Market Design

Who Gets What – and Why

Instructor: Kaniṣka Dam
Semester: Spring, 2023
Instituto Tecnólogico Autónomo de México

ITAM

Preface

The secondary title of the course is inspired by Al Rorth's book *Who Gets What – and Why*. These notes are heavily borrowed from Guillaume Haeringer's excellent book *Market Design* (Haeringer, 2017). Alejandro Robinson (assistant professor of economics at University of Exeter), my ex-student, coauthor and friend, has been enormously generous to let me use his class notes on Market Design (Robinson-Cortés, 2021). In many parts of the notes, I have literally paraphrased texts from Guillaume's book and Alejandro's notes. I made these notes principally for my own use to teach the course. They must be full of errors and typos. Also, interpretations of some results are my own, which may be misconstrued. So, as we go on, we surely will discover a huge room for improvement. I have also used gender neutral pronouns such as "their", "them", etc. to refer to an individual or an agent. I hope this does not create much confusions. Please let me know about all errors or any other comments you have at kaniska.dam@itam.mx.

Mexico City
Spring, 2023

Contents

Part I

Preliminaries of markets and market design

Chapter 1

Markets

1.1 An exchange economy

An exchange economy or market consists of two individuals (traders), denoted by $i = 1, 2$, and two goods, denoted by $j = 1, 2$. Each trader is born with an endowment of each good. Formally, $\omega_{ij} \geq 0$ is trader i's endowment of good j. The aggregate endowment of good j is thus given by $\omega_j = \omega_{1j} + \omega_{2j}$.

Each individual has preferences over the bundles of the two commodities, which is assumed to be rational (complete and transitive). We would assume that preferences are continuous so that they can be represented by a utility function $u : X \to \mathbb{R}$ where $X \subseteq \mathbb{R}^2_+$ is the set of all consumption bundles. Consumption bundles are denoted by x, y, z, etc. where x_{ij} denotes trader i's consumption of good j.

The market with two traders and two goods are represented by the Edgeworth box in Figure 1.1. The two corners represents the origins of the two agents—the southwest corner corresponds to trader 1 and the northeast corner, to trader 2. The endowment is denoted by the point ω. The blue curves are agent 1's indifference curves given their preferences $u(x_{11}, x_{12})$. The red curves are the indifference curves given their preferences $v(x_{21}, x_{22})$. The size of the Edgeworth box is determined by the aggregate endowments of the two goods—the length represents ω_1 and the width represents ω_2.

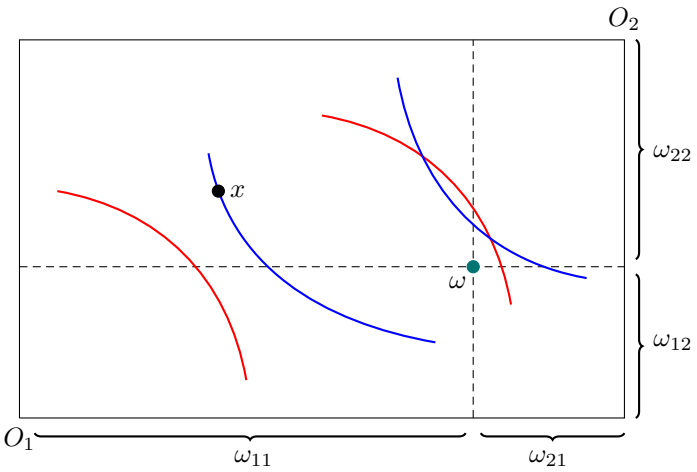


Figure 1.1: A market with two traders.

Now we turn to the concept of an allocation in the Edgeworth box. In Figure 1.1, the point x is an allocation with the property that the aggregate consumption of any good must exhaust its aggregate endowment. Formally,

Definition 1.1: Allocation

A point x in the Edgeworth box is a (consumption) allocation if $x_{1j} + x_{2j} = \omega_j$ for $j = 1, 2$. A set of allocations is given by:

$$A = \{x \mid x_{ij} \geq 0 \text{ for all } i = 1, 2, \ j = 1, 2 \text{ and } x_{1j} + x_{2j} = \omega_j \text{ for } j = 1, 2\}.$$

Definition 1.1 asserts that if a point x represents an allocation, then it cannot be outside the Edgeworth box. In other words, the way an allocation is defined incorporates the notion of feasibility, and in our context, "allocation" is simply a short-hand for "feasible allocation". Also, A, the set of allocations is the entire box. A simpler version of the Edgeworth box economy is obtained when there is only one good (a good without subscript j). Let it's aggregate endowment be $\omega = \omega_1 + \omega_2$. A (consumption) allocation, x of this economy depicted in Figure 1.2. An allocoation $x = (x_1, x_2)$ is simply a point on the line (of length ω) so that $x_1 + x_2 = \omega_1 + \omega_2 = \omega$.

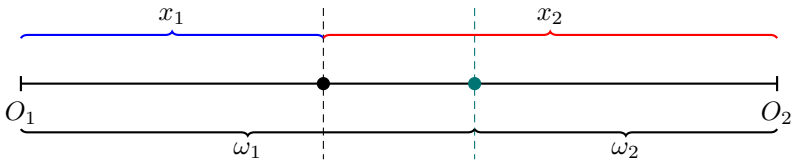


Figure 1.2: A market with two traders and one good.

Every point like x on the line represents an allocation, and, conversely, every allocation can be represented by a point, or division of the line.

In what follows, we would also assume that the traders are self-interested, i.e., the preferences of no trader depends only on their own consumption bundles, and not on that of the other trader. Formally,

$$u(x_{11}, x_{12}, x_{21}, x_{22}) = u(x_{11}, x_{12}),$$
$$v(x_{11}, x_{12}, x_{21}, x_{22}) = v(x_{21}, x_{22}).$$

In other words, we would abstract from consumption externalities.

1.1.1 Pareto efficiency, individual rationality and the core

Pareto efficient allocations. The notion of Pareto efficiency is associated with the idea of improving the situations of the traders in the exchange economy. Consider an allocation x in the Edgeworth box. Can there be a different allocation x' that makes the traders better off? If this is the case for both traders, then allocation x is not a "good" allocation. Formally,

Definition 1.2: Pareto efficient allocation

An allocation x' Pareto dominates another allocation x if $u(x'_{i1}, x'_{i2}) \geq u(x_{i1}, x_{i2})$ for all $i = 1, 2$, and $u(x'_{i1}, x'_{i2}) > u(x_{i1}, x_{i2})$ for at least one i. An allocation x is a Pareto efficient allocation if there is no allocation x' that Pareto dominates x. The set of all Pareto efficient allocations is called the contract curve.

Definition 1.2 asserts that at an alternative allocation, there is no way to make both traders better off and make at least one strictly better off. How do we compute and represent a Pareto efficient allocation in the Edgeworth box? In Figure 1.3, allocations x and y are both Pareto efficient allocations. At these allocations, at

which the indifference curves of traders 1 and 2 are tangent to each other, it is impossible to make one trader better strictly off without making the other strictly worse off. The curve that joins all such points is the contract curve for the exchange economy.

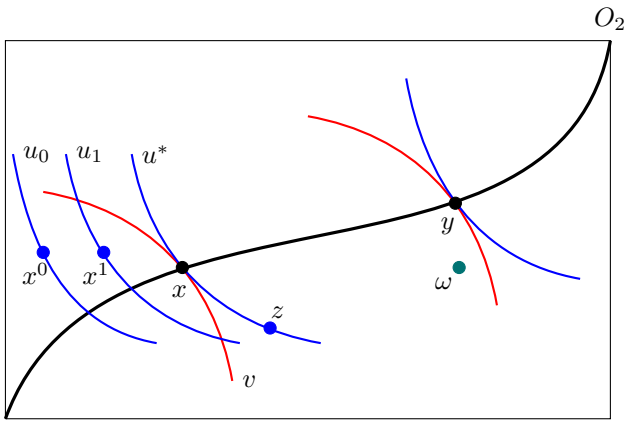


Figure 1.3: Pareto efficient allocations.

The tangency points can be derived in the following way.

- Fix a utility level of trader 2 at v, i.e., $v(x_{21},\, x_{22}) = v$. Think of proposing allocations at all of which we are required to guarantee at least v to trader 2.
- Start from an allocation at which trader 1 obtains u_0. That is, start from allocation such as x^0 (on trader 1's indifference curve labeled u_0).
- Try to improve trader 1's utility. We must move to a higher indifference curve of this trader, say at level u_1, i.e., we must propose an allocation such as x^1.
- However, u_1 is not the best we can do. At an allocation like x^1, trader 2 obtains a utility level that is strictly higher than v, but there is still room for improvement for trader 1 (respecting the minimum utility constraint of trader 2).
- We can go as far as allocation x at which the two indifference curves are tangent to each other. Going further to x will now improve the utility of trader 1, but at the cost of diminishing utility of trader 2. So, x is a Pareto efficient allocation. Note also that a point like z is not Pareto efficient. It yields u^* to trader 1, but z is on the indifference curve of trader 2, which is strictly lower than that gives them v (so, the minimum utility constraint is violated).

So, to find a Pareto efficient allocation we would solve the following problem:

$$\begin{aligned} \max_{\{x_{11},\, x_{12},\, x_{21},\, x_{22}\}} \quad & u(x_{11},\, x_{12}), \\ \text{subject to } & v(x_{21},\, x_{22}) \geq v, \\ & x_{11} + x_{21} = \omega_1, \\ & x_{12} + x_{22} = \omega_2. \end{aligned}$$

The first constraint is the minimum utility constraint of trader 2, and the other two constraints are the feasibility constraints of goods 1 and 2. We would rather solve the following numerical example.

Example 1.1: Finding Pareto efficient allocations

Let the utility functions are given by $u(x_{11}, x_{12}) = x_{11}x_{12}$ and $v(x_{21}, x_{22}) = x_{21}x_{22}$, and the endowments of the two goods are given by $(\omega_{11}, \omega_{12}) = (2, 2)$ and $(\omega_{21}, \omega_{22}) = (2, 1)$. So, the aggregate endowments are $\omega_1 = 2 + 2 = 4$ and $\omega_2 = 2 + 1 = 3$ (i.e., the Edgeworth box has length 4 and width 3). We can write the above maximization problem as follows:

$$\begin{aligned} \max_{\{x_{11}, x_{12}\}} \; & u(x_{11}, x_{12}) = x_{11}x_{12}, \\ \text{subject to } & v(x_{21}, x_{22}) = x_{21}x_{22} \geq v, \\ & x_{11} + x_{21} = \omega_1 = 4, \\ & x_{12} + x_{22} = \omega_2 = 3. \end{aligned}$$

First, note that the first order condition of the above maximization problem is given by:

$$\underbrace{\frac{x_{12}}{x_{11}}}_{-\text{MRS}^1} = \underbrace{\frac{x_{22}}{x_{21}}}_{-\text{MRS}^2} = \alpha,$$

i.e., at any Pareto efficient allocation, the indifference curves of the two traders are tangent to each other. The above equations imply that

$$x_{i2} = \alpha x_{i1} \quad \text{for } i = 1, 2.$$

Substituting the above into the second feasibility constraint, we get

$$\alpha\underbrace{(x_{11} + x_{21})}_{=\omega_1} = \omega_2 \quad \Longleftrightarrow \quad \alpha = \frac{\omega_2}{\omega_1} = \frac{3}{4}.$$

Therefore, the contract curve is given by:

$$x_{12} = \frac{\omega_2}{\omega_1} \cdot x_{11} = \frac{3}{4} \cdot x_{11}.$$

The contract curve is the diagonal of the Edgeworth box. It is easy to see from Figure 1.3 that the minimum utility constraint must hold with equality, i.e., $\alpha(x_{21})^2 = v$. Moreover, $x_{22} = \alpha x_{21}$. Let the Pareto efficient allocations to trader i be denoted by $(\hat{x}_{i1}, \hat{x}_{i2})$. The last two conditions alongwith $\alpha = \omega_2/\omega_1$ yield

$$(\hat{x}_{21}, \hat{x}_{22}) = \left(\sqrt{\frac{\omega_1 v}{\omega_2}}, \sqrt{\frac{\omega_2 v}{\omega_1}}\right) = \left(\sqrt{\frac{4v}{3}}, \sqrt{\frac{3v}{4}}\right).$$

Using the two feasibility constraints, we obtain

$$(\hat{x}_{11}, \hat{x}_{12}) = \left(\omega_1 - \sqrt{\frac{\omega_1 v}{\omega_2}}, \omega_2 - \sqrt{\frac{\omega_2 v}{\omega_1}}\right) = \left(4 - \sqrt{\frac{4v}{3}}, 3 - \sqrt{\frac{3v}{4}}\right).$$

Observations

- The Pareto efficient consumption x_{ij} does not depend on the an individual's endowment of each good, ω_{ij}. It only depends on the aggregate endowment of each good, ω_1 and ω_2 (i.e., the length and width of the Edgeworth box). Thus, if we change the length and width of the box, the Pareto efficient allocations (which lie on the diagonal of the box) change.
- A given Pareto efficient allocation depends on the minimum utility level v of trader 2, and hence, there are a continuum of such allocations, one for each v.

Exercise 1.1

Let the utility functions be given by $u(x_{11}, x_{12}) = \min\{x_{11}, x_{12}\}$ and $v(x_{21}, x_{22}) = \min\{x_{21}, x_{22}\}$. The endowments are $(\omega_{11}, \omega_{12}) = (2, 0)$ and $(\omega_{21}, \omega_{22}) = (0, 3)$. Find the set of Pareto efficient allocations.

Utility possibility frontier. We shall now define two concepts associated with Pareto efficiency. Consider the efficient allocations and compute the maximum value function of the maximization problem in Example 1.1:

$$\begin{aligned} u &= \underbrace{\frac{\omega_2}{\omega_1}}_{\alpha} \cdot (x_{11})^2 = \frac{\omega_2}{\omega_1} \cdot \left(\omega_1 - \sqrt{\frac{\omega_1 v}{\omega_2}}\right)^2 \\ \iff & \sqrt{\frac{\omega_1 u}{\omega_2}} + \sqrt{\frac{\omega_1 v}{\omega_2}} = \omega_1 \\ \iff & \sqrt{u} + \sqrt{v} = \sqrt{\omega_1 \cdot \omega_2} = \sqrt{12}. \end{aligned} \tag{UPF}$$

Condition (UPF) describes what is called the utility possibility frontier (UPF) or the Pareto frontier or the bargaining frontier of the Edgewroth box economy. We can also write (UPF) as

$$u = \phi(v) \equiv (\sqrt{\omega_1 \cdot \omega_2} - \sqrt{v})^2. \tag{UPF$'$}$$

In many contexts, the final utilities that are accrued to the traders, rather than their allocations, are more important and convenient to use. The above discussion is easily generalizable to a market with $n \geq 2$ traders and $m \geq 2$ goods.

Definition 1.3: Utility possibility set

Let $X = \{1, \ldots, m\}$ be the set of m goods that can be potentially traded in an exchange economy or market with the set of traders, $N = \{1, \ldots, n\}$. Let trader i's utility function over m goods be given by $u^i(x_i)$, where $x_i = (x_{i1}, \ldots, x_{im})$ denotes trader i's allocations of m goods, which is continuous and monotonic. The function

$$\phi(u_2, \ldots, u_n) = \max_x \left\{ u^1(x_1) \mid u^i(x_i) \geq u_i \text{ for } i \in N \setminus \{1\} \text{ and } \sum_{i=1}^{n} x_{ij} = \omega_j \text{ for all } j \in X \right\}$$

is called the utility possibility frontier of the exchange economy. The set

$$\mathcal{U} = \{(u_1, \ldots, u_n) \in \mathbb{R}^n_+ \mid u^1 \leq \phi(u_2, \ldots, u_n)\}$$

is called the utility possibility set of the exchange economy.

The Pareto frontier (UPF) and the associated utility possibility set are depicted in Figure 1.4. The utility possibility set is simply the combinations of payoffs that can be reached by the market participants by trading what they initially have (endowments) in a decentralized manner. So, we can see that there is a one-to-one correspondence between the set of (feasible) allocations and the utility possibility set. Also, there is a one-to-one correspondence between the contract curve and the Pareto frontier.

Let us derive the UPF for an exchange economy wherein there are $n \geq 2$ traders and 2 goods, 1 and 2.

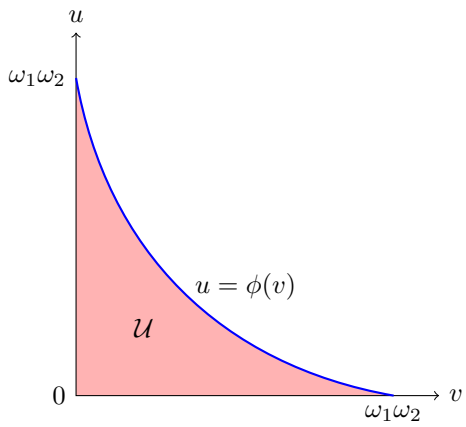


Figure 1.4: The Pareto frontier and the utility possibility set.

Example 1.2: UPF with many traders

Consider the set of traders $N = \{1, \ldots, n\}$ and two goods, 1 and 2. The utility function of trader i is given by $u^i(x_i)$ where $x_i = (x_{i1}, x_{i2})$ denotes trader i's allocations of 2 goods. Trader i's endowments are given by $(\omega_{i1}, \omega_{i2})$, and the aggregate endowment of good $j = 1, 2$ is given by $\sum_{i=1}^{n} \omega_{ij} = \omega_j$. A Pareto efficient allocation, $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ solves

$$\begin{aligned} \max_x \ & u^1(x_1) = x_{11}x_{12}, \\ \text{subject to } & u^2(x_2) = x_{21}x_{22} \geq u_2, && (\text{U}_2) \\ & \vdots \\ & u^n(x_n) = x_{n1}x_{n2} \geq u_n, && (\text{U}_n) \\ & x_{11} + \ldots + x_{n1} = \omega_1, && (\text{F}_1) \\ & x_{12} + \ldots + x_{n2} = \omega_2. && (\text{F}_2) \end{aligned}$$

At the Pareto efficient allocations, the marginal rate of substitution between the two goods (MRS) of all traders must be equal (i.e., the indifference curves of all traders are tangent to each other):

$$\text{MRS}^1 = \ldots = \text{MRS}^n \quad \Longleftrightarrow \quad \frac{x_{12}}{x_{11}} = \ldots = \frac{x_{n2}}{x_{n1}} = \alpha \text{ (say)}.$$

Therefore, we have $x_{i2} = \alpha x_{i1}$ for all $i \in N$. Substituting this into the second feasibility constraint (F_2), we obtain

$$\alpha(\underbrace{x_{11} + \ldots + x_{n1}}_{=\omega_1 \text{ from } (\text{F}_1)}) = \alpha \cdot \omega_1 = \omega_2 \quad \Longleftrightarrow \quad \alpha = \frac{\omega_2}{\omega_1}.$$

Constraints (U_2)-(U_n) would bind at the optimum, and hence, we have

$$\begin{aligned} & \alpha(x_{21})^2 = \frac{\omega_2}{\omega_1} \cdot (x_{21})^2 = u_2, \ \ldots, \ \alpha(x_{n1})^2 = \frac{\omega_2}{\omega_1} \cdot (x_{n1})^2 = u_n \\ \Longleftrightarrow \ & \hat{x}_{21} = \sqrt{\frac{\omega_1 u_2}{\omega_2}}, \ \ldots, \ \hat{x}_{n1} = \sqrt{\frac{\omega_1 u_n}{\omega_2}}. \end{aligned}$$

Let u_1 denote the maximized utility of trader 1, which is given by:

$$u_1 = \alpha(\hat{x}_{11})^2 = \frac{\omega_2}{\omega_1} \cdot (\hat{x}_{11})^2 = \frac{\omega_2}{\omega_1} \cdot (\omega_1 - \hat{x}_{21} - \ldots - \hat{x}_{n1})^2$$

$$\iff \frac{\omega_1 u_1}{\omega_2} = \left(\omega_1 - \sqrt{\frac{\omega_1 u_2}{\omega_2}} - \ldots - \sqrt{\frac{\omega_1 u_n}{\omega_2}}\right)^2$$

$$\iff \sqrt{\frac{\omega_1 u_1}{\omega_2}} + \sqrt{\frac{\omega_1 u_2}{\omega_2}} + \ldots + \sqrt{\frac{\omega_1 u_n}{\omega_2}} = \omega_1$$

$$\iff \sqrt{u_1} + \sqrt{u_2} + \ldots + \sqrt{u_n} = \sqrt{\omega_1 \cdot \omega_2} \iff u_1 = \phi(u_2, \ldots, u_n) \equiv \left(\sqrt{\omega_1 \cdot \omega_2} - \sum_{i=2}^{n} \sqrt{u_i}\right)^2.$$

The above is our desired UPF.

Individually rational allocations. We have seen so far that a Pareto efficient allocation does not depend on individual's endowment of each good. The next property we would like to impose on an allocation is individual rationality. We would require that each trader must do at least as well as what they could with their endowments. Formally,

Definition 1.4: Individually rational allocation

An allocation x in the Edgeworth box is individually rational if $u(x_{11}, x_{12}) \geq u(\omega_{11}, \omega_{12})$ and $v(x_{21}, x_{22}) \geq v(\omega_{21}, \omega_{22})$.

The above definition asserts that, at an individually rational allocation, each trader obtains utility that is at least as large as their utility evaluated at their endowments of the two goods.

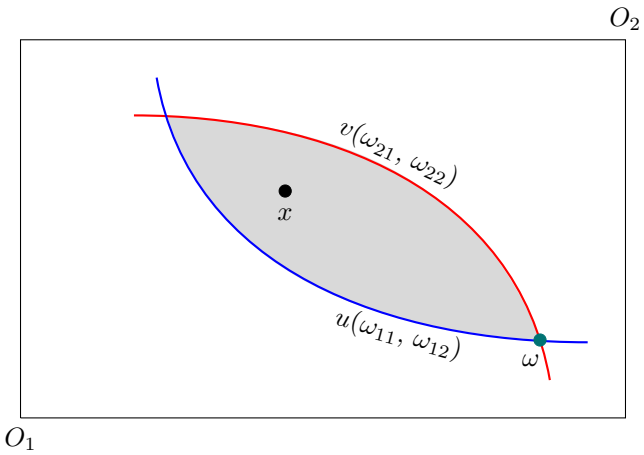


Figure 1.5: Individually rational allocations.

In Figure 1.5, both traders obtain higher utility at a point like x that that evaluated at ω because at x lies at a higher indifference curve of each trader. Therefore, the set of individually rational allocations relative to the endowment point ω is lens-shaped shaded region.

Core allocations. The next property we would like to impose on an allocation is that it is in the core of the exchange economy. This idea will be generalized in Chapter 4. The notion of core is related to the idea that

some allocations can be blocked or objected by the traders. In our 2×2 exchange economy, there are two ways to block a proposed allocation—(a) each trader on their own can object to a proposed allocation, and (b) the two traders together can block a proposed allocation. In a more general model wherein there are $n > 2$ traders, any non-empty subsets of traders, called coalitions, are allowed to block an allocation (there are $2^n - 1$ such coalitions). Formally,

Definition 1.5: Core allocations

An allocation x in the Edgeworth box is blocked by the trader(s) if there is another (feasible) allocation x' such $u(x'_{11}, x'_{12}) \geq u(x_{11}, x_{12})$ and $v(x'_{21}, x'_{22}) \geq u(x_{21}, x_{22})$ with strict inequality for at least one trader. An allocation x is a core allocation if there is no other allocation x' with which x can be blocked.

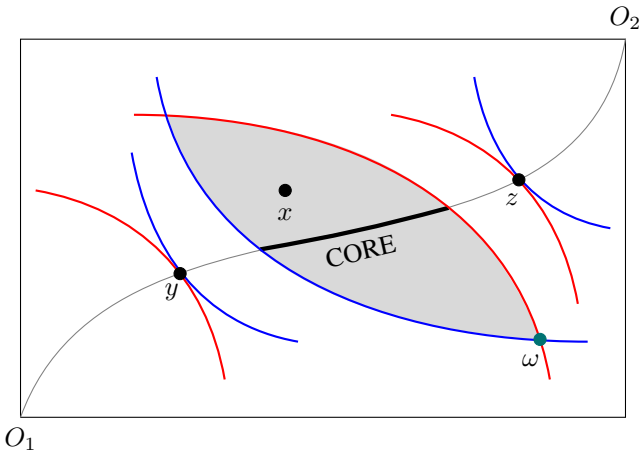


Figure 1.6: The core.

What is crucial is that the traders can object to an allocation by proposing another allocation that is feasible for them. Let us analyze the various possibilities of blocking.

- Consider the case when a single trader can block a proposed allocation. Note that the only feasible allocation that a trader has and can be used to block an allocation their endowments of the two goods, i.e., $(\omega_{i1}, \omega_{i2})$ for each $i = 1, 2$. Therefore, allocations like y and z (in Figure 1.6) cannot be core allocation either (y will blocked by trader 1 and z, by trader 2). Consequently, a core allocation must lie within the lens-shaped shaded region in Figure 1.6.

- Next, consider the possibility of blocking by the two traders together. Proposing an alternative allocation to block an initial one means that the two traders can exchange their endowments to create a "new allocation" that makes both of them better off. In Figure 1.6, an allocation like x will be blocked by the traders together. To see this, create another lens-shaped region with x and ω. Any allocation in the newly-created region yields strictly higher utility of both traders. If we keep on doing this, the only allocations that would not be blocked are those on the contract curve (when no further lens-shaped region of improvement can be created).

From the above discussion we can conclude

Theorem 1.1

An allocation in the Edgeworth box is Pareto efficient and individually rational if and only if it is a core allocation.

The proof is easy for the Edgeworth box economy. In Figure 1.6, the part of contract curve that lies inside the lens-shaped region is the core of our exchange economy. For an exchange economy with more than 2 traders, it is always true that a core allocation is Pareto efficient and individually rational, but the converse is not true in general.

Exercise 1.2: Core is a stronger concept than Pareto efficiency and individual rationality

Consider an exchange economy with three traders, 1, 2 and 3, and two goods, 1 and 2. The utility functions are given by $u^i(x_{i1}, x_{i2}) = x_{i1}x_{i2}$ for $i = 1, 2, 3$. The endowments are given by $(\omega_{11}, \omega_{12}) = (1, 9)$, $(\omega_{21}, \omega_{22}) = (5, 5)$ and $(\omega_{31}, \omega_{32}) = (9, 1)$. Show that there is at least one allocation that is Pareto efficient and individually rational, but it is not a core allocation. Find the core allocations of this market.

The reason for which Theorem 1.1 does not hold for an exchange economy with more than 2 traders is that the possibility of blocking a proposed allocation increases as the number of traders grows. If there are 5 traders, then there are $2^5 - 1 = 31$ coalitions that can potentially block an allocation. However, a market with 20 traders, this number becomes $2^{20} - 1 = 1{,}048{,}575$. One may also wonder what happens if we replicate our Edgeworth box economy.

Exercise 1.3: Core shrinks under replication

Consider an exchange economy with two traders, 1 and 2, and two goods, 1 and 2. The utility functions are given by $u^i(x_{i1}, x_{i2}) = x_{i1}x_{i2}$ for $i = 1, 2$. The endowments are given by $(\omega_{11}, \omega_{12}) = (9, 1)$ and $(\omega_{21}, \omega_{22}) = (1, 9)$. Show that Theorem 1.1 holds for this market.
Next, consider a replication of the Edgeworth box in the following way. Add two more traders, 3 and 4. Trader 3 is the identical twin of trader 1, and trader 4 is the identical twin of trader 2. This is to say that $u^i(x_{i1}, x_{i2}) = x_{i1}x_{i2}$ for $i = 1, 2, 3, 4$, and $(\omega_{31}, \omega_{32}) = (9, 1)$ and $(\omega_{41}, \omega_{42}) = (1, 9)$. Show that Theorem 1.1 does not hold for the replicated economy.

Edgeworth conjectured in 1881 that the core is large in a small market, whereas it is small in a large market. Aumann (1964) proves a striking result that the core of an exchange economy shrinks as the market is replicated. Moreover, he shows that if the market is replicated infinitely, the core converges to a single allocation. This motivates our analysis of a competitive or Walrasian equilibrium.

1.1.2 The Walrasian equilibrium

We now introduce a Walrasian allocation for the Edgeworth box economy. For that we require a price system which the traders would take as given while trading with each other. Let $(p_1, p_2) \in \mathbb{R}_{++}$ be the price vector. The budget set of trader $i = 1, 2$ is given by:

$$B_i(p_1, p_2) = \{(x_{i1}, x_{i2}) \in \mathbb{R}_+ \mid p_1x_{i1} + p_2x_{i2} \leq p_1\omega_{i1} + p_2\omega_{i2}\}.$$

Definition 1.6: Walrasian equilibrium

A Walrasian (or competitive) equilibrium for the Edgeworth box economy is a price vector $p^* = (p_1^*, p_2^*)$ and an allocation $x^* = ((x_{11}^*, x_{12}^*), (x_{21}^*, x_{22}^*))$ in the Edgeworth box such that for each $i = 1, 2$, (x_{i1}^*, x_{i2}^*) solves

$$\max_{\{x_{i1}, x_{i2}\}} u^i(x_{i1}, x_{i2}),$$
$$\text{subject to } p_1x_{i1} + p_2x_{i2} \leq p_1\omega_{i1} + p_2\omega_{i2}.$$

In a Walrasian equilibrium, given the commodity prices, each trader maximizes their utility subject to the budget constraint. Because x^* is an allocation, we must by definition have supply equal to demand for every good, i.e.,

$$x_{1j} + x_{2j} = \omega_{1j} + \omega_{2j} \quad \text{for each good } j = 1, 2.$$

Also, notice that a Walrasian equilibrium depends on the endowments, ω. If we change ω, we change the competitive equilibrium. The maximization problem in Definition 1.6 clearly yields

$$MRS^1_{12} = -\frac{p_1}{p_2} = MRS^2_{12}.$$

In Example 1.3, we shall compute Walrasian equilibrium allocations under specific functional forms of the utility functions. Figure 1.7 depicts a competitive equilibrium where the line joining ω and x^* represents the relative prices with slope $-p_1^*/p_2^*$. As we can clearly see that x^* is a core allocation. We state the following

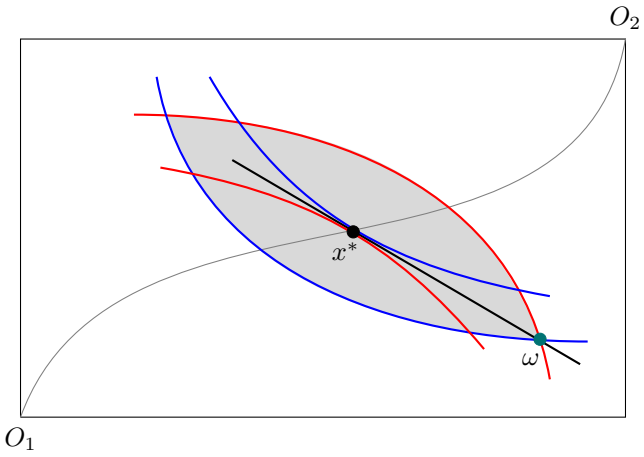


Figure 1.7: A competitive equilibrium.

result (without proof) that describes the welfare properties of a Walrasian allocation.

Theorem 1.2: First theorem of welfare economics

Assume that all traders have monotone utility functions, i.e., $u^i(x_i) > u^i(y_i)$ for $x_i > y_i$ for all $i \in N$. Let (x^*, p) be a Walrasian equilibrium. Then, x^* is a core allocation (and is, therefore, Pareto efficient as well).

Exercise 1.4: Proof of Theorem 1.2

Prove Theorem 1.2. Is the converse true? Give a graphical argument.

1.2 Mechanism design

1.2.1 Implementation of Walrasian allocations

Let us start with the following example of an Edgeworth box economy in order to understand the kind of problem we would analyze in this section.

Example 1.3: Manipulability of Walrasian allocation

Consider the following Edgeworth box economy wherein the endowments are given by $(\omega_{11}, \omega_{12}) = (3, 9)$ and $(\omega_{21}, \omega_{22}) = (9, 3)$. Trader 1's preferences are given by $u(x_{11}, x_{12}) = x_{11}x_{12}$. However, trader 2 has one of the two alternative preferences, $v(x_{21}, x_{22}) = x_{21}+x_{22}$ and $v'(x_{21}, x_{22}) = \sqrt{2}x_{21}+\frac{1}{\sqrt{2}}x_{22}$. The government, who wants to implement the Walrasian allocation (which is Pareto efficient) at prices $(p, 1)$, does not know the true preferences of trader 2. Let us first compute the Walrasian equilibrium under both circumstances.
First, consider the preference profile given by $u(x_{11}, x_{12})$ and $v(x_{21}, x_{22})$. Let (x^*, p^*) be the Walrasian equilibrium which is given by the following set of equations:

$$\begin{aligned}
\text{MRS}^1_{12} &= -\frac{x_{12}}{x_{11}} = -p,\\
px_{11} + x_{12} &= 3p + 9,\\
\text{MRS}^2_{12} &= -1 = -p,\\
px_{21} + x_{22} &= 9p + 3,\\
x_{11} + x_{21} &= 12,\\
x_{12} + x_{22} &= 12.
\end{aligned}$$

The above system yields $(x^*_{11}, x^*_{12}) = (6, 6)$, $(x^*_{21}, x^*_{22}) = (6, 6)$ and $p^* = 1$.

Next, consider the preference profile given by $u(x_{11}, x_{12})$ and $v'(x_{21}, x_{22})$. Let (x', p') be the Walrasian equilibrium which is given by the following set of equations:

$$\begin{aligned}
\text{MRS}^1_{12} &= -\frac{x'_{12}}{x'_{11}} = -p',\\
px'_{11} + x'_{12} &= 3p' + 9,\\
\text{MRS}^2_{12} &= -2 = -p',\\
px'_{21} + x'_{22} &= 9p' + 3,\\
x'_{11} + x'_{21} &= 12,\\
x'_{12} + x'_{22} &= 12.
\end{aligned}$$

The above system yields $(x'_{11}, x'_{12}) = (3.75, 7.5)$, $(x'_{21}, x'_{22}) = (8.25, 4.5)$ and $p' = 2$.

Now suppose that trader 2's true preferences is given by $v(x_{21}, x_{22})$ which is private information. Suppose further that the government asks the traders to report their preferences in an effort to implement the Walrasian allocation $(6, 6), (6, 6))$. Note that if trader 2 falsifies their preferences by reporting $v'(\cdot, \cdot)$, then the government would implement $x' = ((3.75, 7.5), (8.25, 4.5))$. In fact, this trader has incentives to do so because $v(8.25, 4.5) = 12.75 > 12 = v(6, 6)$.

The notion of mis-reporting must be understood carefully. When trader 2 misreport, they do not compare $v'(8.25, 4.5)$ with $v(6, 6)$, but compares $v(8.25, 4.5)$ with $v(6, 6)$. The idea is that a trader mis-report their preferences in an effort to obtain an allocation which is different from the one that would have been implemented if they would have reported truthfully. Then, the trader evaluates their true utility function at the allocation obtained by mis-reporting. If this utility is strictly higher than that at the allocation that would have been implemented under truthful reporting, we say that the trader has incentive to falsify their true preferences.

In Example 1.3, the reason behind the failure to induce trader 1 to report truthfully their preferences is described in the left panel of Figure 1.8. The bundles in the red-shaded region in the left panel were less

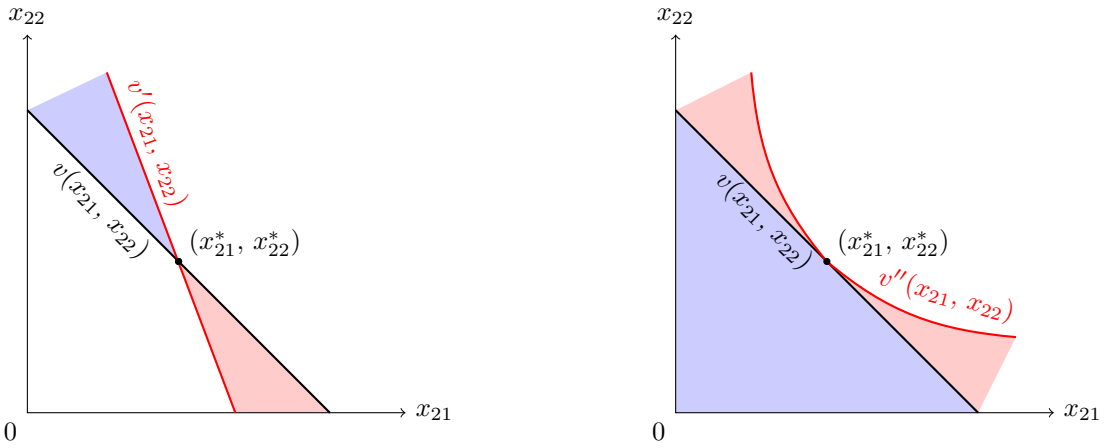


Figure 1.8: Sufficient condition for truthful implementability.

preferred to the Walrasian allocation of trader 2, (x_{21}^*, x_{22}^*) under preferences $v(\cdot, \cdot)$. However, when their preferences have changed to $v'(\cdot, \cdot)$, those bundles became more preferred to (x_{21}^*, x_{22}^*), which induces trader 2 to mis-report their preferences. So, for truthful implementability of the competitive allocation, we require a sort of preference reversal, i.e., for trader 2, (x_{21}^*, x_{22}^*) must be weakly preferred to (x_{21}', x_{22}') under the utility function $v(\cdot, \cdot)$; but (x_{21}', x_{22}') must be weakly preferred to (x_{21}^*, x_{22}^*) under the utility function $v'(\cdot, \cdot)$. This is to say that (x_{21}', x_{22}') must lie in the blue-shaded region in the left panel of Figure 1.8

Suppose that the alternative preferences of trader 2 is given by $v''(x_{21}, x_{22}) = x_{21}x_{22}$ instead of $v'(x_{21}, x_{22})$. In this case we have $p'' = p^*$ and $x'' = x^*$. Under preferences $v''(\cdot, \cdot)$, the set of bundles that are less preferred to (x_{21}^*, x_{22}^*), i.e., the lower contour set of (x_{21}^*, x_{22}^*), expands (the red-shaded region in the right panel of Figure 1.8). In other words, if the bundles that were inferior to (x_{21}^*, x_{22}^*) under preferences $v(\cdot, \cdot)$ are still inferior for trader 2 under the changed preferences $v''(\cdot, \cdot)$, then the same allocation x^* must be chosen under the new preferences. This monotone way of changing preferences is called the Maskin-monotonicity which guarantees truthful implementation of the competitive allocation in the Edgeworth box.

1.2.2 Allocation of objects among several buyers

Consider an economy with one seller (agent S) of single object, and two potential buyers (agents 1 and 2). An allocation of the economy is to allocate the object to a single buyer (think of the sale of a single unit of an indivisible private good). An allocation of the economy is denoted by a. The set of possible allocations is given by:

$$A = \left\{ (x_S, x_1, x_2, t_S, t_1, t_2) \in \{0, 1\}^3 \times \mathbb{R}^3 \mid \sum_i x_i = 1 \text{ and } \sum_i t_i \leq 0 \right\}.$$

In words, in an allocation, $x_i = 1$ if agent $i = S, 1, 2$ gets the object, and $x_i = 0$ if they do not get the object. We allow for monetary transfers in that t_i denotes the transfer received by agent $i = S, 1, 2$. Note that some t_i can be negative. The feasibility restrictions we impose that the aggregate allocation is 1 (as there is a single unit to allocate), and the aggregate transfer must be non-positive. The agents will be characterized by their valuation for the good. In particular, the seller has a valuation θ_S and the buyers have valuations θ_1 and θ_2. The utility function of an agent with valuation or "type" θ_i given by:

$$u_i(a, \theta_i) = \theta_i x_i + t_i.$$

We shall assume that the seller's valuation θ_S is common knowledge, and thus normalize it to 0; however, θ_1 and θ_2 are private information. We assume that both θ_1 and θ_2 are drawn from uniform distribution on $[0, 1]$

Definition 1.7: Ex-post efficiency

An allocation $a \in A$ is ex-post efficient if it allocates the object to the highest-valuation buyer and if it involves no waste of money, i.e., for all $\theta = (\theta_1, \theta_2)$,

$$x_i(\theta)(\theta_i - \max\{\theta_1, \theta_2\}) = 0 \quad \text{for all } i, \quad \text{and} \quad \sum_i t_i = 0.$$

The first part of the above definition reads as (a) if $\max\{\theta_1, \theta_2\} = \theta_1$, then $(x_1, x_2) = (1, 0)$, and hence, $x_1(\theta_1 - \theta_1) = 0$ and $x_2(\theta_2 - \theta_1) = 0$, and (b) if $\max\{\theta_1, \theta_2\} = \theta_2$, then $(x_1, x_2) = (0, 1)$, and hence, $x_1(\theta_1 - \theta_2) = 0$ and $x_2(\theta_2 - \theta_2) = 0$.

Example 1.4: Winner pays the highest valuation

Suppose we want to implement the following allocation $a(\theta)$: For $i, j = 1, 2$, and $i \neq j$,

$$x_i(\theta) = \begin{cases} 1 & \text{if } \theta_i \geq \theta_j, \\ 0 & \text{if } \theta_i < \theta_j. \end{cases} \quad \text{and} \quad t_i(\theta) = -\theta_i x_i(\theta);$$

$$x_S(\theta) = 0, \quad \text{and} \quad t_S(\theta) = -(t_1(\theta) + t_2(\theta)).$$

The above allocation rule allocates the object to the highest-valuation buyer, and the seller receives a transfer equal to the highest valuation. The allocation rule is not only ex-post efficient, but also it is very attractive for the seller in that the seller extracts the entire consumer surplus generated by the trade if the allocation is implemented.

Suppose the buyers are expected utility maximizers. The question is: if buyer 2 always announces their true valuation, will it be optimal for buyer 1 to do the same? For each θ_1, buyer 1's problem is to announce a type $\tilde{\theta}_1$ so as to solve

$$\max_{\tilde{\theta}_1} \mathbb{E}[\theta_1 x_1(\tilde{\theta}_1, \theta_2) + t_1(\tilde{\theta}_1, \theta_2)] = \text{Prob.}(\theta_2 \leq \tilde{\theta}_1)[\theta_1 \cdot 1 - \tilde{\theta}_1 \cdot 1] + \text{Prob.}(\theta_2 > \tilde{\theta}_1)[\theta_1 \cdot 0 - 0) = (\theta_1 - \tilde{\theta}_1)\tilde{\theta}_1.$$

The above maximization yields the optimum announcement $\tilde{\theta}_1 = \theta_1/2$. Likewise, for buyer 2 we have $\tilde{\theta}_2 = \theta_2/2$. The buyers have incentives to under-report their valuation in order to lower the transfer they must make to the seller in case they are the winner. Of course, this increases the probability of not obtaining the object. However, each buyer would exploit this trade-off, at least to some extent.

So, does there exist some other allocation rule that can be truthfully implemented? The answer is yes.

Example 1.5: Winner pays the second-highest valuation

Suppose we want to implement the following allocation $\hat{a}(\theta)$: For $i, j = 1, 2$, and $i \neq j$,

$$x_i(\theta) = \begin{cases} 1 & \text{if } \theta_i \geq \theta_j, \\ 0 & \text{if } \theta_i < \theta_j. \end{cases} \quad \text{and} \quad t_i(\theta) = -\theta_j x_i(\theta);$$

$$x_S(\theta) = 0, \quad \text{and} \quad t_S(\theta) = -(t_1(\theta) + t_2(\theta)).$$

The above rule implies that if buyer i wins the object (i.e., i is the highest-valuation buyer), they pay the lower valuation.

Now, let us analyze the incentive of buyer 1 to reveal truthfully their valuation when buyer 2 announces $\tilde{\theta}_2$ (the case of buyer 2 is symmetric). First, consider the case when $\tilde{\theta}_2 \leq \theta_1$. By announcing

$\tilde{\theta}_1 = \theta_1$ (the true valuation), buyer 1 obtains the object, and their utility is given by $\theta_1 - \tilde{\theta}_2 \geq 0$. If they announce $\tilde{\theta}_1 \neq \theta_1$, they obtain the object as long as $\tilde{\theta}_1 \geq \tilde{\theta}_2$. In this case, buyer 1's utility is $\theta_1 - \tilde{\theta}_2$ which is the same as that under truthful revelation. On the other hand, if $\tilde{\theta}_1 < \tilde{\theta}_2$, buyer 1 does not obtain the object, and consumes 0 utility. So, for $\tilde{\theta}_2 \leq \theta_1$, buyer 1 does not have incentive to falsify their valuation. Next, suppose that $\tilde{\theta}_2 > \theta_1$. In this case, buyer 1 obtains 0 (buyer 2 gets the object) whether or not they report truthfully as long as $\tilde{\theta}_1 < \tilde{\theta}_2$. By contrast, if buyer 1 announces $\tilde{\theta}_1 > \tilde{\theta}_2$, they obtain the object; however, buyer 1 gets a utility equal to $\theta_1 - \tilde{\theta}_2 < 0$. Therefore, in this case too, buyer 1 does not have incentive to falsify their type. So, the optimal announcement of buyer 1 is $\tilde{\theta}_1 = \theta_1$ regardless of what the other buyer announces. Formally, telling truth is a weakly dominant strategy for both buyers. Thus, the allocation rule can be implemented even if buyers' valuations are private information—it suffices to ask each buyer to announce their type, and then choose $\hat{a}(\theta)$.

A mechanism wherein each buyer is asked to report their type is called a direct mechanism. What can we say about the implementability of an allocation rule when agents are asked to announce a function of their type (an indirect mechanism)?

Example 1.6: First-price sealed-bid auction

In a first-price sealed-bid auction, each buyer (bidder) i submits a sealed-bid $b_i \geq 0$. The bids are then opened, and the buyer with the highest bid gets the object and pays their bid to the seller (auctioneer). Because buyer valuations are private information, our solution concept would be the Bayesian Nash equilibrium (BNE). In a Bayesian game, the strategy (bid) of each buyer i is a function of their type θ_i. For the interest of simplicity, suppose buyers follow a linear strategy, $b_i(\theta_i) = \beta_i \theta_i$ with $\beta_i \in [0, 1]$. Consider bidder 1's problem who solves

$$\max_{0 \leq b_1 \leq \beta_2} (\theta_1 - b_1) \cdot \text{Prob.}(b_2(\theta_2) \leq b_1) = (\theta_1 - b_1) \cdot (b_1/\beta_2).$$

The upper limit β_2 on b_1 is because β_2 is buyer 2's maximum bid (when $\theta_2 = 1$), and hence, buyer 1 should never bid more than β_2. Buyer 2 solves a similar problem. So, the optimal bidding functions are given by:

$$b_1(\theta_1) = \min\left\{\frac{1}{2}\theta_1,\, \beta_2\right\} \quad \text{and} \quad b_2(\theta_2) = \min\left\{\frac{1}{2}\theta_2,\, \beta_1\right\}.$$

If $\beta_1 = \beta_2 = \frac{1}{2}$, we have $\min\{\theta_i/2,\, \beta_j\} = \theta_i/2$. In this case, the first-price auction yields the same outcome as Example 1.4.

Example 1.7: Second-price sealed-bid auction

In a second-price sealed-bid auction, each buyer (bidder) i submits a sealed-bid $b_i \geq 0$. The bids are then opened, and the buyer with the highest bid gets the object, but pays the second-highest bid to the seller (auctioneer). Let us solve the equilibrium for more than 2 buyers as the argument is very similar to Example 1.5. Let $N = \{1, \ldots, n\}$ be the set of $n \geq 2$ potential buyers of a single indivisible object. Buyer i's payoff is given by:

$$u_i(\theta_i) = \begin{cases} \theta_i - \max_{j \neq i} b_j & \text{if } b_i > \max_{j \neq i} b_j, \\ 0 & \text{if } b_i < \max_{j \neq i} b_j. \end{cases}$$

We also assume that if there is a tie, i.e., $b_i = \max_{j \neq i} b_j$, the object goes to each winning bidder with equal probability. We now show that $b_i = \theta_i$ for all $i \in N$ is a dominant strategy equilibrium. Consider, say buyer 1, and let $\hat{b}_1 \equiv \max_{j \neq 1} b_j$ be the highest competing bid.

By bidding θ_1, buyer 1 will win if $\theta_1 > \hat{b}_1$, and not if $\theta_1 < \hat{b}_1$ (if $\theta_1 = \hat{b}_1$, buyer 1 is indifferent between winning and losing). Suppose, however, that buyer 1 bids $b_1 < \theta_1$. If $\theta_1 > b_1 \geq \hat{b}_1$, buyer 1 still wins and their utility is still $\theta_1 - \hat{b}_1$. If $\hat{b}_1 > \theta_1 > b_1$, buyer 1 still loses. However, if $\theta_1 > \hat{b}_1 > b_1$, then buyer 1 loses whereas if they had bid $b_1 = \theta_1$, they would have consumed a positive utility. Thus, bidding less than θ_1 can never increase buyer 1's utility but in some circumstances may actually decrease it. A similar argument shows that it is not profitable to bid more than θ_1.

It is not difficult to see that, when there are two buyers, i.e., $n = 2$, the second-price sealed-bid auction yields the same outcome as Example 1.5.

Part II

Two-sided matching without transfers

Chapter 2

The marriage market

2.1 Preferences

This is the classic model of Gale and Shapley (1962) wherein they consider matching of men to women, and hence, the name "marriage market". I shall consider matching between firms and workers instead. Think of small firms hiring workers. We will use feminine pronouns for firms, and masculine pronouns for workers. Make the following assumptions so that the matching is one-to-one:

1. Each firm can hire at most one worker. Think of firms as small shops run by owners. Each shop has a capacity limit which we normalize to 1.
2. No worker can moonlight in other firms.
3. All firms pay the same salary which is exogenously given. A firm-worker pair does not bargain over the salary.

Let $F = \{f_1, \ldots, f_m\}$ and $W = \{w_1, \ldots, w_n\}$ be the sets of firms and workers, respectively. A generic firm will be denoted by $f \in F$, and a generic worker will be denoted by $w \in W$. The preference of any individual, f or w is simply a list of names of individuals on the other side of the market. In general, the preference relation of any individual is assumed to be rational (i.e., reflexive, antisymmetric and transitive) so that she/he cab compare any two individuals on the other side of the market. However, for the marriage market, we consider only strict preferences when an individual compares any two distinct individuals on the other side. Consider a market where Ana, Carolina and Diana are owners of three firms so that we write $F = \{\text{Ana, Carolina, Diana}\}$ to identify them. On the other hand, Alonso, Jorge and Victor are three workers, i.e., $W = \{\text{Alonso, Jorge, Victor}\}$. We would represent the preference relation of Ana over the three workers and that of Jorge over the three owners as follows.

P_{Ana} = Jorge, Alonso, Ana, Victor;
$\mathsf{P}_{\text{Jorge}}$ = Ana, Diana, Carolina, Jorge.

The above representation means the following:

1. For Ana, Jorge and Alonso are the acceptable workers. By contrast, Victor is unacceptable for Ana—she prefers not to hire Victor. In other words, she prefers to remain unmatched or to be matched to herself rather than hiring Victor. Among the acceptable workers, Ana strictly prefers Jorge to Alonso.
2. For Jorge, working for any firm is acceptable, i.e., he prefers to work in one of the three firms rather than

being unemployed or unmatched. Moreover, Jorge's first preference is Ana, second preference is Diana, and Carolina is his least preferred option.

3. We can equivalently represent the preferences of Ana and Jorge respectively as

   Jorge $\succ_{\text{Ana}}$ Alonso $\succ_{\text{Ana}}$ Ana $\succ_{\text{Ana}}$ Victor;
   Ana $\succ_{\text{Jorge}}$ Diana $\succ_{\text{Jorge}}$ Carolina $\succ_{\text{Jorge}}$ Jorge.

Formally, the preferences of a firm $f \in F$ are represented by a strict ordering over the set $W \cup \{f\}$, which is the set of all workers plus herself. Likewise, the preferences of a worker $w \in W$ are represented by a strict ordering over the set $F \cup \{w\}$, which is the set of all firms (or owners) plus himself. We take the preference relations as the primitives for the marriage market.

2.2 One-to-one matching

A matching is a mapping from the set of all individuals to itself. In other words, to each individual, a matching assigns her/his partner. The matching function will be denoted by μ. So, μ(Ana)=Alonso means Alonso is matched with Ana. It is also the case that μ(Alonso)=Ana. On the other hand, μ(Victor)=Victor means Victor is unmatched. Formally,

Definition 2.1

A one-to-one matching is a mapping or function $\mu : F \cup W \to F \cup W$ such that (a) for each firm $f \in F$, $\mu(f) \in W \cup \{f\}$, and for each worker $w \in W$, $\mu(w) \in F \cup \{w\}$; and (b) $\mu(f) = w$ if and only if $\mu(w) = f$.

The first part of the above definition leaves the possibility that any individual can remain unmatched. The second part, on the other hand, asserts that the matching is one-to-one (it is a bit strange to denote a function and its inverse by the same Greek letter!).

We would also like to compare several matchings by defining preferences over matching outcomes. Consider two outcomes μ and μ'. If μ(Jorge)=Ana and μ'(Jorge)=Ana, then Jorge is indifferent between the two matchings. On the other hand, if μ(Jorge)=Carolina and μ'(Jorge)=Diana, then Jorge prefers matching μ' to matching μ because Diana $\succ_{\text{Jorge}}$ Carolina. The crucial assumption here is that, in any matching, Jorge only cares about his partner, and does not care about who else is matched with whom under μ and μ'. If that were the case, we would say that the marriage market is subject to externality, and solving for the market equilibrium (which we have not defined yet) would have much more complicated.

2.3 Stability of the marriage market

The matching function μ defines an outcome of the marriage market. In a very general setting, there may be many outcomes in all of which we may not be interested. We would only be interested in reasonable outcomes. One uninteresting outcome is that nobody is matched with nobody, which is called the empty matching. Another example is μ(Ana)=Victor, i.e., Ana is matched with an unacceptable worker.

Example 2.1

Consider two firm owners Ana and Carolina, and two workers Alonso and Jorge. The preferences are as follows.

P_{Ana} = Jorge, Alonso, Ana ;
P_{Carolina} = Alonso, Jorge, Carolina ;
P_{Alonso} = Ana, Alonso, Carolina ;
P_{Jorge} = Ana, Carolina, Jorge.

Let μ(Alonso)=Carolina and μ(Jorge)=Ana. In this outcome, both Ana and Jorge are happy because each of them is matched with their most preferred partner. Carolina is also happy because she is matched with Alonso, her most preferred partner. However, Alonso is matched with his unacceptable partner, Carolina. Can he be better-off by proposing a new matching? The answer is yes because in a different matching μ' if μ'(Alonso)=Alonso, then he is better-off by remaining unmatched. We would say that matching μ is not individually rational for Alonso.

Example 2.2

Consider two firm owners Ana and Carolina, and two workers Alonso and Jorge. The preferences are as follows.

P_{Ana} = Alonso, Jorge, Ana ;
P_{Carolina} = Jorge, Alonso, Carolina ;
P_{Alonso} = Ana, Carolina, Alonso ;
P_{Jorge} = Ana, Carolina, Jorge.

Let μ(Alonso)=Carolina and μ(Jorge)=Ana. Now consider Ana and Alonso. Under μ, none of them is matched to their most preferred partner, which they would have liked. Formally, Alonso $\succ_{\text{Ana}}$ μ(Ana) and Ana $\succ_{\text{Alonso}}$ μ(Alonso). Thus, if under a different matching μ' we have μ'(Alonso)=Ana, then both of them are strictly better-off. We would say that (Ana, Alonso) pair would block matching μ.

Let us now introduce the concept of stability of the marriage market.

Definition 2.2

A matching μ is stable if (a) it is individually rational, i.e., for each individual $i \in F \cup W$, we have $\mu(i) \succsim_i i$, i.e., individual i weakly prefers to be matched, and (b) there are no blocking pairs, i.e., there is no firm-worker pair (f, w) such that $\mu(f) \neq w$, and $w \succ_f \mu(f)$ and $f \succ_w \mu(w)$.

Note that in defining preferences over individuals, we have only considered strict orderings. However, weak orderings are permissible when individuals compare different matchings. In part (a) of the above definition, we have thus written $\mu(i) \succsim_i i$ because i weakly prefers μ to a different matching μ' wherein $\mu'(i) = i$. Moreover, when an individual market participant blocks a matching (i.e., the matching is not individually rational) or a pair blocks a matching, they do not require to take into account what is happening to the rest of the market. In other words, an individual or a pair (not matched previously) do not require a stable matching to block the initial one.

2.4 The Deferred Acceptance Algorithm: Finding a stable matching

Now, we ask the most relevant question whether a stable matching exists for our marriage market. Because the preferences are ordinal (i.e., no utility function is attached even if preferences are rational), we cannot use the standard equilibrium condition of a market for goods with money, i.e.,

$$\text{marginal rate of substitution between goods } i \text{ and } j = \frac{\text{price of good } i}{\text{price of good } j}.$$

We would instead use algorithms that would lead to a stable allocation. Consider the following example.

Example 2.3

Consider the following preference profiles.

P_{Ana} = Alonso, Victor, Jorge, Ana ;
$P_{Carolina}$ = Victor, Jorge, Alonso, Carolina ;
P_{Diana} = Alonso, Victor, Jorge, Diana ;
P_{Alonso} = Carolina, Ana, Diana, Alonso ;
P_{Jorge} = Ana, Carolina, Diana, Jorge ;
P_{Victor} = Ana, Carolina, Diana, Victor.

Consider now the following worker proposing algorithm. At each stage, each worker proposes to his most preferred firm. Each firm decides whether to accept or reject the offer. A rejected worker proposes in the next stage. The algorithm stops when there are no rejections.

Step 1: Both Jorge and Victor propose to Ana. Ana accepts Victor and rejects Jorge. Alonso proposes to Carolina, which is accepted. Diana receives no proposal. The provisional matching at this step is

μ^1(Alonso)=Carolina, μ^1(Jorge)=Jorge, μ^1(Victor)=Ana and μ^1(Diana)=Diana.

Step 2: Jorge proposes to Carolina because a proposal to Ana would have been rejected as he knows that μ^1(Ana)=Victor $\succ_{Ana}$ Jorge. Carolina rejects Alonso, her provisional partner, and accepts Jorge as Jorge $\succ_{Carolina}$ Victor=μ^1(Carolina). The provisional matching at this step is

μ^2(Alonso)=Alonso, μ^2(Jorge)=Carolina, μ^2(Victor)=Ana and μ^2(Diana)=Diana.

Step 3: Alonso proposes to Ana because Carolina would have rejected him anyway. Ana rejects Victor and accepts Alonso. Diana still does not receive a proposal. The provisional matching at this step is

μ^3(Alonso)=Ana, μ^3(Jorge)=Carolina, μ^3(Victor)=Victor and μ^3(Diana)=Diana.

Step 4: Victor proposes to Carolina because Ana would have rejected his proposal. Carolina rejects Jorge and accepts Victor. Diana is still alone. The provisional matching at this step is

μ^4(Alonso)=Ana, μ^4(Jorge)=Jorge, μ^4(Victor)=Carolina and μ^4(Diana)=Diana.

Step 5: Jorge proposes to Diana because both Ana and Carolina would have rejected his proposal as they have already hired Alonso and Victor who come before Jorge in each firm's preference list. Diana accepts Jorge as Jorge $\succ_{Diana}$ Diana=μ^4(Diana). There is no more rejections, and hence, the algorithm stops here. The final allocation is given by:

μ(Alonso)=Ana, μ(Jorge)=Diana and μ(Victor)=Carolina.

Note that the final outcome is individually rational because no individual has an unacceptable partner. It is also easy to see that there is no blocking pair. Hence, the final matching μ is stable.

The above algorithm is called the Deferred Acceptance algorithm (DA, henceforth) which is due to Gale and Shapley (1962). In this algorithm, individuals on one side of the market propose, and those on the other side accept or reject proposals. Later we shall see that which side proposes matters. In Example 2.3, the stable matching is unique. This may not always be the case. We now formally state the DA.

The Deferred Acceptance algorithm. Consider the workers-proposing version of the algorithm. Initially, all workers are active and no individual is provisionally matched. Proceed in steps as follows.

Step 1: Each worker w proposes to his most preferred firm among his list of acceptable firms (if he has any acceptable choices). Each firm f rejects any unacceptable proposals and, if more than one acceptable proposal is received, holds the most preferred and rejects all others.

$\vdots$

Step k: Any worker rejected at step $k-1$ makes a new proposal to its most preferred acceptable firm who has not yet rejected him. If no acceptable choices remain, he makes no proposal. Each firm holds her most preferred acceptable offer till date, and rejects the rest.

STOP: The algorithm stops when no further proposals are made, and match each firm to the worker (if any) whose proposal she has been holding.

Note: the firms-proposing version of the algorithm is analogous.

We next show that the set of stable matchings is nonempty in every marriage market. A key aspect of the DA algorithm is that both sides of the market go through their ranking lists in opposite directions. In the workers-proposing version, workers propose to firms in the order of their preference rankings from top to bottom. They start proposing to their most preferred firm, and continue to propose to firms in the order of their ranking as long as their proposals are not accepted. By contrast, the provisional matches of firms go from bottom from to top: every time a firm accepts a proposal, it is from a worker who is better than her previous match. In the firms-proposing version of the algorithm, the opposite obtains: firms go from top to bottom, and workers go from bottom to top.

Theorem 2.1: (Gale and Shapley, 1962)

The outcome of the Deferred acceptance algorithm is a stable matching.

Proof. Let μ be the outcome of the workers-proposing DA algorithm. Workers only propose to acceptable firms, and firms only accept offers from acceptable workers. Therefore, μ is individually rational. Let $w \in W$ and $f \in F$ be such that $f \succ_f \mu(w)$. Then, w had proposed to f in some iteration of the algorithm. Because $\mu(w) \neq f$, f accepted the proposal of some worker w' with $w' \succ_f w$. Then, $\mu(f) \succeq_f w' \succ_f w$. Hence, $(w,\ f)$ is not a blocking pair. Therefore, μ is stable. $\square$

Opposing interests in the marriage market. In Example 2.3, it is easy to show that the both the worker-proposing and firms-proposing DA algorithms produce the same stable matching. Thus, the stable matching is unique. However, this is always not the case. Consider the following example.

Example 2.4

Consider the following preference profiles.

P_{Ana} = Alonso, Jorge, Victor, Ana ;
P_{Carolina} = Alonso, Victor, Jorge, Carolina ;
P_{Diana} = Alonso, Jorge, Vivtor, Diana ;
P_{Alonso} = Ana, Carolina, Diana, Alonso ;
P_{Jorge} = Ana, Carolina, Diana, Jorge ;
P_{Victor} = Ana, Diana, Carolina, Victor.

First consider the workers-proposing version of the DA algorithm. In Step 1, Ana receives proposals from all the three workers out of which she keeps Alonso. Carolina, Diana, Jorge and Victor are all unmatched at the end of this step. In Step 2, Carolina receives proposal from Jorge and Diana, from Victor. They accept the respective offers because both workers are acceptable for each of the firm. The algorithm stops here, and it yields

$$\mu_W(\text{Alonso})=\text{Ana},\ \mu_W(\text{Jorge})=\text{Carolina and } \mu_W(\text{Victor})=\text{Diana}\,.$$

Now consider the firms-proposing version. In Step 1, Alonso receives proposals from all the three firms out of which he agrees to work for Ana. Carolina, Diana, Jorge and Victor are all unmatched at the end of this step. In Step 2, Jorge receives offer from Diana and Victor, from Carolina. They accept the respective offers because both firms are acceptable for each of the worker. The algorithm stops here, and it yields

$$\mu_F(\text{Alonso})=\text{Ana},\ \mu_F(\text{Jorge})=\text{Diana and } \mu_F(\text{Victor})=\text{Carolina}\,.$$

So, there are two stable matchings μ_W and μ_F with $\mu_W \neq \mu_F$ depending on which side proposes in the algorithm.

Note that Ana and Alonso are indifferent between μ_W and μ_F because they are matched to each other under both matchings. However, $\mu_W(\text{Jorge}) \succ_{\text{Jorge}} \mu_F(\text{Jorge})$ and $\mu_W(\text{Victor}) \succ_{\text{Victor}} \mu_F(\text{Victor})$. So, μ_W is weakly preferred to μ_F by all workers. By contrast, $\mu_F(\text{Carolina}) \succ_{\text{Carolina}} \mu_W(\text{Carolina})$ and $\mu_F(\text{Diana}) \succ_{\text{Diana}} \mu_W(\text{Diana})$. Therefore, μ_F is weakly preferred to μ_W by all firms.

It follows from Theorem 2.1 and Example 2.4 together that the DA algorithm produces two stable outcomes, μ_W and μ_F. But there may be more stable matchings than μ_W and μ_F which the algorithm does not yield. The following is an interesting result regarding the optimality of stable matchings.

Proposition 2.1: (Gale and Shapley, 1962)

Let $\mathcal{M}^S$ be the set of stable matchings of the marriage market. Further, let the workers-proposing version of the Deferred Acceptance algorithm yields μ_W, while the firms-proposing version yields μ_F. Then, for every matching $\mu \in \mathcal{M}^S$, we have (a) $\mu_W \succeq_w \mu \succeq_w \mu_F$ for all $w \in W$, and (b) $\mu_F \succeq_f \mu \succeq_f \mu_W$ for all $f \in F$.

The above proposition simply asserts that, among all the stable matchings, μ_W is the best outcome and μ_F is the worst outcome for all the workers. On the other hand, among all the stable matchings, μ_F is the best outcome and μ_W is the worst outcome for all the firms. We would call μ_W the worker-optimal stable matching and μ_F the firm-optimal stable matching. For a formal proof of Proposition 2.1, see Robinson-Cortés (2021, Theorem 5.2.1). Proposition 2.1 follows from a more general result:

Theorem 2.2: (Knuth, 1976)

If μ and μ' are stable matchings, then $\mu \succ_w \mu'$ for all $w \in W$ if and only if $\mu' \succ_f \mu$ for all $f \in F$.

Proof. Let μ and μ' be stable matchings such that $\mu \succ_w \mu'$ for all $w \in W$. Towards a contradiction, suppose it is not true that $\mu' \succ_f \mu$ for all $f \in F$. Then, there exists some $f \in F$ with $w = \mu(f)$ such that

$$w = \mu(f) \succ_f \mu'(f). \tag{2.1}$$

On the other hand, because $\mu \succ_w \mu'$ for all $w \in W$, it must be the case that

$$f = \mu(w) \succ_w \mu'(w). \tag{2.2}$$

Thus, it follows from (2.1) and (2.2) that $w \succ_f \mu'(f)$ and $f \succ_w \mu'(w)$, meaning that $(w,\ f)$ blocks μ', which is a contradiction to the fact that μ' is a stable matching. □

We now prove two important results in matching theory.

Proposition 2.2: Decomposition lemma

Let μ and μ' be stable matchings under the same (strict) preference profile. Let $W_{\mu'} \subset W$ be the set of workers who prefers μ' to μ, and $F_\mu \subset F$ be the set of firms who prefer μ to μ'. Then, each worker $w \in W_{\mu'}$ is matched, under both μ and μ', to a firm in F_μ (but not necessarily to the same firm). Likewise, each firm $f \in F_\mu$ is matched, under both μ and μ', to a worker in $W_{\mu'}$.

Proof. For any $w \in W_{\mu'}$, we have $f = \mu'(w) \succ_w \mu(w) \succeq_w w$. If the second relation does not hold, μ would not be individually rational for w. Also, $\mu'(w) \in F$ because $\mu'(w) \succ_w w$, and hence, there is a firm $f \in F$ such that $f = \mu'(w)$. Then for f, it must be that $\mu(f) \succ_f \mu'(f) = w$, which follows from Theorem 2.2. Thus, for any $w \in W_{\mu'}$, we have $\mu'(w) \in F_\mu$, i.e., μ' maps $W_{\mu'}$ to F_μ. Moreover, $f \in \mu'(W_{\mu'})$ means that $\mu'(f) \in W_{\mu'}$, and we just have shown that, in this case, $f = \mu'(w) \in F_\mu$. Therefore, $\mu'(W_{\mu'}) \subseteq F_\mu$, and hence, $\left|W_{\mu'}\right| \leq |F_\mu|$. Now repeat the same argument for any $f \in F_\mu$ to show that $w = \mu(f) \in W_{\mu'}$, i.e., μ maps F_μ to $W_{\mu'}$, and that $|F_\mu| \leq \left|W_{\mu'}\right|$. Therefore, the sets $W_{\mu'}$ and F_μ have the same cardinality, i.e., there are as many workers in $W_{\mu'}$ as there are firms in F_μ. Because both μ and μ' are one-to-one functions, and $W_{\mu'}$ and F_μ are finite with the same cardinality, the result follows. □

The Decomposition lemma, which is obtained as a corollary to Theorem 2.2, asserts that an individual who prefers one stable matching to another is matched at both to a partner with the reverse preferences. We next prove the following result.

Theorem 2.3: Rural hospital theorem

The set of unmatched individuals is the same under every stable matching.

Proof. Take two stable matchings μ and μ'. Suppose that an arbitrary worker w is unmatched under μ, but matched under μ', i.e., $\mu(w) = w$ and $\mu'(w) \in F$. Because, μ' is stable, it is individually rational for w, i.e., $\mu'(w) \succ_w w = \mu(w)$. Thus, $w \in W_{\mu'}$. Because the Decomposition lemma implies μ maps F_μ onto $W_{\mu'}$, we have $\mu(w) \in F_\mu$, i.e., w is matched also under μ, which is a contradiction to the initial supposition that w is unmatched under μ. □

The name "rural hospital theorem" is associated with the model of the medical match which we shall analyze in Chapter 3. It corresponds to the fact that hospitals in remote areas find it difficult to recruit medical interns. The above theorem implies that in no stable matchings this situation may change.

2.5 Incentives in the marriage market

Now, think of the following procedure, which we call the matching mechanism. There is a clearing house (e.g. a giant computer). The preferences of all firms and workers are fed into the computer which runs the DA algorithm, and produces a matching for the marriage market. Clearly, given the preference profiles of firms and workers, if the computer runs the workers-proposing version, it will produce μ_W. On the other hand, if it runs the firms-proposing version, it yields μ_F. The question we ask is:

Would all market participants state their preferences truthfully?

The above question is related to the fact that individuals may engage in strategic behavior.

Definition 2.3

A matching mechanism is strategy-proof if for all individuals stating their true preferences is a dominant strategy.

To understand whether any market participant has incentives for mis-reporting their preferences, consider the following example.

Example 2.5

Consider the following preference profiles.

$\mathsf{P}_{\text{Ana}} =$ Jorge, Alonso, Ana ;
$\mathsf{P}_{\text{Carolina}} =$ Alonso, Jorge, Carolina ;
$\mathsf{P}_{\text{Alonso}} =$ Ana, Carolina, Alonso ;
$\mathsf{P}_{\text{Jorge}} =$ Carolina, Ana, Jorge.
If we run the workers-proposing version of the DA algorithm, then it yields the workers-optimal stable matching in that μ_W(Ana)=Alonso and μ_W(Carolina)=Jorge. Now, consider an alternative preference profile wherein Ana states
$\mathsf{P}'_{\text{Ana}} =$ Jorge, Ana, Alonso,
with others stating preferences P. In Step 1, Alonso proposes to Ana, and Jorge proposes to Carolina. Because Jorge is acceptable for Carolina, she keeps Jorge. However, Alonso is unacceptable for Ana, and hence, she rejects Alonso. This produces a provisional matching: μ^1(Ana)=Ana, μ^1(Alonso)=Alonso and μ^1(Carolina)=Jorge. In Step 2, Alonso proposes to Carolina who is provisionally matched with Jorge. Because Alonso $\succ_{\text{Carolina}}$ Jorge, she accepts Alonso. This step produces a provisional matching: μ^2(Ana)=Ana, μ^2(Carolina)=Alonso and μ^2(Jorge)=Jorge. In Step 3, Jorge proposes to Ana who accepts the proposal. There are no further rejections, and hence, the final matching is given by μ'(Ana)=Jorge and μ'(Carolina)=Alonso.

Note that matching μ' is stable under the stated preference profile (P'_{Ana}, $\mathsf{P}_{\text{Carolina}}$, $\mathsf{P}_{\text{Alonso}}$, $\mathsf{P}_{\text{Jorge}}$) as the DA algorithm produces a stable matching [cf. Theorem 2.1]. However, we see that

$$\text{Jorge}=\mu'(\text{Ana}) \succ_{\text{Ana}} \text{Alonso}=\mu_W(\text{Ana}),$$

i.e., by falsifying her true preferences, Ana achieves to hire her most-preferred worker. The above example presents the stability-incentive tradeoff in the marriage market: it is not incentive compatible for some individuals to falsify their preferences even if the stated preference profile induces a stable matching.

Theorem 2.4

There is no matching mechanism for the marriage market that satisfies the following two properties simultaneously:

(a) The matching is stable with respect to the reported preference profile.

(b) The mechanism is strategy-proof for all individuals.

An interesting question is whether, in a stable matching mechanism, would an individual on any of the two sides of the marriage market mis-report their preferences. In Example 2.5, it is easy to see that no worker has incentives to falsify his preferences. This is in general true for any matching mechanism that uses the DA algorithm.

Theorem 2.5

A matching mechanism that uses the Deferred Acceptance algorithm is strategy-proof for the proposing side (i.e., it is a dominant strategy to state one's true preferences).

Proof. See Roth and Sotomayor (1990, Chapter 4). □

The proof is a bit involved which uses the so-called Blocking Lemma (see Gale and Sotomayor, 1985). However, from Example 2.5 one can understand the intuition. The workers-proposing version of DA produces a stable matching that is the 'best' for all workers. It is thus somewhat natural to expect that workers would state their true preferences (this argument is not that straight forward as it sounds!). By contrast, if the stable matching is the worst outcome for the firm-side of the market, then it is expected that some firms would be better-off by falsifying preferences. Note also that the only strategic concern associated with each market participant in a matching mechanism is to report a preference list, not to name a partner (the clearing house does the job). It is also worth noting that the above results can be generalized in that there is in fact no stable matching mechanism (not only the mechanism that uses DA!) wherein reporting true preferences is a dominant strategy for all market participants (see Roth, 1982a). However, Kojima and Pathak (2009) show that as the marriage market becomes large (i.e., the number of workers and firms increases), on average, there is a little possibility of manipulation because in a large market, the average number of stable matchings tend to be low.

Chapter 3

The college admissions problem

3.1 Preferences

The many-to-one matching model has often been related to real life matching markets wherein a group of individuals on one side of the market are matched with one individuals on the other side such as matching of medical intern (workers) to hospitals (firms)—the medical match, matching students (workers) to colleges (firms)—hence, the name, the college admissions problem, matching workers to firms—job matching, matching students (workers) to schools (firms)—school choice. We would stick to our firm-worker market in order not to change notations. In some examples, we would switch to the example of a doctor-hospital market (the notations will be introduced opportunely). We continue assuming that there is no possibility of bargaining over salaries between firms and workers, i.e., salaries offered by each firm are fixed. A many-to-one matching market is a collection of the set of firms, set of workers, capacities of firms (soon to be elaborated), and their preferences (over the individuals on the other side of the market). Formally,

$$F := \{f_1, \ldots, f_m\} \text{ is the set of firms;}$$
$$W := \{w_1, \ldots, w_n\} \text{ is the set of workers.}$$

To start with, we define the preferences of any individual on one side of the market over the individuals on the other side in a fashion similar to the marriage market. That is, $\succ_w$ represents the preferences over $F \cup \{w\}$ of worker $w \in W$. Likewise, $\succ_f$ represents the preferences over $W \cup \{f\}$ of firm $f \in F$. These preference profiles captures the idea that there are firms unacceptable for a worker, and workers unacceptable for a firm.

A worker, by assumption, can be employed by only one firm, and hence, there is no loss of generality in representing his preferences by $\succ_w$ or P_w as in the case of one-to-one matching market. However, firms now can hire more than one worker, and hence, their preferences must be modified. We would assume that a given firm $f \in F$ can hire a maximum of q_f workers, which we would call its capacity or quota. In the marriage market, clearly all firms have the same capacity, $q_f = 1$ for all $f \in F$. Thus, the third ingredient we require the many-to-one firm-worker market is the capacities of the firms:

$$q = (q_1, \ldots, q_m) \text{ is the vector of capacities of the firms.}$$

Defining firm preferences over a group of workers is not an easy task. However, we shall see that we can confine interests to a special class of preferences, called responsive preferences, which will make our life easy, and responsive preferences are a very general representations of preferences. Suppose Ana and Carolina both have the same capacities, $q_{\text{Ana}} = q_{\text{Carolina}} = 2$. There are three workers, Alonso, Jorge and Victor. Consider the following preference profiles of firms:

P_{Ana} = Jorge, Alonso, Victor, Ana ;
$\mathsf{P}_{\text{Carolina}}$ = Alonso, Victor, Jorge, Carolina.

Ana must compare between sets workers. Preferences of firm $f \in F$ over sets of workers, as opposed to over individual workers, will be denoted by $\mathsf{P}_f^{\#}$ or $\succ_f^{\#}$. Now, consider that Ana is interested in comparing $\{\text{Alonso, Jorge}\}$ and $\{\text{Alonso, Victor}\}$. If the preferences over groups of workers are responsive, then Ana should prefer $\{\text{Alonso, Jorge}\}$ to $\{\text{Alonso, Victor}\}$ because in both groups, Alonso is the common worker, and Ana strictly prefers Jorge to Victor.

Definition 3.1: Responsive preferences

The strict preference relation (of a firm $f \in F$), $\succ_f^{\#}$ over sets of workers is responsive to the preference relation $\succ_f$ over individual workers if we have

$$W' \succ_f^{\#} W' \cup \{w'\} \setminus \{w\} \quad \text{if and only if} \quad w' \succ_f w.$$

Under responsive preferences, a firm compares two distinct individual workers pertaining to two distinct groups, and does not care about the rest of the group members. The following example reveals that such representation is very general.

Example 3.1

Consider firm f and four workers $\{w_1, w_2, w_3, w_4\}$ with $\mathsf{P}_f = w_1, w_2, w_3, w_4, f$. First, note that all workers are acceptable for the firm. Consider first two sets of workers, $W' = \{w_1, w_3, w_4\}$ and $W'' = \{w_1, w_2, w_4\}$. These two sets differ by one worker only: both sets have w_1 and w_4 as common workers; however, W' has w_3 as opposed to w_2 in W''. If preferences over groups are responsive to preferences over individuals, then it is sufficient to see how firm f ranks w_3 and w_2. Because $w_2 \succ_f w_3$, we must have $W'' \succ_f^{\#} W'$.

Now consider the sets $\{w_1, w_3\}$ and $\{w_2, w_4\}$. Note that the two sets are entirely different from each other. Nonetheless, it is possible compare them. Because $w_1 \succ_f w_2$, it is the case that $\{w_1, w_3\} \succ_f^{\#} \{w_2, w_3\}$. On the other hand, because $w_3 \succ_f w_4$, we have $\{w_2, w_3\} \succ_f^{\#} \{w_2, w_4\}$. Therefore, by transitivity, we have $\{w_1, w_3\} \succ_f^{\#} \{w_2, w_4\}$.

Responsive preferences also allow us to compare sets of unequal sizes. Suppose we intend to compare $\{w_2\}$ with $\{w_1, w_3\}$. Here the capacity of f becomes important. Let $q_f = 2$. We can write $\{w_2\}$ as $\{w_2, \varnothing\}$, i.e., firm f hires only w_2, and she has an unfilled vacancy. Because w_3 is acceptable for f, we have $\{w_2, w_3\} \succ_f^{\#} \{w_2, \varnothing\}$. On the other hand, $w_1 \succ_f w_2$ implies that $\{w_1, w_3\} \succ_f^{\#} \{w_2, w_3\}$. Therefore, by transitivity, we have $\{w_1, w_3\} \succ_f^{\#} \{w_2, \varnothing\} = \{w_2\}$.

Clearly, we cannot compare $\{w_1, w_4\}$ and $\{w_2, w_3\}$. However, this non-comparability does not create much of a problem in analyzing a many-to-one matching market.

3.2 Many-to-one matching

The crucial difference between the marriage market and the many-to-one matching market is that, in the latter, each firm is potentially matched to a set (possibly empty) of workers, i.e., the mapping μ assigns to each firm $f \in F$ a subset $\mu(f) \subseteq W$ of workers. Such mapping is called a correspondence, and not a function (we write $\to\!\!\to$ instead of $\to$). Formally,

Definition 3.2: Many-to-one matching

A many-to-one matching is a correspondence $\mu : F \cup W \to\!\!\to F \cup W$ such that (a) $\mu(f) \subseteq W$ with $|\mu(f)| \leq q_f$ for all $f \in F$; (b) $\mu(w) \in F \cup \{w\}$ for all $w \in W$; and (c) $\mu(w) = f$ if and only if

$\mu(f) \in W$.

Example 3.2

Consider $F = \{\text{Ana, Diana}\}$ with $q_{\text{Ana}} = q_{\text{Diana}} = 2$, and $W = \{\text{Alonso, Alex, Jorge, Victor}\}$. Let $\mu(\text{Ana}) = \{\text{Alonso}\}$ and $\mu(\text{Diana}) = \{\text{Jorge, Victor}\}$. Moreover, $\mu(\text{Alex}) = \text{Alex}$, i.e., Alex is not hired by any firm. Clearly, for each $f \in F$, $\mu(f) \subset W$, and for each $w \in W$, $\mu(w) \in F \cup W$. Ana has an unfilled vacancy because $|\mu(\text{Ana})| < q_{\text{Ana}} = 2$.

3.3 Stability in the college admissions problem

The definition of stable matching is very similar to that of the marriage market; however, we require a minor modification because each firm can now hire more than one workers. Consider the following example.

Example 3.3: Stability

Let $W = \{w_1, w_2, w_3, w_4\}$ and $F = \{f_1, f_2\}$. The capacities are given by $q_1 = 2$ and $q_2 = 1$. The preferences are as follows:

$\mathsf{P}_{w_1} = f_1, f_2, w_1$;
$\mathsf{P}_{w_2} = f_1, f_2, w_2$;
$\mathsf{P}_{w_3} = f_1, f_2, w_3$;
$\mathsf{P}_{w_4} = f_2, f_1, w_4$;
$\mathsf{P}_{f_1} = w_1, w_2, w_3, f_1, w_4$;
$\mathsf{P}_{f_2} = w_1, w_3, w_2, f_2, w_4$.

First, consider the matching

$$\mu(f_1) = \{w_1, w_3\} \quad \text{and} \quad \mu(f_2) = \{w_2, w_4\}.$$

Matching μ is not individually rational for firm f_2 because there is a worker, w_4, unacceptable for this firm, yet he is in her matched set.

Consider the following matching

$$\mu'(f_1) = \{w_1, w_3\} \quad \text{and} \quad \mu'(f_2) = \{w_2\}.$$

The above matching is clearly individually rational for all firms and workers. Moreover, there are no vacancies left in any firm. However, w_2 prefers f_1 to f_2, and f_1 prefers w_2 to w_3. Therefore, (f_1, w_2) is a blocking pair for the matching μ'.

Next, consider the matching

$$\mu''(w_1) = f_1, \quad \mu''(w_2) = f_2 \quad \text{and} \quad \mu''(w_3) = w_3.$$

Matching μ'' is wasteful because worker w_3 who is acceptable for f_1 is unmatched and firm f_1 has an unfilled vacancy, i.e., $|\mu''(f_1)| < q_1$.

First, note the difference between no individual rationality and wastefulness. For a firm f, a matching μ is not individually rational if there is an unacceptable worker is in $\mu(f)$. By contrast, a matching μ is wasteful if a firm f has an unfilled quota, and yet there is a worker unmatched in the market who is acceptable for firm f. Second, note the subtle difference between wastefulness and pairwise blocking. If a matching is wasteful (i.e., f_1 has an unfilled vacancy, and w_3, who is acceptable for f_1, is unmatched),

then (f_1, w_3) blocks the matching, but no at the expense of an existing worker of firm f_1 (as in the case of pairwise blocking where f_1 disposes of w_3 in order to hire w_2).
Finally, consider the following matching

$$\mu^*(f_1) = \{w_1, w_2\} \quad \text{and} \quad \mu^*(f_2) = \{w_3\}.$$

The above matching is individually rational, not blocked by any firm-worker pair, and not wasteful, and hence, the matching is stable.

Definition 3.3: Stable matching

A many-to-one matching μ is

(a) individually rational if each worker $w \in W$ weakly prefers $\mu(w)$ (his mate) to w (himself), i.e., $\mu(w) \succeq_w w$, and for each firm $f \in F$, there is no unacceptable worker in her matched set, i.e., there is no $w \in W$ with $w \in \mu(f)$ and $f \succeq_f w$;

(b) immune to pairwise blocking, i.e., if there is no (f, w) with $w \notin \mu(f)$ and $w' \in \mu(f)$ such that $f \succ_w \mu(w)$ and $w \succ_f w'$; and

(c) not wasteful, i.e., there is a worker w and a firm f such that $w \notin \mu(f)$ and w and f are mutually acceptable for each other such that $f \succ_w \mu(w)$ and $|\mu(f)| < q_f$.

A matching μ is stable if it is individually rational, immune to pairwise blocking and not wasteful.

At this juncture, we must be wondering why the above definition of stability considers only blocking by firm-worker pairs, and not by firm-workers coalitions wherein each coalition consists of one firm and a subset of workers if firms are allowed to hire more than one workers apiece. To be more precise, the second part of Definition 3.3 corresponds to the concept of pairwise stability. When a matching is immune to blocking by firm-workers coalitions, it is said to be group stable. Following is an intuitive but surprising result.

Proposition 3.1: Group versus pairwise stability

Under responsive preferences of the firms, a many-to-one matching is group stable if and only if it is (pairwise) stable.

Proof. Proving necessity is trivial. For sufficiency, see Roth and Sotomayor (1990). □

Definition 3.4: Substitutable preferences

Let W' and W'' be two subsets of W, the set of workers, and denote by $Ch_f(W)$, the choice set of firm f when she faces the set of workers W. Formally, $Ch_f(W) = W'$ if $W' \succeq_f W''$ for any $W'' \subset W$. A firm f's preferences over set of workers are said to be substitutable if, for any set S that contains workers w and w', if $w \in Ch_f(S)$, then $w \in Ch_f(S \setminus \{w'\})$.

In words, if a firm has substitutable preferences, then if its preferred set of workers from S includes w, so will its preferred set of workers from any subset of S that still includes w. Let us denote the preference relation of firms that is substitutable by P_f^s or $\succ_f^s$.

Exercise 3.1

Let $F = \{f_1, f_2\}$ and $W = \{w_1, w_2, w_3\}$, and $q_1 = 2$, $q_2 = 1$. The strict preferences are as follows:

$\mathsf{P}_{w_1} = f_1, f_2, w_1;$
$\mathsf{P}_{w_2} = f_1, f_2, w_2;$
$\mathsf{P}_{w_3} = f_1, f_2, w_3;$
$\mathsf{P}_{f_1} = w_3, w_2, w_1, f_1;$
$\mathsf{P}^s_{f_1} = \{w_1, w_2\}, \{w_1, w_3\}, \{w_2, w_3\}, \{w_3\}, \{w_2\}, \{w_1\}, f_1;$
$\mathsf{P}_{f_2} = w_3, f_2;$
$\mathsf{P}^s_{f_2} = \{w_3\}, f_2.$

Note that the preferences of firms, $(\mathsf{P}^s_{f_1}, \mathsf{P}^s_{f_2})$ are substitutable. However, the preferences are not responsive to preferences over individual workers, $(\mathsf{P}_{f_1}, \mathsf{P}_{f_2})$ because $\{w_1, w_2\} \succ^s_{f_1} \{w_1, w_3\}$ but $w_3 \succ_{f_1} w_2$. Find the group stable matchings with respect to preference profiles $(\mathsf{P}^s_{f_1}, \mathsf{P}^s_{f_2})$ and $(\mathsf{P}_{f_1}, \mathsf{P}_{f_2})$ (in the second case, as if the preferences were responsive). ■

3.4 Finding a stable many-to-one matching

A slightly modifies version of the DA algorithm in Chapter **??** induces a stable matching outcome for the college admissions problem. The modification is necessary because in many-to-one matching firms can hire more than one workers. The workers-proposing version of DA is the same as that of the marriage market; however, the firms-proposing version differs. We shall illustrate the differences in terms of the following example.

Example 3.4: Deferred Acceptance algorithm

Let $W = \{w_1, w_2, w_3\}$ and $F = \{f_1, f_2\}$. The capacities are given by $q_1 = 2$ and $q_2 = 1$. The preferences are as follows:

$\mathsf{P}_{w_1} = f_1, f_2, w_1;$
$\mathsf{P}_{w_2} = f_2, f_1, w_2;$
$\mathsf{P}_{w_3} = f_2, f_1, w_3;$
$\mathsf{P}_{f_1} = w_1, w_2, w_3, f_1;$
$\mathsf{P}_{f_2} = w_2, w_3, w_1, f_2.$

Now, we consider the two versions of the DA algorithm for the college admissions problem.

Workers-proposing.

Step 1: Each of the three workers proposes to his most-preferred mate: w_1 to f_1, w_2 to f_2 and w_3 to f_2. Firm f_1 receives only offer form a worker who is acceptable to her, and hence, keeps him provisionally. Firm f_2 receives two offers, but can hire only one worker ($q_2 = 1$). So, this firm keeps w_2 because $w_2 \succ_{f_2} w_3$.

Step 2: Worker f_2 is the only rejected worker in the previous step. So, she proposes to firm f_1 (by whom she has not been rejected earlier). Firm f_1 now has one outstanding offer from w_1, and a new offer from w_3, both of whom are acceptable for f_1, and f_1 can accommodate two workers. Thus, both workers w_1 and w_3 are accepted by firm f_1.

STOP: There are no further rejections, and hence, the algorithm stops. It produces a final matching $\mu(f_1) = \{w_1, w_3\}$ and $\mu(f_2) = w_2$ which is stable.

Firms-proposing.

Step 1: Each of the two firms proposes to his most-preferred set of workers (according to preference rankings P_{f_1} and P_{f_2}: Because $q_1 = 2$, f_1 offers to $\{w_1, w_2\}$. By contrast, f_2 proposes only to w_2. Because $f_2 \succ_{w_2} f_1$, worker w_2 (who has received two offers) keeps the offer of f_2 and rejects f_1. On the other hand, w_1 has received an acceptable offer, and hence, keeps it. The provisional matching is thus, $\mu^1(w_1) = f_1$, $\mu^1(w_2) = f_2$ and $\mu^1(w_3) = w_3$.

Step 2: Firm f_1 has an offer rejected, and hence, she makes an offer to two workers, w_1 and w_3. At this step, firm f_1 must compare the two sets, $\{w_1, w_3\}$ and $\{w_1, \varnothing\}$. Notice the crucial modification of the firms-proposing version of the algorithm over that of the marriage market. The provisionally matched worker of f_1, w_1 receives a new offer from his provisional mate. Because $w_3 \succ_{f_1} f_1$, by the responsiveness criterion, $\{w_1, w_3\} \succ^{\#}_{f_1} \{w_1, \varnothing\} \equiv \{w_1\}$, and hence, f_1 makes an offer to the set of workers $\{w_1, w_3\}$. For both workers this is an acceptable offer. So, the offers are accepted.

STOP: There are no further rejections, and hence, the algorithm stops. It produces a final matching $\mu(f_1) = \{w_1, w_3\}$ and $\mu(f_2) = w_2$ which is stable.

Proposition 3.1 not only allows us to concentrate on small coalitions (each consisting of one firm and one worker), but it also asserts that stable matchings can be identified using only firm's preferences over individual workers. It immediately establishes a one-to-one correspondence between the college admissions problem and the marriage market. Consider any firm f in the college admissions problem with capacity q. We can consider a related marriage market wherein firm f can be thought of as q copies of the same firm: $f^1, f^2, \ldots, f^q$ so that the market participants will be workers and firm positions. The preferences of each firm position will be described by its preferences P over individual workers. However, individual workers will now be indifferent among various positions in the same firm. To avoid complexity, let us assume that worker preferences are strict over many positions in a single firm in that for any worker $w \in W$, we have $f^1 \succ_w f^2 \succ_w \ldots \succ_w f^q$ (a professor in ITAM may prefer an office in the east side to that in the west side) whenever firm f appears to be on their list of acceptable firms. Moreover, if any worker w preferred firm f_i (with capacity q_i) to firm f_j (with capacity q_j) in the original college admissions problem, then we assume that, in the corresponding marriage market, this worker prefers all position in f_i to those in f_j, i.e., $f_i^1 \succ_w \ldots \succ_w f_i^{q_i} \succ_w f_j^1 \succ_w \ldots \succ_w f_j^{q_j}$. With this aforementioned one-to-one correspondence between the two matching problems, we can state the following result.

Proposition 3.2: (Roth and Sotomayor, 1990)

A matching of the college admissions problem is stable if and only if the corresponding matchings of the related marriage market are stable.

From the above proposition, it follows that most of the results regarding stable allocations of the marriage market carry through the college admissions problem. Clearly, the crucial assumption for the college admissions problem is that firm preferences are responsive. However, there will be dissimilarities between many-to-one and one-to-one market, which we discuss in the next subsection.

3.5 Incentives in the college admissions problem

We would be interested to analyze stable matching mechanisms for the college admissions problem. We shall establish that Theorem 2.4 holds true for the college admissions problem, but Theorem 2.5 does not. Consider

the following example.

Example 3.5: Misreporting of preferences

Let $F = \{f_1, f_2\}$ and $W = \{w_1, w_2\}$. Further, let $(q_1, q_2) = (2, 1)$. The preferences are as follows (as a exercise, write down the induced responsive preferences):

$\mathsf{P}_{w_1} = f_1, f_2, w_1;$
$\mathsf{P}_{w_2} = f_2, f_1, w_2;$
$\mathsf{P}_{f_1} = w_2, w_1, f_1;$
$\mathsf{P}_{f_2} = w_1, w_2, f_2.$

We run the firms-proposing version of DA. In step 1, f_1 proposes to $\{w_1, w_2\}$ and f_2 proposes to w_1. Worker w_1 has two offers, out of which he keeps the one from f_1. Worker w_2 keeps the offer from f_1 because it is his most preferred offer. In step 2, firm f_2, who has been rejected by w_1, makes a proposal to w_2. Because w_2 prefers f_2 to f_1, he rejects f_1 and accepts f_2. In step 3, f_1, who has a vacant position stays with w_1 because she has been rejected by w_2 in the previous step. There are no further rejections, and the algorithm stops. It produces a stable matching, $\mu(f_1) = \{w_1\}$ and $\mu(f_2) = \{w_2\}$.

Now, suppose firm f_1 mis-reports her preferences: $\mathsf{P}'_{f_1} = w_2, f_1$ (with all other preferences being truthfully reported). The firms-proposing version of DA produces a stable matching $\mu'(f_1) = w_2$ and $\mu'(f_2) = w_1$. Clearly, firm f_1 is strictly better-off by falsifying her preferences because, under the true preferences, P_{f_1}, she prefers w_2 to w_1.

It follows from the above example that theorem 2.5 does not hold for the college admissions problem even if the firms-proposing version of DA leads to firm-optimal stable matching. Surprisingly, this is not the only way to manipulate the matching outcome in a college admissions problem. Some firms can also lie about their capacities, which is called manipulation via capacities..

Example 3.6: Manipulation via capacities

Let $F = \{f_1, f_2\}$ and $W = \{w_1, w_2\}$. Further, let $(q_1, q_2) = (2, 1)$. The preferences are as follows:

$\mathsf{P}_{w_1} = f_1, f_2, w_1;$
$\mathsf{P}_{w_2} = f_2, f_1, w_2;$
$\mathsf{P}_{f_1} = w_2, w_1, f_1;$
$\mathsf{P}_{f_2} = w_1, w_2, f_2.$

The true preferences are the same as in Example 3.6, and hence, the firms-proposing version of DA produces a stable matching $\mu(f_1) = w_1$ and $\mu(f_2) = \{w_2\}$. Now, suppose firm f_1 reports her capacity to be $q'_1 = 1$. The firms-proposing matching mechanism yields $\mu'(f_1) = w_2$ and $\mu'(f_2) = w_1$. Clearly, firm f_1 is strictly better-off under-reporting her capacity because, under the true preferences, P_{f_1}, she prefers w_2 to w_1.

From the above two examples, we see that the equivalence between the college admissions problem and the marriage market clearly breaks down as far as stable matching mechanisms are concerned—some firms would either misreport their preferences or they would under-report their capacities.

Proposition 3.3: Manipulation in the college admission problem

Let (W, F, P, q) be a college admissions problem. Then,

(a) The firm-optimal stable matching mechanism is not strategy-proof for all firms (Roth, 1985).

(b) Suppose there are at least two firms and two workers. Then the firm-optimal stable matching mechanism is not immune to manipulation via capacities (Sönmez, 1997).

3.6 Application I: The medical match

In this section, we review one of the first applications of market design methods: the assignment of medical interns into residency programs. To this date, more than 40,000 medical interns are allocated to more than 30,000 residency programs every year in the United States. The assignment is done through a centralized clearinghouse, known as the National Resident Matching Program (NRMP), or simply known among doctors as "The Match". The success of the NRMP in the United States has led to the adoption of similar clearinghouses in other countries, such as Canada and the United Kingdom.

The history of the NRMP is both an intellectual delight and an example of how economic theory can guide market design in practice, what Alvin Roth famously calls "economic engineering" (Roth, 2002). First, we will briefly review its history. The three main lessons to draw are: (i) the importance of stability as a condition for the survival of an institutional design; (ii) how real-life markets are shaped by a collection of regulations that are the result of trial-and-error and idiosyncratic factors, and how they can result in desirable institutional designs at times, but also in market inefficiencies at others; (iii) how economists have a lot to learn from looking closely at how real-life markets work. Second, we shall study closely the algorithm underlying the original design of The Match in the 1950s and its relation with the Deferred Acceptance algorithm.

3.6.1 A brief history of unraveling

The system through which medical interns are allocated to medical residency programs in the United States underwent multiple changes in the 1940s. In 1951, it reached a design which persisted for more than four decades, until the end of the 1990s. At that time, prompted for calls for reform, a group led by economists undertook a further redesign of the system.

Until 1945, medical interns were assigned to residency programs in a decentralized fashion. As in typical entry-level labor markets, interns were free to apply to residency programs, which in turn accepted applications from the interns of their preferences. Medical residency programs (hospitals) would also seek interns which they preferred the most and offered them binding agreements to enroll in their programs upon graduation. By 1945, it was clear to the administrators of the Association of American Medical Colleges (AAMC) that the market suffered from what now economists term *unraveling*.

Prior to the mid-1940s, hospitals competed in a typical arms race for the best medical interns. The main way in which this competition took place was through the dates of the binding agreements hospitals offered students to lock them into their residency programs. By offering early binding agreements to students, hospitals tried to guarantee high-quality incoming classes. During the first decades of the twentieth century, hospitals offered binding agreements to students earlier and earlier in their career. Initially, these biding agreements were signed a few months before students in the senior class graduated. As hospitals started undercutting each other's agreements, the dates at which students had to decide which residency program they would enroll upon graduation became all but absurd. By the mid-1940s, agreements were typically signed up to two years before graduation from medical schools. This was clearly inefficient. On the one hand, students did not know which program or specialty they would want to study after graduating. In some cases, they had not even taken the necessary classes to make up their minds. On the other, the earlier the agreements offered by hospitals to students, the less the hospitals knew about the quality and aptitudes of the students. Students who appeared to be very promising after two years of medical school, would turn out to be not so successful by the end of it. Given these clear inefficiencies, in 1945 the AAMC decided to stop the unraveling by imposing an earliest date at which medical schools were allowed to disclose student records to hospitals.

At first, fixing the disclosure date served its purpose in that hospitals were not able to lock in students early on through binding agreements. However, another market inefficiency turned up. Given the chaotic process,

which to this day consists of both applications and interviews, it was common for hospitals to jump the gun and offer so-called exploding offers to students. To lock in students, hospitals would make offers to students with very short deadlines right after the date in which records were released. By forcing a student to decide quickly on whether or not to enroll in a residency program, a hospital minimized the probability that another hospital, which the student might prefer, would also made them an offer. Students thus faced tough career decisions. They could play it safe by accepting an early offer even if it was not from their most preferred hospital, or they could risk it and decline such offers in the hope that a hospital they preferred more would come up with an offer later. Either way, students were likely to end up in a less preferred residency program while another program they liked better had an opening for them. Therefore, it became more and more common for students to back out from offers they had previously accepted, which naturally hospitals found annoying.

Between 1945 and 1951, the AAMC implemented a series of regulations which aimed at solving this problem. They largely consisted on regulating the time at which offers could be made, and the time which they should give interns to make up their minds. After some (unsuccessful) experimentation with different sorts of rules, in 1951, the AAMC resolved to fully centralize the process into a matching clearinghouse. Under the new procedure, students and hospitals would communicate and exchange information as before via applications and interviews, but then both would submit rank-ordered lists of their preferences over the hospitals, and applicants they were considering. The final allocation of interns to residency programs would be decided through a matching algorithm.

In 1951, the AAMC performed a trial-run of the new procedure. It was not meant to actually match students and hospitals, but as a basis for the next year. Despite some caveats, the trial-run was deemed successful, and the AAMC decided to fully implement the matching mechanism with a few tweaks the following year. A key aspect of the market that allowed for this type of organizations was that the salaries and responsibilities for medical interns were mostly standard across all programs and not an important part of contract negotiations. Importantly, the matching procedure was to be voluntary. Students were free to opt out and contract directly with hospitals. The matching algorithm, which was used until the end of the 1990s, is known as the NRMP algorithm (for National Resident Matching Program), which used to be the name of the program at the time.

In a remarkable discovery of the economics discipline, some decades later it was noted that the NRMP and the Gale-Shapley algorithms, though written distinctly, were actually equivalent. Notably, this was unknown to both the administrators of the NRMP, and to David Gale and Lloyd Shapley until the 1970s.[1] The NRMP algorithm was used until the late 1990s. At the time, the mechanism faced strong opposition from students, who claimed, amongst other things, that the mechanism was open to gaming. Furthermore, as years went by, it became more common for couples of medical interns to look for medical residency programs that were geographically close to one another. For this reason, a group of economists led by Alvin Roth undertook a partial redesign of the matching algorithm in the mid-1990s. Though the main aspects of the deferred acceptance algorithm remained in place to date, the redesign focused on (i) changing the algorithm from the programs-proposing to the applicants-proposing version of the DA algorithm, and (ii) the way in which the algorithm deals with couples who have interdependent preferences.

3.6.2 The NRMP algorithm

We now study formally the NRMP algorithm, as well as the algorithm used in the trial run of 1951. There are a finite set of hospitals, $H = \{h_1, \ldots, h_m\}$ and a finite set of interns or residents (students), $S = \{s_1, \ldots, s_n\}$. Generic hospitals will be denoted by h, h_i, etc. and generic students will be denoted by s, s_j etc. For simplicity, we restrict attention to the case in which every hospital has exactly one position, i.e., $q_i = 1$ for all i. As discussed above, the algorithm was modified to incorporate several complaints brought up by students after

[1] According to anecdotal evidence reported by Roth (1984), it was until 1976 when David Gale first heard of the labor market for medical interns, and sent a copy of the Gale and Shapley (1962) paper to an administrator of the NRMP.

the trial run. As we shall see below, the main concern was that the trial run algorithm was not strategy-proof for students. But not only this, it was also not a stable mechanism. After modifying the algorithm, AAMC administrators came up with the NRMP algorithm, which always generates a stable matching. At the time, it was mistakenly claimed that was also strategy-proof for students. Remarkably, this appears to have remained unknown for several decades until economists studied the algorithm formally.

The trial run algorithm of 1951. Students submit a rank ordering of hospitals. Hospitals submit a ranking dividing student into five groups: rank 1, rank 2, ..., rank 5; each group containing as many students as the number of positions the hospital is offering. The algorithm proceeds in consecutive stages as follows.

Step 1.1: Interns and hospitals are matched if they gave each other a rank of 1.

Step 1.2: The remaining students and hospitals are matched if the student has ranked the hospital 1 and the hospital has ranked the student 2.

Step 2.1: Among the remaining students and hospitals, match students who ranked hospitals 2, and hospitals who ranked students 1.

Step 2.2: The remaining students and hospitals are matched if the student has ranked the hospital 2 and the hospital has ranked the student 2.

⋮

Proposition 3.4

The NRMP trial run algorithm is not stable, nor strategy-proof for students.

Proof. Consider the following example with $H = \{h_1, h_2, h_3\}$ and $S = \{s_1, s_2, s_3\}$. The preferences are given as follows:

$$\begin{aligned} \mathsf{P}_{s_1} &= h_1, h_2, h_3, s_1; & \mathsf{P}_{h_1} &= s_2, s_3, s_1, h_1; \\ \mathsf{P}_{s_2} &= h_2, h_3, h_1, s_2; & \mathsf{P}_{h_2} &= s_1, s_2, s_3, h_2; \\ \mathsf{P}_{s_3} &= h_1, h_3, h_2, s_3; & \mathsf{P}_{h_3} &= s_3, s_2, s_1, h_3. \end{aligned}$$

Suppose that everyone submits their true preferences to the NRMP trial run algorithm. In Step 1.1, there are no matches. No one is ranked 1 by whom they ranked first. At Step 1.2, it matches (s_2, h_2) and (s_3, h_1). And, eventually at Step 3.3, it also matches (s_1, h_3). So, final matching is $\mu(s_1) = h_3$, $\mu(s_2) = h_2$ and $\mu(s_3) = h_1$.

First, note that this matching is not stable. Because $h_2 \succ_{s_1} h_3 = \mu(s_1)$ and $s_1 \succ_{h_2} s_2 = \mu(h_2)$, (s_1, h_2) is a blocking pair. Second, note that if s_1 had reported $\mathsf{P}'_{s_1} = h_2, h_1, h_3, s_1$, then in Step 1.1 of the algorithm, s_1 and h_2 would have been matched to each other. Because $h_2 \succ_{s_1} h_3 = \mu(s_1)$ (under his true preference ordering), s_1 would have gained by misreporting. □

The NRMP algorithm. Now, we turn to the NRMP algorithm, which was first used in 1952, and remained without any change until its redesign in the mid 1990s. Instead of writing down the algorithm, we present its main ingredients. (source: NRMP webpage):

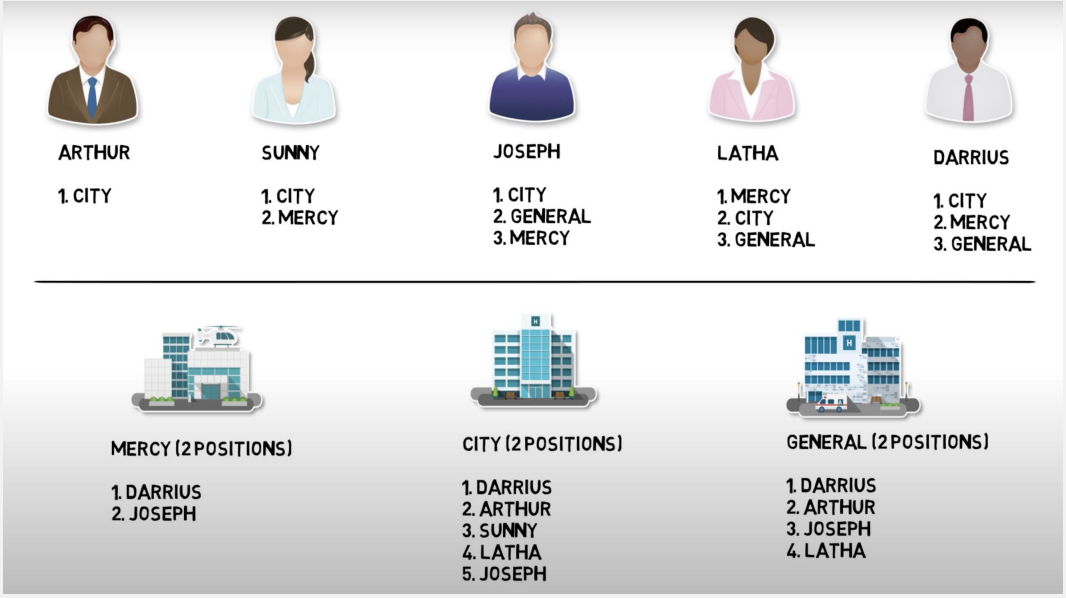
- **How does it work?** The matching algorithm is "applicant-proposing" meaning it attempts to place an applicant (Applicant A) into the program indicated as most preferred on Applicant A's rank order list (ROL). If Applicant A cannot be matched to this first choice program (because the program doesn't also prefer Applicant A), an attempt is then made to place Applicant A into the second choice program, and

so on, until Applicant A obtains a tentative match, or all of Applicant A's choices have been exhausted (meaning Applicant A cannot be tentatively matched to any program on the ROL).

- **What does tentative match mean?** Applicant A will be tentatively matched to a program if the program also ranks Applicant A on its rank order list, and either:
  - the program has an unfilled position (making room for the tentative match to Applicant A) or
  - the program is filled *but* Applicant A is more preferred by the program than another applicant (Applicant B) already tentatively matched. In such a case, Applicant B is "bumped" from the tentative match with the program to make room for Applicant A.
- **What happens to an applicant whose tentative match is "bumped"?** The matching algorithm will return to Applicant B's rank order list and attempt to tentatively match Applicant B at the next most preferred position on Applicant B's list. The attempt to find another tentative match for Applicant B is done in the same manner as outlined for Applicant A.
- **When does a tentative match become final?** When all applicants' rank order lists have been considered, the matching algorithm is complete and all tentative matches become final and **binding** for training.

Example 3.7: DA versus NRMP

Watch the How the NRMP Matching Algorithm Works video. The following figure describes the rank order lists of all market participants.



We first run the applicants-proposing DA algorithm (same as the workers-proposing version in Example 3.4). In Step 1, Latha proposes to Mercy, and all interns but Latha propose to City. Darrius and Arthur are tentatively matched to City because Sunny and Joseph are lower-ranked than Darrius and Arthur by City, and Latha is not ranked by Mercy. In Step 2, the previously rejected interns propose to their second-ranked programs: Sunny to Mercy, Joseph to General, and Latha to City. Sunny is rejected again because Mercy did not rank her. Joseph is acceptable for General, and hence, they are tentatively matched. Latha is rejected by City because both City's tentative matches, Darrius and Arthur are higher-ranked than Latha. In Step 3, Sunny cannot make further proposals because she had exhausted her ROL, and hence, stays unmatched. Latha proposes to her third-ranked program, General. General still has one vacancy, and Latha is ranked by General. There are no more rejections, and the algorithm stops. So, the

applicant-proposing DA algorithm produces the following final match:

$$\mu(\text{CITY}) = \{\text{Darrius, Arthur}\}, \quad \mu(\text{GENERAL}) = \{\text{Joseph, Latha}\},$$
$$\mu(\text{MERCY}) = \{\varnothing, \varnothing\}, \quad \mu(\text{Sunny}) = \text{Sunny}.$$

Note that the DA is different from the NRMP algorithm (as described in the video). However, the NRMP algorithm produces the same stable allocation. In fact, both algorithms are equivalent.

Between 1952 and mid 1990s, the NRMP algorithm (an algorithm that allows voluntary participation) used an equivalent version of hospitals-proposing DA algorithm, which produced hospital-optimal stable matching. However, we have seen in Example 3.5 that this matching mechanism may not be strategy-proof for residency programs. Moreover, the presence of applicant couples made the allocations less likely to be stable, and couples tended more and more to opt out from the residency matching program. Following the suggestions of Roth and Peranson (1999), the current NRMP algorithm has been modified to be applicants-proposing which can also accommodate couples. Two important aspects of the NRMP is that participation is not free (consult the current NRMP fee structure) and once a match occurs, given the ROLs, the contracts are binding.

The couples market. If applicants decide to participate in the couples market, as opposed to the singles market, NRMP treat a couple's primary ROLs as paired ranks (see how couples' ROLs are paired). The rule is the following. Suppose David and Erin decide to participate in the residency matching program as a couple (any two participants such as married couples, friends, siblings can participate in the couples market). In their paired rank order list suppose they have ranked (General, City) as their paired options (as both programs are in San . On the other hand, suppose City has ranked Erin, but General has not ranked David. If this proposal pops up at some stage of the algorithm, then Erin cannot be tentatively matched even if she is an acceptable intern for City. So, apart from the couples fee, participation in the couples market bears the risk of being unmatched.

Example 3.8: Couples in NRMP

Watch how NRMP algorithm with couples works.

3.7 Application II: School choice

In this section we study the problem of assigning public school seats to students. Traditionally, children are assigned to schools according to where they live. However, in several parts of the world this has been deemed unfair in recent years. While wealthier parents can decide to move to a neighborhood with good schools, parents without such means had no choice of school, and had to send their children to schools assigned to them by the district. Today, several states in the United States offer inter-district and intra-district school choice programs. We now study the problem of matching students to schools as a formal two-sided matching problem.

3.7.1 Priorities as opposed to preferences

On the outset, school choice problem is a many-to-one matching problem that looks very similar to the medical match problem. However, there are subtle difference—unlike hospitals, schools are assumed not to have preferences over students. Public schools are treated merely as object who provide educational services. Although schools do not have preferences over students, they nevertheless rank them. A school may give priority to a student whose siblings are already students of that school. Schools may also give priorities to specific social

groups. So, in order to distinguish these rankings over students from the concept of preferences, we shall call such rankings priorities (or priority lists). The main difference between priorities and preferences is that any student is acceptable for a school as long as it has enough capacity to accommodate them.

At this juncture, we are required to impose restrictions on schools' priority lists which is usually a ranking of individual students. However, each school can enroll many student (it has capacity greater than 1). So, we shall assume, similar to the college admission problem, that priorities are responsive. It is in the following sense. For a given school, ranking over any two students who are not yet enrolled is not affected by the students already enrolled in the school. On the other hand, we continue assuming that the preferences of students over schools are strict orders.

3.7.2 Many-to-one matching for school choice

A school choice problem consists of a set of students, $I = \{i_1, \ldots, i_n\}$, a set of schools, $S = \{s_1, \ldots, s_m\}$ with their respective capacities, $q = (q_1, \ldots, q_m)$, the profile of strict preferences of the students, $\mathsf{P}_{i \in I}$ or $\succ_{i \in I}$, and the priorities of the schools over students, $\pi_{s \in S}$.

Definition 3.5: Matching in school choice

A many-to-one matching for the school choice problem is a correspondence $\mu : I \cup S \twoheadrightarrow I \cup S$ such that (a) $\mu(s) \subseteq I$ with $|\mu(s)| \leq q_s$ for all $s \in S$; (b) $\mu(i) \in S \cup \{i\}$ for all $i \in I$; and (c) $\mu(i) = s$ if and only if $\mu(s) \in I$.

The interpretation of the above definition is similar to that [cf. Definition 3.2] of the college admission problem.

3.7.3 Stability and efficiency

The concept of stability in school choice is also very similar to that in the college admission problem.

Definition 3.6: Stability in school choice

Consider a matching μ for the school choice problem. The matching

(a) is individually rational if each student $i \in I$ weakly prefers $\mu(i)$ to being unmatched, i.e., $\mu(i) \succeq_i i$;

(b) is not wasteful, i.e., if there is a student i and a school s such that $i \notin \mu(s)$ and $s \succ_i \mu(i)$, then it must be that $|\mu(s)| = q_s$. In words, if a student prefers a school to their current assignment, then that school must have exhausted its capacity;

(c) eliminates justified envy, i.e., for all i, $i' \in I$ with $\mu(i') = s \in S$, $s \succ_i \mu(i)$ implies $i' \pi_s i$. In words, if a student prefers a school to their assignment, then all students enrolled to that school must have higher priorities than that student.

A matching μ is stable if it is individually rational, not wasteful and eliminates justified envy.

In condition (a), a student being unmatched can be interpreted as the student prefers to opt out of the public school system, and go to a private school. Condition (c) is equivalent to no-blocking by a student-college pair. If a school s has an enrolled student i' who is lower-ranked than a not-enrolled student i who prefers school s to their match, then the pair (i, s) blocks the current matching.

Definition 3.7: Efficient matching

Given a preference profile $(\mathsf{P}_i)_{i\in I}$, a matching μ is efficient if there is no other matching μ' such that $\mu'(i) \succeq_i \mu(i)$ for all $i \in I$ and $\mu(j) \succ_j \mu(j)$ for at least one $j \in I$.

A crucial point to note is that efficiency of a matching is defined only in terms of the preference orderings of the students. Schools having priorities, instead of preferences, means that a school do not derive more utility from enrolling one student over the other although one of them may have higher priority over the other.

Example 3.9: Stability versus efficiency

Consider a school choice problem with $I = \{i_1, i_2, i_3, i_4\}$ and $S = \{s_1, s_2, s_3\}$ with capacities $(q_1, q_2, q_3) = (1, 2, 1)$. The preferences of the students are given by

$\mathsf{P}_{i_1} = s_2, s_1, s_3, i_1;$
$\mathsf{P}_{i_2} = s_1, s_2, s_3, i_2;$
$\mathsf{P}_{i_3} = s_1, s_2, s_3, i_3;$
$\mathsf{P}_{i_4} = s_2, s_3, s_1, i_4.$

On the other hand, schools' priority lists are given by

$\pi_{s_1} = i_1, i_2, i_3, i_4;$
$\pi_{s_2} = i_3, i_4, i_1, i_2;$
$\pi_{s_3} = i_4, i_1, i_2, i_3.$

Consider the following two matching outcomes:

$$\mu : \mu(s_1) = i_1, \quad \mu(s_2) = \{i_3, i_4\}, \quad \mu(s_3) = i_2,$$
$$\mu' : \mu(s_1) = i_3, \quad \mu(s_2) = \{i_1, i_4\}, \quad \mu(s_3) = i_2, .$$

(a) μ is **not efficient** because i_1 and i_3 strictly prefers μ' to μ, whereas i_2 and i_4 are indifferent between μ and μ'.

(b) μ' is **efficient**. In μ', i_1, i_3 and i_4 get their top choices. So to verify the efficiency of μ', we have to guarantee that there is no other matching so that i_2 prefers this matching to μ'. In a different matching, suppose that we match i_2 to s_1. Then, we have to un-assign i_3 from s_1, and reassign them to some other school. Because s_1 is i_3's top choice, they would be strictly worse off by any such re-match. Same is true if we un-assign either i_1 or i_4 from school s_2. Thus, there is no way to make i_2 better off without making one of the other three students worse off.

(c) μ is **stable**. Note that μ is individually rational and not wasteful. So, we have to verify whether μ eliminates justified envy. Because i_4 is assigned to their top choice, only i_1, i_2 and i_3 would like to be enrolled in another school. First, i_2 would like to go to s_2, their top choice. However, both i_3 and i_4 have higher priority over i_1 in school s_2, and they are already assigned to s_2. Similar arguments apply to i_2 and i_3.

(d) μ' is **not stable**. Note first that μ' is individually rational and non-wasteful. So, if μ' is not stable, it must be the case that the matching cannot eliminate justified envy. Students i_1, i_3 and i_4 are enrolled into their top choices. So, if there is a blocking pair, it must involve student i_2. Note that s_1, which is the top choice of i_2, is matched with i_3. However, for s_1, i_2 is ranked higher than i_3. So, i_2 and s_1 form a blocking pair, and hence, μ' is not stable. Now, we can ask the question whether (i_2, s_2) can form a blocking pair. In this case, s_2 must get rid of either i_1 or i_4. This is

to say s_2 has higher priority for $\{i_2, i_4\}$ than $\{i_1, i_4\}$ or for $\{i_1, i_2\}$ than $\{i_1, i_4\}$. But this is not the case with responsive priorities because both i_1 and i_4 are ranked above i_2 by school s_2.

There may be school choice outcomes that are both stable and efficient. However, they can be reached under very restrictive and demanding assumptions, e.g. schools should have very similar priority lists. In general, the trade-off between stability and efficiency would exists in the simple school choice model.

3.7.4 Competing algorithms

We would analyze here two algorithms for school choice—the first one is the Deferred Acceptance algorithm, and the second one known as the Immediate Acceptance algorithm (IA). In Chapter 4 we shall analyze another algorithm, called the Top Trading Cycle algorithm. The DA and IA for school choice are very similar with a little difference which we would discuss opportunely. One important point to make is that, unlike the college admission problem, we would consider only the students-proposing versions of the two algorithms. Schools per se do not derive any utility out of enrolling students although they have priorities over students. So, a schools-proposing version is not so interesting to consider.

The Deferred Acceptance algorithm for school choice. The DA for school choice works as follows.

Step 1: Each student i applies to the school that is ranked first in their preference list (if there is no such school, then i becomes unassigned). Each school s assigns students, one at a time, up to its capacity from the students applying to s, following the priority order π_s. That is, school s first admits the student with the highest priority, then the student with the second-highest priority, and so on until either school s has enrolled q_s students or it has enrolled all the students who applied to s. The remaining students are rejected.

⋮

Step k: Each student rejected in the previous step applies to the most preferred school among those they have not yet proposed to and that are acceptable (if there is no such school, then the student becomes unassigned). Each school receiving applications considers the set of students it accepted at the previous step together with the set of new applicants. From this larger set, the school accepts students up to its capacity, one at a time, following its priority order. The remaining students are rejected.

⋮

STOP: The algorithm halts when no student is rejected or all schools have filled their capacities. Any other student remains unassigned.

The above algorithm yields the student-optimal stable matching. From Theorem 2.5, it follows that a mechanism that uses school choice DA is strategy-proof (for the students, the proposing side). Moreover, Example 3.9 reveals that the matching produced by DA is not necessarily efficient.

Exercise 3.2: DA for school choice

Consider a school choice problem with $I = \{\text{Ana, Belen, Carlos, Daniel}\}$ and $S = \{a, b, c\}$ with capacities $(q_a, q_b, q_c) = (2, 1, 1)$. The preferences of the students are given by

$\mathsf{P}_{\text{Ana}} = a, b, c,$ Ana;
$\mathsf{P}_{\text{Belen}} = a, b, c,$ Belen;

$P_{Carlos} = b,\ a,\ c$, Carlos;
$P_{Daniel} = a,\ c,\ b$, Daniel.

On the other hand, schools' priority lists are given by

π_a = Ana, Carlos, Daniel, Belen;
π_b = Ana, Belen, Daniel, Carlos;
π_c = Belen, Carlos, Daniel, Ana.

Show that DA for school choice yields matching μ wherein $\mu(a) = \{\text{Ana, Carlos}\}$, $\mu(b) = \{\text{Belen}\}$ and $\mu(c) = \{\text{Daniel}\}$ is the student-optimal stable matching, but it is not efficient. Can you suggest an efficient matching?

The Immediate Acceptance algorithm. The difference between DA and IA is that in IA, there are no temporary assignments, i.e., acceptances are immediate, and not deferred till the end. Refer to Haeringer (2017, p. 248-249) to see how IA works.

Example 3.10: IA for school choice

Consider the same set up as in Exercise 3.2 and run IA. In Step 1, Ana, Belen, and Daniel propose to school a. School a can admit at most two students, so it starts accepting the students who applied to it following the order given by its priority order. Ana is the top priority student, so she is accepted. The next student with the highest priority and is applying to a is Daniel. So Daniel is accepted, and Belen is rejected because school a has filled its capacity with Ana and Daniel. The other student, Carlos, proposes to school b. He is the only applicant, so he is accepted. So at the end of the first step, Ana and Daniel are accepted at school a, Carlos is accepted at school b, and Belen is not assigned to any school.

In Step 2, Belen is the rejected student. She applies to school b. But b already filled its capacity in Step 1. This means that in Step 2 the remaining capacity of school b is 0, and hence, it can no longer accept any additional student. So, Belen is rejected.

In Step 3, Belen is the only rejected student from the previous step. She applies to her most preferred school among those she has not yet proposed, i.e., she applies to school c. She is the only applicant, so she is accepted. No student is rejected, so the algorithm halts.

The final matching μ' produced by IA is given by $\mu'(a) = \{\text{Ana, Daniel}\}$, $\mu'(b) = \{\text{Carlos}\}$ and $\mu'(c) = \{\text{Belen}\}$. Notice that μ' is not stable. Observe that the student-school pair (Belen, b) forms a blocking pair: Belen prefers b to her match (school c), and school c gives Belen a higher priority than Carlos (who is enrolled in it). However, μ' is efficient. To see this, observe that Ana, Carlos and Daniel are assigned to their top choices. Only, Belen got her third choice. So, if we want to make her better off by assigning her to school a, then either Ana or Daniel must be un-assigned from school a and be reassigned to some other school. But in the reassignment, one of them is necessarily worse off. Similar argument goes through is we try to assign Belen to school b, which would make Carlos worse off.

However, the matching mechanism that uses A is not strategy-proof. Suppose Belen reports $P'_{Belen} = b,\ a,\ c$, Belen. In Step 1, Ana and Daniel propose to school a, and Belen and Carlos propose to school b. Because school a has two seats, both Ana and Daniel, the only two applicants to this school, are accepted by a. On the other hand, school b has only one seat, and Belen is

ranked higher than Carlos by b. Hence, Belen is accepted and Carlos is rejected by b. In Step 2, Carlos applies to school a which had already exhausted its capacity. So, school a rejects Carlos. In Step 3, Carlos applies to school c; being the only applicant there he is accepted at school c. The final match is given by $\mu''(a) = \{\text{Ana, Daniel}\}$, $\mu''(b) =$ Belen and $\mu''(c) =$ Carlos. Clearly, $b = \mu''(\text{Belen}) \succ_{\text{Belen}} \mu'(\text{Belen}) = c$. Therefore, the mechanism is not strategy-proof for Belen.

To summarize, the Deferred Acceptance algorithm for school choice yields the student-optimal stable matching, and the associated matching mechanism is strategy-proof. However, the matching may not be efficient. By contrast, the Immediate acceptance algorithm produces an efficient matching, but it may yield an unstable matching. Moreover, the associated mechanism may not be strategy-proof. In fact, we have the following general result.

Proposition 3.5: (Kesten, 2010)

There is no strategy-proof mechanism that selects the efficient and stable matching whenever it exists.

As we have discussed earlier that a matching for the school choice problem may be both stable and efficient. However, if such matching exists, there is no mechanism that can truthfully implement this assignment. The intuition lies in the discussion preceding Proposition 3.5. It is worth noting that the Boston mechanism that assigns K-12 students to Boston public schools has been using the Immediate Acceptance algorithm. The theory of school choice matching is due to the seminal paper by Abdulkadiroğlu and Sönmez (2003).

Chapter 4

The housing market

As opposed to the marriage market and the college admission problem, object allocation problem is a two-sided matching market with one-sided preferences. Think of firms as machines to produce a goods in our firm-worker market. Workers have preference orderings over machines, but machines (objects) do not have preferences over workers. Other examples include allocation of houses to individuals (the house allocation problem), allocation of vaccines and medical equipments (e.g. ventilators) to people, kidney transplants, etc. We assume that there is a finite set of agents or individuals, I and a finite set of objects (houses), H. For simplicity, we assume that there are as many houses as individuals, i.e., $|I| = |H| = n$. Workers have strict preferences over firms, denoted by $(\mathsf{P}_i)_{i\in I}$ or $(\succ_i)_{i\in I}$. A housing market will be denoted by $\mathcal{M} = (H, I, (\succ_i)_{i\in I})$. An allocation of the market $\mathcal{M}$ is a matching that assigns a house to each individual. Formally,

Definition 4.1: Matching

A matching is a function $\mu : I \cup H \to I \cup H$ such that (a) for each individual $i \in I$, $\mu(i) \in H \cup \{i\}$, and for each object $h \in H$, $\mu(h) \in I \cup \varnothing$; and (b) $\mu(i) = h$ if and only if $\mu(h) = i$.

A matching specifies which agent is assigned to which house. If $\mu(i) = h$, then house h is assigned to agent i, which is equivalent to $\mu(h) = i$. Note that we allow for agents to remain unmatched by matching them with themselves, and for houses to remain unassigned, which we specify by $\mu(h) = \varnothing$.

4.1 House allocation with public endowments

We start with the simplest possible model of object allocation wherein no individual owns a house, i.e., each individual has zero endowment, and someone (say, the government) outside the set I owns all the houses in H. Such problem will be called house allocation with public endowments. The first "desirable" property we would like to impose on a matching allocation (or simply allocation) is Pareto efficiency.

Definition 4.2: Pareto efficiency

Given a preference profile $(\mathsf{P}_i)_{i\in I}$, a matching μ Pareto efficient if there is no other matching μ' such that $\mu'(i) \succeq_i \mu(i)$ for all $i \in I$ and $\mu(j) \succ_j \mu(j)$ for at least one $j \in I$.

Consider the following example

Example 4.1

Let $I = \{\text{Ana, Belen, Carlos, David}\}$ and $H = \{a, b, c, d\}$. The preferences are given by

$\mathsf{P}_{\text{Ana}} = b, c, d, a$;
$\mathsf{P}_{\text{Belen}} = a, b, c, d$;
$\mathsf{P}_{\text{Carlos}} = a, c, d, b$;
$\mathsf{P}_{\text{David}} = a, d, b, c$.

That is to say, Ana prefers house b over all the houses, followed by house c, house d, etc. Consider the matching μ given by:

$$\mu(\text{Ana}) = d, \quad \mu(\text{Belen}) = a, \quad \mu(\text{Carlos}) = c, \quad \mu(\text{David}) = b.$$

Is matching μ Pareto efficient? The answer is no because their is a matching μ', wherein

$$\mu'(\text{Ana}) = b, \quad \mu'(\text{Belen}) = a, \quad \mu'(\text{Carlos}) = c, \quad \mu'(\text{David}) = d,$$

which Pareto dominates μ. In μ', Ana and David simply trade their houses, and are strictly better off. Is matching μ' Pareto efficient?

So, the question we ask now is whether a Pareto efficient matching exists in house-allocation problems, and, if so, how to find them.

The Serial Dictatorship algorithm. The Serial Dictatorship (SD) algorithm, which is a priority algorithm, is one of the simplest algorithms to find desirable outcomes in the house allocation problem.

Step 0: Fix a priority order of individuals, π. Formally, $\pi : \{1, \ldots, n\} \to I$ that assigns a natural number to each individual form a sequence. That is, in a given order π, $\pi(k)$ is the individual in the k-th position in the sequence. For example, given $I = \{\text{Ana, Belen, Carlos, David}\}$, a priority order is $\pi = (\text{Ana, David, Carlos, Belen})$. Another order is given by $\pi' = (\text{David, Belen, Carlos, Ana})$. How many priority orders are there?

Step 1: All houses are available. Assign agent $\pi(1)$ to their top choice, and remove both this agent and their allocation from the market.

Step 2: Assign agent $\pi(2)$ to their top choice from the set of available houses. Remove both this agent and their allocation from the market.

⋮

Step k: Assign agent $\pi(k)$ to their top choice from the set of available houses. Remove both this agent and their allocation from the market.

⋮

STOP: The algorithm stops when every agent has been assigned to a house or there are no more available houses.

The SD algorithm assigns the agent with the highest priority to their most favorite house, the agent with the second-highest priority to their most favorite house among the remaining ones, and so on, until there is no house left or all agents have been assigned to a house. Clearly, the algorithm is not fair in that it favors agents with higher priority. A common practice to circumvent this issue is to assign priorities randomly. Nonetheless, the most attractive feature of SD is that it always generates a matching that is Pareto efficient. And not only that, it actually characterizes the set of Pareto efficient assignments.

Proposition 4.1: Serial dictatorship is Pareto efficient

Let μ be a matching of the house allocation problem.

(a) If μ is an outcome of the Serial Dictatorship algorithm under a given priority order π, then μ is Pareto efficient under any such priority order;

(b) If μ is Pareto efficient, then there is a priority order π such that μ is a matching generated by the Serial Dictatorship algorithm under π.

Proof. First we show that the outcome of SD is Pareto efficient. Proceed by contradiction. Let μ be an outcome of SD, and assume there exists μ' that Pareto dominates μ. Then, there is a non-empty set $I' = \{i \in I \mid \mu'(i) \succ_i \mu(i)\} \subset I$. Let j (say, Belen) be the individual in I' who gets to choose first under π, and let $\mu(j) = h$ and $\mu'(j) = h'$. Then, each agent who has chosen before j was assigned the same object under both μ and μ' (otherwise, Belen would not be the first individual to be assigned different objects under μ and μ'). This means that h' was available to Belen, yet she had picked h. This contradicts the fact that $h' \succ_j h$ (if this were the case, why had not she chosen h' or some house that is more preferred to h'?). Because the choice of π has been arbitrary, this completes the proof of part (a).

We now prove part (b). Let μ be a Pareto efficient matching. To show that μ is an outcome of the SD for some priority order π, we first claim that under μ, some agent must be getting their top choice. Suppose not. Then, let each agent point at their top choice, and let each house point at its owner under μ. This must lead to a cycle because the number of agents is finite (why?). Move every agent in the cycle to the house they are pointing at. This new allocation Pareto dominates μ, which is a contradiction. Hence, order the $m \geq 1$ agents who get their top choices under μ as $\pi(1), \pi(2), \ldots, \pi(m)$. Repeat the same argument with the remaining $n - m$ agents and the houses that have not been assigned to any of the first m agents. Continue until every agent has been assigned a priority order. Note that, by construction, μ is the resulting matching of the SD under π. □

4.2 House allocation with private endowments

Now think of the housing market wherein each house is owned by some agent, and the house allocation problem is a simple exchange of houses between any pair of owners. Formally, a housing market is represented by the tuple $\mathcal{M} = (H, I, (\omega_i)_{i \in I}, (\succ_i)_{i \in I})$ where ω_i is the endowment of individual $i \in I$. This looks like an Edgeworth box economy with each market participant possessing only one indivisible good. The market works in the following way. All agents gather in the 'marketplace' with the objects they own, and each one of them places their endowment on a huge table. At the end of the day, each one picks (the choice is a function of their preferences) one object (which may be their endowment), which is their allocation. So, how a voluntary participation can be mediated so that everybody is "happy" after the trades? Consider the two following examples.

Example 4.2: A matching that is not individually rational

Let $I = \{\text{Ana, Belen, Carlos, David}\}$ and $H = \{a, b, c, d\}$. The endowments and preferences are given by

$(\omega_{\text{Ana}}, \omega_{\text{Belen}}, \omega_{\text{Carlos}}, \omega_{\text{David}}) = (a, b, c, d)$;
$\mathsf{P}_{\text{Ana}} = b, c, d, a$;
$\mathsf{P}_{\text{Belen}} = a, b, c, d$;
$\mathsf{P}_{\text{Carlos}} = a, c, d, b$;

$P_{\text{David}} = a,\ d,\ b,\ c.$

The Serial Dictatorship may not always give a sensible allocation. Let us fix π =(Carlos, Ana, David, Belen) and run SD. This will produce $\mu(\text{Ana}) = b$, $\mu(\text{Belen}) = c$, $\mu(\text{Carlos}) = a$ and $\mu(\text{David}) = d$. By Proposition 4.1, we know that μ is Pareto efficient. However, Belen does not have incentives to participate in the market because μ made her worse off. She would rather stay with her endowment.

Example 4.3: A matching that is blocked by a coalition

Let $I = \{\text{Ana, Belen, Carlos}\}$ and $H = \{a,\ b,\ c\}$. The endowments and preferences are given by
$(\omega_{\text{Ana}},\ \omega_{\text{Belen}},\ \omega_{\text{Carlos}}) = (a,\ b,\ c)$;
$P_{\text{Ana}} = b,\ c,\ a$;
$P_{\text{Belen}} = a,\ b,\ c$;
$P_{\text{Carlos}} = a,\ b,\ c$.
Let π =(Carlos, Belen, Ana) and run SD. This will produce $\mu(\text{Ana}) = c$, $\mu(\text{Belen}) = b$ and $\mu(\text{Carlos}) = a$. Clearly, μ is Pareto efficient, and assigns to each agent a house they prefer at least as much as the one they initially own. However, note that Ana and Belen would be better off by not participating in the mechanism and trading their endowments amongst themselves. That is, Ana would rather have house b (in exchange for a) than getting c in μ, and Belen would rather have house a (in exchange for b) than keeping her endowment, b in μ.

In Example 4.2, Belen would like not to participate in the market because the (Pareto efficient) assignment μ is not individually rational for her. On the other hand, in Example 4.3, Ana and Belen would be better off by not participating, i.e., they would block μ even if it is Pareto efficient.

Individual rationality and the core. Individual rationality captures the idea of property rights. A matching is individually rational if it assigns to each agent a house that they find at least as good as the one they already own.

Definition 4.3: Individual rationality

Let μ be a matching of the housing market, $(H,\ I,\ (\succ_i)_{i\in I})$. The matching μ is individually rational if $\mu(i) \succeq_i \omega_i$ for every $i \in I$

The reason is the same as in an Edgeworth box economy. The notion of Pareto efficiency does not take agents' endowments into account, whereas the concept of individual rationality does. Notice that any Pareto efficient allocation is trivially individually rational in the housing market with public endowments.

As indicated by Example 4.3, individual rationality however does not guarantee that groups of individuals would want to participate in an exchange. The notion of blocking is the similar to that in the college admission problem with the difference that blocking coalitions are formed by the individuals on the same side of the market. The notion of blocking by coalitions is somewhat vacuous because no one initially owns anything.

Definition 4.4: Core allocation

Let μ be a matching of the housing market, $(H,\ I,\ (\succ_i)_{i\in I})$. The matching μ is blocked by a coalition of agents $C \subseteq I$ if there exists another matching μ' such that $\mu'(i) \succeq_i \mu(i)$ for all $i \in C$ and $\mu'(i) \succ_i \mu(i)$ for some $i \in C$ with the property that for each $i \in C$, $\mu'(i) = \omega_j$ for some $j \in C$ and $j \neq i$. The

matching μ is a core allocation if it is not blocked by any coalitions of agents.

Proposition 4.2

Every matching that is a core allocation is individually rational and Pareto efficient.

Proof. Trivial. If a core allocation is not individually rational, it is blocked by singleton coalitions, i.e., $C = \{i\}$ for some $i \in I$. If it is not Pareto efficient, then it is blocked by the grand coalition, i.e., $C = I$. □

Proposition 4.2 above shows that the property of being in the core implies individual rationality and Pareto efficiency. It is a stronger notion in the sense that the converse is not true (see Example 4.3). Intuitively, the core is an appealing notion when thinking of decentralized exchanges. That is, if agents were to exchange houses on their own, it is natural to think that the resulting matching would lie in the core. Otherwise, the agents in a blocking coalition would further exchange their houses to improve upon their final allocations. This line of reasoning brings up the question of whether the core is nonempty. That is, if left to their own devices to exchange houses, would agents be able to "converge" to an allocation in which no further exchanges are possible? We shall show that in a house allocation problem the we can find a core allocation by an algorithm, called the Top Trading Cycle algorithm (TTC) which was first proposed by David Gale. Prior to its introduction, let us refresh our knowledge of graph theory.[1]

Some concepts of graph theory. Let us start with discussing what is actually meant by a network. To this end, we first concentrate on some basic formal concepts and notations from graph theory. Networks are mathematically known as graphs. In its simplest form, a graph is a collection of vertices that can be connected to each other by means of edges. In particular, each edge of graph joins exactly two vertices. Using a formal notation, a graph is defined as follows.

Definition 4.5: Graphs

A graph G consists of a collection of vertices V, and a collection of edges E, for which we write $G = (V, E)$. Each edge $e \in E$ is said to join two vertices, which are called its end points. If e joins $i, j \in V$, we write $e = \langle i, j \rangle$. Vertex i and j in this case are said to be adjacent. Edge e is said to be incident with vertices i and j, respectively.

In a graph G, two vertices could be connected by one or more edges. An edge has been represented by an unordered pair of vertices. However, having no ordering is not always convenient. Consider the following examples:

- Suppose we want to model a street plan as a network. This is naturally done by representing a junction as a vertex and a street as an edge connecting two junctions. However, we need a notion of edge direction if we want to represent one-way streets.
- In social relations it is often convenient to represent the fact that Alice knows Bob, but that the opposite is not the case. In a social network this is done by representing people by vertices, and the "who knows whom" relation by means of directed edge.
- In computer networks, and notably wireless networks, links between two different nodes are often not symmetric in the sense that messages can generally be successfully sent from station A to B, but not the other way around. Modeling such a computer network is more conveniently done using directed edges.

[1] The discussion on graphs and networks is borrowed from van Steen (2010)

What we are thus seeking is a way to extend graphs that we will be able to model these and similar situations. The need for associating a direction with the edges of a graph leads to the notion of a directed graph, or simply digraph:

Definition 4.6: Directed graphs

A directed graph or digraph D consists of a collection of vertices V, and a collection of arcs Ω, for which we write $D = (V, \Omega)$. Each arc $\omega \in \Omega$ is said to join vertex $i \in V$ to another (not necessarily distinct) vertex $j \in V$, we write $\omega = \langle \overrightarrow{i, j} \rangle$ (graphically, represented by an arrow starting at vertex i and ending at vertex j). Vertex i is called the tail of ω, whereas j is its head. A loop $\langle \overrightarrow{i, i} \rangle$ is an arc where vertex i is both head and tail of the arc. For a vertex i of digraph D, the number of arcs with head i is called the indegree $\delta_{in}(i)$ of i. Likewise, the outdegree $\delta_{out}(i)$ is the number of arcs having i as their tail.

The underlying graph $G(D)$ of a digraph D is obtained by replacing each arc $\omega = \langle \overrightarrow{i, j} \rangle$ by its undirected counterpart, i.e., the edge $e = \langle i, j \rangle$. Figure 4.1 depicts a directed graph.

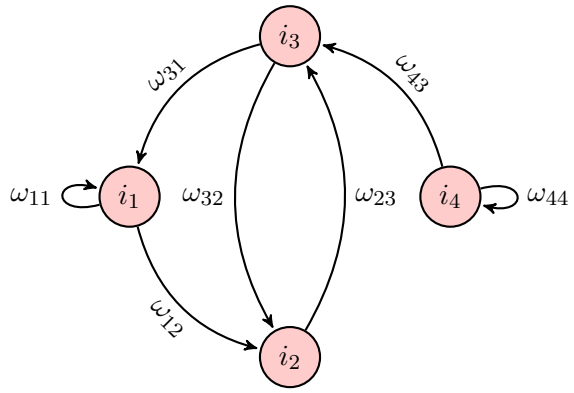


Figure 4.1: A directed graph with loops and cycles.

Definition 4.7: Cycle

A directed path (i_1, i_k) is an alternating sequence $[i_1, \omega_1, i_2, \omega_2, \ldots, i_{k-1}, \omega_{k-1}, i_k]$ of vertices and arcs with $\omega_l = \langle \overrightarrow{i_l, i_{l+1}} \rangle$ such that all vertices and all arcs are distinct. A directed cycle or simply, a cycle is a directed path (i_1, i_k) with $i_1 = i_k$.

In Figure 4.1, $[i_2, \omega_{23}, i_3, \omega_{32}, i_2]$ depicts a cycle. The arcs ω_{11} and ω_{44} are loops, which we would call self-cycles.

The Top Trading Cycle algorithm. We now introduce the Top Trading Cycle algorithm (TTC) which was first proposed by David Gale. It is a simple algorithm that leads to a non-empty core of the housing market with private endowments. The algorithm works as follows.

Step 1: Construct a digraph, D whose vertices are the individuals in I, i.e., $V = I = \{i_1, \ldots, i_n\}$. Each individual points at the agent who owns their most preferred house. Suppose ω_5 is the most preferred house of individual i_9. Then, draw the arc $\omega_5 = \langle \overrightarrow{i_9, i_5} \rangle$. If an agent's, say i_7, most preferred house is already owned by them, then the arc $\langle \overrightarrow{i_7, i_7} \rangle$ is a loop or self-cycle.

Because H is finite, and each vertex has outdegree exactly equal to 1, there must be at least one cycle (which can be a loop!). Moreover, each agent can be part of at most one cycle. Each agent in the cycle

is assigned the house they have been pointing at. The entire cycle, i.e., the agents (vertices) and the assigned houses (arcs) are removed from the market.

Step 2: Let $H_2 \subset H_1 \equiv H$ be the set of available houses after those removed from the market at the end of Step 1. Perform the same procedure on H_2. Each remaining agent now points at the owner of their most preferred house in H_2.

$\vdots$

Step k: Let $H_k \subset H_{k-1}$ be the set of available houses after those removed from the market at the end of Step $k-1$. Perform the same procedure on H_k. Each remaining agent now points at the owner of their most preferred house in H_k.

$\vdots$

STOP: The algorithm stops when every agent has been assigned to a house or there are no more available houses. Clearly, it halts in a finite number of steps.

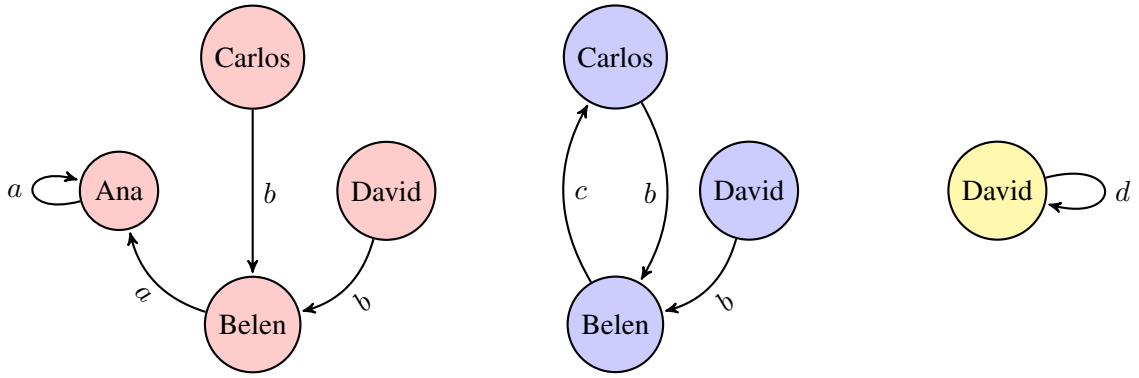
Consider first the following example to see how TTC works.

Example 4.4: TTC

Let $I = \{\text{Ana, Belen, Carlos, David}\}$ and $H = \{a, b, c, d\}$. The endowments and preferences are given by

$(\omega_{\text{Ana}}, \omega_{\text{Belen}}, \omega_{\text{Carlos}}, \omega_{\text{David}}) = (a, b, c, d)$;
$\mathsf{P}_{\text{Ana}} = a, c, d, b$;
$\mathsf{P}_{\text{Belen}} = a, c, d, b$;
$\mathsf{P}_{\text{Carlos}} = b, d, c, a$;
$\mathsf{P}_{\text{David}} = b, c, d, a$.

The algorithm yields the allocation $\mu(\text{Ana}) = a$, $\mu(\text{Belen}) = b$, $\mu(\text{Carlos}) = c$, $\mu(\text{David}) = d$. Note that μ is individually rational because all individuals consume at least their endowments. It is also Pareto efficient because each agent gets their top choice among the available houses.



Step 1: $H_1 = \{a, b, c, d\}$
Remove (Ana, a)

Step 2: $H_2 = \{b, c, d\}$
Remove (Belen, c) and (Carlos, b)

Step 3: $H_3 = \{d\}$
Remove (David, d)

Figure 4.2: TTC with four individuals and four houses.

Proposition 4.3

The Top Trading Cycle algorithm yields a matching that is Pareto efficient and individually rational.

Proof. The proof of Pareto efficiency is very similar to the one for Serial Dictatorship (Proposition 4.1). By contradiction, suppose there is a matching μ' that Pareto dominates μ, the outcome of TTC. All agents leaving the market in Step 1 of the algorithm obtain their top choices (among all houses); hence, their assignment in μ' must be the same as in μ. Because all agents leaving in Step 1 get their top choices in both μ and μ', agents leaving in Step 2 cannot get in μ', houses that, under μ, were assigned to agents leaving in Step 1. Hence, they must also be getting the same houses under μ' and μ. The same argument applies inductively for every step, yielding a contradiction.

To show that the resulting matching is individually rational it suffices to note that houses never leave the market prior to their owners. Because agents always leave the market with their most preferred house among the available ones, which includes their endowments, the house to which they are assigned will be at least as preferred as what they own. □

Next, we show that TTC not only produces Pareto efficient and individually rational matchings, but also that the matching belongs to the core. Furthermore, we shall show that the core contains a single matching, which is precisely the one generated by TTC. Therefore, TTC gives us both a way to prove that the core of a housing market is non-empty, and an algorithm to find the unique matching in the core.

Theorem 4.1: (Shapley and Scarf, 1974; Roth and Postlewaite, 1977)

The core of a housing market is nonempty and contains a unique matching, the one generated by the Top Trading Cycle algorithm.

Proof. Let μ be the matching obtained from TTC. We want to show that μ is in the core of the housing market. Let I_k generically denote the set of agents who leave the market in Step k of TTC, i.e., I_k is the cycle generated in Step k. Suppose on the contrary that there is a coalition C of agents that blocks μ, i.e., there exists a matching $\mu' \neq \mu$ such that $\mu'(i) \succeq_i \mu(i)$ for all $i \in C$ and $\mu'(i) \succ_i \mu(i)$ for some $i \in C$. Let $C' = \{i \in C \mid \mu'(i) \succ_i \mu(i)\} \subset C$, i.e., C' contains the agents in C, each of whom strictly prefers the house assigned to them under μ' to that under μ. Now let $j \in C'$ be the first one among the members of C' who is matched and leaves the market (with the house $\mu(j)$) at some step k of TTC. Clearly, the house $\mu'(j)$ has been removed from the market strictly before Step k, say $k^* < k$. Let $\mu'(j) = \omega_{i_1}$ so that $i_1 \in C$. Agent i_1 also has left the market in Step k^*. Furthermore, $i_1 \notin C'$, i.e., $\mu'(i_1) = \mu(i_1)$. Because i_1 has left the market in Step k^*, they belonged to the cycle

$$I_{k^*} = [i_1, \omega_{i_2}, i_2, \omega_{i_3}, \ldots, \omega_{i_m}, i_m, \omega_{i_1}, i_1].$$

The key of the proof is to show that agents $i_2, \ldots, i_m$ are also in $C \setminus C'$ (the same as i_1), which implies that everyone leaving the market in Step k^* is getting the same house under μ and μ'. In particular, because i_1 and i_m are the first and the last agents of cycle I_{k^*}, respectively, we have $\mu'(i_m) = \mu(i_m) = \omega_{i_1}$. So we have shown that $\omega_{i_1} = \mu'(i_m)$ where $i_m \in (C \setminus C') \cap I_{k^*}$, and at the same time, $\omega_{i_1} = \mu'(j)$ where $j \in C' \cap I_k$. The two statements contradict each other.

To show the uniqueness of μ, suppose on the contrary that there is a matching $\mu' \neq \mu$, and μ' is also in the core of the housing market. Let i be the first agent who leaves the market in TTC with $\mu(i)$, which is not the same as $\mu'(i)$. Suppose, without loss of generality, that $i \in I_k$. Hence, every agent in $I_1, \ldots, I_{k-1}$ gets the same house under μ and μ'. This implies that, under μ', every agent in I_k obtains a house of an agent who leaves the market at Step k or afterwards. Because under μ, agent i obtains their most favorite house among all

these houses, and $\mu(i) \neq \mu'(i)$, μ' must make agent i worse-off. So, agents in I_k would block μ' via μ, which is a contradiction to the fact that μ' is a core allocation. □

The proof of Theorem 4.1 is a bit involved, but the intuition is rather simple. It suffices to look at the execution of TTC. Consider any cycle that is obtained in the first step. All the individuals involved in this cycle will obtain their most preferred object. Therefore, any assignment that does not give these individuals their most preferred object cannot be in the core. All of these individuals will prefer the assignment they obtain through the cycle. Observe now that, in the cycle, any individual obtains the endowment of another individual who is also in the cycle.

To summarize, if an assignment is in the core, it must be that all the individuals assigned in the first step (i.e., in the first cycle) are assigned the same object as the one they get with TTC. It now suffices to repeat this reasoning with the agents who are assigned in a cycle found in step 2 of TTC. Those individuals can only be better off if we assign them an object that was assigned in step 1. This is not possible, so the best they can get is the assignment they obtain in step 2. It is not difficult to see that if the step 2 individuals get something different from their assignment under TTC (but individuals from step 1 are assigned their TTC assignment), then the assignment they obtain in step 2 satisfies conditions in the definition of core (cf. Definition 4.4). Repeating the procedure with the assignment found in steps 3, 4, ... we end up with a well-defined assignment that is in the core and such that any other assignment cannot be in the core.

4.3 Incentives in the housing market

We now analyze the incentive issues in the housing market. In particular we analyze mechanisms that use the Serial Dictatorship algorithm and the Top Trading Cycle algorithm.

Implementability of SD. Recall Example 4.1. The SD algorithm generates μ' under the priority order, $\pi =$ (Ana, Belen, Carlos, David). Then by Proposition 4.1, μ' is Pareto efficient. However, the proposition assumes that we know the preferences of the agents; otherwise, how would we be able to run the algorithm? We now analyze the incentive issues, i.e., if the agents participate in a mechanism that uses the SD algorithm, do they voluntarily report their true preferences? In other words, is the Serial Dictatorship mechanism strategy-proof?

Proposition 4.4

The Serial Dictatorship mechanism is strategy-proof.

Intuition. The formal proof is left as an exercise. The intuition is very simple. Consider Example 4.1, and fix the priority order, $\pi =$ (Ana, Carlos, Belen, David). Note that when it is Belen's turn to choose, the set of available houses to her is $\{c, d\}$, and this set of available alternatives does not depend on her preferences (but on those of Ana and Carlos). So, Belen cannot do better by mis-reporting her preferences. □

Implementability of TTC. Next, we analyze the incentive properties of TTC, i.e., whether in a mechanism that uses TTC all individuals report their preferences truthfully. *Why this question is important?* If we leave TTC to run in a decentralized fashion, i.e., individuals (with endowments) participate in voluntary exchanges, producing a core matching allocation requires a lot of coordination among the individuals, especially in a market with many objects. To understand this, suppose Ana's most preferred house is b which is owned by Belen. So, Ana is first required to find Belen. A trade can immediately take place if Ana owns Belen's most preferred house, say a. However, if Belen's most favorite house is $c \neq a$, then they have to look for the owner of c. As you can see that finding the owner of c is not enough to construct a trading cycle as the owner of c may

have their most favorite house which is different from c.

In an abstract model of house allocation, this delay caused by searching for appropriate trading partners may not be that outrageous. However, there are real-life situations wherein such delays may turn out to be exorbitantly costly. Think of the case of kidney transplants. As we shall see in details later (it is well-known, though) that a successful kidney transplant depends crucially on blood group and tissue group compatibility. For simplicity, think of an economy consisting of many households, each with only two members—household i comprises of A_i and B_i who may be wife and husband, mother and daughter, two siblings, etc. Suppose further that in each household, member A requires a kidney transplant and member B is a potential donor; however, kidneys of A member and B member are not compatible. So, a household i has to search for another household j so that the kidneys of A_i and B_j, and those of A_j and B_i are compatible. What are the odds in finding such compatible households? The possibility of such exchanges gets complicated very quickly if A_i can receive a kidney from B_j, but A_j cannot from B_i (Thinking of side payments from household i to household j? This is strictly prohibited in every country in the world except Iran!). In this case, households i and j require to search for another household k so that a three-way exchange (A_i receives from B_j, A_j receives from B_k, and A_k receives from B_i) is possible. It requires a lot of coordination and information to identify a trading cycle involving three or more households in a decentralized market. Hence, the solution is a *centralized clearinghouse*.

In the house (or any other object) allocation problem (similar to the college admission problem), a central clearinghouse in general runs algorithms that yield matching allocations (with some desirable properties, e.g. efficiency, individual rationality, fairness). However, to run an algorithm, as we have already seen, it requires to feed the preference list of each individual into a giant computer. What guarantees that individuals report their preferences truthfully? With any reported preference profiles, it is true that the giant computer will produce a matching, but it may be not the right one. Think of the very costly situation wherein the clearinghouse algorithm suggests, according to the reported preferences, a two-way kidney exchange between households i and j. However, B_i's kidney will not function in A_j's body (moreover, IMSS had incurred a huge cost, which implies a burden on the tax payers!). The Top Trading Cycle algorithm (or a modified version of it) is very compelling in situations starting from room allocation in student dormitories to kidney transplants—it is dominant strategy for all participants in a mechanism that uses TTC to report their preferences truthfully. We first prove the following result.

Lemma 4.1

Take an agent i and fix a profile of preferences P_{-i} for the other agents. Consider two preference relations of i, P_i and P'_i. Let k and k' be the steps in TTC at which agent i leaves the market while reporting P_i and P'_i, respectively. At step $k^* = \min\{k, k'\}$, the houses and agents remaining in the market are the same under both preferences.

Proof. The key to show this lemma is to note that whether i reports P_i or P'_i does not affect any of the cycles formed prior to i leaves the market. Assume, without loss of generality, that $k' \geq k > 1$, i.e., there is at least one step before i leaves the market while reporting P_i. The cycle formed in round 1 depends on the preferences reported by other agents, P_{-i}. Even if i points at an agent in the cycle, the only way for i to be part of the cycle (and leave the market) is that someone in the cycle also point at i, which does not occur (otherwise i would have left the market in Step 1). Therefore, whatever preferences i reports, P_i or P'_i, at the start of round k, as the set of cycles leaving the market in prior rounds is the same, the sets of remaining houses and agents are also the same. □

Theorem 4.2: (Roth, 1982b)

The Top Trading Cycle mechanism is strategy-proof.

Proof. Consider an agent i with true preferences P_i, a fixed profile P_{-i} of other agents, and alternative preferences P'_i of i. Let k and k' be the steps in TTC at which i leaves the market while reporting P_i and P'_i, respectively. We consider two cases.

First, assume that $k' \leq k$, i.e., the case in which i would leave the market at the same time or before by misreporting their preferences. At the beginning of round k', by Lemma 4.1, the sets of agents and houses in the market are the same under both P_i and P'_i. Note that under P'_i, agent i leaves the market with some house h', which is part of a cycle:

$$I_{k'} = [i_1, \omega_2, i_2, \omega_3, \ldots, \omega_i, i, h' = \omega_1, i_1],$$

in which $i_1, i_2, \ldots$ are all pointing at their favorite houses, under P_{-i}. The key is to note that the chain $(h' = \omega_1, i_1, \omega_2, , i_2, \omega_3, \ldots, \omega_i)$ will remain in the market until i chooses to close off the cycle, either by pointing at h' or by pointing somewhere else that eventually ends in h' (or any other house in the cycle). Hence, by reporting truthfully, i will point at their top choice in every subsequent rounds, and might get something better or eventually pick h' if it is the most preferred remaining house. In other words, i has no incentive to "close" the cycle before, and leave the market with h'.

Second, assume that $k < k'$, i.e., the case in which i would leave the market afterwards by misreporting preferences. Note that by reporting truthfully, i leaves the market at step k with the best house among all the remaining ones at the start of step k. Because the houses remaining at the start of round k' is a subset of the ones at round k, i has no incentive to misreport their preferences to leave afterwards. □

The next Theorem goes further and shows that, actually, TTC is the unique mechanism that satisfies the above properties. That is, there exists no other mechanism that is also strategy-proof, Pareto efficient and individually rational. We omit the technical proof which can be consulted in Robinson-Cortés (2021).

Theorem 4.3: (Ma, 1994)

An allocation mechanism is strategy-proof, Pareto efficient and individually rational if and only if it is the Top Trading Cycle mechanism.

4.4 House allocation with mixed endowments

In the housing market, there are situations where some individuals have endowments (private endowments) and some do not (public endowments). A common situation is that of student dorms on campuses. Many colleges and universities offer housing to their students, and because each year some students graduate and leave their dorms and new students arrive, we have a situation with a mix of private and public endowments:

- Private endowments: the students who were already on campus the previous academic year (existing tenants). The room they occupied the previous year is their private endowments (occupied houses).
- Public endowments: the rooms that are left vacant by the students who just graduated (vacant houses). The newly arrived students do not have any endowments (new applicants).

Abdulkadiroğlu and Sönmez (1999) were the first to analyze house allocation problems with mixed private-public endowments. In particular, they identify a few popularly-used algorithms that turn out to be inefficient. They then propose solutions to the inefficiency problem. We shall analyze such matching mechanisms in terms of examples (without going into much of their details).

4.4.1 Inefficient algorithms

Random Serial Dictatorship with Squatting Rights. The first matching mechanism that is extremely common is the so-called Random Serial Dictatorship with Squatting Rights (henceforth, RSDSR). This mechanism is or was used for undergraduate housing at Carnegie-Mellon, Duke, and Harvard among others. The idea is to run SD with random priority orders. However, prior to that, the existing tenants are given a choice whether they want to participate in the market or not. If an existing tenant chooses to opt out, they keep their house; otherwise they participate in the RSD mechanism but loses their endowment. The algorithm works as follows.

Step 1: Existing tenants decide whether they want to keep their houses. If they do so, they are matched with their endowments. Otherwise, the houses are added to the pool of vacant houses.

Step 2: The Serial Dictatorship algorithm is run under a random priority of the market participants (new applicants and existing tenants who have decided to participate).

The main concern with this algorithm is the following. If an existing tenant decides to participate in the mechanism, they run the risk of ending up with a house assignment that is less preferred to the one they are initially endowed with. Consequently, risk-averse existing tenants may prefer to opt out, which can lead to inefficiency.

Example 4.5

Let $I = \{\text{Ana, Belen, Carlos}\}$ and $H = \{a, b, c\}$. The endowments are given by $(\omega_{\text{Ana}}, \omega_{\text{Belen}}, \omega_{\text{Carlos}}) = (a, \varnothing, \varnothing)$. That is, Ana is the only existing tenant who owns house a (which is the occupied house), Belen and Carlos are the new applicants, and b and c are vacant houses. The preferences are given by:

$\mathsf{P}_{\text{Ana}} = b, a, c$;
$\mathsf{P}_{\text{Belen}} = b, c, a$;
$\mathsf{P}_{\text{Carlos}} = a, b, c$.

Run SD with priority order π =(Carlos, Belen, Ana). First suppose that Ana participates in the market. First Carlos gets to choose, and he will choose house a. Then, Belen chooses b, her most preferred one among the remaining houses $\{b, c\}$. So, Ana is left with house c. The matching is not individually rational. Clearly, Ana would be better-off keeping her endowment, house a. If Ana opts out of the market, then there are only two available houses, b and c. Carlos, first in the priority order, takes b. So, Belen is assigned house c. This matching is not Pareto efficient because it is dominated by the matching μ(Ana)$= b$, μ(Belen)$= c$ and μ(Carlos)$= a$.

Two popular algorithms, which are used for graduate housing at the University of Rochester and MIT, try to correct the shortcomings (that existing tenants may end up being worse-off) of RSDSR

Random Serial Dictatorship with Waiting List. The Rochester solution is called the Random Serial Dictatorship with Waiting List (henceforth, RSDWL). It runs SD with a random priority order, but at any step an individual can only take a house that is available. A house is available if it is either a vacant house in the public endowment or if it is left vacant by an existing tenant. For a new applicant, all houses are obtainable, and for an existing tenant, a house is obtainable if it is their house (i.e., endowment) or a house they prefer to their house. The algorithm works as follows.

Step 1: Draw a random priority order of all market participants (new applicants and existing tenants).

Step 2: The set of available houses is the set of vacant houses (i.e., those not currently owned by an existing tenant). The agent with the highest priority among those who have at least one obtainable house is assigned their most preferred available house. This agent is then removed from the market along with their newly assigned house. If the individual is an existing tenant, their endowment is added to the pool of available houses (if the endowment is not the most preferred one).

⋮

Step k: The set of available houses is constructed at the end of step $k-1$. The agent with the highest priority among the remaining individuals with at least one obtainable house is assigned their most preferred available house (and is removed from the procedure along with the newly matched house). If the individual is an existing tenant, their endowment is added to the pool of available houses (if the endowment is not the most preferred one).

⋮

STOP: The algorithm halts when there is either no remaining individual or when there is no available house left.

The algorithm is clearly individually rational. However, it may yield a Pareto inefficient matching.

Example 4.6

Let $I = \{\text{Ana, Belen, Carlos}\}$ and $H = \{a, b, c, d\}$. The endowments are given by $(\omega_{\text{Ana}}, \omega_{\text{Belen}}, \omega_{\text{Carlos}}) = (a, b, c)$. That is, there are three existing tenants, no new applicants, three occupied houses and one vacant house. The preferences are given by:

$\mathsf{P}_{\text{Ana}} = b, c, a, d$;
$\mathsf{P}_{\text{Belen}} = c, a, b, d$;
$\mathsf{P}_{\text{Carlos}} = a, d, c, b$.
Run SD with priority order π =(Ana, Belen, Carlos). In the first step, only the vacant house, d can be occupied. Neither Ana nor Belen wants it. Carlos takes d, and both Carlos and d are removed from the market. Moreover, now c becomes available as a vacant house. In the next step, we consider the remaining individuals, Ana and Belen. The only available house is house c, which is obtainable for both of them. Ana has the highest priority, so she takes c, now making her endowment, a, available. Also, Ana and c are removed from the market. In the final step, the available house is house a, which is obtainable for Belen. So she takes it. The final matching is given by μ(Ana)= c, μ(Belen)= a and μ(Carlos)= d.

It is not difficult to see that the above matching μ is individually rational. However, it is not Pareto efficient as it is Pareto dominated by another matching μ' wherein μ'(Ana)= b, μ'(Belen)= c and μ'(Carlos)= a.

MIT-NH4. The New House 4 (NH4) algorithm, which has been in use at MIT since the 1980s is another attempt not to make existing tenants worse-off relative to their endowments. The algorithm works as follows.

Step 1: Draw a random priority order of all market participants (new applicants and existing tenants).

Step 2: The first individual is tentatively assigned to their most preferred house among all houses, the second individual is assigned to their most preferred house among the remaining houses, and so on, until a squatting conflict occurs. A squatting conflict occurs if the requested house has an existing tenant for whom all the remaining houses are less preferred to their endowment. Thus, there is a conflicting individual who chose (earlier) the tenant's house.

Step 3: If a squatting conflict emerges, then:

- The existing tenant (the one with the conflict) is assigned their house.
- The tentative assignment of the conflicting individual is canceled as well as those of all the individuals who chose after the conflicting individual.
- The process starts again with the conflicting individual as first in the priority order.

STOP: The algorithm halts when there is no house or individual left. At this point, all tentative matchings are the final ones.

Example 4.7

Let $I = \{\text{Ana, Belen, Carlos, Diana, Eduardo}\}$ and $H = \{a, b, c, d, e\}$. The endowments are given by $(\omega_{\text{Ana}}, \omega_{\text{Belen}}, \omega_{\text{Carlos}}, \omega_{\text{Diana}}, \omega_{\text{Eduardo}}) = (a, b, c, d, \varnothing)$. That is, there are four existing tenants, one new applicant, four occupied houses and one vacant house. The preferences are given by:

$\mathsf{P}_{\text{Ana}} = c, d, e, a, b$;
$\mathsf{P}_{\text{Belen}} = d, e, b, c, a$;
$\mathsf{P}_{\text{Carlos}} = e, c, d, b, a$;
$\mathsf{P}_{\text{Diana}} = c, e, d, b, a$;
$\mathsf{P}_{\text{Eduardo}} = d, e, c, a, b$.

Take the priority order π =(Ana, Belen, Carlos, Diana, Eduardo). In Step 1, Ana is tentatively assigned house c (a is vacant), Belen is tentatively assigned house d (b is vacant), and then Carlos is tentatively assigned house e (c is vacant). When it is Diana's turn, there is a conflict: her house, d, and all the houses she prefers to her endowment, c and e, are taken. The person who is tentatively assigned her house is Belen; so Belen is the conflicting individual. So we cancel Belen's as well as Carlos' tentative matches (because Carlos decided after Belen), and house d is assigned to Diana. Both Diana and house d are removed from the market.

In Step 2, we start again with Belen. The next best alternative for her among the available houses $\{a, b, e\}$ for her is house e, which she gets (temporarily). When it is Carlos' turn, again we have a conflict: Ana is the conflicting individual. So, we cancel Ana's temporary assignment (as well as Belen's), and house c is assigned to Carlos. Both Carlos and house c are removed from the market.

In Step 3, we start with the conflicting individual, Ana. She gets e, her best choice among the available houses $\{a, b, e\}$. Belen is assigned to her endowment, b, and Eduardo, to a. The algorithm halts. The final matching is given by μ(Ana)$= e$, μ(Belen)$= b$, μ(Carlos)$= c$, μ(Diana)$= d$ and μ(Eduardo)$= a$.

Although the algorithm yields an individually rational matching μ, it is Pareto dominated by another matching μ' (not necessarily Pareto efficient) wherein μ'(Ana)$= c$, μ'(Belen)$= b$, μ'(Carlos)$= e$, μ'(Diana)$= d$ and μ'(Eduardo)$= a$.

4.4.2 Efficient mechanisms

You Request My House-I Get Your Turn. The problem with the MIT-NH4 algorithm is that whenever there is a squatting conflict, the existing tenant does not get anything better than their endowment. Abdulkadiroğlu and Sönmez (1999) propose a modification of the MIT-NH4 algorithm—termed as You Request My House-I Get Your Turn (YRMH-IGYT). The difference between MIT-NH4 and YRMH-IGYT lies in that how a squatting conflict is resolved in the latter. In MIT-NH4, a conflict emerges when some agent has chosen the house of an existing tenant, but the existing tenant did not yet get the chance to choose. So, they are stuck with their endowment. Instead of writing the YRMH-IGYT algorithm, we point out the main modifications over MIT-NH4.

- When a squatting conflict emerges, if the existing tenant has already been temporarily assigned a house (different from their endowment), do not disturb the process, and proceed. By contrast, if the existing tenant did not yet get the chance to choose, then cancel the temporary assignment of the conflicting individual and all the subsequent ones. Insert the existing tenant before the conflicting individual in the priority order, and proceed. Why is it called YRMH-IGYT?
- A cycle, say $[i_1, \omega_2, i_2, \ldots, \omega_k, i_k, \omega_1, i_1]$ (including a loop) may form. In this case, remove all the individuals in the cycle after assigning them the house they pointing at, and we go ahead with the procedure.

Example 4.8

Consider the set up of Example 4.7. Run YRMH-IGYT with the priority order π =(Ana, Belen, Carlos, Diana, Eduardo).

In Step 1, Ana, first in the priority order, chooses her most preferred house, c. But c has an existing tenant, Carlos, who has not yet chosen any house. This is a squatting conflict, so we cancel Ana's tentative match, and move Carlos above Ana. The order of individuals is now π'=(Carlos, Ana, Belen, Diana, Eduardo).
In Step 2, according to the priority order, Carlos chooses e which is a vacant house, then Ana chooses c. When it is Belen's turn, there is a conflict. She demands Diana's endowment d, but Diana has not chosen yet. So, we cancel Belen's assignment, and move Diana ahead of Belen. The new priority order is π'=(Carlos, Ana, Diana, Belen Eduardo).
In Step 3, we start with Diana. She chooses her endowment d. This is a loop. So, Diana is assigned d, and both of them are removed from the market.
In Step 4, now it is Belen's turn. She picks b. Again it is a loop. So, Belen is given her own house, and both of them are removed from the market.
In Step 5, it is now Eduardo's turn to pick, and he chooses a. Ana is a's tenant, but she already has a tentative assignment, c. So, there is no conflict.
There are mo more claims. The algorithm halts. The final matching is μ' in Example 4.7, i.e., μ'(Ana)= c, μ'(Belen)= b, μ'(Carlos)= e, μ'(Diana)= d and μ'(Eduardo)= a.

Theorem 4.4: (Abdulkadiroğlu and Sönmez, 1999)

A matching mechanism that uses the YRMH-IGYT algorithm is Pareto efficient, individually rational, and strategy-proof.

In fact, Abdulkadiroğlu and Sönmez (1999) show that YRMH-IGYT is equivalent to a version of TTC (modified for mixed endowments). A matching mechanism that uses the modified TTC is also Pareto efficient, individually rational, and strategy-proof (see Abdulkadiroğlu and Sönmez, 1999, Propositions 1, 2, Theorem 2).

Modified Top Trading Cycle. The modified TTC is constructed in the following way. All individuals and all houses (occupied and vacant) are the vertices of a directed graph. Fix a priority order π. Each agent points at their most preferred house (draw an arc with an individual as its tail, and with their most preferred house as its head), each occupied house points at its existing tenant (draw an arc with an occupied house as its tail, and the owner as its head), and each vacant house points at the first individual in the priority order (draw an arc with each vacant house as its tail, and with the first individual in the priority order as its head). Remove from the market the cycles (there must be at least one) by assigning each agent in the each cycle the house they are pointing at. Repeat the same process with the remaining individuals and houses.

Example 4.9

Consider the set up of Example 4.7. Run the modified TTC with the priority order π =(Ana, Belen, Carlos, Diana, Eduardo). If you draw so many arcs (from agents to houses and vice versa), the figure looks messy, and hence, identifying a cycle is visually difficult (see Haeringer, 2017, Figures 11.5, 11.6, 11.7). Instead, do the following equivalent process. Start with the first agent in the priority order to find a cycle. Remove the cycle. Then start with the second individual to find a cycle; remove the cycle, and so on until the algorithm halts.
In the first round, Start with Ana to find a cycle:
Ana $\rightarrow$ c $\rightarrow$ Carlos $\rightarrow$ e $\rightarrow$ Ana
House e, the vacant house pointed at Ana, the first agent in π. Assign house c to Ana and house e to Carlos, and remove them from the market.
In the second round, Start with Belen to find a cycle:
Belen $\rightarrow$ d $\rightarrow$ Diana $\rightarrow$ d
Assign house d to Diana, and remove them from the market
In the third round, Start again with Belen to find a cycle:
Belen $\rightarrow$ b $\rightarrow$ Belen
Finally in the fourth round, only Eduardo and house a are left. Match them. The process ends here. The final matching is μ' in Example 4.8, i.e., μ'(Ana)= c, μ'(Belen)= b, μ'(Carlos)= e, μ'(Diana)= d and μ'(Eduardo)= a.

Given a priority order, YRMH-IGYT and the modified TTC produce the same matching outcome which is Pareto efficient, individually rational and strategy-proof. Note that the final matching depends on the priority order (i.e., two distinct orders may yield two distinct allocations) because a vacant house at the beginning of any round points at the first agent in the priority order.

4.5 Application I: Kidney exchange

Getting a kidney transplant is the preferred treatment for people suffering from acute kidney failure. Transplanted kidneys come from both deceased and living donors. Of the 17,107 kidney transplants that took place in the United States in 2014, 11,570 (67.6%) came from deceased donors and 5,537 (32.4%) came from living donors. However, there is a worldwide shortage of kidneys. In 2016, over 120,000 people were waiting for a lifesaving organ transplant in the United States. Of these, more than 100,000 were waiting for kidneys. The median patient waits over 3.5 years to receive a kidney. In every country in the world, with the exception of Iran, it is illegal to buy and sell human kidneys.

Typically, living donors are relatives or closely-related people who are willing to donate one of their kidneys to a loved one. However, despite the good intentions, wishing to donate a kidney is sometimes not enough. For a kidney donation to be successful, the blood and tissue types of the donor and the recipient need to be compatible. One of the most successful applications of market design to date has been to increase the supply of kidneys from living donors by performing kidney exchanges. The first kidney exchange in the world was made in 1991 in South Korea. In Europe, the first kidney exchange was made in Switzerland in 1999. In the United States, it was in 2000 in Rhode Island.

As of January, 2016 in the United States, there are a bit more than 100,000 people waiting for a kidney transplant. Each year, nearly 4,000 patients become too sick to receive a transplant, and almost 5,000 patients die waiting for a kidney. An alternative (not close substitute of transplant!) is dialysis which costs more than 80,000 USD per year. Hence, maximizing the number of transplants can save lives (and money).

In a two-way pairwise kidney exchange, two patients who are incompatible with their respective donor exchange kidney donors. That is, Ana's donor Albert gives his kidney to Bernardo; in exchange, Ana receives the kidney Belen, Bernardo's donor. In this section, we shall apply some of the tools that we have learnt from the house allocation problem to the problem of how to design kidney exchanges. We shall also learn additional tools to design optimal pairwise exchanges.

4.5.1 Blood and tissue type compatibility

Humans may have one of four different ABO blood-types: O, A, B, or AB. As far as blood-types are concerned, everyone can donate or receive a kidney from someone with the same blood-type, but not necessarily so across blood-types. Figure 4.3 illustrates blood-type compatibilities. People with blood-type O may donate a kidney to anyone, but cannot receive a kidney from someone with a different blood-type. People with blood-type A or B may donate a kidney to AB's, but may receive a kidney only from O-types. And people with blood-type AB cannot donate a kidney to someone with a distinct blood-type, but may receive a kidney from anyone. Around 41.2% of the worldwide population has blood-type O, 29.4% has A, 23.12% has B, and 6.2% has AB. Interestingly, the distribution of ABO blood-types varies across countries and ethnicities.

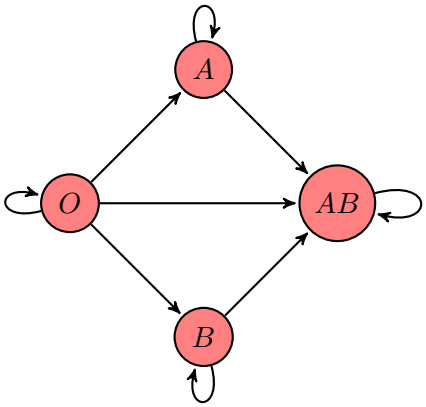


Figure 4.3: Blood-type compatibility.

If one person wishes to donate a kidney to another person, in addition to their blood-types being compatible, it is required to have tissue type compatibility. When two people share the same **H**uman **L**eukocyte **A**ntigens (abbreviated as HLA), they are said to be a "match", that is, their tissues are immunologically compatible with each other. HLA are proteins that are located on the surface of the white blood cells and other tissues in the body. There are three general groups of HLA, they are HLA-A, HLA-B and HLA-DR. There are many different specific HLA proteins within each of these three groups (e.g. 59 different HLA-A proteins, 118 different HLA-B and 124 different HLA-DR). Each of these HLA has a different numerical designation, for example, you may have HLA-A1, while some one else might have HLA-A10.[2] A set of HLA-A, HLA-B and HLA-DR is called a **haplotype**. For example, a haplotype is $\{\text{A3}, \text{B14}, \text{DR10}\}$. Each human in general has two HLA haplotypes, and each human being inherits one haplotype from each of the two parents. So, two siblings have a probability of 25% of being an exact "HLA match". After an HLA match, a patient-donor pair is required to undergo a "crossmatch" test. The crossmatch is a test which determines if the recipient has antibody against the potential donor. Antibody is a protein, present in the blood serum, which could injure the donor's cells by attacking the HLA. The antibody will only injure the donor's cells if it is specific for the donor's particular HLA. Not everyone has antibody against HLA. However, with the advancement of medicines in this field, a small degree of tissue-type incompatibility can be rectified by *immunosuppressants*, which are drugs that diminish the body's

[2]Source: *Kidney Transplantation: Past, Present, and Future*.

ability to reject the foreign organs.

Blood-type and tissue-type compatibilities have been used in the design of mechanisms to allocate donated kidneys. For instance, when a cadaveric kidney (donated by a deceased patient) becomes available for transplantation, the priority of each patient in the waiting list is typically determined by factors including the blood-type, HLA type-compatibility, time spent on the waiting list, etc. Similarly, the effects of distinct design choices may depend on the structure of the ABO blood-type compatibilities. For example, a proposed allocation rule known as an indirect exchange program aims to increase the amount of kidneys by exchanging kidneys from living donors for priorities in the waiting list. That is, a living donor who is incompatible with their intended recipient donates their kidney, and, in exchange, their intended recipient receives a higher priority for the next compatible kidney. However, it has been observed that such indirect exchange programs can harm type O patients who have no living donors and are currently in the waiting list.

4.5.2 Kidneys as houses, patients as owners

Pairwise kidney exchange may be seen as an application of the *house allocation problem with mixed endowments* the we have analyzed in Section 4.4. We can view kidneys as "houses". Incompatible patient-donor pair as "existing tenants", each of whom *owns a house*. And patients who require kidneys and have no donor are "new applicants" who do not *own a house* initially. Finally, cadaveric kidneys and kidneys donated from altruistic donors are "unoccupied houses". Patients have a preference ranking over kidneys which may depend on the blood and tissue-type compatibilities, location of the kidney, and any other factor that may affect the probability of a successful transplant, such as donor age, kidney size, medical history, etc. Importantly, patients may have heterogeneous preferences over otherwise compatible kidneys.

We would thus use an algorithm similar to Top Trading Cycle for the pairwise kidney exchange problem. However, there are important differences with the problem of house allocation with existing tenants which are pointed out by Roth, Sönmez, and Ünver (2004). In a kidney exchange problem, there are two *types* of kidneys and two *types* of patients.

1. Some patients come with a donor (and thus a kidney). In other words, they have "private endowments". But there are also patients who do not have a living donor. Those patients are theoretically only eligible for cadaveric kidneys, as they cannot theoretically participate in a pairwise exchange.
2. Some kidneys are "held" by some patients, namely the patients who have a living donor. The other kidneys are those coming from cadaveric donors. However, there are uncertainties regarding the availability of such kidneys. Those that come from cadaveric donors differ from those coming from living donors in that they cannot be considered kidneys that are available (or not yet assigned). In other words, kidneys from deceased donors are not part of the public endowments. So, for a patient, there are two options:
   (a) Obtain a kidney from a living donor.
   (b) Join the waiting list (or remain on it).

To sum up, the problem of designing an exchange procedure for kidneys is to combine, at the same time,

- trades among patient-donor pairs, and
- the management of the waiting list.

4.5.3 Trading cycles and chains

Consider the following example.

Example 4.10

Consider the set of patients, $I = \{\text{Angel, Carlos}\}$ and the set of donors, $N = \{\text{Ana, Carolina}\}$. The patient-donor pairs are $A =$(Angel, Ana) and $C =$(Carlos, Carolina). This means Angel and Ana are not compatible with each other, and so are Carlos and Carolina.

$\mathsf{P}_{\text{Angel}}$= Carolina, w;
$\mathsf{P}_{\text{Carlos}}$= Ana, w.
Each patient has a (strict) preference relation over the available kidneys; otherwise, he joins the "waiting list", denoted by w. Now each donor *points at* her paired patient (as if houses pointing at their owners, but the owners cannot live in the houses they own), and each patient *points at* his most preferred kidney. In this case, we have the following cycle:
Ana $\rightarrow$ Angel $\rightarrow$ Carolina$\rightarrow$ Carlos $\rightarrow$ Ana
Now suppose an alternative preference relation of Carlos, $\mathsf{P}'_{\text{Carlos}}$= w, Ana, i.e., Carlos prefer to joint the waiting list rather than having Ana's kidney. In this case, we have a chain:
Ana $\rightarrow$ Angel $\rightarrow$ Carolina$\rightarrow$ Carlos $\rightarrow$ w
Thus, TTC cannot be applied here in a straight forward way. Moreover, as we shall see later that a patient-donor pair can be part of multiple chains (as several patients can be assigned places in the same waiting list) as opposed to the fact that it can form part of at most one cycle.

The Top Trading Cycle and Chains (TTCC) algorithm. Roth et al. (2004) analyze a modified version of TTC in which they allow for patients to have preferences over the set of currently available kidneys and a place in the waiting list, denoted by w. The algorithm is a multi-step procedure, where all steps are identical. For each $h \geq 1$, proceed as follows.

Step h.1: Each patient points at their most preferred among acceptable kidney among the available kidneys. If, for a patient, none of the available kidneys is acceptable, then they points at the waiting list option. Each kidney points at its paired patient (e.g., if Angel and Ana are a patient-donor pair, then the kidney of Ana points at Angel).

Step h.2: If there is one or more cycles, proceed to the exchange as follows. For each cycle, each patient in the cycle is allocated the kidney they are pointing at. All the patients and kidneys involved in a cycle are removed from the market. Then proceed to Step $h + 1$. If there is no cycle (i.e., there is a chain of the form $[i_1, k_2, i_2, \ldots, k_m, i_m, w]$ in that patient i_m is pointing at the waiting list rather than the kidney of donor 1, k_1), go to Step h.3.

Step h.3: Select one chain, and allocate the kidneys in the following way:

- The last patient in the chain is added to the waiting list.
- The other patients in the chain (if any) are assigned the kidney they are pointing at.

For all patients involved in the selected chain, the assignment is final. A *chain selection rule* determines whether the selected chain is removed from the problem. Then go to Step $h + 1$.

STOP: The algorithm halts when all patients have either been assigned a kidney or added to the waiting list.

A couple of comments on TTCC are called for. Roth et al. (2004) show that, as long as there is a finite number of patient-donor pairs, there must always exist either a cycle or a chain. Whether there is a cycle or a chain, at each step of the algorithm, some patient-donor pair(s) will be removed from the market. Therefore, at each step, the pool of remaining patients shrinks. Eventually, we shall have all patients removed and TTCC halts. Second, if there are multiple chains, the algorithm does not specify how a particular chain is selected, and how it is removed from the market. In what follows, we analyze several *chain selection rules*.

4.5.4 Chain selection rules

Consider the following example.

Example 4.11

Consider the set of patients, $I = \{\text{Angel, Bernal, Carlos}\}$ and the set of donors, $N = \{\text{Ana, Belen, Carolina}\}$. The patient-donor pairs are A =(Angel, Ana), B =(Bernal, Belen) and C =(Carlos, Carolina). This means Angel and Ana are not compatible with each other, and so on.

P_{Angel}= w, Belen, Carolina;
P_{Bernal}= Ana, w, Carolina;
P_{Carlos}= Ana, w, Belen.
Under the above preferences, as depicted in Figure 4.4, we have two chains:
[Belen, Bernal, Ana, Angel, w] and [Carolina, Carlos, Ana, Angel, w].
So, patient-donor pair A is clearly part of two different chains. Thus, the algorithm must select one of the two chains.

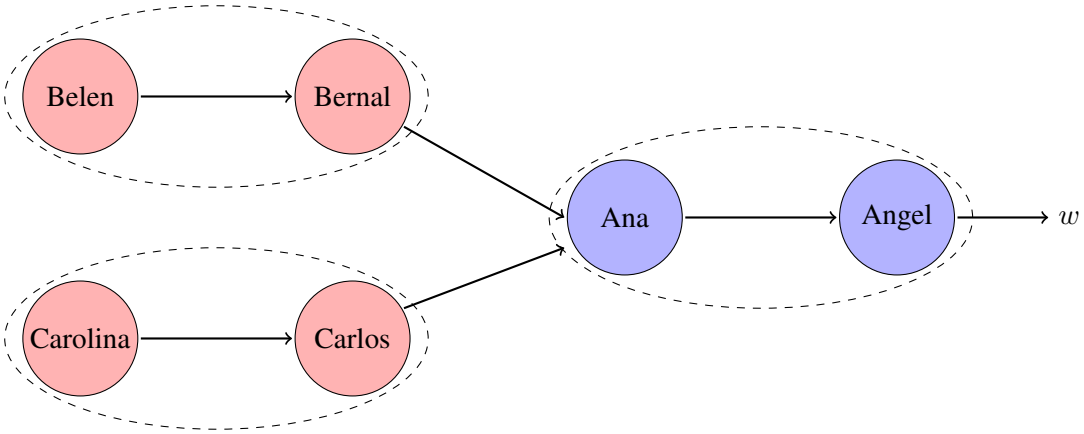


Figure 4.4: Two overlapping chains.

Another question we have to address when treating with chains is whether we should remove from the problem the patients and the kidneys involved in a selected chain. For example, consider the chain starting with Belen in Figure 4.4, and suppose that the algorithm selected it. The algorithm states that once this chain is selected, Bernal will obtain Ana's kidney and Angel will join the waiting list, independently of what happens at the following steps (i.e., the algorithm states that this assignment is final for Bernal and Angel). However, note that in this matching we do not specify what to do with Belen's kidney. There are two solutions:

1. It is assigned to someone on the waiting list. In this case, because Belen's kidney is no longer available (i.e., for any patient that is still in the pool of patients without an assigned kidney), we can remove Belen and all the other patient-donor pairs involved in the chain.

2. We keep Belen's kidney available. In this case, we do not remove the chain from the problem.

To see the difference between the two possible solutions, consider the following example.

Example 4.12

We keep the same specifications of Example 4.11, except that we change the preference relation of Carlos to P'_{Carlos}= Ana, Belen, w. The "arrow drawing" phase will yield the same figure as Figure 4.4. If

we select the chain staring from Belen, and remove everybody involved in this chain, then Carlos does not have any possibility of having a transplant. By contrast, if we keep the chain in the problem, then in the second step, Carlos will point at Belen's Kidney (which is now available to him). As depicted in Figure 4.5, there will be new chain
[Carolina, Carlos, Belen, Bernal, Ana, Angel, w]
which would open up the possibility of a transplant for Carlos.

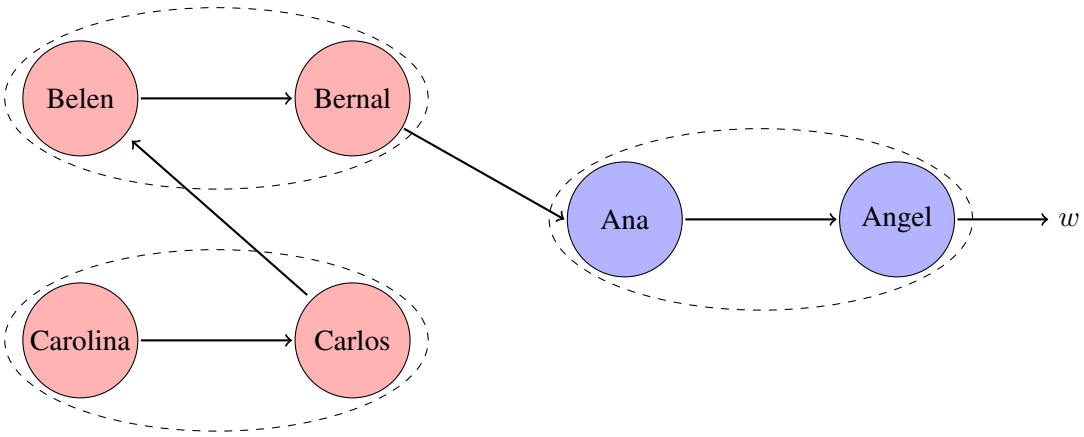


Figure 4.5: Two overlapping chains.

Roth et al. (2004) propose and analyze different selection rules:

(a) Choose the smallest chains, and remove from the problem the patients and kidneys in those chains once the assignment is determined.

(b) Choose the longest chain, and remove from the problem the patients and kidneys in that chain (pick one chain at random if it is not unique).

(c) Choose the longest chain, and keep in the problem the patients and kidneys in that chain (pick one at random if it is not unique).

(d) Choose the chain that starts with the highest-priority patient, and remove from the problem the patients and kidneys in that chain.

(e) Choose the chain that starts with the highest-priority patient, and keep in the problem the patients and kidneys in that chain.

(f) Prioritize patient-donor pairs so that pairs with a type O donor have a higher priority (than the pairs whose donor is not of type O). Then choose the chain that starts with the highest-priority pair. If the starting pair in the chain has a type O donor, then remove from the problem the patients and kidneys in that chain. Otherwise keep in the problem all the patients and donors that are in the chain.

Given a pairwise kidney exchange problem, a matching is Pareto efficient if there is no other matching that is weakly preferred by all patients and donors and strictly preferred by at least one patient-donor pair. Roth et al. (2004) analyzes whether TTCC selects a *Pareto efficient* matching.

Theorem 4.5: (Roth, Sönmez, and Ünver, 2004)

Consider a chain selection rule such that any chain selected at a nonterminal round remains in the procedure, and thus the kidney at its tail remains available for the next round. The Top Trading Cycle and Chains algorithm yields a Pareto efficient matching.

4.6 Application II: School choice revisited

In Chapter 3, we have analyzed the school choice problem as an application to the college admission problem. Because schools are assumed not to have preferences over students, they can as well be treated as objects (that provide educational services). Therefore, the school choice problem can also be treated as an application to the housing market. In order to obtain an efficient matching, we can also use a modified version of the Top Trading Cycle algorithm (school choice TTC).

The principal difference with the housing market is that schools can enroll many students up to their capacities, and modifications to the standard TTC must be made accordingly. We do the following. Suppose a school s has capacities $q_s > 1$. We would treat this school as q_s copies of the same school. In the pointing phase of the algorithm if a student i and a school s are in a cycle, then we shall remove the student and one copy of the school from the market. This is equivalent to keeping the school s in the market, but reducing its capacity by 1, i.e., $q_s - 1$ for the subsequent pointing phases. Otherwise, the algorithm is identical to the studied in Section 4.2. The school choice TTC halts when all students have been removed from the market or when all schools have remained with 0 capacity.

Two equivalent versions of the algorithm will be considered. We only sketch how the pointing phases work. For details, see Haeringer (2017, Section 13.2.4). The main ingredients of the first version are the following.

- Each student points at their most favorite school. Each school points at its highest-priority students.
- Once a student and a school are in a cycle (there is at least one at each step of the algorithm), remove the student from the market after assigning them to the school. By contrast, let the school remain in the market after reducing its capacity by 1.

We reconsider the set up of Exercise 3.2 in Chapter 3.

Example 4.13: School choice TTC: version 1

Consider a school choice problem with $I = \{\text{Ana, Belen, Carlos, Daniel}\}$ and $S = \{a, b, c\}$ with capacities $(q_a, q_b, q_c) = (2, 1, 1)$. The preferences of the students are given by

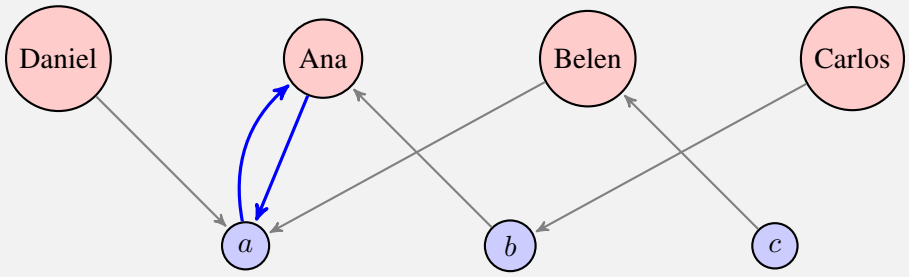
$\mathsf{P}_{\text{Ana}} = a, b, c,$ Ana;
$\mathsf{P}_{\text{Belen}} = a, b, c,$ Belen;
$\mathsf{P}_{\text{Carlos}} = b, a, c,$ Carlos;
$\mathsf{P}_{\text{Daniel}} = a, c, b,$ Daniel.

On the other hand, schools' priority lists are given by

$\pi_a =$ Ana, Carlos, Daniel, Belen;
$\pi_b =$ Ana, Belen, Daniel, Carlos;
$\pi_c =$ Belen, Carlos, Daniel, Ana.

In Step 1, there is one cycle

$$\text{Ana} \to a \to \text{Ana}.$$



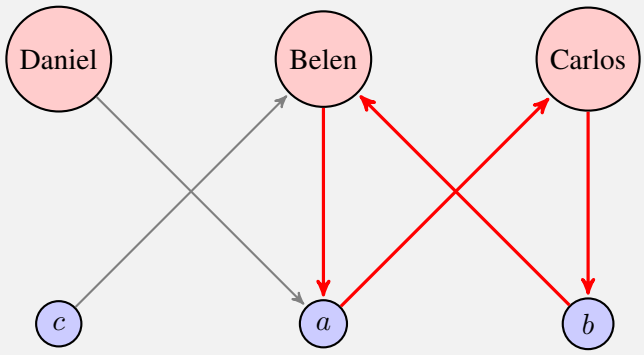
So, remove Ana after assigning her to school a.

At the beginning of Step 2, the remaining set of students is $I' = I \setminus \{\text{Ana}\} = \{\text{Belen, Carlos, Daniel}\}$. The set of schools remains the same, but the capacity vector is updated to $q' = (1, 1, 1)$. The priority lists are now updated to

$\pi'_a =$ Carlos, Daniel, Belen;
$\pi'_b =$ Belen, Daniel, Carlos;
$\pi'_c =$ Belen, Carlos, Daniel.

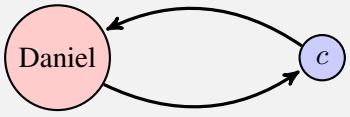
In this step, there is one cycle

$$\text{Carlos} \to b \to \text{Belen} \to a \to \text{Carlos}.$$



Both schools a and b now have exhausted their capacities. So, assign Carlos to school b and Belen to school a, and remove them from the market.

In Step 3, There remain only one student, Daniel and only one school, c with capacity 1. So, they are assigned to each other.



The final match is given by $\mu(a) = \{\text{Ana, Belen}\}$, $\mu(b) =$ Carlos and $\mu(c) =$ Daniel.

Note that Ana, Belen and Carlos are matched to their top choices, whereas Daniel assigned to his second choice. Thus, the matching is efficient. It is also easy to verify that the mechanism is also strategy-proof. However, note the the above matching is not stable. Clearly, it is individually rational and not wasteful, but it does not eliminate justified envy. Notice that Daniel prefers school a to school c (his current match). On the

other hand, school a has Belen who is lower-ranked than Daniel in this school. So, Daniel and school a form a blocking pair.

In TTC, drawing arrows may be confusing in that it may mean that both sides of the market have preferences. But schools do not have preferences over the students, they merely have priority lists. An equivalent version of school choice TTC is similar to TTC in the case of house allocation with existing tenants. So, we shall consider a school choice TTC wherein only the students point to the schools. The highest-priority students in a school are considered as if they own one seat apiece in that school (as existing tenants). So, in the arrow drawing phase, if student i wants a seat in a given school which is owned by student j, we would draw an arrow from i to j. The subtle difference here is that if a particular student is the top priority student of more than one schools, say of three schools, then this student initially owns three seats in three different schools. In the algorithm, whenever cycles are removed, the ownerships are updated among the remaining students and schools.

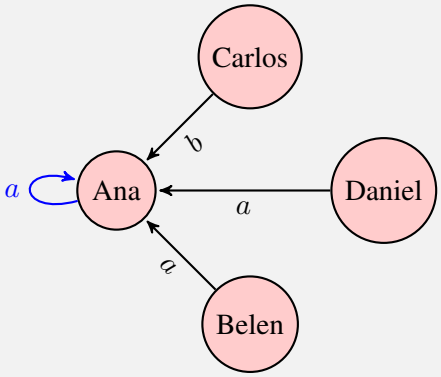
Example 4.14: School choice TTC: version 2

Consider the same specifications of Example 4.13. The ownership structure along with the preferences of the students are given by

$\mathsf{P}_{\text{Ana}} = a,\ b,\ c,$ Ana;
$\mathsf{P}_{\text{Belen}} = a,\ b,\ c,$ Belen;
$\mathsf{P}_{\text{Carlos}} = b,\ a,\ c,$ Carlos;
$\mathsf{P}_{\text{Daniel}} = a,\ c,\ b,$ Daniel.

Because Ana is the highest-priority student in both schools a and b, she owns one seat in each school. On the other hand, Belen owns one seat in school c. The ownerships are indicated by red letters.

In Step 1, Ana points at herself because she owns a seat in her most favorite school (a loop), and all the rest point at Ana.



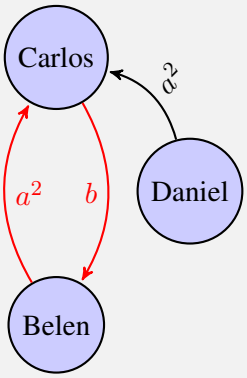
Step 1: $S_1 = \{a^1,\ a^2,\ b,\ c\}$
Remove (Ana, a^1)

Call the two copies of school a, a^1 and a^2. because there is a cycle: [Ana, a, Ana], assign Ana to a^1, and remove them from the market.

At the beginning of Step 2, update the ownership and capacity structures.

$\mathsf{P}_{\text{Belen}} = a^2,\ b,\ c,$ Belen;
$\mathsf{P}_{\text{Carlos}} = b,\ a^2,\ c,$ Carlos;
$\mathsf{P}_{\text{Daniel}} = a^2,\ c,\ b,$ Daniel;
$(q_{a^2},\ q_b,\ q_c) = (1,\ 1,\ 1)$.

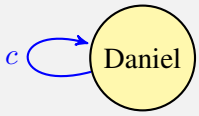
Belen is second-ranked by school a after Ana, and hence, Belen now owns a seat in a as well. Belen points at Carlos who owns a seat in school a, her top choice, Carlos points at Belen who owns a seat in school b, her top choice, and Daniel points at Carlos who owns a seat in a, Daniel's top choice.



Step 2: $S_2 = \{a^2, b, c\}$
Remove (Belen, a^2) and (Carlos, b)

There is a cycle: [Carlos, b, Belen, a^2, Carlos]. Assign Carlos to school b and Belen to school a, and remove them from the market because schools a and b have now exhausted their respective capacities.

In Step 3, There is only one student, Daniel and one school, c with capacity $q_c = 1$.



Step 3: $S_3 = \{c\}$
Remove (Daniel, c)

There is a trivial loop: [Daniel, c, Daniel]. Assign Daniel to school c and remove them from the market. There are no more students left and there are no schools with excess capacity. The algorithm stops. The final match produced is the same as in Example 4.13.

To summarize, the school choice TTC yields efficient matching, but the matching may be unstable. Moreover, the associated mechanism is strategy-proof.

4.6.1 The Boston school match

Boston's school district consists of over 60,000 students between kindergarten and twelfth grade in almost 140 schools. School assignments essentially take place in grades K, 1, 6, and 9. Each year, there are on average about 4000 students entering each of those grades. In Boston, the priorities for each school are constructed as follows, in this order:

1. The students with the highest priority at a school are those who have an older sibling attending that school.

2. Next are the students who live within walking distance of the school, the walk zone, and these zones are defined by the Boston Public Schools (BPS). Those students have priority over half of the seats offered by that school. For instance, if a school can admit 100 students, the students in the walk zone have higher

priority than the other students for only 50 seats. For the remaining 50 seats, there is no higher priority granted to the student living near the school compared to the other students.

3. Last are all the other students.

Traditionally, the Boston school match has used the Immediate acceptance algorithm. Clearly, truthful reporting of preferences has been an issue. Similar mechanisms have been used in other U.S. cities, e.g. Minneapolis. Glazerman and Meyer (1994) document that

> *It may be optimal for some families to be strategic in listing their school choices. For example, if a parent thinks that their favorite school is oversubscribed and they have a close second favorite, they may try to avoid "wasting" their first choice on a very popular school and instead list their number two school first.*

It has been noted that, In Boston, some parents understood the matching mechanism well and could manipulate at low costs, whereas other parents found it difficult to strategize over the mechanism. When a reform of the Boston school match mechanism was suggested by Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005b), the BPS sought to modify the match mechanism to make it strategy-proof with the objective of "leveling the playing field" in order to shed advantages for the sophisticated group of parents. In 2004, the choice then boiled down to choosing between stable and efficient school matches. Once this is decided the choice of algorithm was clear—it was either the Deferred Acceptance for school choice or the school choice TTC.

One interpretation of the school choice TTC is that, given the priority listings of schools, students can "trade" priorities among themselves which is similar to trading houses in the problem of house allocation with existing tenants. If students are allowed to do so, efficiency is the natural choice. By contrast, if priorities are not tradable, i.e., students have no ownership of their enrollments, then stability is the natural choice. Although the task force of the BPS advocated for the Top Trading Cycle algorithm, it was the Deferred Acceptance algorithm that was eventually adopted, in 2007.

4.6.2 The New York City school match

The case of New York City is quite different from that of Boston for several reasons. First, the problem is of a much larger scale. There are over a million students attending public schools in New York, with about 90,000 students entering one of the 500 different academic programs offered by public high schools. Unlike Boston, the reform of school choice in New York did not start at the same time the BPS started the reform. Instead, several people at the New York City Department of Education (NYCDOE) were aware of the National Resident Matching Program, and wondered if it could be adapted to New York City.

The procedure in place in New York was quite different from that of Boston. Without entering into much detail, the assignment mechanism in New York was decentralized. Students would apply to schools, and the schools had to decide which students to admit, reject, or place on a waiting list. Students were restricted in the number of applications they could send (they also had to establish a preference ordering over schools that could be observed by the schools). Some schools also faced constraints that others did not. For instance, schools offering unscreened programs admitted students by lottery, whereas schools that had the status of zone schools had to give priority to students from the neighborhood, and schools offering screened programs evaluated students individually.

Students would receive decision letters from schools, and successful applicants could accept at most one offer (and be on one waiting list). After this, schools with vacant positions would make new offers to students. There were three such rounds of processing. This clearly was not enough, as about one-third of the students

were ending up without any school and had to be assigned to a school by the authorities (and of course that school was not on their preference list). In other words, the assignment procedure in New York City suffered from congestion.

The fact that many schools could decide which students to admit had a profound impact on the design of the new system. Some schools were not only choosing which students to accept or reject but were also strategically concealing their capacities from the central administration. By not revealing their exact capacities, schools can reserve places that would be assigned later (and thus have more choice over which students to admit). If the assignment is stable, the incentives for schools to conceal their capacities are minimized. Also, unlike in Boston, many schools in New York City have special academic programs that target students with specific needs and skills. In other words, unlike in Boston, many schools in New York City do have preferences among students. Finally, the fact that many schools were strategic in their decisions (trying to game the system) convinced the team of market designers (see Abdulkadiroğlu, Pathak, and Roth, 2005a) that the situation in New York was more akin to the college admission problem, rather than the house allocation problem. The experience and the lessons from the medical match (see chapter 10) made it clear that stability of the matching was the key property that was needed in New York.

Once it was clear that an algorithm that produces stable matching was needed, there came another question: should we use the students-proposing or the schools-proposing version of DA? The choice of the student-proposing version was quickly considered the best option for the reasons we have already analyzed in Chapter 3:

1. It is strategy-proof for the students and produces the student-optimal matching.
2. Also, there is no algorithm that produces stable matchings such that it is dominant strategy for the schools to reveal their true preferences and their true capacities [cf. Proposition 3.3].

In the first year of operation of the new matching mechanism, over 70,000 students managed to be matched to a school on their initial preference list (20,000 more than under the previous system). Students who would not be assigned to a school (and did not withdraw from the New York City public schools) could submit a secondary preference list containing schools that still had vacancies. At the end of the matching process, there were about 3000 students who were administratively matched to a school that was not on their preference list (compared to 30,000 students under the previous system).

Chapter 5

Concluding remarks

In Part II, we have considered markets wherein trades do not occur through monetary transfers (as opposed to Part I). There are either no prices or prices (if any) do not influence market outcomes. Consider (private) schools in Mexico City. Each schools in general charges same fee to all its students (at least of the same cohort), and the parents do not bargain with the schools over fees. So matching of students to schools is determined by attributes other than prices. All relevant information are summarized in the preference orderings of each market participant.

We have analyzed two-sided markets where transactions take place in the absence of monetary transfers among agents. A two-sided matching market consists of two disjoint sets of individuals (e.g. firms and workers, schools and student, houses and tenants). Two categories have been considered—two-sided markets with two-sided preferences and two-sided markets with one-sided preferences. In the first category fall the marriage market and the college admission problem. In contrast, the housing market and the kidney exchange fall in the second category. School choice may fall in both categories. It is akin to the college admission problem, but instead of having preferences over students, schools may have priority lists. On the other hand, because schools do not have preferences, school choice also looks similar to the housing market. For this reason, we have analyzed school choice both in Chapters 3 and 4.

The distinguishing features of the two types of markets (with two-sided and one-sided preferences) are as follows. As representative of each category, consider the marriage market and the housing market. In both types of markets, in general, the set of core allocations is non-empty (for the marriage market core coincides with the set of stable matching outcomes), and core allocations can be reached through distinct algorithms such as the Deferred Acceptance and the Top Trading Cycle algorithms. However, what makes these two types of markets different is in terms of their incentive properties. In the marriage market, there are no mechanisms that can truthfully implement the core allocations [cf. Theorem 2.4]. By contrast, in the housing market, the core allocation (it is unique) can be reached by a strategy-proof mechanism. The intuition behind this contrasting feature regarding stability-incentive tradeoff is simple. In the marriage market, although the core is non-empty, it consists of allocations that generate conflicting interests between the participants on opposite sides. Hence, any mechanism fails to align preferences in the marriage market. A mechanism is strategy-proof only for the proposing side (when it uses the Deferred Acceptance algorithm). With one-sided preferences in the housing market, on the other hand, conflicting interests do not arise, and hence, any mechanism that selects a core allocation is also dominant strategy implementable.

Intermediate between these two sorts of problem is the roommate problem wherein there is only one set of agents, each of whom has preferences over the rest. Such markets may be called a one-sided matching market. The core of such one-sided market can be empty. The reason is that when two agents strongly prefer to be together, the matching problem presents complementarity. If such complementarity is strong, a core allocation

may fail to exist. There are two-sided matching markets that may present features of the one-sided markets. For example, the medical match in the presence of couples, who may decide to accept an inferior allocation but prefer to be together. In job matching, a worker may have preferences over potential co-workers. Analysis of such models have been omitted for the interest of time. For a detailed discussion on similarities and differences among different types of market structures, see (Roth, 1982b).

Part III

Two-sided matching with transfers

Chapter 6

Matching with transfers

Consider the housing market with public endowments in Chapter 4. Now, we allow for monetary transactions between sellers and buyers of houses. In particular, the seller of any house will set a price at which trade may occur. Because there are monetary transfers between one market participant to the other, we shall represent preferences in terms of utility functions. The setup will be the following. The economy comprises a set of sellers or houses, H and a set of buyers, I. Each potential buyer can consume only one unit, i.e., can buy only one house. On the other hand, each house can have only one buyer. Thus, the present framework is a one-to-one matching market (similar to the marriage market). Formally,

Definition 6.1: One-to-one matching

A one-to-one matching of the housing market with transfers is a mapping $\mu : H \to I$ such that (i) $\mu(h) \in I \cup \{h\}$ for each $h \in H$; (ii) $\mu(i) \in H \cup \{i\}$ for each $i \in I$; and (iii) $\mu(h) = i$ if and only if $\mu(i) = h$ for all house-buyer pairs $(h,\, i) \in H \times I$.

The definition of one-to-one matching is the same as that of the marriage market. However, we shall introduce money in a very particular way. To fix ideas, we would first consider a discrete market with two houses and two buyers.

6.1 House quality and buyer valuation in a discrete housing market

Let the set of sellers be $S = \{s_1,\, s_2\}$. Because each one has a unit of house to sell, S and H are isomorphic, and hence, we shall treat S and H in an equivalent fashion. Although houses are sold in unit quantities, they differ in quality. Let q_j be the quality of house j with $q_1 > q_2 > 0$, i.e., house 1 is a better-quality house. House j can be traded at price p_j, and its seller has a reservation price $r_j > 0$. So, seller j derives utility

$$v_j = \max\{p_j - r_j,\, 0\}.$$

On the other hand, the two buyers, $B = \{b_1,\, b_2\}$ are distinct with respect to their valuation for a house. Formally, let θ_i be the valuation of buyer i with $\theta_1 > \theta_2 > 0$, i.e., buyer 1 is the high-valuation buyer. Buyer i derives utility from consuming quality q_j

$$U(\theta_i,\, q_j) = \theta_i u(q_j) - p_j,$$

where $u(\cdot)$ is the utility from the consumption of quality. We immediately establish the following result.

Proposition 6.1: Who gets what?

If $u(q)$ is an increasing function, then the optimal matching involves assigning house 1 to buyer 1, and house 2 to buyer 2, i.e., $\mu(h_1) = b_1$ and $\mu(h_2) = b_2$.

Proof. Suppose on the contrary that b_1 is assigned h_2 and b_2 is assigned h_1. Then, it must be the case that

$$\theta_1 u(q_2) - p_2 \geq \theta_1 u(q_1) - p_1, \tag{6.1}$$
$$\theta_2 u(q_1) - p_1 \geq \theta_2 u(q_2) - p_2. \tag{6.2}$$

Summing (6.1) and (6.2), and rearranging, we obtain

$$(\theta_1 - \theta_2)[u(q_1) - u(q_2)] \leq 0.$$

Because $\theta_1 > \theta_2$, the above inequality implies that $u(q_1) \leq u(q_2)$ which is a contradiction to the fact that $u'(q) > 0$ because $q_1 > q_2$. □

The above proposition asserts that the high-valuation buyer must obtain the better-quality house. This is easily generalizable to $n \geq 2$ buyers and $n \geq 2$ houses. We have ignored the analysis that in equilibrium each buyer is indeed assigned one house. The proof of such existence result can be found in Shapley and Shubik (1971).

6.2 A continuum housing market and positive assortative matching

Let us now generalize the house allocation with transfers to the housing market with many houses and many buyers. In particular, Let $Q \equiv [q_{min}, q_{max}] \subset \mathbb{R}_+$ be the set of house qualities, and $\Theta \equiv [\theta_{min}, \theta_{max}] \subset \mathbb{R}_+$ be the set of valuations. Let $p(q)$ be the price of a house of quality q, and $r(q)$ be its reservation price. On the other hand, a buyer with valuation θ derives utility from consuming a house of quality q, which is given by:

$$u(\theta, q) - p(q) \quad \text{with } u_\theta(\theta, q),\ u_q(\theta, q) > 0.$$

We impose the following property on the function $u(\theta, q)$.

Definition 6.2: Increasing differences

A function $u(\theta, q)$ is said to have increasing differences (ID) in (θ, q) if for any two q' and q'' with $q'' > q'$, the function

$$u(\theta, q'') - u(\theta, q')$$

is an increasing function of θ.

Because each house must be sold in unit quantity, what matters for the housing market is the house quality. Likewise, for the buyer-side of the market, the crucial aspect is the differences in buyer valuation. Thus, we can define a type-type matching or assignment.

Definition 6.3: Positive assortative matching

Let $\mu : Q \to \Theta$ be a type-type matching rule that assigns a house of quality q to a buyer of valuation θ, i.e., for each house $q \in Q$, there is a buyer $\mu(q) \in \Theta$. A matching is a positive assortative matching if $\mu'(q) \geq 0$, i.e., for any two q' and q'' with $q'' > q'$, and any two θ' and θ'' with $\theta'' > \theta'$, we have $\mu(q') = \theta'$ and $\mu(q'') = \theta''$.

The following result provides sufficient condition for an optimal assignment.

Theorem 6.1: Who gets what?

If $u(\theta, q)$ has increasing differences in (θ, q), then the optimal matching is positive assortative.

Proof. Note first that the increasing differences property of $u(\theta, q)$ is equivalent to

$$u(\theta'', q'') - u(\theta'', q') \geq u(\theta', q'') - u(\theta', q') \tag{6.3}$$

for any $\theta'' > \theta'$ and $q'' > q'$. In a way of contradiction, suppose $u(\theta, q)$ has increasing differences in all (θ, q), but there are (θ', θ'') with $\theta'' > \theta'$, and (q', q'') with $q'' > q'$ such that $\mu(q'') = \theta'$ and $\mu(q') = \theta''$. Then it must be the case that

$$u(\theta'', q') - p(q') \geq u(\theta'', q'') - p(q''), \tag{6.4}$$
$$u(\theta', q'') - p(q'') > u(\theta', q') - p(q'). \tag{6.5}$$

Summing (6.4) and (6.5), we obtain

$$u(\theta'', q'') - u(\theta'', q') < u(\theta', q'') - u(\theta', q')$$

which contradicts the fact that $u(\theta, q)$ has increasing differences for all (θ, q), i.e., the inequality in (6.3). □

So far, we have not defined the equilibrium of the housing market with transfers. The equilibrium not only consists of the optimal allocation, but also the house prices, $p(q)$. The equilibrium of the market is the standard Walrasian equilibrium of a discrete good economy. We do not define the equilibrium formally in order to avoid technicalities. The following result characterizes the Walrasian prices.

Proposition 6.2: Equilibrium prices

The equilibrium price of a house with quality q, $p(q)$ is a strictly increasing function.

Proof. Take any two houses with qualities q' and q'' with $q'' > q'$. Then, by Theorem 6.1, there are θ' and θ'' with $\theta'' > \theta'$ such that $\mu(q') = \theta'$ and $\mu(q'') = \theta''$. Then, it must be that

$$u(\theta', q') - p(q') \geq u(\theta', q'') - p(q'') \iff p(q'') - p(q') \geq u(\theta', q'') - u(\theta', q').$$

The right-hand-side of the above inequality is strictly positive because $u_q(\theta, q) > 0$. Therefore, $p(q'') > p(q')$, i.e., $p(q)$ is strictly increasing in q. □

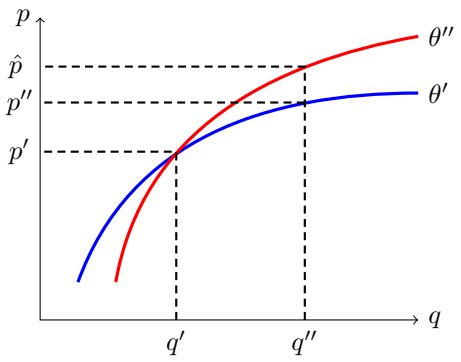


Figure 6.1: Increasing differences is a single-crossing condition.

Why increasing differences implies positive sorting has a simple intuition. Note that Definition 6.2 reads as the marginal utility of consuming greater quality is increasing in buyer valuation which is depicted in Figure

6.1. Let us now consider the utility of a type-θ buyer who consumes quality q and pays p, which is given by

$$U(q,\, p;\, \theta) \equiv u(\theta,\, q) - p$$

If we set a constant utility level, $\bar{u}$, i.e., $U(q,\, p;\, \theta) = \bar{u}$, we get the indifference curve of each θ in the q-p space, whose slope is given by

$$\frac{dp}{dq} = u_q(\theta,\, q).$$

The increasing differences condition then assets that the indifference curve is steeper for higher θ. As a consequence, any two indifference curve may cross only once (single-crossing).

What is the implications of single-crossing for the equilibrium? In Figure 6.1, let two quality-price combinations, $(q',\, p')$ and $(q'',\, p'')$, with $q'' > q'$ and $p'' > p'$, be on the same indifference curve of the buyer with valuation θ'. This means that θ' is willing to pay p'' for buying a higher-quality house, q''. Then, the buyer of type-θ'' (higher-valuation buyer) is willing to pay more, i.e., $\hat{p} > p''$ for consuming quality q''. As a consequence, θ'' obtains a house of higher quality and ends up paying more in equilibrium.

Part IV

Auctions

Chapter 7

Auctions

So far, we have seen (matching) markets or institutions that comprise a set of rules in order to allocate "individuals" or "objects" among "individuals" or "organizations". On the one hand, [in Part II] there are no monetary transactions or even if there are, they do not play any role in the equilibrium allocations. On the other, [in Part III] introduction of monetary transfers makes equilibrium price of an object depend on the willingness-to-pay, which are public knowledge. We now relax the assumption of public information regarding willingness-to-pay. In such environments, prices will still depend on willingness-to-pay; however, we have to elicit willingness-to-pay in order to elicit prices. Thus, Auctions are institutions that are not only meant to set allocation rules, but also they comprise payment rules, i.e., how much market participants must pay to obtain a single object or several objects. Moreover, auctions are indirect mechanisms that are used to reveal private information of the market participants.

We shall consider a simple problem of market design wherein a single object to be allocated to one of many potential buyers whose valuations for the object are private information. This is a special case of the housing market where more than one objects are allocated among the individuals. However, the objective of a mechanism would be to elicit the true willingness-to-pay of the market participants as we have seen Chapter 1. Because there are monetary transfers between buyer(s) and the auctioneer, we shall represent preferences in terms of utility functions. Allocating multiple objects or multiple units of a single objets are also of great interest; however, such models are considerably more complicated.

Auctions are designed to accomplish several goals. First, we would like to understand the behavior of bidders. For this, we will adopt an appropriate notion of equilibrium and analyze equilibrium behavior of bidders. Second, we would like to compare auction formats in terms of their equilibrium outcomes. When comparing auction formats, we usually take two criteria into account—(a) efficiency and (b) expected revenue to the seller. The notion of efficiency is the standard one of ex-post Pareto efficiency that we have analyzed for the housing market in Chapter 4. In the context of single-object auction, the concept of efficiency is simple—an auction is efficient if the object is allocated to the highest-valuation bidder. Expected revenue, on the other hand, reflects an ex-ante objective of the seller, which is to maximize expected revenue across auction formats. Under reasonable conditions, we are able to rank standard auctions in terms of efficiency and expected revenue.

7.1 Auction formats

There are two main popular auction formats—open-bid auction and sealed-bid auction. In an open-bid auction, the bids are announced publicly, and hence, all bids are observable. In a sealed-bid auction, on the other hand, potential buyers of an object submit their bids in sealed envelopes, and thus, bids are not observed by others. Examples of open-bid format includes

- Ascending price auction or English auction. In this format, the auctioneer announces the price starting from a given price (called the reserve price) which increases as the auction advances. The potential buyers announces whether they want the object at the last announced price. The auction stops when only one bidder is left who wins the object and pays their bid. Artworks, antiques, etc. are sold in this format.
- Descending price auction or Dutch auction. The price starts at a high level, and starts to drop as the auction goes on. The potential buyers call out "mine". The first individual to call out gets the object, and pays their bid. Fresh food items, flowers, etc. are sold in this format.

On the other hand, examples of sealed-bid auction are

- First-price auction. In this format, all bidders simultaneously submit bids in sealed envelopes. The highest bid wins and the winner pays their bid.
- Second-price auction or Vickrey auction. Similar to the first-price auction. However, the winner pays the highest losing bid or the second-highest bid.
- Third-price auction. Winner pays the third-highest bid.
- All-pay auction. Everybody pays their bid irrespective of winning or losing.

Another categorization of auctions is based on informational aspects—private value auction and common value auction. In a private value auction, bidder valuation is private information. For example, the willingness-to-pay for a valuable painting is highly subjective. Bidders do not know each others' valuations. By contrast, in a common value auction, all bidders have similar objective valuation for an object. For example, in a corporate takeover bid, all potential buyers of a company value the target more or less the same as the target company's financial statement is publicly available.

Example 7.1: First- and second-price auctions

There are four potential buyers of a car, $i = 1, 2, 3, 4$. The bids are $b_1 = \$7$, $b_2 = \$9$, $b_3 = \$4$ and $b_4 = \$3$. In the first-price auction, bidder 2, the highest bidder wins the object, and pays \$9. In the second price auction also bidder 2 wins, but they pay \$7.

The revenue for the auctioneer in the second price-auction is clearly lower as they obtain \$7 as opposed to \$9 in the first-price auction. Then, why does the auctioneer settle for the second-highest bid? As we shall see that bidding strategies change according to the payment rule. However, the expected revenues of first- and second-price auctions are the same, and hence, the auctioneer is indifferent between the two formats. In most of the auction formats, the allocation rules are the same—the highest bidder wins the object. The formats mostly differ in payment rules. Of course, there are exceptions. In Chinese auction, the object is allocated probabilistically. In particular,

$$\text{Prob. } \{\text{bidder } i \text{ wins by bidding } b_i\} = \frac{b_i}{\sum_{j=1}^{n} b_j}.$$

Example 7.2: English auction

There are four potential buyers of a car, $i = 1, 2, 3, 4$. $\theta_1 = \$8$, $\theta_2 = \$12$, $\theta_3 = \$5$ and $\theta_5 = \$2$. Price starts at 0 and increases. Bidders interested in purchasing at current price press button to indicate. Price stops when all but one bidder (say bidder 2) drops out. Bidder 2 wins at the stopped price.

We shall later learn that the English auction and the second price auctions are strategically equivalent. On the other hand, the Dutch and the first-price auctions are strategically equivalent.

7.2 A formal model of independent private value auction

We describe a formal canonical model of private value auctions. There is a set $I = \{1, \ldots, n\}$ bidders with $n \geq 2$. There is a single object to be sold. Bidder i's valuation or willingness-to-pay, θ_i is a random variable which is distributed according to the cumulative distribution function $F_i(\theta_i)$ on a support $[0, v]$. Valuations of any two bidders are independent of each other. We shall also assume that valuations are identically distributed, i.e., $F_i(\cdot) = F(\cdot)$ for all $i \in I$. In other words, all bidders draw valuations from the same probability distribution. Assume that $F(\theta_i)$ is continuous with the pdf f. Also, all bidders are risk neutral.

7.2.1 First-price sealed-bid auction

Recall from Chapter 1 that we have solved for first-price auction for 2 bidders with linear bidding strategies where valuations are drawn from a uniform distribution. Our objective here is to generalize to $n \geq 2$ bidders who can submit any non-linear bidding function and valuations are drawn from the distribution function $F(\cdot)$. The game is as follows. All n bidders submit simultaneously (in sealed envelopes) their bids, $b_1, \ldots, b_n$. The highest bidder wins and pays their valuation.

We shall analyze a symmetric Bayesian Nash equilibrium (BNE) of the bidding game, i.e., $b_i(\theta_i) = b(\theta_i)$ for all $i \in I$.[1] Recall that in a Bayesian game, strategy of any player is a function of their type. We shall assume that $b(\theta_i)$ is increasing, continuous and differentiable on $[0, v]$. To solve for the symmetric BNE, let all bidders j, different from i, submit the identical bidding function, i.e., $b_j(\theta_j) = b(\theta_j)$ for all $j \neq i$. Then, bidder i's expected payoff (as a function of their bid b_i and valuation θ_i) is given by

$$u_i(b_i, b_{-i}, \theta_i) = (\theta_i - b_i) \times \text{Prob.}\{b(\theta_j) \leq b_i, \text{ for all } j \neq i\} = (\theta_i - b_i)[F(b^{-1}(b_i))]^{n-1}.$$

Thus, bidder i chooses b_i to solve

$$\max_{b_i} \ (\theta_i - b_i)[F(b^{-1}(b_i))]^{n-1}.$$

The first-order condition of the above maximization problem is given by

$$-(F(b^{-1}(b_i)))^{n-1} + (\theta_i - b_i)(n-1)(F(b^{-1}(b_i)))^{n-2} f(b^{-1}(b_i))(b^{-1})'(b_i) = 0.$$

Because we want to show that $b_i = b(\theta_i)$ is the best response of bidder i against $b_j = b(\theta_j)$ for all $j \neq i$, replace b_i by $b(\theta_i)$ in the above expression. Note that $(b^{-1})'(b(\theta_i)) = 1/b'(\theta_i)$. Finally, ignoring the subscript i, the above first-order condition gives rise to the following linear differential equation

$$b'(\theta) = (n-1)(\theta - b(\theta)) \cdot \frac{f(\theta)}{F(\theta)}. \tag{7.1}$$

There are several ways to solve the above differential equation. One way to do is the following. Write (7.1) as

$$\underbrace{b'(\theta)(F(\theta))^{n-1} + b(\theta)\{(n-1)(F(\theta))^{n-2} f(\theta)\}}_{\frac{d}{d\theta}[b(\theta)(F(\theta))^{n-1}]} = \underbrace{\theta(n-1)(F(\theta))^{n-2} f(\theta)}_{\theta \cdot \frac{d}{d\theta}[F(\theta)^{n-1}]}. \tag{7.2}$$

Taking integrals on both sides of (7.2), we obtain

$$\int_0^\theta d[b(x)(F(x))^{n-1}] = \int_0^\theta x d[(F(x))^{n-1}]. \tag{7.3}$$

The left-hand-side of (7.3) is $b(\theta)(F(\theta))^{n-1}$ because $F(0) = 0$. On the other hand, the right-hand-side is given by $\theta(F(\theta))^{n-1} - \int_0^\theta (F(x))^{n-1} dx$. Then, it follows from (7.3) that

[1] See Krishna (2010, Appendix G) for the general existence result for first-price auction.

Proposition 7.1: Equilibrium bidding in first-price auction

In first-price sealed-bid auction with n bidders, and identically and independently distributed bidder valuations with cdf $F(\cdot)$, the symmetric BNE bidding strategy is given by

$$b(\theta) = \theta - \frac{\int_0^\theta (F(x))^{n-1} dx}{(F(\theta))^{n-1}}. \tag{7.4}$$

Clearly, all bidders bid below their true valuation, i.e., $b(\theta) < \theta$. The reason is simple. If a bidder bids their valuation and wins the auction, their payoff would be zero which is exactly equal to that if they do not win. Thus, the objective is to obtain a strictly positive expected payoff in case of winning the object. This is in contrast with the second price auction wherein the highest bidder wins but pays the highest losing bid, and hence, bidding own valuation does not induces zero payoffs in case of winning. Note that the degree of bid shading is given by

$$\theta - b(\theta) = \int_0^\theta \left[\frac{F(x)}{F(\theta)}\right]^{n-1} dx,$$

which depends on the number of competing bidders in that as n increases the above quantity approaches zero, and hence, $b(\theta)$ approaches θ. However, the above expression cannot be computed without knowing the exact functional form of $F(\theta)$. The following example analyzes the equilibrium bidding function for uniform distribution.

Example 7.3: First-price auction with uniform distribution

Let $\theta \sim U[0, 1]$. Then $F(\theta) = \theta$. In this case, we have

$$\theta - b(\theta) = \int_0^\theta \left(\frac{x}{\theta}\right)^{n-1} dx = \frac{\theta}{n} \quad \Longleftrightarrow \quad b(\theta) = \left(1 - \frac{1}{n}\right)\theta.$$

Thus, all bidders adopt a linearly increasing bidding strategy. When, $n = 2$, we have $b(\theta) = \theta/2$.

Note in first-price auction (as well as in many "standard" auctions) that the highest-bidder wins the object. So, it is important to analyze the second-highest bid. Fix a bidder, say i, and let the random variable $Y_i = \max\{\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, i.e., Y_i be the second-highest valuation among the remaining $n-1$ bidders, $I \setminus \{i\}$. Let G denote the distribution function of Y_i, i.e., $G(y) = \text{Prob.}\{Y_i \leq y\}$. Note that

$$\begin{aligned} G(y) &= \text{Prob.}\{Y_i \leq y\} \\ &= \underbrace{\text{Prob.}\{\theta_1 \leq y\}}_{F(y)} \times \ldots \underbrace{\text{Prob.}\{\theta_{i-1} \leq y\}}_{F(y)} \times \underbrace{\text{Prob.}\{\theta_{i+1}}_{F(y)} \leq y\} \times \ldots \times \underbrace{\text{Prob.}\{\theta_n \leq y\}}_{F(y)} \\ &= (F(y))^{n-1}. \end{aligned}$$

If $g(y)$ is the associated density function, it is given by

$$g(y) = (n-1)(F(y))^{n-2} f(y).$$

Note that bidder i with valuation θ wins a first-price auction if $b(Y_i) \leq b(\theta)$. Because $b(\cdot)$ is an increasing function, the winning probability in first-price auction is given by

$$\text{Prob.}\{b(Y_i) \leq b(\theta)\} = \text{Prob.}\{Y_i \leq \theta\} = G(\theta) = (F(\theta))^{n-1}.$$

With the above modification, (7.2) can be written as

$$b'(\theta)G(\theta) + b(\theta)g(\theta) = \theta g(\theta). \tag{7.5}$$

Integrating the above, we obtain

$$b(\theta) = \frac{\int_0^\theta y g(y) dy}{G(\theta)} = \mathbb{E}(Y_i \mid Y_i \leq \theta). \tag{7.6}$$

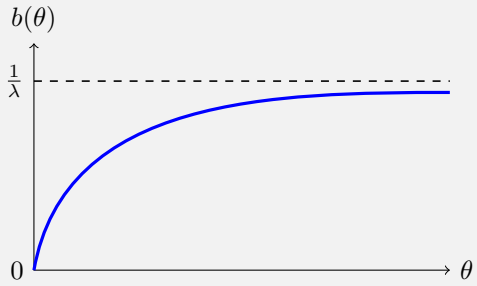
The symmetric BNE bidding strategy is thus a conditional expectation of the highest competing valuation. Now, consider the following example.

Example 7.4: First price auction with exponential distribution

Let $n = 2$ and valuations be exponentially distributed on $[0, \infty)$ with parameter $\lambda > 0$, i.e., $F(\theta) = 1 - e^{-\lambda\theta}$. In this case, the symmetric BNE bidding functions are given by

$$b(\theta) = \frac{1}{\lambda} - \frac{\theta e^{-\lambda\theta}}{1 - e^{-\lambda\theta}}.$$

The equilibrium bidding function is increasing and concave with $\lim_{\theta\to\infty} b(\theta) = \frac{1}{\lambda} = \mathbb{E}(\theta)$. The following figure depicts $b(\theta)$ for $\lambda = 1$.



An interesting point to note is that the bidders bid less aggressively as their valuation grows. For example, when $\theta = \$1$, the symmetric equilibrium bid is given by $b(\$1) = \$0.42$. However, a bidder with valuation \$100 would not bid more than \$1, i.e., $b(\$100) \leq \$1$. In other words, a 9,900% increase in valuation explains a maximum of 19.048% increase in the equilibrium bids. The reason is that there is a very small chance that a bidder with high valuation would lose in equilibrium. In fact, for $\lambda = 1$, a bidder with valuation \$1 does not win with probability $e^{-1} = 0.367879$, whereas a bidder with valuation \$100,000 loses with probability $e^{-100} = 3.72008^{-44}$.

Now, we would relax the assumption of risk-neutral bidders. Let all bidders have identical utility functions $v(\cdot)$ with $v'(\cdot) > 0$ and $v''(\cdot) < 0$. We continue analyzing a symmetric BNE wherein $b_i(\theta_i) = b(\theta_i)$ for all $i \in I$. Thus, bidder i chooses b_i to solve

$$\max_{b_i} v(\theta_i - b_i) \times \text{Prob.}\{b(\theta_j) \leq b_i, \text{ for all } j \neq i\} = v(\theta_i - b_i)[F(b^{-1}(b_i))]^{n-1}.$$

Notice that, relative to the case of risk-neutral bidders, the ex-ante probability that bidder i wins does not change under risk-aversion. The only difference arises from the utility function of each bidder, $v(\cdot)$. Solving the following exercise would give us some idea about the behavior of risk-averse bidders in first-price auction.

Exercise 7.1: Risk-averse bidders in first-price auction

Consider a first-price auction with two bidders, i.e., $n = 2$ wherein each bidder has CRRA utility function, $v(\omega) = \omega^\alpha$ with $0 < \alpha < 1$, and valuations are distributed according to the cdf $F(\theta_i)$ on $[0, v]$ for all $i \in I$. The two extreme cases correspond to $\alpha = 1$ that implies risk-neutral bidders, and $\alpha = 0$ meaning extreme risk aversion. Find the symmetric BNE bidding function $b(\theta)$. Now, assume that valuations are uniformly distributed on $[0, 1]$. Show that the bidder bid more aggressively (i.e., lesser bid shading) as they become more risk-averse (i.e., α decreases) with the limiting case that

$\lim_{\alpha \to 0} b(\theta) = \theta$.

Finally, we discuss the relation between first-price auction and other auction formats, in particular, the Dutch auction.

Proposition 7.2: Strategic equivalence

The first-price sealed-bid auction is strategically equivalent to the descending-price or the Dutch auction.

The intuition behind the above result is simple. Descending price auction is a dynamic process where the price starts at a higher level, and starts dropping over time. If a bidder calls out "mine" at any prevailing price, the auction stops, and that bidder gets the object and pays the prevailing price. Therefore, the bids are not history-dependent in that the prevailing price contains all the information. If the auction is still on at some price, say \$50, this means that potential buyer has not called out when the price was higher than \$50. Because a bidder just needs to decide at what price he will shout "mine", he can decide it just before the auction starts. This implies that we can perfectly study bidders' behavior in a Dutch auction by assuming that they indeed choose their bids (i.e., their strategies) before the auction starts. But then, because the winner is the bidder with the highest bid, and the price is the winner's bid, the Dutch auction is equivalent to the first-price sealed-bid auction.

7.2.2 Second-price sealed-bid auction

The equilibrium bidding behavior in a second-price sealed-bid auction has already been analyzed in Chapter 1 [see Example 1.7]. In what follows, we shall compare second-price auction with other auction formats.

Proposition 7.3: Strategic equivalence

The second-price sealed-bid auction is strategically equivalent to the ascending-price or the English auction.

The English or ascending-price auction is a dynamic auction, and a bidder must decide at every point whether to continue or not. A potential buyer's bidding strategy is to pick a price at which to drop out (dynamically). Let bidder i has valuation θ_i. Should the bidder be willing to buy the object when the price exceeds valuation? If $p > \theta_i$, bidder i drops out of the auction. By contrast, if $p < \theta_i$, then bidder i stays in the auction until the price reaches their valuation. If θ_i is the highest valuation, then they pay $p = \theta_i$. In other words, the optimal bidding strategy in English auction is $b_{EA}(\theta_i) = \theta_i$ for all $i \in I$. Thus, English auction is strategically equivalent to second-price sealed-bid auction.

7.3 Revenue equivalence

In this section, we analyze a striking result in auction theory—under very general conditions, an auctioneer is indifferent among all "standard" auction formats. An auction is said to be standard if the allocation rule dictates that the highest-bidder is awarded the object.

Theorem 7.1: The revenue equivalence theorem

Suppose that valuations are independently and identically distributed, and all bidders are risk-neutral. Then, any symmetric and increasing equilibrium of any standard auction, such that the lowest-valuation

bidder makes zero payment in expectation, yields the same expected revenue to the seller.

Proof. Consider a standard auction format, and denote it by a. Fix a symmetric equilibrium increasing bidding function $b(\cdot)$ of a. Let $m^a(\theta)$ denote the equilibrium expected payment a bidder with valuation θ makes in the auction. Suppose that $b(\cdot)$ is such that $m^a(0) = 0$. Consider a given bidder, say bidder i with valuation θ who bids $b(\theta')$ instead of the equilibrium bid $b(\theta)$. Bidder i wins the auction if $b(Y_i) \leq b(\theta')$ where Y_i is the second-highest valuation. This is equivalent to $Y_i \leq \theta'$. Bidder i's expected payoff is

$$\pi^a(\theta', \theta) = G(\theta')\theta - m^a(\theta'), \tag{7.7}$$

where $G(y) = (F(y))^{n-1}$ is the distribution of the second-highest valuation, Y_i. Note that $m^a(\theta')$ depends on the other players' strategy b through $G(\theta')$ and θ', but not on the true valuation θ. The first-order condition associated with (7.7) is

$$\frac{\partial \pi^a(\theta', \theta)}{\partial \theta'} = g(\theta')\theta - \frac{dm^a(\theta')}{d\theta'} = 0.$$

At an equilibrium, it is optimal to report $\theta' = \theta$, and hence, the above first-order condition becomes

$$\frac{dm^a(\theta)}{d\theta} = g(\theta)\theta$$

for all $\theta \in [0, v]$. Thus,

$$m^a(\theta) = m^a(0) + \int_0^\theta yg(y)dy = \int_0^\theta yg(y)dy = G(\theta) \times \mathbb{E}(Y_i \mid Y_i \leq \theta) \tag{7.8}$$

because, by assumption, $m^a(0) = 0$. Note that $m^a(\theta)$ is a random variable because θ is a random variable. The ex-ante expected payment of any given bidder (before they picked their valuation from $F(\cdot)$, and hence, the terminology, *ex-ante*) is given by

$$\mathbb{E}[m^a(\theta)] = \int_0^v m^a(\theta)f(\theta)d\theta = \int_0^v \left(\int_0^\theta yg(y)dy \right) f(\theta)d\theta.$$

The above is how much the seller will receive in expectation from a given bidder. Because there are n bidders, the expected revenue for the seller from auction a is

$$\mathbb{E}[R^a] = n \times \mathbb{E}[m^a(\theta)] = n \times \mathbb{E}[G(\theta) \times \mathbb{E}(Y_i \mid Y_i \leq \theta)]. \tag{7.9}$$

Because the right-hand-side of (7.9) is independent of the auction format, a, the theorem holds. □

In the following example, we compute the expected revenue of a standard auction for a specific distribution function of valuations.

Example 7.5: Expected revenue under power distribution

Let valuations follow a power distribution with parameter $\alpha > 0$ on $[0, 1]$, i.e., $F(\theta) = \theta^\alpha$. The corresponding density function is $f(\theta) = \alpha\theta^{\alpha-1}$. Thus, $G(\theta) = \theta^{\alpha(n-1)}$, and $g(\theta) = \alpha(n-1)\theta^{\alpha(n-1)-1}$ which implies $\theta g(\theta) = \alpha(n-1)\theta^{\alpha(n-1)}$. The expected payment of a bidder with valuation θ is given by

$$m^a(\theta) = \int_0^\theta yg(y)dy = \int_0^\theta \alpha(n-1)y^{\alpha(n-1)}dy = \frac{\alpha(n-1)\theta^{\alpha(n-1)+1}}{\alpha(n-1)+1}.$$

Therefore,

$$\mathbb{E}[m^a(\theta); \alpha] = \int_0^1 m^a(\theta)f(\theta)d\theta = \frac{\alpha^2(n-1)}{\alpha(n-1)+1} \int_0^1 \theta^{\alpha n}d\theta = \frac{\alpha^2(n-1)}{(\alpha(n-1)+1)(\alpha n+1)}.$$

The expected revenue thus is

$$\mathbb{E}[R^a; \alpha] = n \times \mathbb{E}[m^a(\theta)] = \frac{\alpha^2 n(n-1)}{(\alpha(n-1)+1)(\alpha n+1)}\,.$$

When $\alpha = 1$, we have uniform distribution on $[0,\ 1]$. The expected revenue is thus given by

$$\mathbb{E}[R^a; 1] = \frac{n-1}{n+1}\,.$$

From the Revenue equivalence theorem 7.1, it follows that first- and second-price auctions are revenue-equivalent. We should also be careful in interpreting the revenue equivalence principle. This result holds under several assumptions—valuations are private and they are independently and identically distributed, and bidders are risk-neutral. The revenue equivalence theorem thus serves as a benchmark; relaxing some of the assumptions would allow us to rank various auction formats in terms of seller's expected revenues. For example, if bidders were risk-averse, first-price auction yields higher expected revenue than second-price auction (why?). We have so far also ignored the analysis of efficiency. We have already mentioned that in a single-object auction wherein the highest-bidder wins, the auction is ex-post efficient. So, all auction formats we have studied above are efficient.

Bibliography

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2005a), "The new york city high school match." *The American Economic Review, Papers and Proceedings*, 95, 346–367.

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez (2005b), "The boston public school match." *The American Economic Review, Papers and Proceedings*, 95, 368–371.

Abdulkadiroğlu, Atila and Tayfun Sönmez (1999), "House allocation with existing tenants." *Journal of Economic Theory*, 88, 233–260.

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School choice: A mechanism design approach." *The American Economic Review*, 93, 729–747.

Aumann, Robert J. (1964), "Markets with a continuum of traders." *Econometrica*, 32, 39–50.

Gale, David and Lloyd S. Shapley (1962), "College admissions and the stability of marriage." *The American Mathematical Monthly*, 69, 9–15.

Gale, David and Marilda O. Sotomayor (1985), "Some remarks on the stable matching problem." *Discrete Applied Mathematica*, 11, 223–232.

Glazerman, Steven and Robert H. Meyer (1994), "Public school choice in minneapolis." In *Midwest approaches to school reform* (T.A. Downes and W. A. Testa, eds.), 110–26.

Haeringer, Guillaume (2017), *Market Design: Auctions and Matching*. The MIT Press.

Kesten, Onur (2010), "School choice with consent." *The Quarterly Journal of Economics*, 125, 1297–1348.

Knuth, Donald E. (1976), *Marriages Stables*. Les Presses de l'Université de Montreal, Montreal.

Kojima, Fuhito and Parag Pathak (2009), "Incentives and stability in large two-sided matching markets." *The American Economic Review*, 99, 608–627.

Krishna, Vijay (2010), *Auction Theory*, 2nd edition. Academic Press, San Diego, CA.

Ma, Jinpeng (1994), "Strategy-proofness and the strict core in a market with indivisibilities." *International Journal of Game Theory*, 23, 75 – 83.

Robinson-Cortés, Alejandro (2021), "Topics in Microeconomic Theory II: Matching and Market Design.", URL https://drive.google.com/file/d/1tEh1sYJ4BQwLNB5BbvFG5nRCVsjkaoW9/view. Lecture notes, University of Exeter.

Roth, Alvin E. (1982a), "The economics of matching: Stability and incentives." *Mathematics of Operations Research*, 7, 617–628.

Roth, Alvin E. (1982b), "Incentive compatibility in a market with indivisible goods." *Economics Letters*, 9, 127–132.

Roth, Alvin E. (1984), "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92, 991–1016.

Roth, Alvin E. (1985), "The college admissions problem is not equivalent to the marriage problem." *Journal of Economic Theory*, 36, 277–288.

Roth, Alvin E. (2002), "The economist as engineer: Game theory, experimentation, and computation as tools for design economics." *Econometrica*, 70, 1341–1378.

Roth, Alvin E. and Elliott Peranson (1999), "The redesign of the matching market for american physicians: Some engineering aspects of economic design." *The American Economic Review*, 89, 748–780.

Roth, Alvin E. and Andrew Postlewaite (1977), "Weak versus strong domination in a market with indivisible goods." *Journal of Mathematical Economics*, 4, 131–137.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver (2004), "Kidney exchange." *The Quarterly Journal of Economics*, 119, 457–488.

Roth, Alvin E. and Marilda A. Oliveira Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monographs, Cambridge University Press.

Shapley, Lloyd and Herbert Scarf (1974), "On cores and indivisibility." *Journal of Mathematical Economics*, 1, 23 – 37.

Shapley, Lloyd S. and Martin Shubik (1971), "The assignment game I: The core." *International Journal of Game Theory*, 1, 111–130.

Sönmez, Tayfun (1997), "Manipulation via capacities in two-sided matching markets." *Journal of Economic Theory*, 77, 197–204.

van Steen, Maarten (2010), *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen.