth 1984b; 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 $\succsim$ iff 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 finite a set of colleges, I is a finite 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 $\succ_c, \succ_i$ denotes strict preferences derived from $\succsim_c, \succsim_i$.

Throughout this section, we assume that whether or not a student is acceptable for a college does not depend on other students in his or 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 $|\mu(c)| \leq q_c$ for all $c \in C$.