0.0.22 Matching

We’re jumping now to Chapter 9 and the topic of matching. This is a very hot topic in economic research right now. The problem in general can be described as follows. There are two sets, A and B. The elements of set A must be assigned to the elements of the set B. The elements of the set A are people who have strict preferences over the elements of B to which they may be assigned. Examples include:

- assigning students to dormitory rooms;
- admitting students to colleges and universities;
- pairing a graduate student with an academic advisor.

Campbell distinguishes between assignment, admissions and the marriage problems:

- an assignment problem is distinguished by the fact that the elements of the set B do not have preferences over the elements of the set A. Dormitory rooms, for instance, do not have preferences over the students who may be assigned to live in them.
- an admissions problem is distinguished by the fact that multiple elements of the set A may be assigned to a single element of the set B.
- a marriage problem is distinguished by the fact that elements of the two sets are either paired with each other (one to one), or they are not matched at all.

With reference to the marriage problem, it is often common in a matching problem to have a default option of being unmatched (i.e., remaining single, or not going to a college or university). We will discuss the procedures by which matchings are made. We typically want a matching to be individually rational in the sense that no one is matched with an alternative that he deems worse than being unmatched. If the matching is to be based upon the preferences of the individuals, then we will also want the matching to be incentive compatible in the sense that no individual can benefit by misrepresenting his true preferences.

Here are some current and topical matching problems:

- The matching of new physicians to medical residency programs of hospitals. New physicians (the elements of set A) typically spend 2-3 years in a medical residency program of a hospital (set B) during which time they receive additional training in a specialty. The hospitals are dependent on their residents as a form of relatively cheap labor that works long hours. Both sides have preferences over the other, and hospitals typically hire a number of doctors for each of its specialties. The National Residency Matching Program is a system implemented by the hospitals to organize the matching process; we’ll provide more details its operation later on. One might imagine instead have a free and competitive market in which doctors apply to hospitals and the hospitals hire who they want. This is in fact how the process worked before 1945. A problem, however, was unraveling of the market: because of competitive pressures, hospitals began signing up doctors earlier and earlier during their medical
training (e.g., in the middle of their medical school education, not in the final year). The properties of the resulting matching were questionable. Economists are always wary of market intervention, but in this case it seems to have improved the situation.

• The matching of new lawyers to judicial clerkships. There is no structure to this process. Judges receive applicants from new lawyers and they make offers. Offers are sometimes even "exploding" in the sense a judge may offer a new lawyer a position in his office on the condition that the offer expires if it is not accepted immediately. This can cause regret among the lawyers if they sense that they did not receive the best possible clerkship.

• Kidney exchange for transplant: matching donors to recipients. One possible donor if one needs a kidney for transplant is a cadaver. Each of us has two kidneys, however; if one is removed, then the remaining kidney can easily handle the role previously played by the pair. This leaves open the possibility of a living person donating a kidney to someone who needs a transplant. Typically, a loved one or relative may volunteer. The problem, however, is that there are many medical "markers" that determine the suitability of a particular kidney for transplant (e.g., the blood types of the potential donor and recipient). The person who volunteers his kidney for another may not be a suitable match. This creates the possibility of exchange: person A1 volunteers to give a kidney to person B1, but the match is poor. Person A2 volunteers to give a kidney to person B2, but the match is also poor. Suppose that A1 is a good match for B2 and A2 is a good match for B1. A trade could be arranged. In fact, you can imagine such trades occurring among larger numbers of donor-recipient pairs; these are called "chains".

• School choice. It is increasingly common for school districts to allow students to have some ability to choose among the schools in the district. A student, for instance, may want the school closest to his home, or his parent may want the student a school close to his or her place of work. A student may also wish to attend a school with particular honors programs or areas of expertise (e.g., math and science, or the arts). Unlike universities and colleges, schools typically do not allow schools to express preferences over students (except for matters of qualification, such as minimal test scores). The school district instead typically specifies a priority system for each school in determining which students receive the available slots. A student typically receives priority, for instance, if he lives within walking distance of the school or if he has a sibling who already attends the school. A variety of procedures are used in practice around the United States to match students to schools, including the "Boston mechanism" and the "Student Optimal Stable Mechanism."

A salient property of each of these four examples is that money is either immaterial or second order in its importance in comparison with the match: doctors have regional preferences over hospitals (e.g., "I want to be matched near my spouse’s match"), but their priority is typically the quality and prestige of the residency program. The point of the residency is not to make money; it is to gain worthwhile experience that may in turn provide a lifetime of opportunities and benefits. The same issues apply in judicial clerkships. While money may change hands in gaining admission to private schools, it
Remark 2 As a side note, here’s a counter argument. Suppose I’m dying of kidney disease and I’m wealthy. A high quality match for me is identified in an underdeveloped country. I pay this person $100,000 for his kidney. After several days in bed, he recovers and now has a range of opportunities (e.g., business, education, travel) for his life that he didn’t have before he denoted his kidney. Is this wrong? Why?

We can model these problems as two distinct sets such that we wish to match elements of one set to elements of the other. The elements of the two sets have preferences and so the matching creates welfare. It is clear that some matchings may be better from a welfare perspective than others. Another feature of these problems is money is either insignificant or can play no role in allocation of welfare.

Roth and Niederle have identified three types of market failure in which a matching algorithm or clearinghouse may be an appropriate remedy:
1. Unraveling so that offers were being made at earlier and at dispersed times. This is a form of "thinness" in the market.
2. Congestion so that employers found that they did not have the time to make all of the offers that they wished to make.
3. Strategic behavior in the sense that participants are concerned that they cannot act straightforwardly based upon their true preferences.

The use of a matching algorithm is a form of market intervention. Economists are generally wary of intervening in markets, though the problems that Roth and Niederle motivate intervention. What are the attributes of a market in which a matching algorithm may be helpful? I can list four possible attributes:
1. The goods that are traded are extremely heterogeneous. For instance, doctors vary by both specialities and their intellects.
2. There is extreme excess demand for some goods, which is what causes the market unraveling.
3. Money is "second-order" in preferences. Lawyers who seek judicial clerkships, for instance, bear the opportunity cost of extremely large salaries in corporate law. The point of a clerkship is both the high-level experience and its signaling value on one’s resume throughout the remainder of one’s career. Perhaps the new lawyers even have lexicographic preferences where the quality of the clerkship counts first and salary counts second (i.e., there is no trade-off between the quality of the clerkship and salary). It is difficult for the price mechanism to successfully allocate goods in a situation in which the participants are not particularly concerned about the prices.
4. Roth emphasizes that traders on both sides of many of these markets have strong preferences over who they are matched with. Buyers typically care only about the good and sellers only about the price. A person selling a house, for instance, is concerned mainly about the price and not who buys the house.

We understand how externalities cause market failure and thus motivate market intervention, and we understand how the attributes of a public good cause underprovision of that good. The four attributes above are perhaps a step towards similarly understanding in a formal theoretical sense when a matching algorithm may be needed to improve upon a market’s allocation.

0.0.23 A Simple Model of a Marriage Market

We consider a simple model of a marriage market. There are two finite and disjoint sets \( M \) (men) and \( W \) (women),

\[
M = \{m_1, m_2, ..., m_n\}
\]

and

\[
W = \{w_1, w_2, ..., w_p\}.
\]

We assume that each man \( m_i \) has preferences over the set \( W \cup \{m_i\} \). For example,

\[
P(m_i) = w_1, w_5, m_i, w_2, ...
\]

indicates that man \( m_i \) prefers woman \( w_1 \) to woman \( w_5 \), but prefers staying single (i.e., matching with himself) to all other women. We use brackets to indicate indifference, i.e.,

\[
P(m_j) = [w_1, w_2], w_4, w_5, ...
\]

means that man \( m_j \) is indifferent between women \( w_1 \) and \( w_2 \), but prefers each strictly to \( w_4 \),... We assume that each woman similarly has preferences over the set of men. We often abbreviate preferences by omitting every person after oneself on one’s list. Marriage is voluntary and those of the opposite sex that one ranks below staying single are typically irrelevant in the discussion. A member \( B \) of the opposite sex from person \( A \) is acceptable to person \( A \) if person \( A \) does not strictly prefer being single to being matched with \( B \).

**Definition 2** A matching \( \mu \) is a one-to-one function from the set \( M \cup W \) to itself such that:

1. \( \mu^2(x) = x \) for all \( x \in M \cup W \).
2. \( \mu(m) \in W \cup \{m\} \) for all \( m \in M \) and \( \mu(w) \in M \cup \{w\} \) for all \( w \in W \).

The matching \( \mu \) identifies the mate of each man \( m \) and woman \( w \). If a person’s mate is him or her self, then the person stays single (or is unmatched). Condition 1. states that if person \( A \) is matched to person \( B \), then person \( B \) is matched to person \( A \). Condition 2. states that a person is either matched to a person of the opposite sex or remains single (there is no "same sex"marriage in this model).

An individual blocks a matching \( \mu \) if it matches him to someone he/she deems unacceptable. A pair \((m, w)\) block \( \mu \) if they are not matched in \( \mu \) but each strictly prefers the other to their matches under \( \mu \). A matching \( \mu \) is stable if it is not blocked by any individual or couple \((m, w)\).
It is desirable that a matching be stable. Consider a procedure by which doctors are matched to hospitals for medical residencies, or lawyers are matched to judges for judicial clerkships. If the procedure produces an unstable matching, then either an individual participant will wish to opt out of his assignment or some pair will wish to form their own agreement outside of the procedure. This literally makes the matching and the procedure unstable.

**Theorem 13 (Gale and Shapley)** A stable matching exists for every marriage market.

The proof of the theorem is important in its own right because it is constructive: a procedure or algorithm is presented that finds a stable matching. This is the deferred acceptance algorithm and it turns out to be very useful. For the moment, we assume that men and women honestly report their preferences. We’ll deal with the issue of incentive compatibility in this algorithm below.

**Proof.** Assume for the moment that each man/woman has strict preferences. The algorithm works as follows:

- **Stage 1:** Each man proposes to his top-ranked woman; if he prefers being single to all women, then he makes no proposals in this or any future stages and "exits" the marriage market. A woman rejects any unacceptable man who proposes; if at least one proposal is acceptable, then she rejects all but her favorite and keeps him as "engaged".

  ;

- **Stage i:** Each man who has not exited and who is currently not engaged proposes to the top-ranked woman among those who have not already rejected him. If no such woman exists, then he exits the marriage market and makes no further proposals. A woman rejects all unacceptable men and becomes engaged to her favorite in the set consisting of those men who propose to her in stage i together with anyone with whom she is currently engaged.

The algorithm stops at a stage in which there are no men who wish to make proposals, i.e., all men have either left the marriage market or are engaged. Such a stage is reached because the number of people is finite. At this stage, each engaged man is matched to the woman he is engaged with and all other people remain single.

The algorithm extends to the case in which preferences are not necessarily strict by introducing a tie-breaking rule for each person (e.g., in the case of indifference between two prospective mates, an individual ranks them alphabetically, or chooses the younger, etc.). This effectively makes each person’s preferences strict.

Is the matching that results from this algorithm stable? We first consider individuals. A man who is matched would not prefer to be single, for he would never have proposed to a woman who is unacceptable to him. Similarly, a woman only becomes engaged to men who are acceptable to her. Therefore, no individual would block this matching.

Suppose there is a couple \((m, w)\) that blocks \(\mu\). This means that male \(m\) strictly prefers \(w\) to being single or his mate \(\mu(m)\). Consequently, he must have proposed to
Example 50

To compare to the stable matching obtained by having the women make the proposals, let's consider the stable matching derived by having men make the proposals. The roles of men and women are symmetric in this model, though not in the algorithm we stated to this point. If we reverse the roles of the sexes in this algorithm and allow women to make the proposals, then the same argument shows that the resulting matching is stable. How does the stable matching derived by having men make the proposals compare to the stable matching obtained by having the women make the proposals?

The fact that he was not matched with \( w \) means that she rejected him in favor of a man she prefers, and so \( \mu(w) > w \). This contradicts \((m, w)\) blocking \( \mu \). We conclude that \( \mu \) is stable.

### Example 49

Let's compare the two matchings:

\[
\begin{align*}
P(m_1) &= w_1, w_2, w_3, w_4 \quad P(w_1) = m_2, m_3, m_1, m_4, m_5 \\
P(m_2) &= w_4, w_2, w_3, w_1 \\
P(m_3) &= w_4, w_3, w_1, w_2 \\
P(m_4) &= w_1, w_4, w_3, w_2 \\
P(m_5) &= w_1, w_2, w_4
\end{align*}
\]

**Stage 1:** Proposals: \( m_1 \rightarrow w_1, m_2 \rightarrow w_4, m_3 \rightarrow w_4, m_4 \rightarrow w_1, m_5 \rightarrow w_1 \), which results in the engagements

\[
\begin{align*}
w_1 &\rightarrow m_1 & w_2 &\rightarrow m_2 & w_3 &\rightarrow w_4 & w_4 &\rightarrow w_1 & m_5 &\rightarrow w_1
\end{align*}
\]

**Stage 2:** Proposals: \( m_3 \rightarrow w_3, m_4 \rightarrow w_4, m_5 \rightarrow w_2 \), which results in the engagements

\[
\begin{align*}
w_1 &\rightarrow m_1 & w_2 &\rightarrow m_2 & w_3 &\rightarrow w_3 & w_4 &\rightarrow w_2 & m_5 &\rightarrow m_5
\end{align*}
\]

**Stage 3:** Proposals: \( m_2 \rightarrow w_2 \), which results in the engagements

\[
\begin{align*}
w_1 &\rightarrow m_1 & w_2 &\rightarrow m_2 & w_3 &\rightarrow w_4 & m_5 &\rightarrow w_5
\end{align*}
\]

**Stage 4:** Proposals: \( m_5 \rightarrow w_4 \), which \( w_4 \) rejects, leaving

\[
\begin{align*}
w_1 &\rightarrow m_1 & w_2 &\rightarrow w_2 & w_3 &\rightarrow w_4 & m_5 &\rightarrow w_4
\end{align*}
\]

At this point, no more proposals will be made, and so the matching is determined.

The roles of men and women are symmetric in this model, though not in the algorithm we stated to this point. If we reverse the roles of the sexes in this algorithm and allow women to make the proposals, then the same argument shows that the resulting matching is stable. How does the stable matching derived by having men make the proposals compare to the stable matching obtained by having the women make the proposals?

The fact that he was not matched with \( w \) means that she rejected him in favor of a man she prefers, and so \( \mu(w) > w \). This contradicts \((m, w)\) blocking \( \mu \). We conclude that \( \mu \) is stable.

### Example 50

\[
\begin{align*}
P(m_1) &= w_1, w_2, w_3, w_4 \quad P(w_1) = m_2, m_3, m_1, m_4, m_5 \\
P(m_2) &= w_4, w_2, w_3, w_1 \\
P(m_3) &= w_4, w_3, w_1, w_2 \\
P(m_4) &= w_1, w_4, w_3, w_2 \\
P(m_5) &= w_1, w_2, w_4
\end{align*}
\]

**Stage 1:** Proposals: \( w_1 \rightarrow m_2, w_2 \rightarrow m_3, w_3 \rightarrow w_5, w_4 \rightarrow m_1 \), which results in the engagements

\[
\begin{align*}
m_1 &\rightarrow m_2 & m_2 &\rightarrow m_3 & w_3 &\rightarrow w_4 & w_4 &\rightarrow m_1
\end{align*}
\]

At this point, no more proposals will be made, and so the matching is determined.

Let's compare the two matchings:

\[
\begin{align*}
\text{men propose:} & \quad w_1 \quad w_2 \quad w_3 \quad w_4 \\
\text{women propose:} & \quad m_1 \quad m_2 \quad m_3 \quad m_4
\end{align*}
\]

114
women propose: $m_1 \ m_2 \ m_3 \ m_4 \ m_5$

In this example, each woman is matched with her favorite man in the algorithm in which women make the proposals. There clearly cannot be a better stable matching for the women (or even a better unstable matching).

When the men propose, the matching is:

$w_1 \ w_2 \ w_3 \ w_4$

The number in the bottom row indicates where each woman ranks in the man’s list. There is a sense in which this is the best possible stable matching for men. We’ll discuss this point below.

0.0.24 Incentives

We will not prove the following results, but merely list them for the sake of completeness:

- Impossibility Theorem (Roth; R&S, Thm. 4.4): No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.

- It is a dominant strategy for each man to report his true preferences in the deferred acceptance algorithm in which men make the proposals, and similarly for women in the deferred acceptance algorithm in which women make the proposals. In applications such as the matching of students to schools, the problem of misrepresentation can be solved for at least the students and their parents. The problem of misrepresentation of preferences may be less of an issue when the schools are all part of the same public school system.

Example 51 Let’s go back to the first example that we considered, where the men make the proposals:

$P(m_1) = w_1, w_2, w_3, w_4 \quad P(w_1) = m_2, m_3, m_1, m_4, m_5$

$P(m_2) = w_1, w_2, w_3, w_4 \quad P(w_2) = m_3, m_1, m_2, m_4, m_5$

$P(m_3) = w_1, w_3, w_1, w_2 \quad P(w_3) = m_5, m_4, m_1, m_2, m_3$

$P(m_4) = w_1, w_4, w_3, w_2 \quad P(w_4) = m_1, m_4, m_5, m_2, m_3$

$P(m_5) = w_1, w_2, w_4$

Stage 1: proposals: $m_1 \rightarrow w_1, m_2 \rightarrow w_4, m_3 \rightarrow w_4, m_4 \rightarrow w_1, m_5 \rightarrow w_1$, which results in the engagements

$w_1 \ w_2 \ w_3 \ w_4$

$w_1 \ w_2 \ w_3 \ w_4$

Stage 2: proposals: $m_3 \rightarrow w_3, m_4 \rightarrow w_4, m_5 \rightarrow w_2$, which results in the engagements

$w_1 \ w_2 \ w_3 \ w_4$

m1 m5 m3 m4

Stage 3: proposals: $m_2 \rightarrow w_2$, which results in the engagements

w1 w2 w3 w4

m1 m2 m3 m4
Stage 4: proposals: $m_5 \rightarrow w_4$, which $w_4$ rejects, leaving

\begin{align*}
& w_1 \\ & w_2 \\ & w_3 \\ & w_4 \\ & m_1 \\ & m_2 \\ & m_3 \\ & m_4
\end{align*}

At this point, no more proposals will be made, and so the matching is determined. Notice that $w_3$ is matched with $m_3$, who she ranks as worst among the 5 men. Suppose she rejects $m_3$ at stage 2, leaving

\begin{align*}
& w_1 \\ & w_2 \\ & w_3 \\ & w_4 \\ & m_1 \\ & m_2 \\ & m_4
\end{align*}

Stage 3$: proposals: $m_2 \rightarrow w_2$, and $m_3 \rightarrow w_1$, which results in the engagements

\begin{align*}
& w_1 \\ & w_2 \\ & w_3 \\ & w_4 \\ & m_3 \\ & m_2 \\ & m_4
\end{align*}

Stage 4$: proposals: $m_1 \rightarrow w_2$, $m_5 \rightarrow w_4$, which $w_4$ rejects, leaving

\begin{align*}
& w_1 \\ & w_2 \\ & w_3 \\ & w_4 \\ & m_3 \\ & m_1 \\ & m_4
\end{align*}

Stage 5$: proposals: $m_2 \rightarrow w_3$, leaving

\begin{align*}
& w_1 \\ & w_2 \\ & w_3 \\ & w_4 \\ & m_3 \\ & m_1 \\ & m_2 \\ & m_4
\end{align*}

Thus, $w_3$ can improve her match by rejecting an acceptable man.

**Definition 3** Man $m$ and woman $w$ are **achievable** for each other in a marriage market if they are matched together in some stable matching.

**Definition 4** A stable matching $\mu$ is **M-optimal** if every man weakly prefers his match in $\mu$ to his match in any other stable matching. A stable matching $\mu$ is **W-optimal** if every woman prefers her match in $\mu$ to her match in any other stable matching.

**Theorem 14 (Gale and Shapley)** If the preferences of each man and each woman are strict, then the matching produced by the deferred acceptance algorithm in which men make the proposals is M-optimal and the matching that is produced in this algorithm when women make the proposals is W-optimal.

**Theorem 15** If the rankings of each man and woman are strict, then a stable matching is efficient.

**Theorem 16** Impossibility Theorem (Roth; R&S, Thm. 4.4): No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.