22 Stable matching

Chapter summary

In this chapter we present the subject of *stable matching*. Introduced in 1962 by David Gale and Lloyd Shapley, stable matching became the starting point of a rich literature on matching problems in two-sided markets (e.g., workers and employers, interns and hospitals, students and universities), and remains one of the most applied areas in game theory to date.

We present Gale and Shapley's basic model of matching *men* to *women*, the concept of stable matching, and an algorithm for finding it. It is proved that the set of stable matchings has a lattice structure based on the preferences of women and men. We then study several variations of the model: the case in which there are more men than women; the case in which bachelorhood is not the worst outcome; the case of many-to-one matchings (e.g., many students to one college); and matchings in a single-gender population. It is also shown that the Gale–Shapley algorithm is not immune to strategic manipulations.

The study of the subject of this chapter began at the end of the nineteenth century, with the introduction of residency requirements for recent medical school graduates. Fresh medical school graduates needed to find a hospital in which to pursue their medical internships. Over the years, the residents played an increasingly important role in the staffs of hospitals, and hospitals began competing with each other for the best medical school graduates. To get a jump on the competition, hospitals kept moving up the dates on which they granted medical students residency positions. By 1944, medical students beginning their third year of medical school (out of four years) were interviewed for residency positions. This state of affairs did not serve the interests of either students or hospitals, and as the situation only deteriorated further, medical schools and hospitals agreed in 1951 to adopt a formal system for matching graduating medical students with hospitals, beginning the following year. The system works as follows: after interviews are conducted, the fourth-year medical students rank the hospitals at which they were interviewed, while at the same time the hospitals rank the students that they interviewed. Each hospital also announces the maximum number of residents that it can hire. The data are collected by a special national resident matching program, which then inputs the data into an algorithm composed for the purpose of matching hospitals and residents. The algorithm takes into account the preferences of both the hospitals and the medical students with an attempt to arrive at a "best fit."

Participation in this central mechanism, on the part of both hospitals and medical students, is voluntary. The satisfaction that all parties derive from a good suggested "matching" between residents and hospitals is therefore a significant factor in determining the extent of participation in the matching system and its success.

It is not easy to arrive at matches that please everyone. For example, if Mark prefers Massachusetts General Hospital to Johns Hopkins Hospital, and Massachusetts General Hospital prefers Mark to Andrew, it would be unwise to send Mark to Johns Hopkins and Andrew to Massachusetts General. Such a matching will generate dissatisfaction for both Mark and Massachusetts General, and both of them will then have an incentive to disregard the matching program. If too many such cases multiply, the entire system could be abandoned (this actually happened in the United Kingdom, where the system matching house officer posts at British hospitals with medical students led to so many unsatisfactory matches that it fell into disuse).

The beauty of the algorithm used by the American national resident matching program is that it leads to matches in which no pair is dissatisfied: if Mark is matched with Johns Hopkins Hospital, then Massachusetts General Hospital must have been matched with residents whom it preferred to Mark, and Mark has no justified complaint, because his preferred hospital simply preferred others to him; no injustice was involved, nor any inefficiency in the algorithm.

The problem of matching elements from two different populations is not limited to the example of hospitals and potential residents. Two additional examples that may be adduced are matching workers and employers, and matching men and women in couples.

In 1962, David Gale and Lloyd Shapley published a paper defining the matching problem and the concept of "stable matching." In that paper, they also proved that stable matchings always exist, and spelled out an algorithm for computing stable matchings. Several years later, Alvin Roth connected the Gale–Shapley result to the algorithms used to match residents and hospitals in the United States, by showing that the algorithms used in the resident matching system created stable matchings according to Gale and Shapley's definition of the term.

The subject of matching raises many natural questions: What is the best definition of a stable placement of candidates for residency with open positions? Do such stable placements always exist? If so, how can they be found? Can a hospital (or a candidate for residency) obtain a more satisfactory placement by submitting a preference ordering over candidates that differs from the honest preference ordering?

In this chapter, we will first consider the simple case in which the number of residency candidates equals the number of hospitals, and each hospital is seeking only one resident. Such a situation better describes the matching of men and women in married couples, and we will therefore use the language of that metaphor in analyzing this problem. The matching situation we consider then consists of n men and n women. Each man orders the women in decreasing order, from the woman he most prefers to the woman he least prefers as a mate, and each woman similarly orders the men in decreasing order of preference. The goal is to match each man to one woman in such a way that no complaints will be registered: if Julius is matched to Cornelia and Mark is matched to Cleopatra, then Julius and Cleopatra should not leave their spouses for each other: either Julius prefers Cornelia to Cleopatra or Cleopatra prefers Mark to Julius (or both).

Later in the chapter we will study extensions of this basic model.

22.1 The model

We begin by recalling the definition of a preference relation. In words, a preference relation enables us to compare any two elements of a set, and state which of the two is more preferred.

Definition 22.1 *Let X be a set. A* preference relation[1] *over X is a binary relation $\succ$ satisfying the following properties:*

- *For every $x \neq y$, either $x \succ y$ or $y \succ x$ (the relation is* complete*; i.e., every pair of distinct elements can be compared).*
- *$x \not\succ x$ (the relation is* irreflexive*; i.e., x is not preferred to itself).*
- *If $x \succ y$, and $y \succ z$, then $x \succ z$ (the relation is* transitive*; i.e., if x is preferred to y, and y is preferred to z, then x is also preferred to z).*

Every complete, irreflexive, and transitive relation is *asymmetric*: if $x \neq y$, then $x \succ y$ if and only if $y \not\succ x$ (Exercise 22.1).

Note that a preference relation, as we have defined it, represents strict preferences; we are not allowing for the possibility of indifference. We are assuming this for the sake of simplifying the analysis. Some of the results presented in this chapter can be generalized to preference relations with indifference (see Exercise 22.35).

We are now ready for the formal definition of a matching problem.

Definition 22.2 *A* matching problem *is given by:*

- *A natural number n representing the number of men and the number of women in a population (thus, we assume that the number of women equals the number of men).*
- *Every woman has a preference relation over the set of men.*
- *Every man has a preference relation over the set of women.*

The set of women will be denoted by W, and an element in that set is denoted by w. The set of men will be denoted by M, and an element in that set is denoted by m. The fact that a woman w prefers a man m_1 to a man m_2 is denoted

$$w : m_1 \succ m_2. \tag{22.1}$$

For example, "Cleopatra prefers Julius to Mark" is denoted as:

$$\text{Cleopatra: Julius} \succ \text{Mark.}$$

Definition 22.3 *A* matching *is a bijection from the set of men to the set of women.*

Equivalently, a matching is a collection of n pairs $\{(w_1, m_1), (w_2, m_2), \ldots, (w_n, m_n)\}$ such that $\{m_1, m_2, \ldots, m_n\} = M$ and $\{w_1, w_2, \ldots, w_n\} = W$. If a pair (w, m) is included in a matching, then we say that the man m is matched to the woman w (or that the woman w

1 The definition appearing here is of a complete and strict ordering relation (meaning that there is no possibility of indifference between any two elements). In other places in this book (Chapters 2 and 21) we have presented definitions of ordering relations that satisfy other properties.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

is matched to the man m). We will henceforth denote matchings using the letters A, B, or C, etc.

Definition 22.4 *A man and a woman* object *to a matching, if they prefer each other to the mates to whom they are matched under the matching. A matching is* stable *if there is no pair consisting of a man and a woman who have an objection to the matching.*

The following definition is equivalent to that of a stable matching (Exercise 22.4).

Definition 22.5 *A matching A is* stable *if in every case that a man prefers another woman to the woman to whom he is matched under A, that woman prefers the man to whom she is matched to him.*

The definition may be similarly phrased by interchanging the roles of the men and the women: matching A is *stable* if in every case that a woman prefers another man to the man to whom she is matched under A, that man prefers the woman to whom he is matched to her.

Example 22.6 Consider the following example, where $n = 4$: the set of men is {Adam, Ben, Charles, Dean}, and the set of women is {Anne, Bess, Carol, Donna}. The preferences of these men and women are presented in Figure 22.1. The preferences of the women appear in the lower right-hand side of each cell (read vertically) and the preferences of the men appear in the upper left-hand side of each cell (read horizontally). For example, in the upper left cell of the table, corresponding to the pair Anne – Adam, the numbers 2 and 4 are listed: Adam is second on Anne's preference list, and Anne is fourth on Adam's preference list.

The matching depicted in Figure 22.1 by stars, and detailed again in Figure 22.2, is not stable. This is because Carol and Adam have an objection to the matching: Carol prefers Adam (number 2 on her list) to Charles (number 3 on her list), and Adam prefers Carol (number 2 on his list) to Anne (number 4 on his list).

	Anne	Bess	Carol	Donna
Adam	4 ★ 2	3 2	2 ♣ 2	1 4
Ben	2 ♣ 1	1 ★ 4	3 4	4 3
Charles	4 3	1 ♣ 1	3 ★ 3	2 2
Dean	1 4	4 3	3 1	2 ★ ♣ 1

Figure 22.1 The preference relations of the men and the women, along with two matchings (one denoted by ★, and the other by ♣)

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

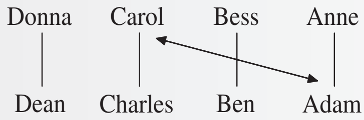

Figure 22.2 An unstable matching (denoted by ★ in Figure 22.1) and an objection to that matching

In contrast, the matching

$$(\text{Adam} - \text{Carol}, \quad \text{Ben} - \text{Anne}, \quad \text{Charles} - \text{Bess}, \quad \text{Dean} - \text{Donna}),$$

depicted in Figure 22.1 by ♣, is stable. To see this, note that Anne, Bess, and Donna are all matched with the men who are number 1 on their lists, so that none of them will object to the matching with any man. Carol is matched with Adam, who is number 2 on her list, and therefore the only possible objection she may have is with Dean, who is number 1 on her list. But Dean prefers Donna (number 2 on his list) to Carol (number 3 on his list). This matching is thus stable, because no pair consisting of a man and a woman has an objection. ◀

22.2 Existence of stable matching: the men's courtship algorithm

The first theorem we prove states that there always exists a stable matching. The proof is attained by presenting an algorithm that leads to a stable matching. The algorithm is due to Gale and Shapley [1962].

Theorem 22.7 *To every matching problem there exists a stable matching.*

Proof:
Step 1: Description of the algorithm.
Consider the following algorithm:

1. Stage 1(a): Every man goes to stand in front of the house of the woman he most prefers.
2. Stage 1(b): Every woman asks the man whom she most prefers from among the men standing in front of her house, if there are any, to wait, and dismisses all the other men.
3. Stage 2(a): Every man who was dismissed by a woman in the first stage goes to stand in front of the house of the woman he most prefers from among the women who have not previously dismissed him (i.e., the woman who is second on his list).
4. Stage 2(b): Every woman asks the man whom she most prefers from among the men standing in front of her house, if there are any (including the man whom she told to wait in the previous stage), to wait, and dismisses all the other men. In general:
5. Stage k(a): Every man who was dismissed by a woman in the previous stage goes to stand in front of the house of the woman he most prefers from among the women who have not previously dismissed him.
6. Stage k(b): Every woman asks the man whom she most prefers from among the men standing in front of her house, if there are any, to wait, and dismisses all the other men.
7. The algorithm terminates when there is one man standing in front of every woman's house.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

It is possible, in principle, that a particular man will be dismissed by every woman. We will show that this cannot happen. The algorithm will always terminate, and every woman will have one man standing in front of her house. The algorithm therefore always terminates by finding a matching. We will further prove that the algorithm always terminates by finding a stable matching.

Before proceeding to the proof of the theorem, we note that the algorithm satisfies the following three properties that will later be needed. The reader can readily ascertain that the first and third properties are satisfied. The reader is further asked to show that the second property is also satisfied, in Exercise 22.2.

(1) The preferred women have been courted in the past: If, in stage k, Henry stands in front of the house of Anne, but prefers Catherine to Anne, then it must be the case that he previously courted Catherine by standing in front of her house, and was dismissed. In other words, the men go down their preference list along the course of the algorithm.

(2) The preferred men will come courting in the future: If, in stage k, Cleopatra asks Mark to wait, and in a later stage asks Julius to wait, then she prefers Julius to Mark. In other words, the women go up their preference list along the course of the algorithm.

(3) Once a woman is courted, she will always be courted: If, in stage k, Mark stands in front of Cleopatra's house, then from stage k onwards, there will always be at least one man courting Cleopatra by standing in front of her house.

Step 2: The algorithm terminates in a finite number of stages and produces a matching.
We first show that there is a stage at which no man is dismissed. A man who is dismissed by a women never returns to court her again. This means that each woman dismisses at most $n-1$ men. It follows that after at most $n(n-1) = n^2 - n$ stages, there are no more rejections, and therefore after at most $n^2 - n + 1$ stages we arrive at a stage at which no woman dismisses any man.[2]

We next show that a man cannot be dismissed by all women. Assume by way of contradiction that there exists a man, let's call him Joe, who was dismissed by all the women. By the algorithm's construction, Joe must have paid a visit to the house of every woman. By the "once a women is courted, she will always be courted" property, after Joe has courted every woman, every woman has a man standing in front of her house. Since the number of women equals the number of men, and Joe is not standing in front of any woman's house, there must be a woman who has no man in front of her house, a contradiction. Hence there cannot be a man who has been dismissed by every woman.

It follows that the algorithm terminates by producing a matching. Indeed, if at the end of any particular stage no man is sent home, it follows that in front of each woman's house there is one and only one man left standing, who is then her mate. A matching has been attained.

Step 3: The resulting matching is stable.
Suppose by contradiction that the resulting matching is unstable. Then there exist at least one man and one woman, let's call them Julius and Cleopatra, who prefer each other to the

2 In fact, the algorithm terminates after at most $(n-1)^2 + 1$ stages (see Exercise 22.13).

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

mates to which they have been matched under the algorithm. Suppose that the algorithm has matched Julius to Cornelia, and Cleopatra to Mark.

Since Julius prefers Cleopatra to Cornelia, by "the preferred women were courted in the past" property, Julius must have courted Cleopatra before he courted Cornelia. Since Cleopatra has been matched by the algorithm to Mark, by "the preferred men will come courting in the future" property, she prefers Mark to Julius. This contradicts the assumption that Julius and Cleopatra have an objection to the matching. □

Example 22.6 (*Continued*) We apply the matching algorithm to Example 22.6. The table in Figure 22.3 describes the run of the algorithm.

	Anne	Bess	Carol	Donna	Dismissed men
Stage 1(a)	Dean(1)	Ben(1), Charles(1)		Adam(1)	
Stage 1(b)	Dean(1)	Charles(1)		Adam(1)	Ben
Stage 2(a)	Dean(1), Ben(2)	Charles(1)		Adam(1)	
Stage 2(b)	Ben	Charles(1)		Adam(1)	Dean
Stage 3(a)	Ben(2)	Charles(1)		Adam(1), Dean(2)	
Stage 3(b)	Ben(2)	Charles(1)		Dean(2)	Adam
Stage 4(a)	Ben(2)	Charles(1)	Adam(2)	Dean(2)	

Figure 22.3 Men's courtship algorithm

Each stage of the run of the algorithm is described by two consecutive rows. In the row corresponding to part (a) we note the men who are standing in front of women's houses after the application of part (a) of that stage (prior to the dismissals of the women), and in the row corresponding to part (b) we note the men who are standing in front of women's houses after the application of part (b) of that stage (after the women have announced their dismissals). The number appearing by the name of each man represents the ranking of the woman in front of whose house he is standing, in his preference relation.

The algorithm ends with the matching:

$$\text{(Adam – Carol,} \quad \text{Ben – Anne,} \quad \text{Charles – Bess,} \quad \text{Dean – Donna),}$$

which is the stable matching previously mentioned. ◀

22.3 The women's courtship algorithm

In the algorithm presented in the previous section, the men courted the women, and the women kept the men waiting or dismissed them. We call the resulting matching a "*men's courtship matching*," and denote it by O^m. The roles of the men and women, however, may be reversed, with women taking the courting initiative, going to the mens' houses, with each man keeping only the most preferred woman from among those standing in front of his house and dismissing all the others. By the proof of Theorem 22.7, this algorithm

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

also leads to a stable matching, which we call the "*women's courtship matching*," denoted by O^w.

Example 22.6 (*Continued*) We apply the women's courtship algorithm to Example 22.6 (see Figure 22.4).

	Adam	Ben	Charles	Dean	Dismissed women
Stage 1(a)		Anne(1)	Bess(1)	Carol(1), Donna(1)	
Stage 1(b)		Anne(1)	Bess(1)	Donna(1)	Carol
Stage 2(a)	Carol(2)	Anne(1)	Bess(1)	Donna(1)	

Figure 22.4 Women's courtship algorithm

The algorithm ends with the matching:

(Adam – Carol, Ben – Anne, Charles – Bess, Dean – Donna),

which is exactly the same matching that the men's courtship algorithm produced. Do the two different algorithms always lead to the same matching? The answer to that question is negative: as the next example shows, the men's courtship algorithm may lead to a different matching from the women's courtship algorithm, which indicates that there may be more than one stable matching. In addition, there may be stable matchings that are different from those attained by applying either the men's courtship or the women's courtship algorithms described above. ◀

Example 22.8 Consider the matching problem depicted in Figure 22.5, with three men and three women.

	Anne	Bess	Carol
Adam	1 ■ 3	2 ★ 2	3 ♣ 1
Ben	3 ♣ 1	1 ■ 3	2 ★ 2
Charles	2 ★ 2	3 ♣ 1	1 ■ 3

Figure 22.5 The preferences of the men and women in Example 22.8, and three stable matchings

The matching indicated by darkened squares is the men's courtship matching (attained by each man matched to the woman he most prefers), the matching indicated by clubs is the women's courtship matching (attained by each woman matched to the man she most prefers), and the matching indicated by stars is a third stable matching that differs from the previous two matchings (the reader is asked to ascertain that each of these three matchings is indeed stable). ◀

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

The number of possible matchings equals $n!$, the number of permutations of n elements. This number rapidly grows large as n grows. In principle the number of stable matchings could still be rather small. However, Gusfield and Irving [1989] show that matching problems may have a very large number of stable matchings. Using their method, one can construct a matching problem with 8 men and 8 women, with 268 stable matchings; a matching problem with 16 men and 16 women, with 195,472 stable matchings; and a matching problem with 32 men and 32 women, with 104,310,534,400 stable matchings.

22.4 Comparing matchings

So far we have found ways to attain two possible matchings: the men's courtship algorithm and the women's courtship algorithm. As shown in Example 22.8, there may be more stable matchings. Since these algorithms may lead to the creation of different pairs, if we compare two stable matchings there may be men and women who will prefer the application of one matching, and others who will prefer the application of the other matching.

Example 22.9 Consider a matching problem with four women and four men, with the preference relations depicted in Figure 22.6. This matching problem has four stable matchings: (check that this is true):

A^1 : Adam – Anne, Ben – Bess, Charles – Carol, Dean – Donna
A^2 : Adam – Bess, Ben – Anne, Charles – Carol, Dean – Donna
A^3 : Adam – Anne, Ben – Bess, Charles – Donna, Dean – Carol
A^4 : Adam – Bess, Ben – Anne, Charles – Donna, Dean – Carol

	Anne	Bess	Carol	Donna
Adam	1, 2	2, 1	3, 3	4, 3
Ben	2, 1	1, 2	3, 4	4, 4
Charles	3, 3	4, 3	1, 2	2, 1
Dean	3, 4	4, 4	2, 1	1, 2

Figure 22.6 The preference relations of the men and women in Example 22.9

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

Matching A^1 is the men's courtship matching, and matching A^4 is the women's courtship matching. Matching A^1 is the best from the perspective of the men, because each man is matched with the woman ranked highest on his preference list, and matching A^4 is the best from the perspective of the women, because each woman is matched with the man ranked highest on her list. From among these four matchings, the men's courtship matching A^1 is the "worst" for the women, and the women's courtship matching A^4 is the "worst" for the men.

Comparing the matchings A^2 and A^3, Anne and Bess prefer A^2 to A^3 (because under A^2 each is matched with the man she ranks highest), while Donna and Carol prefer A^3 to A^2 (because under A^3 each is matched with the man she ranks highest). Similarly, Charles and Dean prefer A^2 to A^3, and Adam and Ben prefer A^3 to A^2. ◀

The last example raises several broad questions regarding the preferences of men and women over stable matchings. Is there a stable matching that is the worst for all the men, among all the stable matchings? Which matching algorithm is better for the men, the men's courtship algorithm or the women's courtship algorithm?

Definition 22.10 *Let A and B be two matchings. Denote $A \succsim^m B$ if every man who is matched under A and B to different women prefers the woman to whom he is matched under A to the woman to whom he is matched under B. Denote $A \succsim^w B$ if and only if every woman who is matched under A and B to different men prefers the man to whom she is matched under A to the man to whom she is matched under B.*

This is a definition of two ordering relations $\succsim^m$ and $\succsim^w$ over the set of matchings: $A \succsim^m B$ if and only if every man matched to two different women under matchings A and B prefers the woman to whom he is matched under A. Equivalently, no man prefers matching B to matching A. It follows that $A \succsim^m A$ for every stable matching, and therefore $\succsim^m$ (and similarly $\succsim^w$) is a reflexive relation. It can be ascertained that the relations $\succsim^m$ and $\succsim^w$ are also transitive relations (Exercise 22.18).

Despite the fact that these ordering relations are defined for all matchings, we will mostly be interested in using them to order stable matchings. As we will show, in stable matchings these orderings induce a special structure. First of all, note that even if we restrict attention to stable matchings, these orderings are not complete orderings; in Example 22.9 $A^2 \not\succsim^w A^3$ and $A^3 \not\succsim^w A^2$ and similarly $A^2 \not\succsim^m A^3$ and $A^3 \not\succsim^m A^2$.

The next theorem states that if one stable matching is better for the men than another stable matching, then it is worse for the women.

Theorem 22.11 *For every pair of stable matchings A and B, $A \succsim^m B$ if and only if $B \succsim^w A$.*

Proof: We will prove that if $A \succsim^m B$ then $B \succsim^w A$. The other direction of the statement of the theorem is then proved by reversing the roles of the women and the men.

Let A and B be two stable matchings, and suppose that $A \succsim^m B$. We need to show that $B \succsim^w A$, i.e., every woman who is matched to different men under A and under B prefers her mate under matching B to her mate under matching A. Let Lena be such a woman, and suppose that she is matched with Aaron under A, and with Benjamin under B. Suppose that under matching B, Aaron is matched with Mandy. Since $A \succsim^m B$, Aaron prefers the

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

woman with whom he is matched under A, Lena, to the woman with whom he is matched under B, Mandy:

$$\text{Aaron: Lena} \succ \text{Mandy}.$$

Since B is a stable matching, Aaron and Lena do not have an objection under B. Since Aaron prefers Lena to Mandy, his mate under matching B, it follows that Lena prefers the man with whom she is matched under B, Benjamin, to Aaron, with whom she is matched under A, which is what we wanted to show. □

The following theorem states that the phenomenon we saw in Example 22.9, in which the men's courtship matching O^m is the best stable matching from the perspective of all the men, and the worst from the perspective of all the women, and in which the women's courtship matching O^w is the best stable matching from the perspective of all the women, and the worst from the perspective of all the men, holds true for every matching problem.

Theorem 22.12 *For every stable matching A, one has $O^m \succsim^m A \succsim^m O^w$ and $O^w \succsim^w A \succsim^w O^m$.*

Proof: To prove the theorem, it suffices to prove that $O^m \succsim^m A$ holds for every stable matching A. To see this, note that by Theorem 22.11 this would imply that $A \succsim^w O^m$. By reversing the roles of the men and the women, we also get $O^w \succsim^w A$ and $A \succsim^m O^w$.

We will say that a woman w is *possible* for a man m if there exists a stable matching under which they are matched to each other.

Step 1: Any woman who dismissed a man under the men's courtship algorithm is not possible for him.

The proof will be by induction over the stages k in which the man is dismissed. Start with the first stage, $k = 1$. We will prove that if Adam is dismissed by Bess in the first stage of the men's courtship algorithm then Bess is not possible for Adam. To see this, suppose that Adam and Bill stand in front of Bess's house in the first stage and that Bess dismisses Adam in that first stage while telling Bill to stay. This means that Bess prefers Bill to Adam and that Bill prefers Bess to any other woman, because he went to her house in the first stage of the men's courtship algorithm. It follows that any matching A that matches Bess to Adam is unstable, because the pair (Bill, Bess) objects to it, since Bess prefers Bill to Adam and Bill prefers Bess to any other woman, and in particular to any women to whom he is matched under A. We deduce that Bess is not possible for Adam.

Let $k \geq 1$ and suppose by induction that every woman who dismisses a man in the first k stages of the algorithm is not possible for him. Suppose that Adam is dismissed by Bess in stage $k + 1$ of the algorithm. We will prove that she is not possible for him; in other words, every matching A in which Bess is matched to Adam is unstable. To see this, suppose that Adam and Ben stand in from of Bess's house in stage $k + 1$ of the algorithm and that Bess dismisses Adam while telling Ben to stay. Then Bess prefers Ben to Adam. Suppose that Ben's mate under the matching A is Abigail. If the matching A were stable, then since under A Abigail is matched to Ben, by the inductive hypothesis she could not have dismissed Ben in the first k stages of the algorithm. It follows that Ben did not go to Abigail's house before he went to Bess's house (which he does in stage $k + 1$), and

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

therefore he prefers Bess to Abigail. Hence, the pair (Ben, Bess) objects to the matching A, and A is unstable. This completes the inductive stage.

Step 2: $O^m \succsim^m A$ for every stable matching A.
Suppose that under the men's courtship algorithm a particular man, Adam, is matched to Bess. We wish to show that for every stable matching A in which Adam is matched to a woman who is not Bess, say Betty, he prefers Bess to Betty. If he were to prefer Betty to Bess then that means that under the men's courtship algorithm Adam visits Betty before he visits Bess and that Betty dismissed him (since he is matched to Bess under the men's courtship algorithm). But then by Step 1, Betty is not possible for him, contradicting the assumption that she is matched to him in a stable matching. □

For two stable matchings A and B define a rule under which every woman chooses a mate: if the woman is matched to the same man under A and B, then she chooses that man. If she is matched to two different men under the two matchings, then she chooses the man whom she most prefers from among those two. Is the resulting outcome a matching, or can this lead to a situation in which two women choose the same man? If this process leads to a matching, is it stable? If the answer to these questions is affirmative, clearly this matching is at least as good for the women as the original two matchings.

Definition 22.13 *Let A and B be two matchings. Denote by $A \vee^w B$ the set of all n pairs $\{(m_1, w_1), (m_2, w_2), \ldots, (m_n, w_n)\}$ that satisfies $\{w_1, w_2, \ldots, w_n\} = W$ and for all $i = 1, 2, \ldots, n$, m_i is the one man whom woman w_i prefers from among the men to whom she is matched under A and B.*

Theorem 22.14 *If A and B are stable matchings, then $A \vee^w B$ is also a stable matching.*

Proof: The theorem is proved in two steps. We first prove that $A \vee^w B$ is a matching, and then that it is a stable matching. Denote $C := A \vee^w B$.

Step 1: C is a matching.
Suppose by contradiction that C is not a matching. In other words, suppose that after each woman chooses the man whom she most prefers among those matched to her under A and B, there is a man who is chosen by two women. For example, suppose that in C Adam is chosen by both Anne and Bess. It then follows that Adam is matched to one of them under A, and matched to the other one under B. Suppose in particular that under A Adam is matched to Anne and Elton is matched to Bess, and that under B Adam is matched to Bess and Dan is matched to Anne.

$$\begin{aligned} &\text{Matching } A\text{: Adam – Anne,} \quad \text{Elton – Bess} \\ &\text{Matching } B\text{: Adam – Bess,} \quad \text{Dan – Anne.} \end{aligned}$$

(Note that Elton and Dan may be the same person.) Since both Anne and Bess choose Adam in C:

$$\begin{aligned} \text{Anne: Adam} &\succ \text{Dan,} \\ \text{Bess: Adam} &\succ \text{Elton.} \end{aligned}$$

If

$$\text{Adam : Anne} \succ \text{Bess}$$

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

then the pair (Adam, Anne) has an objection to matching B.

If

$$\text{Adam: Bess} \succ \text{Anne}$$

then the pair (Adam, Bess) has an objection to matching A, contradicting the fact that both A and B are stable matchings. This contradiction proves that C is a matching.

Step 2: C is a stable matching.
Suppose by contradiction that C is not stable. Then there is a pair, say Adam and Claire, who are not matched to each other under C and have an objection.

Suppose that under the matching C:

$$\text{Matching } C\text{: Adam} - \text{Bess}, \quad \text{Elton} - \text{Claire}.$$

Since Adam and Claire have an objection, it must be true that

$$\text{Claire: Adam} \succ \text{Elton},$$
$$\text{Adam : Claire} \succ \text{Bess}.$$

Since under the matching C Claire is matched to Elton, she must be matched to him under either A or B. Suppose without loss of generality that she is matched to him under A, and suppose that she is matched to Frank under matching B. (Our claims still hold if Claire is also matched to Elton under matching B.)

$$\text{Matching } A\text{: Elton} - \text{Claire}, \qquad \text{Matching } B\text{: Frank} - \text{Claire}.$$

Claire prefers Adam to Elton, and Elton to Frank (since under matching C she chooses Elton and not Frank), and therefore she prefers Adam to Frank:

$$\text{Claire: Adam} \succ \text{Elton} \succ \text{Frank}.$$

We claim that Bess is not matched to Adam under matching A. Since matching A is a stable matching, Adam and Claire have no objection to it. Since Claire prefers Adam to Elton, Adam must prefer the woman to whom he is matched under matching A to Claire. Since Adam prefers Claire to Bess, it is impossible for Bess to be matched to Adam under A.

Finally, we show that Bess is not matched to Adam under matching B. Since B is a stable matching, Adam and Claire have no objection to it. Since Claire prefers Adam to Frank, Adam must prefer the woman to whom he is matched under B to Claire. Since Adam prefers Claire to Bess, it is impossible for Bess to be matched to Adam under B.

In other words, Bess is not matched to Adam under either A or B. If so, how could she be matched to Adam under the matching C? The contradiction establishes that C is a stable matching. □

Clearly, the matching $C := A \vee^w B$ is, for every woman, at least as good as A, and at least as good as B; that is, $C \succsim^w A$ and $C \succsim^w B$. By Theorem 22.11, the matching C is, for every man, worse than (or equally preferred to) both A and B; that is, $A \succsim^m C$ and $B \succsim^m C$.

Similarly, we can define, for every pair of matchings A and B, the collection of n pairs $D := A \vee^m B$, in which every man chooses the woman who is most preferred by him from among the women to whom he is matched under A and B. By reversing the roles of the

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

men and the women in Theorem 22.14, we deduce that D is a stable matching, and that $D \succsim^m A$, $D \succsim^m B$, $A \succsim^w D$, and $B \succsim^w D$.

22.4.1 The lattice structure of the set of stable matchings

A partial ordering $\succsim$ is a reflexive and transitive binary relation.[3] In words, a partial ordering relation enables us to compare the members of some pairs of elements in a given set. A *partially ordered set* is a set with a partial ordering relation defined over it. An example of a partially ordered set is the pair $(X, \succsim)$, where X is the collection of all subsets of a set S, and for every two subsets U and V of S one has $U \succsim V$ if $U \supseteq V$. This is not a complete ordering when S contains at least two elements, because in this case there are two subsets of S that cannot be compared by this ordering (that is, two sets neither of which is a subset of the other).

Using the relation $\succsim$, we defined in Chapter 2 the concept of a *strict preference relation* $\succ$:

$$x \succ y \iff x \succsim y \text{ and } y \not\succsim x. \tag{22.2}$$

We also defined the *indifference* relation $\approx$:

$$x \approx y \iff x \succsim y \text{ and } y \succsim x. \tag{22.3}$$

Definition 22.15 *Let X be a finite set, let $\succsim$ be a partial ordering over X, and let $x_1, x_2 \in X$ be two elements in X. The element $y \in X$ is called the* maximum *of x_1 and x_2, and denoted $y = \max\{x_1, x_2\}$, if the following two conditions hold:*

1. *$y \succsim x_1$ and $y \succsim x_2$.*
2. *If $z \succsim x_1$ and $z \succsim x_2$, then $z \succsim y$.*

If X is the collection of subsets of a set S, and the relation $\succsim$ is the set inclusion relation, then the maximum of a pair of subsets of S is their union. The next example shows that there are partial orderings for which the maximum does not exist.

Example 22.16 Let $X = \{x_1, x_2, x_3, x_4\}$, and let the ordering relation be given by

$$x_1 \succ x_3, \quad x_2 \succ x_3, \quad x_1 \succ x_4, \quad x_2 \succ x_4. \tag{22.4}$$

The elements x_3 and x_4 have no maximum. Indeed, both x_1 and x_2 are greater than x_3 and greater than x_4, but one cannot compare them to each other, and therefore neither of them is a maximum for x_3 and x_4. The elements x_1 and x_2 also do not have a maximum, because there is no element greater than both of them. ◀

The minimum of a set is defined analogously to the definition of the maximum of a set.

Definition 22.17 *Let X be a finite set, $\succsim$ be a partial ordering relation over X, and let $x_1, x_2 \in X$ be two elements of X. An element $y \in X$ is a* minimum *of x_1 and x_2, denoted $y = \min\{x_1, x_2\}$, if the following conditions are satisfied:*

3 We previously encountered, earlier in this chapter and in Chapter 2, the concept of a complete ordering in which all elements are comparable. In contrast, under a partial ordering it is possible for two elements to be incomparable.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

1. $x_1 \succsim y$ and $x_2 \succsim y$.
2. If $x_1 \succsim z$ and $x_2 \succsim z$, then $y \succsim z$.

Definition 22.18 A lattice *is a partially ordered set satisfying the property that any pair of elements in the set has a minimum and a maximum.*

For the partial ordering $\succsim^w$ over the set of stable matchings, the maximum of a pair of stable matchings A and B is the matching $A \vee^w B$; in other words, $\max\{A, B\} = A \vee^w B$, and the minimum of this pair of stable matchings is the matching $\min\{A, B\} = A \vee^m B$ (Exercise 22.20). Similarly, for the partial ordering $\succsim^m$ over a set of stable matchings, $\max\{A, B\} = A \vee^m B$ and $\min\{A, B\} = A \vee^w B$.

The above discussion leads to the following theorems.

Theorem 22.19 *Under the ordering relation $\succsim^w$, the set of stable matchings is a lattice.*

Similarly,

Theorem 22.20 *Under the ordering relation $\succsim^m$, the set of stable matchings is a lattice.*

Theorem 22.11 (page 893) states that in effect the lattices described by Theorems 22.19 and 22.20 are the same lattice: both are defined over the same set of elements (the set of stable matchings) and the maximum under the ordering relation $\succsim^w$ is the minimum under the ordering relation $\succsim^m$ (and vice versa).

22.5 Extensions

In this section we consider several extensions of the basic model that has been studied up to now.

22.5.1 When the number of men does not equal the number of women

We have so far dealt only with models in which the number of men equals the number of women. Suppose instead that n_m, the number of men, is greater than n_w, the number of women. The case in which the number of women is greater than the number of men is analyzed similarly. Since $n_m > n_w$, under every stable matching there must remain $n_m - n_w$ men who are not matched to a woman.

To fit this new situation, we need to change the definition of a matching. Recall that we denote by M the set of men, and by W the set of women.

Definition 22.21 A matching *is a function associating every man with an element of the set $W \cup \{single\}$, such that every woman is associated under this function with one man.*

We assume here that being single is considered by every man to be a worse outcome than being matched to any of the women.

Definition 22.22 *Suppose that a matching is given. A man m and a woman w* object *to the matching if the following conditions hold: (a) the man m is single, or is matched to a*

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

woman $\widehat{w}$ and he prefers w to $\widehat{w}$; and (b) the woman w prefers m to the man to whom she is matched under the matching.

The Gale–Shapley algorithm that is presented in the previous sections is applicable to this case, and the proof of Theorem 22.7 goes through with minor modifications. In particular, the algorithm is guaranteed to terminate with a stable matching (Exercise 22.24). The next theorem, whose proof is left to the reader (Exercise 22.25), establishes that under all stable matchings, the set of singles is the same.

Theorem 22.23 *Suppose that the number of men is greater than the number of women. If a particular man is not matched to any woman under some stable matching, then he is not matched to a woman under any stable matching.*

22.5.2 Strategic considerations

In the men's courtship algorithm presented above, in every stage, every man goes to the house of the woman whom he ranks highest among all the women who have not previously dismissed him, and every woman asks the man she ranks highest from among the men standing in front of her house to stay, while dismissing all the rest. Does this algorithm leave any room for "strategic behavior"? In other words, is it possible that by pretending to have a preference relation that differs from her true preference relation, a woman can obtain a better match for herself than by being honest? The next example shows that this may indeed be possible.

Example 22.24 **An example of strategic behavior** Consider an example with three men and three women whose preference relations are given by the tables shown in Figure 22.7, where 1, 2, and 3 denote placement within the preference list, with 1 representing highest preference.

	1	2	3
Hector:	Helena	Andromache	Lavinia
Aeneas:	Helena	Lavinia	Andromache
Paris:	Andromache	Helena	Lavinia

	1	2	3
Helena:	Paris	Hector	Aeneas
Andromache:	Hector	Paris	Aeneas
Lavinia:	Aeneas	Hector	Paris

Figure 22.7 The preferences of the men and the women

We trace the men's courtship algorithm (Figure 22.8).

	Helena	Andromache	Lavinia	Dismissed men
Stage 1(1a)	Aeneas(1), Hector(1)	Paris(1)		
Stage 1(b)	Hector(1)	Paris(1)		Aeneas
Stage 2(a)	Hector(1)	Paris(1)	Aeneas(2)	

Figure 22.8 Men's courtship algorithm

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

The resulting matching is

(Hector – Helena, Aeneas – Lavinia, Paris – Andromache).

Suppose, instead, that Helena were to act as if her preference relation is:

	1	2	3
Helena:	Paris	Aeneas	Hector

Then the men's courtship algorithm would look as shown in Figure 22.9.

	Helena	Andromache	Lavinia	Dismissed men
Stage 1(a)	Hector(1), Aeneas(1)	Paris(1)		
Stage 1(b)	Aeneas(1)	Paris(1)		Hector
Stage 2(a)	Aeneas(1)	Paris(1), Hector(2)		
Stage 2(b)	Aeneas(1)	Hector(2)		Paris
Stage 3(a)	Aeneas(1), Paris(2)	Hector(2)		
Stage 3(b)	Paris(2)	Hector(2)		Aeneas
Stage 4(a)	Paris(2)	Hector(2)	Aeneas(2)	

Figure 22.9 The men's courtship algorithm when Helena behaves strategically

The resulting matching is

(Hector – Andromache, Aeneas – Lavinia, Paris – Helena).

As can be seen, Helena has improved her situation by pretending that her preference relation is different from her true one: under this matching, she is matched to Paris, who is ranked first in her (true) preference relation, instead of to Hector, who is ranked second in her (true) preference relation. The matching that results from Helena's strategic behavior is a stable matching relative to the true preferences of the participants (check that this is true). This is no coincidence; in Exercise 22.29 we present conditions relating to a woman's strategic behavior that guarantee that the men's courtship algorithm terminates with a matching that is stable relative to the true preferences of all participants. ◀

The last example shows that a woman may sometimes obtain a more preferred mate under the men's courtship algorithm if she does not dismiss men according to her true preferences. It turns out that this is not possible for the men: under the men's courtship algorithm, a man cannot obtain a more preferred mate if he courts the women in an ordering that is different from his true preference ordering (Exercise 22.28). It follows in particular that when there are more men than women, a man who remains single at the end the men's courtship algorithm will also remain single if he courts women under a different ordering than his true preference ordering.

22.5.3 The desire to remain single, or: getting married, but not at any price

In the model we have regarded so far, we assumed that every man prefers being with any woman to remaining single, and that every woman prefers being with any man to remaining single. How does our analysis change if some participants prefer the single life

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

to being in a match that they dislike? We first need to define a model that enables this possibility.

Definition 22.25 *A* matching problem *is defined by:*

- *A set of men M and a set of women W.*
- *For every woman w, a preference relation over the set $M \cup \{(w\text{-single})\}$.*
- *For every man, a preference relation over the set $W \cup \{(m\text{-single})\}$.*

The elements "m-single" and "w-single" enable the participants to rank the single life in their list of preferences. For example, if

$$\text{Juliet: Romeo} \succ \text{Juliet-single},$$

then Juliet prefers a match with Romeo to remaining single, but if

$$\text{Juliet: Juliet-single} \succ \text{Benvolio},$$

then she prefers remaining single to a match with Benvolio.

The definition of a matching in this case is:

Definition 22.26 *A* matching *is a function associating every man $m \in M$ with an element of the set $W \cup \{(m\text{-single})\}$ and every woman $w \in W$ with an element of the set $M \cup \{(w\text{-single})\}$, such that if a man m is matched to a woman w, the woman w is matched to the man m.*

We also update the definitions of an objection and a stable matching. In contrast to Definition 22.4, where only pairs could object to a matching, in this model an objection to a matching can be raised by a single woman or a single man, in case one of them prefers remaining single to the person to whom he or she is matched under the matching. In addition, we also allow objections to be raised by pairs who are not necessarily matched under the matching, because one, or both of them, is single under the matching.

Definition 22.27 *Given a matching,*

1. *A man* objects *to the matching, if he is matched to a woman but prefers remaining single to the woman to whom he is matched.*
2. *A woman* objects *to the matching if she is matched to a man but prefers remaining single to the man to whom she is matched.*
3. *A man m and a woman w* object *to the matching if (a) the man prefers w to the woman to whom he is matched under the matching, or prefers w to remaining single, if he is not matched to any woman under the matching; and (b) the woman prefers m to the man to whom she is matched under the matching, or prefers m to remaining single, if she is not matched to any man under the matching.*

A matching is stable *if no man, no woman, and no pair of a man and a woman object to it.*

So far, we have changed our model by adding fictitious elements of the form (w-single) and (m-single); every woman w has a preference relation over the set $M \cup \{(w\text{-single})\}$ and every man m has a preference relation over the set $W \cup \{(m\text{-single})\}$. To generalize

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

the Gale–Shapley algorithm to this model, expand the set of women and the set of men as follows:

- The set of men is $M' := M \cup \{(w\text{-single})\colon w \in W\}$.
- The set of women is $W' := W \cup \{(m\text{-single})\colon m \in M\}$.

We also expand the preference relation of every woman to a preference relation over M', and the preference relation of every man to a preference relation over W'. Define the preference relations of the fictitious elements, as follows:

- Every woman w prefers the element (w-single) to the element ($\widehat{w}$-single), for every woman $\widehat{w} \neq w$. The precise ordering under which the elements $\{(\widehat{w}\text{-single}), \widehat{w} \in W \setminus \{w\}\}$ are ordered is immaterial.
- Every man m prefers the element (m-single) to the element ($\widehat{m}$-single), for every man $\widehat{m} \neq m$. The precise ordering under which the elements $\{(\widehat{m}\text{-single}), \widehat{w} \in W \setminus \{w\}\}$ are ordered is immaterial.
- The element (w-single) prefers w to every other element in W'. The precise ordering under which the other elements are ordered in the preference relation of this element is immaterial.
- The element (m-single) prefers m to every other element in M'. The precise ordering under which the other elements are ordered in the preference relation of this element is immaterial.

In the men's courtship algorithm, every man courts the women in descending order according to his preference list. If m gets far enough down his list to reach the fictitious (m-single), then he will remain "at the door" of the element (m-single) when the algorithm terminates. Indeed, since the fictitious element (m-single) prefers m to all the other men, the man m will not be dismissed by (m-single). The algorithm terminates when at most one man stands in front of the house of each woman $w \in W$, i.e., when there can be no more dismissals.

Example 22.28 Consider a matching problem with the set of men $M = \{\text{Alan, Basil, Colin}\}$ and the set of women $W = \{\text{Rose, Sara}\}$. The preference relations of the men and women are depicted in the following tables, where 1, 2, and 3 denote placement within the preference list, with 1 representing highest preference.

	1	2	3
Alan:	Sara	Rose	Alan-single
Basil:	Rose	Basil-single	Sara
Colin:	Rose	Sara	Colin-single

	1	2	3	4
Rose:	Alan	Basil	Rose-single	Colin
Sara:	Basil	Colin	Sara-single	Alan

Two elements are now added to the set of men: (Rose-single) and (Sara-single), and similarly three elements are added to the set of women: (Alan-single), (Basil-single), and (Colin-single).

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

One possible definition of a preference relation following the addition of the fictitious elements is:

	1	2	3	4	5
Alan:	Sara	Rose	Alan-single	Colin-single	Basil-single
Basil:	Rose	Basil-single	Sara	Colin-single	Alan-single
Colin:	Rose	Sara	Colin-single	Alan-single	Basil-single
Rose-single:	Rose	Sara	Colin-single	Alan-single	Basil-single
Sara-single:	Sara	Rose	Colin-single	Alan-single	Basil-single

	1	2	3	4	5
Rose:	Alan	Basil	Rose-single	Colin	Sara-single
Sara:	Basil	Colin	Sara-single	Alan	Sara-single
Alan-single:	Alan	Basil	Colin	Sara-single	Rose-single
Basil-single:	Basil	Alan	Colin	Sara-single	Rose-single
Colin-single:	Colin	Basil	Alan	Sara-single	Rose-single

The men's courtship matching and the women's courtship matching of the expanded problem is as follows (check that this is true):

Basil – (Basil-single), Alan – Rose, Colin – Sara,
(Rose-single) – (Alan-single), (Sara-single) – (Colin-single).

It follows that the only stable matching to the original problem is

(Basil – Basil-single, Alan – Rose, Colin – Sara). ◀

In Exercise 22.30, the reader is asked to spell out in detail the Gale-Shapley algorithm for this model and to prove that the algorithm always terminates after finding a stable matching.

22.5.4 Polygamous matching: placement of students in universities

We have so far assumed that every man is permitted to marry only one woman, and each woman is permitted to marry only one man. In some cases, however, such as matching medical residents and hospitals, universities and students, and corporations and employees, there is an asymmetry between the two sides of the market: hospitals, universities, and corporations are interested in more than one resident, student, or employee, while each resident, student, and employee generally chooses only one hospital, university, or employing corporation. A model that can accommodate such situations is as follows (for convenience, in the definition we have adopted terms used in student placement in universities).

Definition 22.29 *A* polygamous matching problem *is given by:*

- *A finite set of universities U and a finite set of students S.*
- *For each university $u \in U$ a quota $q_u \in \mathbb{N}$ that represents the maximal number of students it will accept.*

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

- *For each student $s \in S$, a preference relation over the set U.*
- *For each university $u \in U$, a preference relation over the set S.*

Definition 22.30 *A* matching *is a function assigning, to each university $u \in U$, a subset S containing between 0 and q_u students, such that each student in S is associated with at most one university. In other words, the sets of students associated to two different universities are disjoint sets.*

The concept of an objection is defined as follows.

Definition 22.31 *A university u and a student s* object *to a matching if both of the following conditions are met:*

- *The student s prefers university u to the university to which he is matched.*
- *University u is matched to fewer than q_u students, or it is matched with q_u students but prefers s to one of the students that are matched to it.*

A matching is stable *if there is no pair consisting of a university and a student who have an objection to it.*

To show that there always exists a stable matching, consider the following "student courtship" algorithm, under which students apply to universities and each university u asks the q_u students that it most prefers from among the students gathered at the university's gate to remain (instead of asking only one) and rejecting all the rest; if less than q_u students have applied, it asks all of them to remain. Every rejected student then goes to the university next down on his or her preference list. The proof that this algorithm always terminates after finding a stable matching is similar to the proof in the monogamous matching case (Exercise 22.31).

22.5.5 Unisexual matchings

We have so far matched men to women, thus assuming that pairs must be composed of at least one member of the two different sexes. But there are cases in which pairs need to be matched from a homogeneous population: students being paired for dormitory rooms, police officers paired in patrol cars, and so on. Interestingly, while heterosexual stable matchings are guaranteed always to exist, unisexual stable matchings may not exist, as the next example shows.

Example 22.32 **A unisexual population without a stable matching** Consider an example with four men, Alex, Benjamin, Chris, and Franklin, who are to be partitioned into two pairings. Each man has a preference relation over the other men given by the table shown in Figure 22.10.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

	1	2	3
Alex:	Benjamin	Chris	Franklin
Benjamin:	Chris	Alex	Franklin
Chris:	Alex	Benjamin	Franklin
Franklin:	Alex	Chris	Benjamin

Figure 22.10 Preference relations in a single-sex example with no stable matching

We will now show that given these preference relations, there is no stable matching. Suppose, for example, that Franklin is paired with Alex, which then means that Benjamin is paired with Chris:

(Franklin – Alex, Benjamin – Chris).

Then Alex and Chris have an objection: Alex prefers Chris (number 2 on his list) to Franklin (number 3 on his list), and Chris prefers Alex (number 1 on his list) to Benjamin (number 2 on his list).

It can similarly be shown that there is no stable matching under which Franklin is paired with either Benjamin or Chris (Exercise 22.38). ◀

22.6 Remarks

The authors wish to thank Dov Samet for his assistance in the composition of this chapter. The introduction to this chapter is based on Roth [2005]. The reader interested in further study of the material in this chapter is directed to Roth and Sotomayer [1990] and Gusfield and Irving [1989], both of which contain a wealth of information on matching theory.

Exercise 22.17 is based on Dubins and Freedman [1981]. Exercises 22.27 and 22.28 are based on Gale and Sotomayor [1985].

22.7 Exercises

22.1 Prove that every complete, irreflexive, and transitive relation is asymmetric: if $x \neq y$, then $x \succ y$ if and only if $y \not\succ x$.

22.2 Prove that the Gale–Shapley algorithm for finding a stable matching satisfies the following property: if Cleopatra asks Mark to stay in front of her house at stage k, and at a later stage she asks Julius to stay in front of her house, then she prefers Julius to Mark.

22.3 Consider the following system of preferences (recall that the preferences of the women appear in the lower right side of each cell (read vertically) and the preferences of the men appear on the upper left side (read horizontally)).

	Anne	Betty	Claire	Donna
Alfredo	1 4	2 3	3 1	4 3
Ben	3 3	4 1	1 3	2 4
Chris	2 2	4 2	1 4	3 1
Dean	4 1	3 4	2 2	1 2

Check whether the following matchings are stable. Justify your answers.

(Chris – Anne, Alfredo – Betty, Dean – Claire, Ben – Donna),
(Chris – Anne, Alfredo – Betty, Ben – Claire, Dean – Donna).

22.4 Prove that Definitions 22.4 and 22.5 on page 887 are equivalent.

22.5 In each of the following systems of preferences, find the stable matching that is obtained by the men's courtship algorithm, and the stable matching that is obtained by the women's courtship algorithm:

(a)

	1	2	3
Andre:	Anne	Barbara	Claire
Boris:	Barbara	Claire	Anne
Chris:	Claire	Anne	Barbara

	1	2	3
Anne:	Boris	Chris	Andre
Barbara:	Chris	Andre	Boris
Claire:	Chris	Boris	Andre

(b)

	1	2	3	4
Alex:	Ellen	Flora	Gail	Hillary
Bill:	Ellen	Hillary	Gail	Flora
Colin:	Flora	Ellen	Gail	Hillary
David:	Hillary	Flora	Gail	Ellen

	1	2	3	4
Ellen:	David	Colin	Alex	Bill
Flora:	Bill	David	Alex	Colin
Gail:	David	Alex	Bill	Colin
Hillary:	Colin	Bill	Alex	David

(c)

	1	2	3	4
Peter:	Olivia	Patty	Mary	Netty
Jacob:	Patty	Mary	Netty	Olivia
Kevin:	Mary	Netty	Olivia	Patty
Larry:	Patty	Olivia	Mary	Netty

	1	2	3	4
Mary:	Peter	Jacob	Kevin	Larry
Netty:	Peter	Jacob	Kevin	Larry
Olivia:	Jacob	Kevin	Peter	Larry
Patty:	Kevin	Peter	Jacob	Larry

(d)

	1	2	3	4
Ernest:	Felicia	Emma	Donna	Carol
Felix:	Emma	Donna	Carol	Felicia
George:	Donna	Carol	Felicia	Emma
Henry:	Emma	Felicia	Donna	Carol

	1	2	3	4
Carol:	Ernest	Felix	George	Henry
Donna:	Ernest	Felix	George	Henry
Emma:	George	Ernest	Felix	Henry
Felicia:	Felix	George	Ernest	Henry

(e)

	1	2	3	4
Peter:	Lisa	Melissa	Natasha	Octavia
Quentin:	Lisa	Melissa	Natasha	Octavia
Ron:	Melissa	Natasha	Lisa	Octavia
Sam:	Natasha	Lisa	Melissa	Octavia

	1	2	3	4
Lisa:	Ron	Sam	Peter	Quentin
Melissa:	Sam	Peter	Quentin	Ron
Natasha:	Peter	Quentin	Ron	Sam
Octavia:	Sam	Ron	Peter	Quentin

22.6 Suppose that the number of men equals the number of women. Prove the following claims, or provide counterexamples:

(a) For every pair of matchings there exist preference relations for which these are two stable matchings.
(b) For every three matchings there exist preference relations for which these are three stable matchings.
(c) For every four matchings there exist preference relations for which these are four stable matchings.

22.7 Gary is at the top of Gail's preference list, and Gail is at the top of Gary's preference list. Prove that in every stable matching Gary and Gail are matched to each other.

22.8 Dan is at the bottom of Donna's preference list, and Donna is at the bottom of Dan's preference list. Is it possible that there is a stable matching that matches Dan to Donna? Justify your answer.

22.9 Prove that if Romeo and Juliet are matched to each other under both the men's courtship and the women's courtship algorithms, then they are matched to each other under any stable matching.

22.10 Prove that if the result of the men's courtship algorithm yields the same result as the women's courtship algorithm, then this resulting matching is the unique stable matching.

22.11 Given a stable matching of n men and n women,

(a) Is it possible to find three pairs such that if the matching among them is changed, each man will be matched to a woman whom he prefers, and each woman will be matched to a man whom she prefers?
(b) Generalize this conclusion to a subset of k pairs, for every $4 \leq k \leq n$.

22.12 In Julius's list of preferences, Agrippina appears first, Messalina appears second, and Cleopatra appears third. Suppose there is a stable matching under which Julius is matched to Agrippina, and that there is a stable matching under which Julius is matched to Cleopatra. Is there necessarily a stable matching under which Julius is matched to Messalina? Either prove this statement or provide a counterexample.

22.13 In this exercise we consider a situation with n men and n women.

(a) Prove that if in stage t of the men's courtship algorithm, a particular man is dismissed for the $(n-1)$-th time, then the algorithm terminates at stage $(t+1)$.
(b) Prove that the men's courtship algorithm terminates after at most $(n-1)^2+1$ stages.
(c) Find preference relations under which the algorithm terminates after precisely $(n-1)^2+1$ stages (hence this is the lowest bound for the length of the algorithm).

22.14 Suppose that Fara is preferred by every man to all the other women. Prove that under every stable matching Fara is matched to the same man. Who is the lucky guy?

22.15 Suppose that the number of men equals the number of women, and that Vera is last on the preference list of every man. Prove that under every stable matching, Vera is matched to the same man. Who is the unlucky guy?

22.16 Suppose that every man has the same preference relation over the set of women. Prove that there exists only one stable matching.

22.17 In this exercise, we present a family of algorithms, each of which produces the men's courtship matching, and contains the Gale–Shapley algorithm.

Consider the following algorithm for matching men and women. At the start of the algorithm, all the men leave the room, while all the women remain in the room. In every stage of the algorithm, one of the men who is outside the room enters, and goes directly to the woman whom he most prefers from among the women who have not previously dismissed him. If another man is already standing next to that woman, the woman asks the man whom she prefers from among those two who are now next to her to stay, and dismisses the other one, who then leaves the room. If, however, a man entering the room goes to a woman who is standing alone, she asks him to stay. The algorithm terminates when every woman has exactly one man standing next to her.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

(a) Prove that this algorithm satisfies the three properties of the Gale–Shapley algorithm (specified on page 889).
(b) Prove that this algorithm terminates, and that it always yields a stable matching.
(c) Prove that the algorithm always produces the men's courtship matching O^m.

22.18 Prove that the preference relations on matchings $\succsim^m$ and $\succsim^w$ over the set of stable matchings are transitive relations.

22.19 Show that Theorem 22.11 on page 893 does not hold if the matchings A and B are not stable. In other words, find two matchings A and B satisfying $A \succsim^m B$ but $B \not\succsim^w A$.

22.20 Prove that when the partial ordering over the set of stable matchings is $\succsim^w$, then $\max\{A, B\} = A \vee^w B$ and $\min\{A, B\} = A \vee^m B$.

22.21 Show by example that if A and B are two matchings (not necessarily stable), then $A \vee^w B$ is not necessarily a matching.

22.22 Prove that when the partial ordering over the set of stable matchings is $\succsim^w$, the minimum of the stable matchings A and B is $A \vee^m B$.

22.23 In this exercise we generalize the definition of the maximum of two matchings to a definition of the maximum of any finite set with two or more matchings. Suppose that the number of women equals the number of men. Let $A_1, A_2, \ldots, A_K$ be stable matchings. Define a function B from the set of men to the set of women as follows. Under the function B, a man m is matched to the woman he most prefers from among the women to whom he is matched under the matchings $A_1, A_2, \ldots, A_K$.

(a) Prove that the function B is a matching.
(b) Prove that the function B is a stable matching.
(c) Prove that for each stable matching C, if $C \succsim^m A_k$ for every $k \in \{1, 2, \ldots, K\}$ then $C \succsim^m B$. Matching C is the maximum of $A_1, A_2, \ldots, A_K$ according to the preference relations of the men.

22.24 Suppose that the matching problem is to match n_m men to n_w women, where $n_w < n_m$.

(a) Describe in detail the generalization of the Gale-Shapley algorithm for this case. Prove that the algorithm terminates with a stable matching.
(b) What is the maximal number of stages in the men's courtship algorithm?
(c) Construct an example where $n_m > n_w$ such that the men's courtship algorithm runs through the maximal number of stages.

22.25 Prove Theorem 22.23 on page 899: when the number of men n_m is greater than the number of women n_w, if a particular man is not matched to any woman under some stable matching, then he is not matched to any woman under any stable matching.

22.26 In each of the following pairs of preference relation systems, find the men's courtship matching and the women's courtship matching:

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

(a)

	1	2	3
Oscar:	Kara	Lilly	Mary
Peter:	Kara	Lilly	Mary
Quinn:	Kara	Mary	Lilly
Ralph:	Lilly	Kara	Mary

	1	2	3	4
Kara:	Ralph	Quinn	Peter	Oscar
Lilly:	Ralph	Oscar	Quinn	Peter
Mary:	Oscar	Ralph	Peter	Quinn

(b) The notation ϕ signifies a preference for remaining single.

	1	2	3	4
Grant:	Gloria	Hanna	ϕ	Ida
Howard:	Hanna	Gloria	Ida	ϕ
Isaac:	Gloria	Ida	ϕ	Hanna
Jack:	Hanna	Gloria	ϕ	Ida

	1	2	3	4	5
Gloria:	Jack	Isaac	ϕ	Howard	Grant
Hanna:	Jack	Howard	Grant	ϕ	Isaac
Ida:	Isaac	Grant	ϕ	Howard	Jack

22.27 Let A be a matching (not necessarily stable), and let M_{A,O^m} be the set of men who prefer the women to whom they are matched under A to the women to whom they are matched under the men's courtship matching O^m.

(a) Prove that if A is a stable matching then $M_{A,O^m} = \emptyset$.

In this exercise we prove that if $M_{A,O^m} \neq \emptyset$, then there exists a pair (m, w) who object to the matching A, and $m \notin M_{A,O^m}$.

Denote by W_1 the set of the women who are matched to men in M_{A,O^m} under the matching A, and by W_2 the set of the women matched to men in M_{A,O^m} under O^m.

(b) Prove that if $W_1 \neq W_2$, then $W_1 \setminus W_2 \neq \emptyset$.
(c) Prove that if $W_1 \neq W_2$, every woman $w \in W_1 \setminus W_2$ objects to A along with the man to whom she is matched under O^m.
(d) From here to the end of the exercise, assume that $W_1 = W_2$. Prove that every woman in W_1 dismisses at least one man under the men's courtship algorithm (her match under the matching A).
(e) Consider the woman $w^* \in W_1$ who was the last woman approached by a man m^* from the set M_{A,O^m} under the men's courtship algorithm. Prove that when w^* receives an offer from m^*, there was another man at her doorstep, call him m', whom she dismissed in favor of m^*.
(f) Use the fact that w^* is the last woman to get an offer from a man in M_{A,O^m} to show that m' is not in M_{A,O^m}.
(g) Prove that the pair (m', w^*) object to the matching A.

22.28 Suppose we are given a matching problem G. Let O^m be the men's courtship matching of this problem. In this exercise we prove the claim on page 900: in the course of the Gale–Shapley algorithm, a man cannot obtain a better result by pretending to have a preference relation that is different from his true preference relation. We will, in fact, prove a stronger claim: even if a set of men all pretend to have preference relations that are different from their true preference relations, under any stable matching of the new problem that is different from O^m, at least one of these men loses out; in other words, one of these men will be matched to a woman whom he prefers less than the woman to whom he has been matched under O^m.

Let $\widehat{G}$ be the matching problem derived from G by changing the preference relations of some of the men: let $\widehat{M} \subseteq M$ be a set of men whom we term "dishonest," whose preference relations in matching problem $\widehat{G}$ differ from their preference relations in matching problem G.

Using Exercise 22.27, prove that there is no stable matching $\widehat{A}$ for matching problem $\widehat{G}$ satisfying the following properties: every man in $\widehat{M}$ prefers (according to his true preference relation) the woman to whom he is matched under $\widehat{A}$ to the woman to whom he is matched under O^m.

22.29 We saw in Example 22.24 that under the men's courtship algorithm a woman may be matched to a man whom she prefers to the man to whom she is matched under the men's courtship matching O^m by pretending that her preference relation is different from her true preference relation. In this exercise we present a condition that guarantees that a matching resulting from such strategic behavior on the part of a woman is a stable matching.

Suppose that under the men's courtship matching O^m Messalina is matched to Claudius. Consider the matching A that results from the men's courtship algorithm when Messalina pretends that her preference relation is different from her true preference relation.

(a) Prove that the matching A is a stable matching under the true preferences of the men and women if and only if the man to whom Messalina is matched under A is the man she most prefers from among the men that came to her door throughout the algorithm; that is, she does not regret any rejection she made.

(b) Conclude that if Messalina improved her result by this behavior then she necessarily makes at least one man worse off; i.e., there is at least one man who is matched under A to a woman whom he prefers less than the woman to whom he is matched under O^m.

22.30 Describe in detail the generalization of the Gale–Shapley algorithm for the case in which the single life is not universally considered the worst possible outcome (see Section 22.5.3 on page 900), and prove that the algorithm always terminates in finding a stable matching.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

22.31 Describe in detail the generalization of the Gale–Shapley algorithm (the students' courtship algorithm) in the case in which each university has a quota for the maximal number of students that it can accept (see Section 22.5.4 on page 903), and prove that the algorithm always terminates in finding a stable matching. Describe the universities' courtship algorithm in this case. Which algorithm is more preferred by the students? Explain.

22.32 In this exercise, we work with polygamous matchings (see Section 22.5.4 on page 903).

(a) Prove that in the students' courtship algorithm, every student is matched to the university that is most preferred by him from among all the universities to which he is matched under all possible stable matchings.
(b) Given a stable matching, prove that every student who is not matched by any university under this matching is not matched to any university under any stable matching.
(c) Prove that if under one stable matching there is a university that does not fill its student quota, then that university does not fill its student quota under any stable matching.

22.33 In this exercise we generalize the model of polygamous matchings (see Section 22.5.4 on page 903).

Suppose that every university $u \in U$ is given a quota $q_u \geq 1$, a subset $S_u \subseteq S$ of the set of students, and a preference relation over the set S_u. The set S_u is interpreted as the set of students that university u is willing to accept: university u will not accept any student who is not in S_u, even if that means that it will not fill its quota of students. Suppose that each student $s \in S$ is given a subset $U_s \subseteq U$ of the set of universities, and a preference relation over U_s. The set U_s is interpreted as the set of universities that the student is willing to attend. If a student s is rejected by all the universities in S_u, then he prefers not attending any university.

(a) Generalize the definition of stable matching to this case.
(b) Generalize the Gale–Shapley algorithm to this case, and prove that it terminates after finding a stable matching.
(c) Prove that a student who is not accepted by any university under a particular stable matching will not be accepted into any university under any stable matching.

22.34 The following tables depict the preferences of three universities regarding a set of applicants, and the preferences of the applicants regarding the universities. Every university also has a specified quota for the number of students that it can accept. These preferences are strict preferences, i.e., there are no instances of indifference. Find a stable matching A between the universities and the applicants that is preferred by all the applicants to any other stable matching; i.e., for any stable matching B, an applicant who is matched to two different universities under

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

A and B prefers the university to which he is matched under A to the university to which he is matched under B.

(a) The preferences of the universities (from left to right) and each university's quota of students (between the round brackets) are:
University X (4): $j\ b\ a\ k\ g\ d\ c\ e\ f\ h\ i$.
University Y (2): $d\ b\ h\ a\ j\ f\ k\ e\ c\ i\ g$.
University Z (3): $f\ c\ j\ b\ h\ d\ e\ g\ i\ k\ a$.

The preferences of the applicants (from left to right):

$a : ZYX$	$d : ZXY$	$g : XZY$	$j : XYZ$
$b : ZYX$	$e : ZXY$	$h : YXZ$	$k : XYZ$
$c : YZX$	$f : XZY$	$i : YXZ$	

(b) The preferences of the universities (from left to right) and each university's quota of students (between the round brackets):
University X (6): $a\ b\ c\ d\ e\ f\ g\ h\ i\ j\ k\ l\ m\ n$.
University Y (6): $n\ m\ l\ k\ j\ i\ h\ g\ f\ e\ d\ c\ b\ a$.
The preferences of the applicants: Applicants $a, b, c, d, e, f, g, h, i, j$ prefer University X to University Y, and applicants k, l, m, n prefer University Y to University X.

22.35 In this exercise, we will assume a given matching problem in which the preference relations are weak preference relations, meaning that the preference relations may include indifference; a woman may be indifferent to being matched to any of several different men, and a man may be indifferent to being matched to any of several different women. For example, Yoko may be indifferent between John, Paul, George, and Ringo, and she may be indifferent between Tony and Allan, and she may prefer each of John, Paul, George, and Ringo to Tony and to Allan. That is,

$$\text{Yoko} : \text{Ringo} \approx \text{George} \approx \text{Paul} \approx \text{John} \succ \text{Allan} \approx \text{Tony}.$$

A man and a woman object to a matching if they both (strictly) prefer each other to the woman and man to whom they are respectively matched. A matching is stable if there are no men and women objecting to it.

(a) Given each man's weak preference relation, construct a strict preference relation by ordering the women to whom he is indifferent in an arbitrary way, while maintaining transitivity of preferences, and do the same with the preference relation of each woman. Prove that a stable matching under this constructed set of preference relations is also a stable matching in the original matching problem with weak preferences.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

(b) Find at least two stable matchings for the following preference relations:

Tina : Albert $\succ$ Boris $\succ$ Chester $\approx$ David
Ursula : Chester $\approx$ Albert $\succ$ Boris $\succ$ David
Victoria : Chester $\approx$ David $\succ$ Boris $\approx$ Albert

Albert : Tina $\approx$ Victoria $\succ$ Ursula
Boris : Victoria $\succ$ Ursula $\approx$ Tina
Chester : Tina $\succ$ Victoria $\approx$ Ursula
David : Victoria $\approx$ Tina $\approx$ Ursula

22.36 As in Exercise 22.35, consider a matching problem in which the preference relations of the men and women may include instances of indifference. Given a stable matching to such a matching problem, is it always possible to replace the weak preference relations (i.e., with instances of indifference) with strict preference relations in such a way that the stable matching with respect to the weak preference relations is also a stable matching with respect to the strict preference relations? If your answer is yes, prove it; if your answer is no, provide a counterexample.

22.37 Theorem 22.23 is proved on page 899 under the assumption that all preference relations are strict. Show by example that the theorem does not hold if there may be instances of indifference in the preference relations of the men and the women.

22.38 Example 22.32 (page 904) has no stable matching. Complete the proof of this statement.

22.39 Suppose that a population of size $3n$ is partitioned into three subsets: n contractors, n carpenters, and n plumbers. Each person in this population has two preference relations: a preference relation over each one of the two subsets listed above to which he does not belong. For example, each contractor has a preference relation over the set of carpenters, and a preference relation over the set of plumbers, and so on. A matching A in this case is a partition of the population into n triples, each composed of a contractor, a carpenter, and a plumber. A trio composed of a contractor x, a carpenter y, and a plumber z has an objection to A if (a) the matching A does not contain the set $\{x, y, z\}$, and (b) every pair of workers within this trio who are not matched to each other under A prefer each other to the corresponding workers to whom they have been matched under A. In other words, x prefers y to the carpenter to whom he is matched under A (if he is matched to a carpenter other than y), x prefers z to the plumber to whom he is matched under A (if he is matched to a plumber other than z), y prefers x to the contractor to whom he is matched under A (if he is matched to a contractor other than x), and so on. A matching is stable if there is no trio composed of a contractor, a carpenter, and a plumber who object to it.

Does there always exist a stable matching in this model? If your answer is no, provide a counterexample. If your answer is yes, provide an algorithm for finding a

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press

stable matching, and prove that this algorithm always terminates in finding a stable matching.

22.40 Repeat Exercise 22.39, but this time assume that each member of the population has a preference relation over pairs of potential co-workers, i.e., every contractor has a preference relation over the set of pairs composed of a carpenter and a plumber, every carpenter has a preference relation over the set of pairs composed of a contractor and a plumber, and every plumber has a preference relation over the set of pairs composed of a contractor and a carpenter. A contractor x, carpenter y, and plumber z have an objection to a matching A if (a) A does not contain the set $\{x, y, z\}$, and (b) each member of the trio $\{x, y, z\}$ ranks the other two above the pair to which he is matched under A. A matching is stable if there is not a trio composed of a contractor, a carpenter, and a plumber who have an objection to it.

https://doi.org/10.1017/CBO9780511794216.023 Published online by Cambridge University Press