Matching and Market Design

Theory and Practice

Xiang Sun

August 21, 2015

Contents

Acknowledgement

This note was initiated when I gave lectures at Wuhan University in spring 2014. I thank Wei He, Gaoji Hu and Qianfeng Tang for helpful discussion.

This note is much influenced by the following books, surveys, and lecture notes

- Atila Abdulkadiroğlu and Tayfun Sönmez, Matching Markets: Theory and Practice, in *Advances in Economics and Econometrics Theory and Applications, Volume II*, Tenth World Congress.
- Fuhito Kojima, *Lecture notes on Market Design*, 2015. Available at Kojima's homepage.
- Alvin E. Roth's blog on Matching and Market Design.
- Alvin E. Roth and Marilda A. Oliveira Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press, 1992.
- Tayfun Sönmez, *Mini-Course on Matching*. Available at Sönmez's homepage.
- Tayfun Sönmez and M. Utku Ünver, Matching, Allocation, and Exchange of Discrete Resources, in *Handbook of Social Economics, Volume 1A* (Jess Benhabib, Alberto Bisin and Matthew O. Jackson Eds.), Elsevier B.V., 2010.
- Qianfeng Tang and Yongchao Zhang, *Lecture notes on matching*, 2015.
- Jerusalem Summer School in Matching and Market Design (with recorded lectures), 2014. Available at `http://www.as.huji.ac.il/schools/econ25`.

I am grateful for the following students who provide lots of helpful suggestions and comments: Xiaoguang Chen (陈晓光), Yichen Chen (陈奕辰), Yongkang Hu (胡永康), Liucong Huang (黄柳淙), Zhijiang Huang (黄志疆), Jia'en Li (李佳恩), Qin Li (李擎), Xiao Lin (林潇), Yuting Liu (刘雨婷), Jie Lu (陆劼), Xiaojie Mao (毛小介), Tianchen Song (宋天辰), Xunkang Tian (田珣康), Rui Wang (汪瑞), Xiao Wang (王筱), Ya Wen (文雅), Yue Wu (伍玥), Weihao Yan (闫玮昊), Chen Zhang (张晨), Ying Zhou (周颖).

Chapter 1

Introduction

Contents

1.1 Matching and market design

1.1 Matching theory, a name referring to several loosely related research areas concerning matching, allocation, and exchange of indivisible resources, such as jobs, school seats, houses, *etc.*, lies at the intersection of game theory, social choice theory, and mechanism design.

1.2 Matching can involve two-sided matching, in markets with two sides, such as firms and workers, students and schools, or men and women, that need to be matched with each other. Or matching can involve the allocation or exchange of indivisible objects, such as dormitory rooms, transplant organs, courses, summer houses, *etc.*

Recently, matching theory and its application to market design have emerged as one of the success stories of economic theory and applied mechanism design.

1.3 The economics of "matching and market design" analyzes and designs real-life institutions. A lot of emphasis is placed on concrete markets and details so that we can offer practical solutions.

1.4 Labor markets: the case of American hospital-intern markets:

- Medical students in many countries work as residents (interns) at hospitals.
- In the U.S. more than 20,000 medical students and 4,000 hospitals are matched through a clearinghouse, called NRMP (National Resident Matching Program).
- Doctors and hospitals submit preference rankings to the clearinghouse, and the clearinghouse uses a specified rule (computer program) to decide who works where.
- Some markets succeeded while others failed. What is a "good way" to match doctors and hospitals?

1.5 Kidney exchange:

- Kidney Exchange is a preferred method to save kidney-disease patients.

- There are lots of kidney shortages, and willing donor may be incompatible with the donor.
- Kidney Exchange tries to solve this by matching donor-patient pairs.
- What is a "good way" to match donor-patient pairs?

1.6 School choice:

- In many countries, especially in the past, children were automatically sent to a school in their neighborhoods.
- Recently, more and more cities in the United States and in other countries employ school choice programs: school authorities take into account preferences of children and their parents.

1.7 Targets: Efficiency, fairness, incentives.

1.2 Time line of the main evolution of matching and market design

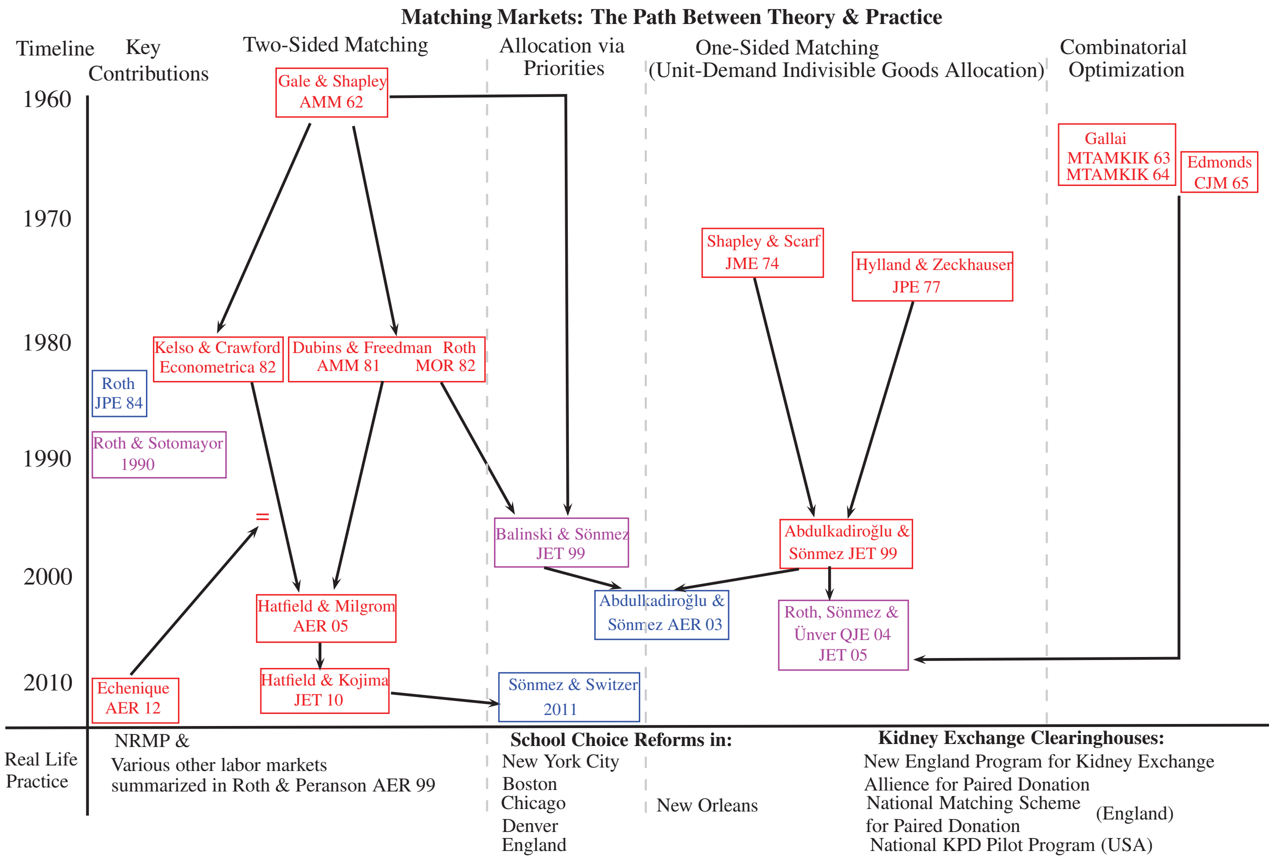


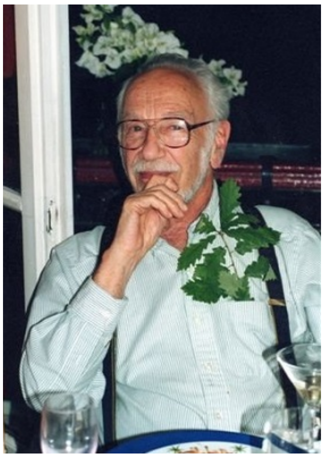
Figure 1.1: Overview (Taken from Sönmez's lecture notes).

Two-sided matching

1.8 In 1962, deferred-acceptance algorithm by David Gale and Lloyd Shapley.

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

(a) Lloyd Stowell Shapley.



(b) David Gale.

Figure 1.2

Gale and Shapley asked whether it is possible to match m women with m men so that there is no pair consisting of a woman and a man who prefer each other to the partners with whom they are currently matched. They proved not only non-emptiness but also provided an algorithm for finding a point in it.

1.9 Shapley and Shubik (1972) and Kelso and Crawford (1982) introduced variants of the two-sided matching model where monetary transfers are also possible between matching sides.

- Lloyd Shapley and Martin Shubik, The assignment game I: the core, *International Journal of Game Theory* **1** (1972), 111–130.

- Alexander S. Kelso and Vincent P. Crawford, Job matchings, coalition formation, and gross substitutes, *Econometrica* **50:6** (1982), 1483–1504.

(a) Martin Shubik.

(b) Vincent Crawford.

Figure 1.3

1.10 In 1982, impossibility theorem by Alvin Roth.

- Alvin Roth, The economics of matching: stability and incentives, *Mathematics of Operations Research* **7:4** (1982), 671–628.

Figure 1.4: Alvin Roth.

Roth proved that no stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.

1.11 Gale and Shapley's short note was almost forgotten until 1984, when Roth showed that the same algorithm was independently discovered by the National Residency Matching Program (NRMP) in the United States (US).

Alvin Roth, The evolution of the labor market for medical interns and residents: a case study in game theory, *Journal of Political Economy* **92** (1984), 991–1016.

1.12 Recently, new links between auctions, two-sided matching, and lattice theory were discovered; for example, matching with contracts by Hatfield and Milgrom in 2005.

J. W. Hatfield, P. R. Milgrom, Matching with contracts, *American Economic Review* **95** (2005), 913–935.

(a) Paul Milgrom. (b) John Hatfield.

Figure 1.5

One-sided matching

1.13 In 1974, top trading cycles algorithm by David Gale, Herbert Scarf and Lloyd Shapley.

Lloyd Shapley and Herbert Scarf, On cores and indivisibility, *Journal of Mathematical Economics* **1** (1974), 23–28.

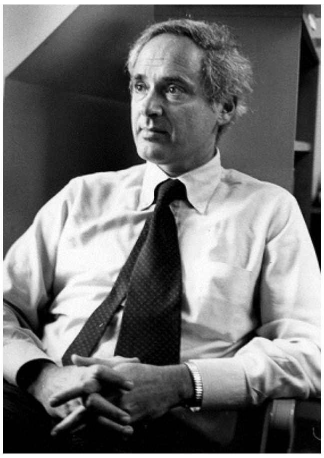


Figure 1.6: Herbert Scarf.

In the other branch of matching theory, allocation and exchange of indivisible goods, the basic model, referred to as the housing market, consists of agents each of whom owns an object, *e.g.*, a house. They have preferences over all houses including their own. The agents are allowed to exchange the houses in an exchange economy. Shapley and Scarf showed that such a market always has a (strict) core matching, which is also a competitive equilibrium allocation. They also noted that a simple algorithm suggested by David Gale, now commonly referred to as Gale's top trading cycles algorithm, also finds this particular core outcome.

1.14 In 1979, Hylland and Zeckhauser proposed the house allocation problem.

Aanund Hylland and Richard Zeckhauser, The efficient allocation of individuals to positions, *Journal of Political Economy* **87:2** (1979), 293–314.

(a) Aanund Hylland.

(b) Richard Zeckhauser.

Figure 1.7

1.15 In 1999, Atila Abdulkadiroğlu and Tayfun Sönmez proposed YQMH-IGYT (you request my house—I get your turn) algorithm for the house allocation problem with existing tenants.

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

(a) Atila Abdulkadiroğlu. (b) Tayfun Sönmez.

Figure 1.8

1.16 In 2003, Atila Abdulkadiroğlu and Tayfun Sönmez proposed school choice problem.

Atila Abdulkadiroğlu and Tayfun Sönmez, School choice: a mechanism design approach, *American Economic Review* **93:3** (2003), 729–747.

1.17 In 2004, Alvin Roth, Tayfun Sönmez and M. Utku Ünver proposed kidney exchange problem.

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

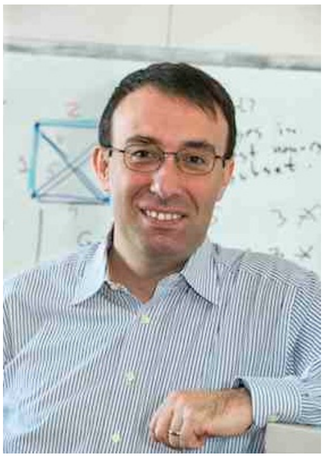


Figure 1.9: M. Utku Ünver.

Part I

Two-sided matching

Chapter 2

Marriage

Contents

2.1 The formal model

☞ 2.1 A marriage problem is a triple $\Gamma = \langle M, W, \succsim\rangle$, where

- M is a finite set of men,
- W is a finite set of women,
- $\succsim = (\succsim_i)_{i\in M\cup W}$ is a list of preferences. Here
  - $\succsim_m$ denotes the preference of man m over $W \cup \{m\}$,
  - $\succsim_w$ denotes the preference of woman w over $M \cup \{w\}$,
  - $\succ_i$ denotes the strict preference derived from $\succsim_i$ for each $i \in M \cup W$.

2.2 For man m:

- $w \succ_m w'$ means that man m prefers woman w to woman w'.
- $w \succ_m m$ means that man m prefers woman w to remaining single.
- $m \succ_m w$ means that woman w is unacceptable to man m.

We use similar notation for women.

2.3 If an individual is not indifferent between any two distinct acceptable alternatives, he has strict preferences. Unless otherwise mentioned all preferences are strict.

☞ 2.4 A matching in a marriage problem $\Gamma = \langle M, W, \succsim \rangle$ is a function $\mu\colon M \cup W \to M \cup W$ such that

- for all $m \in M$, if $\mu(m) \neq m$ then $\mu(m) \in W$,
- for all $w \in W$, if $\mu(w) \neq w$ then $\mu(w) \in M$,
- for all $m \in M$ and $w \in W$, $\mu(m) = w$ if and only if $\mu(w) = m$.

We refer to $\mu(i)$ as the mate of i, and $\mu(i) = i$ means that agent i remains single under the matching μ.

2.5 A matching will sometimes be represented as a set of matched pairs. Thus, for example, the matching

$$\mu = \begin{bmatrix} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{bmatrix}$$

has m_1 married to w_4 and m_5 remaining single.

2.2 Stability and optimality

Let us focus on a marriage problem $\Gamma = \langle M, W, \succsim \rangle$.

2.6 For two matchings μ and ν, an individual i prefers μ to ν if and only if i prefers $\mu(i)$ to $\nu(i)$.

Let $\mu \succ_M \nu$ if $\mu(m) \succsim_m \nu(m)$ for all $m \in M$, and $\mu(m) \succ_m \nu(m)$ for at least one man m.

Let $\mu \succsim_M \nu$ denote that either $\mu \succ_M \nu$ or that all men are indifferent between μ and ν.

☞ 2.7 A matching μ is Pareto efficient if there is no other matching ν such that

- $\nu(i) \succsim_i \mu(i)$ for all $i \in M \cup W$,
- $\nu(i_0) \succ_{i_0} \mu(i_0)$ for some $i_0 \in M \cup W$.

☞ 2.8 A matching μ is blocked by an individual $i \in M \cup W$ if $i \succ_i \mu(i)$.

A matching is individually rational if it is not blocked by any individual.

2.9 A matching μ is blocked by a pair $(m, w) \in M \cup W$ if they both prefer each other to their partners under μ, *i.e.*,

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

☞ 2.10 A matching μ is stable if it is not blocked by any individual or any pair.

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

m_1	m_2	m_3	w_1	w_2	w_3
w_2	w_1	w_1	m_1	m_3	m_1
w_1	w_3	w_2	m_3	m_1	m_3
w_3	w_2	w_3	m_2	m_2	m_2

Table 2.1

All possible matchings are individually rational, since all pairs (m, w) are mutually acceptable.

The matching μ given below is unstable, since (m_1, w_2) is a blocking pair.

$$\mu = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_1 & m_2 & m_3 \end{bmatrix}.$$

The matching μ' is stable.

$$\mu' = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_1 & m_3 & m_2 \end{bmatrix}.$$

2.12 Proposition: Stability implies Pareto efficiency.

Proof. (1) Suppose the matching μ is not Pareto efficient, that is, there exists a matching ν such that $\nu(i) \succsim_i \mu(i)$ for all $i \in M \cup W$ and $\nu(i_0) \succ_{i_0} \mu(i_0)$ for some $i_0 \in M \cup W$.

(2) If $\nu(i_0) = i_0$, then μ is blocked by the individual i_0. Contradiction.

(3) Suppose $\nu(i_0) \neq i_0$, without loss of generality, denote i_0 by m, and $\nu(i_0) = \nu(m)$ by w. Hence we have $w \succ_m \mu(m)$.

(4) Since $\nu(i) \succsim_i \mu(i)$ holds for all i, we have $m = \nu(w) \succsim_w \mu(w)$.

(5) Since all preferences are strict, $m \succsim_w \mu(w)$ if and only if $m \succ_w \mu(w)$ or $m = \mu(w)$.

(6) If $m = \mu(w)$, then $\mu(m) = w$, which contradicts to $w \succ_m \mu(m)$. Hence we have $m \succ_w \mu(w)$. Therefore μ is blocked by the pair (m, w). Contradiction.

□

2.13 Example: Stability can not be implied by Pareto efficiency.

Exercise!

2.3 Deferred acceptance algorithm

2.14 Men-proposing deferred acceptance algorithm.

Step 1: Each man m proposes to his first choice (if he has any acceptable choices). Each woman rejects any offer except the best acceptable proposal and "holds" the most-preferred acceptable proposal (if any). Note that she does not accept him yet, but keeps him on a string to allow for the possibility that someone better may come along later.

Step k: Any man who was rejected at Step $k-1$ makes a new proposal to his most-preferred acceptable potential mate who has not yet rejected him (If no acceptable choices remain, he makes no proposal). Each woman receiving proposals chooses her most-preferred acceptable proposal from the group consisting of the new proposers and the man on her string, if any. She rejects all the rest and again keeps the best-preferred in suspense.

End: The algorithm terminates when there are no more rejections. Each woman is matched with the man she has been holding in the last step. Any woman who has not been holding an offer or any man who was rejected by all acceptable women remains single.

2.15 Theorem on stability (Theorem 1 in Gale and Shapley (1962)): The men-proposing deferred acceptance algorithm gives a stable matching for each marriage problem.

Proof. (1) It suffices to show that the matching μ determined by the men-proposing deferred acceptance algorithm is not blocked by any pair (m, w).

(2) Suppose that there is a pair (m, w), such that $m \neq \mu(w)$ and $w \succ_m \mu(m)$.

(3) Then m must have proposed to w at some step and subsequently been rejected in favor of someone that w likes better.

(4) It is now clear that w must prefer her mate $\mu(w)$ to m and there is no instability.

(5) Similar discussion applies to the pair (m, w) with $m \neq \mu(w)$ and $m \succ_w \mu(w)$.

□

2.16 Quotation from Roth (2008): At his birthday celebration in Stony Brook on 12 July 2007, David Gale related the story of his collaboration with Shapley to produce deferred acceptance algorithm by saying that he (Gale) had proposed the model and definition of stability, and had sent to a number of colleagues the conjecture that a stable matching always existed. By return mail, Shapley proposed the deferred acceptance algorithm and the corresponding proof.

2.17 Example of men-proposing deferred acceptance algorithm: There are five men and four women, and their preferences are as follows:

m_1	m_2	m_3	m_4	m_5	w_1	w_2	w_3	w_4
w_1	w_4	w_4	w_1	w_1	m_2	m_3	m_5	m_1
w_2	w_2	w_3	w_4	w_2	m_3	m_1	m_4	m_4
w_3	w_3	w_1	w_3	w_4	m_1	m_2	m_1	m_5
w_4	w_1	w_2	w_2		m_4	m_4	m_2	m_2
					m_5	m_5	m_3	m_3

Table 2.2

Step 1: m_1, m_4, and m_5 propose to w_1, and m_2 and m_3 propose to w_4; w_1 rejects m_4 and m_5 and keeps m_1 engaged; w_4 rejects m_3 and keeps m_2 engaged. That is,

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ m_1, \cancel{m_4}, \cancel{m_5} & & & m_2, \cancel{m_3} \end{bmatrix}.$$

Step 2: m_3, m_4 and m_5 propose to their second choice, that is, to w_3, w_4 and w_2 respectively; w_4 rejects m_2 and keeps m_4 engaged:

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ m_1 & m_5 & m_3 & m_4, \cancel{m_2} \end{bmatrix}.$$

Step 3: m_2 proposes to his second choice, w_2, who rejects m_5 and keeps m_2 engaged:

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ m_1 & m_2, \cancel{m_5} & m_3 & m_4 \end{bmatrix}.$$

Step 4: m_5 proposes to his third choice, w_4, who rejects m_5 and continues with m_4 engaged. Since m_5 has been rejected by every woman on his list of acceptable women, he stays single, and the matching is:

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{bmatrix}.$$

2.18 Theorem on optimality (Theorem 2 in Gale and Shapley (1962)): The matching determined by men-proposing deferred acceptance algorithm is at least good as any other stable matching for all men.

Proof. Let us call a woman "achievable" for a particular man if there is a stable matching that sends him to her. The proof is by induction.

(1) Assume that up to a given step in the procedure no man has yet been turned away from a woman that is achievable for him. At this step suppose that women w holds m_1 and rejects m. We will show that w is not achievable for m.

(2) Since m_1 prefers w to all others, except for those that have previously rejected him, and hence (by assumption) these women are not achievable for m_1.

(3) Consider a hypothetical matching that sends m to w, and everyone else to women that are achievable for them.

(4) Then m_1 will have to match with a less desirable woman than w.

(5) This matching is unstable, since m_1 and w could upset it to the benefit of both.

(6) Therefore, w is not achievable for m.

(7) The conclusion is that the men-proposing deferred acceptance algorithm only rejects men from women which they could not possibly be matched to in any stable matching. The resulting matching is therefore optimal.

□

2.19 For $\Gamma = \langle M, W, \succsim \rangle$, we refer to the outcome of the men-proposing deferred acceptance algorithm as the man-optimal stable matching and denote it by $\mu^M[\Gamma]$ or $\mu^M[\succsim]$ (when M and W are fixed) or μ^M (when M, W and $\succsim$ are fixed).

The algorithm where the roles of men and women are reversed is known as the women-proposing deferred acceptance algorithm and we refer to its outcome $\mu^W[\Gamma]$ or $\mu^W[\succsim]$ (when M and W are fixed) or μ^W (when M, W and $\succsim$ are fixed) as the woman-optimal stable matching.

2.20 These two matchings will not typically be the same. For Example 2.17, the matching obtained when the women propose to the men is

$$\begin{bmatrix} w_4 & w_1 & w_2 & w_3 & (m_5) \\ m_1 & m_2 & m_3 & m_4 & m_5 \end{bmatrix}.$$

2.21 If some individuals may be indifferent between possible mates, Theorem 2.18 need not hold.

Example: There are three men and three women, and their preferences are as follows:

m_1	m_2	m_3	w_1	w_2	w_3
w_2, w_3	w_2	w_3	m_1	m_1	m_1
w_1	w_1	w_1	m_2	m_2	m_3
			m_3		

Table 2.3

The stable matchings are

$$\mu_1 = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_2 & m_1 & m_3 \end{bmatrix} \text{ and } \mu_2 = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_3 & m_2 & m_1 \end{bmatrix},$$

but there are no optimal stable matchings since

- $\mu_1(m_3) \succ_{m_3} \mu_2(m_3)$ and $\mu_2(m_2) \succ_{m_2} \mu_1(m_2)$;
- $\mu_1(w_2) \succ_{w_2} \mu_2(w_2)$ and $\mu_2(w_3) \succ_{w_3} \mu_1(w_3)$.

2.4 Properties of stable matchings

✌ 2.22 Decomposition theorem (Knuth (1976)): Let μ and μ' be stable matchings in $\langle M, W, \succsim\rangle$, where all preferences are strict. Let $M(\mu)$ be the set of men who prefers μ to μ' and $W(\mu)$ the set of women who prefer μ to μ'. Analogously define $M(\mu')$ and $W(\mu')$. Then μ and μ' map $M(\mu')$ onto $W(\mu)$ and $M(\mu)$ onto $W(\mu')$.

Proof. (1) For any $m \in M(\mu')$, we have $\mu'(m) \succ_m \mu(m) \succsim_m m$, where the second equation holds since μ is stable and not blocked by any individual.

(2) Then $\mu'(m) \neq m$, and hence $\mu'(m) \in W$, denoted by w.

(3) Since μ is a stable matching in $\langle M, W, \succsim\rangle$, $\mu(w) \succsim_w \mu'(w)$; otherwise the pair (m, w) blocks μ.

(4) Furthermore, $\mu(w) \succ_w \mu'(w)$ otherwise $\mu'(m) = w = \mu(m)$.

(5) We have $\mu'(m) = w \in W(\mu)$, and hence $\mu'(M(\mu')) \subseteq W(\mu)$.

(6) For any $w \in W(\mu)$, we have $\mu(w) \succ_w \mu'(w) \succsim_w w$, where the second equation holds since μ is stable and not blocked by any individual.

(7) Then $\mu(w) \in M$, denoted by m.

(8) Since μ' is a stable matching in $\langle M, W, \succsim\rangle$, $\mu'(m) \succ_m \mu(m)$; otherwise the pair (m, w) blocks μ'.

(9) We have $\mu'(m) \succ_m \mu(m) = w$ and $\mu(m) \succ_m m$, then $\mu'(m) \succ_m \mu(m) = w$.

(10) We have $m \in M(\mu')$ and hence $\mu(W(\mu)) \subseteq M(\mu')$.

(11) Since μ and μ' are one-to-one and $M(\mu')$ and $W(\mu)$ are finite, the conclusion follows.

□

2.23 Remark: Decomposition theorem (Theorem 2.22) implies that if m prefers μ to μ' and $\mu(m) = w$ and $\mu'(m) = w'$, then both w and w' will prefer μ' to μ. That is, both μ and μ' decompose the men and women as illustrated in Figure 2.1:

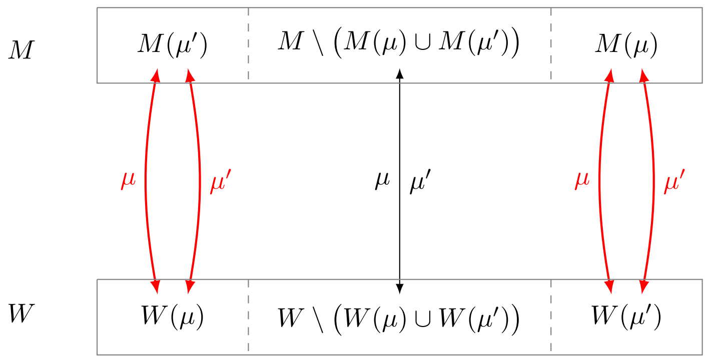


Figure 2.1: Decomposition theorem

✌ 2.24 Theorem (Knuth (1976)): When all agents have strict preferences, if μ and μ' are stable matchings, then $\mu' \succ_M \mu$ if and only $\mu \succ_W \mu'$.

Proof. (1) $\mu' \succ_M \mu$ if and only if $M(\mu) = \emptyset$ and $M(\mu') \neq \emptyset$.

(2) This is equivalent to $W(\mu') = \emptyset$ and $W(\mu) \neq \emptyset$.

(3) This is equivalent to $\mu \succ_W \mu'$.

□

2.25 Corollary: When all agents have strict preferences, the man-optimal stable matching is the worst matching for the women; that is, it matches each woman with her least-preferred achievable mate.

Similarly, the woman-optimal stable matching matches each man with his least-preferred achievable mate.

✌ 2.26 Rural hospital theorem[1] (Theorem in McVitie and Wilson (1970), Theorem 1 in Gale and Sotomayor (1985)): The set of individuals who are matched is the same for all stable matchings.

Proof. (1) Suppose m is matched under μ' but not under μ. Then $m \in M(\mu')$.

(2) By decomposition theorem (Theorem 2.22), μ maps $M(\mu')$ to $W(\mu)$.

(3) So m is also matched under μ. Contradiction.

□

2.27 In $\langle M, W, \succsim\rangle$, when preferences are strict, for any two matchings μ and μ', define the following function on $M \cup W$:

$$\mu \vee_M \mu'(m) = \begin{cases} \mu(m), & \text{if } \mu(m) \succ_m \mu'(m) \\ \mu'(m), & \text{otherwise} \end{cases}, \quad \mu \vee_M \mu'(w) = \begin{cases} \mu(w), & \text{if } \mu'(w) \succ_w \mu(w) \\ \mu'(w), & \text{otherwise} \end{cases}.$$

This function assigns each man his more preferred mate from μ and μ', and it assigns each woman her less preferred mate.

Similarly, we can define the function $\mu \wedge_M \mu'$, which gives each man his less preferred mate and each woman her more preferred mate.

2.28 Remark: $\mu \vee_M \mu'$ may fail to be matchings due to the following two ways.

- $\mu \vee_M \mu'$ might assign the same woman to two different men.
- $\mu \vee_M \mu'$ might be that giving each man the more preferred of his mates at μ and μ' is not identical to giving each woman the less preferred of her mates.

Even when $\mu \vee_M \mu'$ and $\mu \wedge_M \mu'$ are matchings, they might not be stable.

2.29 Lattice theorem (Conway): When all preferences are strict, if μ and μ' are stable matchings for $\langle M, W, \succsim\rangle$, then the functions $\lambda = \mu \vee_M \mu'$ and $\nu = \mu \wedge_M \mu'$ are both stable matchings.

Proof. We only prove the statements for λ.

(1) By definition, $\mu \vee_M \mu'$ agrees with μ' on $M(\mu')$ and $W(\mu)$, and with μ otherwise.

(2) By decomposition theorem (Theorem 2.22), λ is therefore a matching.

(3) It is trivial that λ is not blocked by any individual in $\langle M, W, \succsim\rangle$.

(4) Suppose that some pair (m, w) blocks λ.

[1] This theorem is renamed as "屌丝孤独终身定理" by Xiaoguang Chen and Tianchen Song for fun.

(5) If $m \in M(\mu')$, then $w \succ_m \lambda(m) = \mu'(m) \succ_m \mu(m)$.

- If $w \in W(\mu)$, then $m \succ_w \lambda(w) = \mu'(w)$, and hence μ' is blocked by (m, w).
- If $w \in W \setminus W(\mu)$, then $m \succ_w \lambda(w) = \mu(w)$, and hence μ is blocked by (m, w).

(6) If $m \in M \setminus M(\mu')$, then $w \succ_m \lambda(m) = \mu(m) \succsim_m \mu'(m)$.

- If $w \in W(\mu)$, then $m \succ_w \lambda(w) = \mu'(w)$, and hence μ' is blocked by (m, w).
- If $w \in W \setminus W(\mu)$, then $m \succ_w \lambda(w) = \mu(w)$, and hence μ is blocked by (m, w).

(7) Therefore, λ is a stable matching.

□

2.30 Remark: The existence of man-optimal and woman-optimal stable matchings can be deduced from the lattice theorem.

A lattice is a partially ordered set in which every two elements have a supremum (also called a least upper bound or join) and an infimum (also called a greatest lower bound or meet). Lattice theorem (Theorem 2.29) implies that the set of stable matchings is a lattice under $\succsim_M$ (defined in 2.6), dual to $\succsim_W$.

✌ 2.31 Theorem on weak Pareto optimality for the men (Theorem 6 in Roth (1982b)): In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, there is no individually rational matching μ (stable or not) such that $\mu(m) \succ_m \mu^M(m)$ for all $m \in M$, where μ^M is the matching obtained by the men-proposing deferred acceptance algorithm.

Proof. (1) Suppose there exists such a matching μ.

(2) μ matches every man m to some woman $w \triangleq \mu(m)$ who has rejected him in the men-proposing deferred acceptance algorithm, so

$$\mu(m) \succ_m \mu^M(m) \succsim_m m$$

holds for every m, and hence $\mu(m) \in W$ for every m.

(3) Since μ^M is a stable matching, $\mu^M(w) \succ_w m = \mu(w)$.

(4) Since μ is individually rational, $\mu(w) \succsim_w w$, and hence

$$\mu^M(w) \succ_w m = \mu(w) \succsim_w w.$$

(5) Therefore, $\mu^M(w) \in M$ for every w with the form $w = \mu(m)$.

(6) Hence $\mu(M)$ have been matched under μ^M. That is, $\mu^M(\mu(M)) \subseteq M$.

(7) Since μ and μ^M are one-to-one and $\mu(M) \subseteq W$, we have $|\mu^M(\mu(M))| = |M|$, and hence $\mu^M(\mu(M)) = M$.

(8) Hence all of M have been matched under μ^M and $\mu^M(M) = \mu(M)$.

(9) Since all of M are matched under μ^M, any woman w who gets a proposal at the last step of the algorithm at which proposals were issued has not rejected any acceptable man.

(10) That is, the algorithm stops as soon as every woman in $\mu^M(M)$ has an acceptable proposal.

(11) Since every man prefers μ to μ^M, such a woman w must be single under μ, which contradicts the fact that $\mu^M(M) = \mu(M)$.

□

m_1	m_2	m_3	w_1	w_2	w_3
w_2	w_1	w_1	m_1	m_3	m_1
w_1	w_2	w_2	m_2	m_1	m_2
w_3	w_3	w_3	m_3	m_2	m_3

Table 2.4

2.32 Example: μ^M is not strongly Pareto optimal, that is, there exists an individually rational matching μ, such that $\mu(m) \succsim_m \mu^M(m)$ for all m, and $\mu(m_0) \succ_{m_0} \mu^M(m_0)$ for some $m_0 \in M$.

There are three men and three women, and their preferences are as follows:

Then

$$\mu^M = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_1 & m_2 & m_3 \end{bmatrix}.$$

Nevertheless

$$\mu = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_3 & m_1 & m_2 \end{bmatrix}$$

leaves m_2 no worse than under μ^M, but benefits m_1 and m_3.

2.33 Definition: In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, a matching μ' dominates another matching μ if and only if there exists a coalition $\emptyset \neq A \subseteq M \cup W$, such that $\mu'(i) \succ_i \mu(i)$ and $\mu'(i) \in A$ for any $i \in A$.

A matching μ is in the core defined via strict domination if there exists no matching μ' which dominates μ.

2.34 Theorem: In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, the core defined via strict domination equals to the set of stable matchings.

Proof. "$\Rightarrow$": (1) If μ is individually irrational, then it is dominated via a singleton coalition.

(2) If μ is blocked by the pair (m, w), then it is dominated via the coalition $\{m, w\}$ by any matching μ' with $\mu'(m) = w$.

"$\Leftarrow$": (1) If μ is not in the core, then μ is dominated by some matching μ' via a coalition A.

(2) If μ is not individually irrational, this implies $\mu'(w) \in M$ for all $w \in A$, since every woman $w \in A$ prefers $\mu'(w)$ to $\mu(w)$.

(3) Let $w \in A$ and $m = \mu'(w) \in A$. Then m prefers $w = \mu'(m)$ to $\mu(m)$ and μ is blocked by (m, w).

□

2.35 Theorem on strong stability property (Demange, Gale and Sotomayor (1987)): If μ is an unstable matching, then either there exists a blocking pair (m, w) and a stable matching $\bar{\mu}$ such that

$$\bar{\mu}(m) \succsim_m \mu(m) \text{ and } \bar{\mu}(w) \succsim_w \mu(w),$$

or μ is not individually rational.

2.36 Blocking lemma (Hwang (0000), Gale and Sotomayor (1985)): Let μ be any individually rational matching with respect to strict preferences $\succsim$ and let M' be all men who prefer μ to μ^M. If M' is non-empty, there is a pair (m, w) that blocked μ such that $m \in M \setminus M'$ and $w \in \mu(M')$.

Proof. Case 1: Suppose $\mu^M(M') \neq \mu(M')$.

(1) Choose $w \in \mu(M') \setminus \mu^M(M')$, say, $w = \mu(m')$.

(2) Then m' prefers μ to μ^M, that is, $w = \mu(m') \succ_{m'} \mu^M(m')$.

(3) Since μ^M is stable, we have $m \triangleq \mu^M(w) \succsim_w \mu(w) = m'$.

(4) Furthermore, $m = \mu^M(w) \succ_w \mu(w) = m'$; otherwise $m = \mu^M(w) = \mu(w) = m'$ contradicts with the fact $w \in \mu(M') \setminus \mu^M(M')$.

(5) Since $\mu^M(m) = w \notin \mu^M(M')$, m is not in M'.

(6) Hence, $\mu^M(m) \succsim_m \mu(m)$.

(7) Furthermore, $\mu^M(m) \succ_m \mu(m)$; otherwise $\mu(m') = w = \mu^M(m) = \mu(m)$.

(8) Hence (m, w) blocks μ.

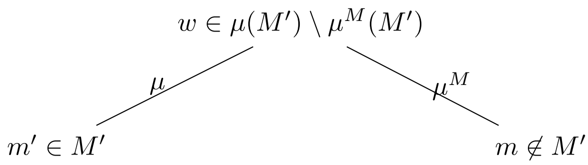


Figure 2.2

Case 2: Suppose $\mu^M(M') = \mu(M') \triangleq W'$.

(1) Let w be the last woman in W' to receive a proposal from an acceptable member of M' in the deferred acceptance algorithm.

(2) Since $\mu^M(M') = \mu(M')$ and each $m \in M'$ prefers $\mu(m)$ to $\mu^M(m)$, all $w \in W'$ have rejects acceptable men from M', and hence w has some man m engaged when she received this last proposal.

(3) We claim (m, w) is the desirable blocking pair.

- m is not in M'; otherwise, after being rejected by w, he will propose again to a member of W', contradicting the fact that w received the last such proposal.
- Since m is rejected by w, m prefers w to his mate $\mu^M(m)$ under μ^M. Since $m \notin M'$, m is not better off under μ than under μ^M, and hence m prefers w to $\mu(m)$.
- In the algorithm, m is the last man to be rejected by w, so she must have rejected her mate $\mu(m)$ under μ before she rejected m. Hence, she prefers m to $\mu(w)$.

□

2.37 Remark: Since $m \in M \setminus M'$, we have $\mu^M(m) \succsim_m \mu(m)$.

Since $w \in \mu(M')$, we have $w \triangleq \mu(m') \succ_{m'} \mu^M(m')$. Then by stability of μ^M we have $\mu^M(w) \succsim_w \mu(w)$.

2.38 *Proof of Theorem 2.35.* (1) If $\mu^M[\succsim] \succsim_M \mu$ is not satisfied, the set M' would be non-empty and the blocking pair (m, w) will satisfy

$$\mu^M[\succsim](m) \succsim_m \mu(m) \text{ and } \mu^M[\succsim](w) \succsim_w \mu(w),$$

so Theorem will be true with (m, w) and $\bar{\mu} = \mu^M$.

(2) Henceforth, we therefore assume

$$\mu^M[\succsim] \succsim_M \mu \text{ and symmetrically } \mu^W[\succsim] \succsim_W \mu.$$

(3) The set of stable matchings μ' such that $\mu' \succsim_M \mu$ is non-empty since it contains $\mu^M[\succsim]$, and it has a smallest element μ^*, since the set of stable matchings is a lattice under the partial order $\succsim_M$.

(4) If $\mu^*(w) \succ_w \mu(w)$ for some w, then Theorem holds with $(\mu^*(w), w)$ and μ^*. We can now restrict our consideration to the case where

$$\mu \succsim_W \mu^*.$$

(5) Define a new preference profiles $\succsim'$ by modifying $\succsim$ as follows:

- Each w who is matched under the stable matchings deletes from her preference list of acceptable men all m such that $\mu^*(w) \succ_w m$.
- If $\mu(w) \succ_w \mu^*(w)$, then $\mu^*(w)$ is also deleted.

Clearly the second item must hold for some w; otherwise $\mu = \mu^*$.

(6) Let $\mu^M[\succsim']$ be the man-optimal stable matching for $\langle M, W, \succsim' \rangle$. We will show that $\mu^M[\succsim']$ is the matching $\bar{\mu}$ of the Theorem.

(7) First we claim $\mu^M[\succsim']$ is stable under $\succsim$.

(i) Since $\mu^W[\succsim] \succsim_W \mu \succsim_W \mu^*$, $\mu^W[\succsim](w)$ is acceptable for w under $\succsim'$, and hence the woman-optimal stable matching $\mu^W[\succsim']$ in $\langle M, W, \succsim' \rangle$ is still $\mu^W[\succsim]$.

(ii) Since $\mu^W[\succsim]$ and $\mu^M[\succsim']$ are two stable matchings in $\langle M, W, \succsim' \rangle$, we have $\mu^M[\succsim'] \succsim'_M \mu^W[\succsim]$, which is equivalent to $\mu^M[\succsim'] \succsim_M \mu^W[\succsim]$ due to every man use the same preference in $\succsim$ and $\succsim'$.

(iii) Suppose w is single under $\mu^M[\succsim']$.

- Then w is also single under $\mu^W[\succsim]$, since both are stable matchings in $\langle M, W, \succsim' \rangle$.
- If w is part of a blocking pair for $\mu^M[\succsim']$ under $\succsim$, that is, there exists m, such that (m, w) blocks $\mu^M[\succsim']$ under $\succsim$.
- We have

$$m \succ_w \mu^M[\succsim'](w) = w, \text{and } w \succ_m \mu^M[\succsim'](m) \succsim_m \mu^W[\succsim](m).$$

- Since $\mu^W[\succsim]$ is stable in $\langle M, W, \succsim \rangle$, we have

$$w = \mu^W[\succsim](w) \succsim_w m,$$

which contradicts the fact $m \succ_w w$.

- Therefore, w can not be part of a blocking pair for $\mu^M[\succsim']$ under $\succsim$.

(iv) Suppose w is matched under $\mu^M[\succsim']$.

- Then she prefers her mate to the men she has deleted.
- Hence she can not block with any deleted man and hence she belongs to no blocking pair.

(8) Next we show that $\mu^* \succsim_M \mu^M[\succsim']$.

(i) If not, we have $w \triangleq \mu^M[\succsim'](m) \succ_m \mu^*(m)$.

(ii) Then by stability of μ^* we have $\mu^*(w) \succ_w m$.

(iii) By the definition of $\succsim'$, m is deleted by w, so $w = \mu^M[\succsim'](m)$ is impossible.

(9) It follows that $\mu(m) \succ_m \mu^M[\succsim'](m)$ for at least one m.

(i) If not we have $\mu^* \succsim_M \mu^M[\succsim'] \succsim_M \mu$.

(ii) By the definition of $\succsim'$, $\mu^M[\succsim'] \neq \mu^*$.

(iii) It contradicts that μ^* is the smallest stable matching preferred by M to μ.

(10) Finally, we apply the blocking lemma to the preference profile $\succsim'$ for which $\mu^M[\succsim']$ is man-optimal.

(11) Then there is a blocking pair (m_0, w_0) for μ under $\succsim'$ and hence under $\succsim$.

(12) The proof is complete with $\bar{\mu} = \mu^M[\succsim']$ as claimed, under the assumption that preferences are strict, by Remark 2.37.

(13) To prove the theorem without the assumption that preferences are strict, we need the following additional observation. Let μ be an unstable matching under non-strict preferences $\succsim$. Then there exists a way to break ties so that the strict preferences $\succsim'$ correspond to $\succsim$, and every pair (m, w) that blocks μ under $\succsim'$ also blocks μ under $\succsim$: If any agent x is indifferent under $\succsim$ between $\mu(x)$ and some other alternative, then under $\succsim'$, x prefers $\mu(x)$. Then the theorem applied to the case of the strict preferences $\succsim'$ gives the desired result.

□

2.5 Extension: Extending the men's preferences

2.39 Example: The effect of extending the men's preferences.

In the marriage problem $\Gamma = \langle M, W, \succsim \rangle$, there are six men and five women, and their preferences are given as follows:

m_1	m_2	m_3	m_4	m_5	m_6	w_1	w_2	w_3	w_4	w_5
w_1	w_2	w_4	w_3	w_5	w_1	m_2	m_6	m_3	m_4	m_5
w_3	w_4	w_3	w_4		w_4	m_1	m_1	m_4	m_3	
						m_6	m_2	m_1	m_2	
								m_2		

The man-optimal and woman-optimal stable matchings are given by:

$$\mu^M[\succsim] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_6) \\ m_1 & m_2 & m_4 & m_3 & m_5 & m_6 \end{bmatrix}, \quad \mu^W[\succsim] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_6) \\ m_1 & m_2 & m_3 & m_4 & m_5 & m_6 \end{bmatrix}.$$

Consider a new marriage problem $\Gamma' = \langle M, W, \succsim' \rangle$ some of men decide to extend their lists of acceptable women yielding the new preference profile $\succsim'$:

m_1	m_2	m_3	m_4	m_5	m_6	w_1	w_2	w_3	w_4	w_5
w_1	w_2	w_4	w_3	w_5	w_1	m_2	m_6	m_3	m_4	m_5
w_3	w_4	w_3	w_4	w_3	w_4	m_1	m_1	m_4	m_3	
w_2	w_1	w_2			w_2	m_6	m_2	m_1	m_2	
								m_2		

In this case the man-optimal and woman-optimal stable matchings are:

$$\mu^M[\succsim'] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_1) \\ m_2 & m_6 & m_4 & m_3 & m_5 & m_1 \end{bmatrix}, \quad \mu^W[\succsim'] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_1) \\ m_2 & m_6 & m_3 & m_4 & m_5 & m_1 \end{bmatrix}.$$

Under the original preferences $\succsim$, no man is worse off, and no woman is better off at $\mu^M[\succsim]$ (resp. $\mu^W[\succsim]$) than at $\mu^M[\succsim']$ (resp. $\mu^W[\succsim']$).

2.40 Notation: We will write $\succsim'_m \rhd \succsim_m$ if $\succsim'_m$ is an extension of $\succsim_m$ by adding people to the end of the original list of acceptable people. Similarly, we will write $\succsim'_w \rhd \succsim_w$ and finally we will write $\succsim' \rhd_M \succsim$ if $\succsim'_m \rhd \succsim_m$ for all $m \in M$.

Note that for any woman w, her preferences in $\succsim'$ and $\succsim$ are same when $\succsim' \rhd_M \succsim$.

2.41 Decomposition lemma (Lemma 1 in Gale and Sotomayor (1985)): Let μ and μ' be, respectively, stable matchings in $\langle M, W, \succsim\rangle$ and $\langle M, W, \succsim'\rangle$ with $\succsim' \rhd_M \succsim$, and all preferences are strict. Let $M(\mu')$ be the set of men who prefers μ' to μ under $\succsim$ and let $W(\mu)$ be the set of women who prefer μ to μ'. Then μ' and μ are bijections from $M(\mu')$ to $W(\mu)$. (That is, both μ' and μ match any man who prefers μ' to a woman who prefers μ, and vice versa.)

Proof. (1) For any $m \in M(\mu')$, we have $\mu'(m) \succ_m \mu(m) \succsim_m m$, where the second equation holds since μ is stable and not blocked by any individual.

(2) Then $\mu'(m) \neq m$, and hence $\mu'(m) \in W$, denoted by w. So we have $w = \mu'(m) \succ_m \mu(m)$.

(3) Since μ is a stable matching in $\langle M, W, \succsim\rangle$, $\mu(w) \succsim_w m = \mu'(w)$; otherwise the pair (m, w) blocks μ.

(4) Furthermore, $\mu(w) \succ_w \mu'(w)$ otherwise $\mu'(m) = w = \mu(m)$.

(5) We have $\mu'(m) = w \in W(\mu)$, and hence $\mu'(M(\mu')) \subseteq W(\mu)$.

(6) For any $w \in W(\mu)$, we have $\mu(w) \succ_w \mu'(w) \succsim_w w$, where the second equation holds since μ' is stable and not blocked by any individual.

(7) Then $\mu(w) \in M$, denoted by m.

(8) Since μ' is a stable matching in $\langle M, W, \succsim'\rangle$, $\mu'(m) \succ'_m \mu(m)$; otherwise the pair (m, w) blocks μ'.

(9) We have $\mu'(m) \succ'_m \mu(m) = w$ and $\mu(m) \succ_m m$, then $\mu'(m) \succ'_m \mu(m) \succ_m m$, and hence $\mu'(m) \succ_m \mu(m) = w$.

(10) We have $m \in M(\mu')$ and hence $\mu(W(\mu)) \subseteq M(\mu')$.

(11) Since μ and μ' are one-to-one and $M(\mu')$ and $W(\mu)$ are finite, the conclusion follows.

□

2.42 Remark: μ and μ' are not bijections from $M(\mu)$ to $W(\mu')$.

Consider the Example 2.39. Let

$$\mu \triangleq \mu^M[\succsim] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_6) \\ m_1 & m_2 & m_4 & m_3 & m_5 & m_6 \end{bmatrix}, \quad \mu' \triangleq \mu^M[\succsim'] = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & w_5 & (m_1) \\ m_2 & m_6 & m_4 & m_3 & m_5 & m_1 \end{bmatrix}.$$

Then it is clear that there is no bijection between $M(\mu)$ and $W(\mu')$, where

$$M(\mu) = \{m_1, m_2, m_6\} \text{ and } W(\mu') = \{w_1, w_2\}.$$

2.43 Lattice lemma: Let μ and μ' be, respectively, stable matchings in $\langle M, W, \succsim\rangle$ and $\langle M, W, \succsim'\rangle$ with $\succsim' \rhd_M \succsim$, and all preferences are strict. Then we have

- $\lambda = \mu \vee_M \mu'$, under $\succsim$, is a matching and is stable for $\langle M, W, \succsim\rangle$.
- $\nu = \mu \wedge_M \mu'$, under $\succsim$, is a matching and is stable for $\langle M, W, \succsim'\rangle$.

Proof. We only prove the first statement.

(1) By definition, $\mu \vee_M \mu'$ agrees with μ' on $M(\mu')$ and $W(\mu)$, and with μ otherwise.

(2) By decomposition lemma, λ is therefore a matching.

(3) For $m \in M(\mu')$, we have $\mu'(m) \succ_m \mu(m) \succsim_m m$ so $\mu'(m)$ is acceptable to m under $\succsim$, and hence λ is not blocked by any individual in $\langle M, W, \succsim\rangle$.

(4) Suppose that some pair (m, w) blocks λ.

(5) If $m \in M(\mu')$, then $w \succ_m \lambda(m) = \mu'(m) \succ_m \mu(m)$.

- If $w \in W(\mu)$, then $m \succ_w \lambda(w) = \mu'(w)$, and hence μ' is blocked by (m, w).
- If $w \in W \setminus W(\mu)$, then $m \succ_w \lambda(w) = \mu(w)$, and hence μ is blocked by (m, w).

(6) If $m \in M \setminus M(\mu')$, then $w \succ_m \lambda(m) = \mu(m) \succsim_m \mu'(m)$.

- If $w \in W(\mu)$, then $m \succ_w \lambda(w) = \mu'(w)$, and hence μ' is blocked by (m, w).
- If $w \in W \setminus W(\mu)$, then $m \succ_w \lambda(w) = \mu(w)$, and hence μ is blocked by (m, w).

(7) Therefore, λ is a stable matching.

□

2.44 Theorem (Gale and Sotomayor (1985)): Suppose $\succsim' \rhd_M \succsim$, and let $\mu^M[\succsim']$, $\mu^M[\succsim]$, $\mu^W[\succsim']$ and $\mu^W[\succsim]$ be the corresponding optimal matchings. Then under the preference $\succsim$ the men are not worse off and the women are not better off in $\langle M, W, \succsim\rangle$ than in $\langle M, W, \succsim'\rangle$, no matter which of the two optimal matchings are considered. That is,

$$\mu^M[\succsim] \succsim_M \mu^M[\succsim'], \text{ and } \mu^W[\succsim'] \succsim_W \mu^W[\succsim].$$

Proof. (1) By lattice lemma (Lemma 2.43), $\mu^M[\succsim] \vee_M \mu^M[\succsim']$ under $\succsim$ is stable for $\langle M, W, \succsim\rangle$.

(2) Then by optimality we have $\mu^M[\succsim] \succsim_M \left(\mu^M[\succsim] \vee_M \mu^M[\succsim']\right) \succsim_M \mu^M[\succsim']$.

(3) Also by lattice lemma (Lemma 2.43), $\mu^W[\succsim] \vee_W \mu^W[\succsim']$ under $\succsim$ is stable for $\langle M, W, \succsim'\rangle$.

(4) Then by optimality we have $\mu^W[\succsim'] \succsim_W \left(\mu^W[\succsim] \vee_W \mu^W[\succsim']\right) \succsim_W \mu^W[\succsim]$.

□

2.45 Corollary: $\mu^M[\succsim'] \succsim_W \mu^M[\succsim]$ by the stability of $\mu^M[\succsim']$ and $\mu^W[\succsim] \succsim_M \mu^W[\succsim']$ by the stability of $\mu^W[\succsim]$.

2.6 Extension: Adding another woman

2.46 Example: Effect of adding another woman.

In the marriage problem $\Gamma = \langle M, W, \succsim\rangle$, where there are three men and three women, and their preferences are as follows:

m_1	m_2	m_3	w_1	w_2	w_3
w_1	w_3	w_1	m_1	m_2	m_3
w_3	w_2	w_3	m_3		m_2

Table 2.5

There is a single stable matching in this example:

$$\mu^M[\Gamma] = \mu^W[\Gamma] = \begin{bmatrix} w_1 & w_2 & w_3 \\ m_1 & m_2 & m_3 \end{bmatrix}.$$

Suppose woman w_4 now enters, and the new marriage problem $\Gamma' = \langle M, W', \succsim' \rangle$ is given by $W' = \{w_1, w_2, w_3, w_4\}$, and $\succsim'$ given by:

m_1	m_2	m_3	w_1	w_2	w_3	w_4
w_4	w_3	w_1	m_1	m_2	m_3	m_2
w_1	w_2	w_3	m_3		m_2	m_1
w_3						

Table 2.6

Again there is a single stable matching under $\succsim'$;

$$\mu^M(\Gamma') = \mu^W(\Gamma') = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ m_3 & (w_2) & m_2 & m_1 \end{bmatrix}$$

Under the preferences $\succsim'$, all the men are better off under $\mu^M[\Gamma']$ than under $\mu^M[\Gamma]$.

2.47 Theorem (Gale and Sotomayor (1985)): Suppose $W \subseteq W'$ and $\mu^M[\Gamma]$ and $\mu^W[\Gamma]$ are the man-optimal and woman-optimal matchings, respectively, for $\Gamma = \langle M, W, \succsim \rangle$. Let $\mu^M[\Gamma']$ and $\mu^W[\Gamma']$ be the man-optimal and woman-optimal matchings, respectively, for $\Gamma' = \langle M, W', \succsim' \rangle$, where $\succsim'$ agrees with $\succsim$ on M and W. Then

$$\mu^W[\Gamma] \succsim_W \mu^W[\Gamma'],\ \mu^W[\Gamma'] \succsim'_M \mu^W[\Gamma],\ \mu^M[\Gamma'] \succsim'_M \mu^M[\Gamma],\ \mu^M[\Gamma] \succsim_W \mu^M[\Gamma'].$$

Proof. (1) Denote by $\succsim''$ the set of preferences on $M \cup W'$ such that $\succsim''$ agrees with $\succsim'$ on $M \cup W$, and for each $w \in W' \setminus W$, w has no acceptable man under $\succsim''$.

(2) Let $\mu^M[\Gamma'']$ and $\mu^W[\Gamma'']$ be the man-optimal and woman-optimal stable matchings for $\Gamma'' = \langle M, W', \succsim'' \rangle$.

(3) Since no man is acceptable to any woman in $W' \setminus W$ under $\succsim''$, $\mu^M[\Gamma'']$ agrees with $\mu^M[\Gamma]$ on $M \cup W$, and $\mu^W[\Gamma'']$ agrees with $\mu^W[\Gamma]$ on $M \cup W$.

(4) Note that $\succsim' \rhd_W \succsim''$.

(5) So we can apply Theorem 2.44 and obtain that

$$\mu^W[\Gamma''] \succsim''_{W'} \mu^W[\Gamma'],$$

so $\mu^W[\Gamma] \succsim_W \mu^W[\Gamma']$.

(6) Similarly, $\mu^W[\Gamma'] \succsim'_M \mu^W[\Gamma'']$ so $\mu^W[\Gamma'] \succsim'_M \mu^W[\Gamma]$.

(7) Similarly, $\mu^M[\Gamma'] \succsim'_M \mu^M[\Gamma'']$ so $\mu^M[\Gamma'] \succsim'_M \mu^M[\Gamma]$.

(8) Finally, $\mu^M[\Gamma''] \succsim''_{W'} \mu^M[\Gamma']$ so $\mu^M[\Gamma] \succsim_W \mu^M[\Gamma']$.

□

2.48 Remark: Theorem 2.47 states that when new women enter, no man is hurt under the man-optimal matchings.

2.49 Theorem: Suppose a woman w_0 is added and let $\mu^W[\Gamma']$ be the woman-optimal stable matching for $\Gamma' = \langle M, W' = W \cup \{w_0\}, \succsim' \rangle$, where $\succsim'$ agrees with $\succsim$ on W. Let $\mu^M[\Gamma]$ be the man-optimal stable matching for $\Gamma = \langle M, W, \succsim \rangle$. If w_0 is not single under $\mu^W[\Gamma']$, then there exists a non-empty subset of men, S, such that if a man is in S he is better off, and if a woman is in $\mu^M[\Gamma](S)$ she is worse off under any stable matching for the new marriage problem than under any stable matching for the original marriage problem, under the new (strict) preferences $\succsim'$.

Proof. (1) Let $\mu^W[\Gamma'](w_0) = m_0$.

(2) If m_0 is single under $\mu^M[\Gamma]$, then Theorem holds by taking $S = \{m_0\}$.

(3) So suppose m_0 is matched to $w_1 \in W$ under $\mu^M[\Gamma]$.

(4) It suffices to show that there exists a set of men S such that

$$\mu^W[\Gamma'](m) \succ'_m \mu^M[\Gamma] \text{ for all } m \in S, \text{ and } \mu^M[\Gamma](w) \succ_w \mu^W[\Gamma'] \text{ for any } w \in \mu^M[\Gamma](S).$$

(5) Construct a directed graph whose vertices are $M \cup W$. There are two type of arcs.

- If $m \in M$ and $\mu^M[\Gamma](m) = w \in W$, there is an arc from m to w.
- If $w \in W$ and $\mu^W[\Gamma'](w) = m \in M$, there is an arc from w to m.

(6) Let $\bar{M} \cup \bar{W}$ be all vertices that can be reached by a directed path starting from m_0.

(7) Case 1: The path starting from m_0 ends at w_{k+1}, that is,

$$\begin{tikzcd}
m_0 \arrow[r, "\mu^M[\Gamma]"'] & w_1 \arrow[r, "\mu^W[\Gamma']"] & m_1 \arrow[r, "\mu^M[\Gamma]"'] & w_2 \arrow[r, "\mu^W[\Gamma']"] & m_2 \arrow[d, dashed] \\
w_{i+1} \arrow[d, dashed] & m_i \arrow[l, "\mu^M[\Gamma]"] & w_i \arrow[l, "\mu^W[\Gamma']"'] & m_{i-1} \arrow[l, "\mu^M[\Gamma]"] & w_{i-1} \arrow[l, "\mu^W[\Gamma']"'] \\
m_{k-1} \arrow[r, "\mu^M[\Gamma]"'] & w_k \arrow[r, "\mu^W[\Gamma']"] & m_k \arrow[r, "\mu^M[\Gamma]"'] & w_{k+1} \arrow[r, "\mu^W[\Gamma']"] & w_{k+1}
\end{tikzcd}$$

Figure 2.3

(i) We claim that $S = \{m_0, m_1, \ldots, m_k\}$ has the desired property. $\mu^M[\Gamma](S) = \{w_1, w_2, \ldots, w_{k+1}\}$

(ii) $m_k = \mu^M[\Gamma](w_{k+1}) \succ_{w_{k+1}} w_{k+1} = \mu^W[\Gamma'](w_{k+1})$ implies

$$w_k = \mu^W[\Gamma'](m_k) \succ_{m_k} w_{k+1} = \mu^M[\Gamma](m_k).$$

(iii) Then $m_{k-1} = \mu^M[\Gamma](w_k) \succ_{w_k} m_k = \mu^W[\Gamma'](w_k)$.

(iv) By induction, we have

$$\mu^W[\Gamma'](m_i) \succ_{m_i} \mu^M[\Gamma](m_i),\ i = 0, 1, \ldots, k$$
$$\mu^M[\Gamma](w_j) \succ_{w_j} \mu^W[\Gamma'](w_j),\ j = 1, 2, \ldots, k+1.$$

(8) Case 2: The path starting from m_0 ends at m_k, that is,

(i) We claim that $S = \{m_0, m_1, \ldots, m_k\}$ has the desired property. $\mu(S) = \{w_1, w_2, \ldots, w_k\}$.

(ii) $w_k = \mu^W[\Gamma'](m_k) \succ_{m_k} \mu^M[\Gamma](m_k) = m_k$ implies

$$m_{k-1} = \mu^M[\Gamma](w_k) \succ_{w_k} m_k = \mu^W[\Gamma'](w_k).$$

(iii) Then $w_{k-1} = \mu^W[\Gamma'](m_{k-1}) \succ_{m_{k-1}} w_k = \mu^M[\Gamma](m_{k-1})$.

$$
\begin{tikzcd}
m_0 \arrow[r, "\mu^M[\Gamma]"'] & w_1 \arrow[r, "\mu^W[\Gamma']"] & m_1 \arrow[r, "\mu^M[\Gamma]"'] & w_2 \arrow[r, "\mu^W[\Gamma']"] & m_2 \arrow[d, dashed] \\
w_{i+1} \arrow[d, dashed] & m_i \arrow[l, "\mu^M[\Gamma]"] & w_i \arrow[l, "\mu^W[\Gamma']"'] & m_{i-1} \arrow[l, "\mu^M[\Gamma]"] & w_{i-1} \arrow[l, "\mu^W[\Gamma']"'] \\
m_{k-1} \arrow[r, "\mu^M[\Gamma]"'] & w_k \arrow[r, "\mu^W[\Gamma']"] & m_k \arrow[r, "\mu^M[\Gamma]"'] & m_k &
\end{tikzcd}
$$

Figure 2.4

(iv) By induction, we have

$$\mu^W[\Gamma'](m_i) \succ_{m_i} \mu^M[\Gamma](m_i),\ i = 0, 1, \ldots, k$$
$$\mu^M[\Gamma](w_j) \succ_{w_j} \mu^W[\Gamma'](w_j),\ j = 1, 2, \ldots, k.$$

□

2.50 Remark:

2.7 Incentive compatibility

2.51 A (direct) mechanism φ is a systematic procedure that determines a matching for each marriage problem $\langle M, W, \succsim \rangle$. Note that M, W and $\succsim$ are all allowed to vary.

2.52 A mechanism φ is stable if it is always selects a stable matching.

A mechanism φ is Pareto efficient if it is always selects a Pareto efficient matching.

A mechanism φ is individually rational if it is always selects an individually rational matching.

2.53 Let φ^M and φ^W be the mechanisms that selects the man-optimal and woman-optimal stable matchings respectively.

2.54 Let $M = \{m_1, \ldots, m_p\}$ and $W = \{w_1, \ldots, w_q\}$ be the sets of men and women. Let $\mathcal{P}_m$ denote the set of all preferences for man m, $\mathcal{P}_w$ denote the set of all preferences for woman w, $\mathcal{P} = (\mathcal{P}_m)^p \times (\mathcal{P}_w)^q$ denote the set of all preference profiles, and $\mathcal{P}_{-i}$ denote the set of all preference profiles for all individuals except i. Let $\mathcal{M}$ denote the set of all matchings.

2.55 In a marriage problem $\langle M, W, \succsim \rangle$, we assume that everything is known except $\succsim$. Therefore, people are the only strategic agents in the problem and can manipulate the mechanism by misreporting their preferences.

When other components of the problem are clear, we represent the problem just by $\succ$, represent the outcome of the mechanism by $\varphi[\succ]$, and a mechanism becomes a function $\varphi\colon \mathcal{P} \to \mathcal{M}$.

2.56 A mechanism φ is strategy-proof if for any M and W, for each $i \in M \cup W$, for each $\succsim_i, \succsim_i' \in \mathcal{P}_i$, for each $\succsim_{-i} \in \mathcal{P}_{-i}$,

$$\varphi[\succsim_{-i}, \succsim_i](i) \succsim_i \varphi[\succsim_{-i}, \succsim_i'](i).$$

2.57 Example: A strategy-proof and Pareto efficient mechanism.

For any marriage problem $\langle M, W, \succsim\rangle$, let the men be placed in some order, $\{m_1, \ldots, m_p\}$. Consider the mechanism that for any stated preference profile $\succsim'$ yields the matching $\mu = \varphi[\succsim']$ that matches m_1 to his stated first choice, m_2 to his stated first choice of possible mates remaining after $\mu(m_1)$ has been removed from the market, and any m_k to his stated first choice after $\mu(m_1)$ through $\mu(m_{k-1})$.

- It is clearly a dominant strategy for each man to state his true preferences, since each man is married to whomever he indicates is his first choice among those remaining when his turn comes. It is also (degenerately) a dominant strategy for each woman to state her true preferences, since the preferences stated by the women have no influence.
- The mechanism φ is Pareto efficient, since at any other matching some man would do no better.
- However φ is not a stable matching mechanism, since it might happen, for example, that woman $w = \varphi[\succsim](m_1)$, who is the (draft) choice of man m_1 would prefer to be matched with someone else, who would also prefer to be matched to her. That is, φ is not a stable matching mechanism because there are some sets of preferences for which it will produce unstable outcomes.

✌ 2.58 Impossibility theorem (Theorem 3 in Roth (1982b)): There exists no mechanism that is both stable and strategy-proof.

Proof. (1) Consider the following marriage problem with two men and two women with preferences $\succsim$ given by:

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
w_2	w_1	m_1	m_2

Table 2.7

(2) In this problem there are only two stable matchings:

$$\mu^M = \begin{bmatrix} m_1 & m_2 \\ w_1 & w_2 \end{bmatrix} \text{ and } \mu^W = \begin{bmatrix} m_1 & m_2 \\ w_2 & w_1 \end{bmatrix}.$$

(3) Let φ be any stable mechanism. Then $\varphi[\succsim] = \mu^M$ or $\varphi[\succsim] = \mu^W$.

(4) If $\varphi[\succsim] = \mu^M$ then woman w_1 can report a fake preference $\succsim'_{w_1}$ where only her top choice m_2 is acceptable and force her favorite stable matching μ^W to be selected by φ since it is the only stable matching for the marriage problem $(\succsim_{-w_1}, \succsim'_{w_1})$.

(5) If, on the other hand, $\varphi[\succsim] = \mu^W$, then man m_1 can report a fake preference $\succsim'_{m_1}$ where only his top choice w_1 is acceptable and force his favorite stable matching μ^M to be selected by φ since it is the only stable matching for the marriage problem $(\succsim_{-m_1}, \succsim'_{m_1})$.

□

2.59 Corollary: No stable mechanism exists for which stating the true preferences is always a best response for every individual when all other individuals state their true preferences.

2.60 Theorem: When any stable mechanism is applied to a marriage problem in which preferences are strict and there is more than one stable matching, then at least one individual can profitably misreport his or her preference, assuming the others tell the truth.

Proof. (1) By hypothesis $\mu^M \neq \mu^W$.

(2) Without loss of generality, suppose that when all individuals state their true preferences, the mechanism selects a stable matching $\mu \neq \mu^W$.

(3) Let w be any woman such that $\mu^W(w) \succ_w \mu(w)$. Note that w is not single at μ^W.

(4) Let w misreport her preference by removing from her stated preference list of acceptable men all men who rank below $\mu^W(w)$.

(5) Clearly the matching μ^W will still be stable under this preference profile.

(6) Letting μ' be the stable matching selected by the mechanism for these new preference profile.

(7) It follows from rural hospital theorem (Theorem 2.26) that w is not single under μ'.

(8) Hence she is matched with someone she likes at least as well as $\mu^W(w)$, since all other men have been removed from her list of acceptable men.

(9) It is clear that μ' is also stable for the original preference profile.

(10) Then $\mu^W(w) \succsim_w \mu'(w)$ due to the woman-optimality of μ^W.

(11) It follows that $\mu^W(w) = \mu'(w)$, and hence $\mu'(w) \succ_w \mu(w)$.

(12) Therefore, w prefers matching μ' to μ.

(13) If the mechanism originally selects the matching μ^W, then the symmetric argument can be made for any man m who strictly prefers μ^M.

□

✌ 2.61 Theorem (Proposition 1 in Alcalde and Barberà (1994)): There exists no mechanism that is Pareto efficient, individually rational, and strategy-proof.

Proof. (1) Consider the following marriage problem with two men and two women with preferences $\succsim^1$ given by:

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
w_2	w_1	m_1	m_2

Table 2.8

(2) In this problem there are only two individually rational, Pareto efficient matchings:

$$\mu_1^1 = \begin{bmatrix} m_1 & m_2 \\ w_1 & w_2 \end{bmatrix} \text{ and } \mu_2^1 = \begin{bmatrix} m_1 & m_2 \\ w_2 & w_1 \end{bmatrix}.$$

(3) Let φ be any individually rational, and Pareto efficient mechanism. Then $\varphi[\succsim^1] = \mu_1^1$ or $\varphi[\succsim^1] = \mu_2^1$.

(4) If $\varphi[\succsim^1] = \mu_1^1$. Then consider the marriage problem with two men and two women with preferences $\succsim^2$ given by:

In this problem there are only two individually rational, Pareto efficient matchings:

$$\mu_1^2 = \begin{bmatrix} m_1 & m_2 & (w_1) \\ (m_1) & w_2 & w_1 \end{bmatrix} \text{ and } \mu_2^2 = \begin{bmatrix} m_1 & m_2 \\ w_2 & w_1 \end{bmatrix}.$$

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
w_2	w_1		m_2

Table 2.9

- If $\varphi[\succsim^2] = \mu_2^2$, w_1 can manipulate φ at $\succsim^1$ via $\succsim^2_{w_1}$: w_1 will get m_1 if reporting true preference $\succsim^1_{w_1}$, and get m_2 if misreporting $\succsim^2_{w_1}$.
- If $\varphi[\succsim^2] = \mu_1^2$, then consider the marriage problem with two men and two women with preferences $\succsim^3$ given by:

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
w_2	w_1		

Table 2.10

In this problem there is only one individually rational, Pareto efficient matching:

$$\mu^3 = \begin{bmatrix} m_1 & m_2 \\ w_2 & w_1 \end{bmatrix}.$$

w_2 can manipulate at $\succsim^2$ via $\succsim^3_{w_2}$: w_2 will get m_2 if reporting the true preference $\succsim^2_{w_2}$, and get w_1 if misreporting $\succsim^3_{w_2}$.

(5) If $\varphi[\succsim] = \mu_2^1$. Then consider the marriage problem with two men and two women with preferences $\succsim^4$ given by:

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
	w_1	m_1	m_2

Table 2.11

In this problem there are only two individually rational, Pareto efficient matchings:

$$\mu_1^4 = \begin{bmatrix} m_1 & m_2 \\ w_1 & w_2 \end{bmatrix} \text{ and } \mu_2^4 = \begin{bmatrix} m_1 & m_2 & (w_2) \\ (m_1) & w_1 & w_2 \end{bmatrix}.$$

- If $\varphi[\succsim^4] = \mu_1^4$, m_1 can manipulate φ at $\succsim^1$ via $\succsim^4_{m_1}$: m_1 will get w_2 if reporting true preference $\succsim^1_{m_1}$, and get w_1 if misreporting $\succsim^4_{m_1}$.
- If $\varphi[\succsim^4] = \mu_2^4$, then consider the marriage problem with two men and two women with preferences $\succsim^5$ given by:
  In this problem there is only one individually rational, Pareto efficient matching:

$$\mu^5 = \begin{bmatrix} m_1 & m_2 \\ w_1 & w_2 \end{bmatrix}.$$

m_2 can manipulate at $\succsim^4$ via $\succsim^5_{m_2}$: m_2 will get w_1 if reporting the true preference $\succsim^5_{m_2}$, and get w_2 if misreporting $\succsim^5_{m_2}$.

$\square$

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
		m_1	m_2

Table 2.12

2.62 Theorem (Theorem 9 in Dubins and Freedman (1981), Theorem 5 in Roth (1982b)): Truth-telling is a weakly dominant strategy for any man under the man-optimal stable mechanism. Similarly truth-telling is a weakly dominant strategy for any woman under the woman-optimal stable mechanism.

Proof. It is a corollary of theorem of limits on successful manipulation (Theorem 2.70). □

2.63 Simple misreport manipulation lemma (Lemma 1 in Roth (1982b)): Let m be in M. Let $\mu^M[\succsim']$ and $\mu^M[\succsim'']$ be the corresponding man-optimal stable matchings for $\langle M, W, \succsim'\rangle$ and $\langle M, W, \succsim''\rangle$, where $\succsim'_i = \succsim''_i$ for all agents i other than m, and $\mu^M[\succsim'](m)$ is the first choice for m in $\succsim''_m$. Then $\mu^M[\succsim''](m) = \mu^M[\succsim'](m)$.

Proof. (1) Clearly the matching $\mu^M[\succsim']$ is stable under the preference profile $\succsim''$.

(2) Since $\mu^M[\succsim'']$ is man-optimal in $\langle M, W, \succsim''\rangle$ and $\mu^M[\succsim'](m)$ is the first choice of $\succsim''_m$, we have $\mu^M[\succsim'](m) = \mu^M[\succsim''](m)$.

□

2.64 Remark: There are of course many ways in which a man m might report a preference ordering $\succsim'_m$ different from $\succsim_m$, but this lemma shows that, in considering man m's incentives to misreport his preferences, we can confine our attention to certain kinds of simple misreport.

Suppose by reporting some preference $\succsim'_m$, man m can change his mate from $\mu^M[\succsim](m)$ to $\mu^M[\succsim'](m)$. Then he can get the same result—that is, he can be matched to $\mu^M[\succsim'](m)$—by reporting a preference $\succsim''_m$ in which $\mu^M[\succsim'](m)$ is his first choice. So, if there is any way for m to be matched to $\mu^M[\succsim'](m)$ by reporting some appropriate preference, then there is a simple way—he can just list her as his first choice.

2.65 Lemma (Lemma 2 in Roth (1982b)): Let m be in M. Let $\mu^M[\succsim']$ be the man-optimal stable matching for $\langle M, W, \succsim'\rangle$. If $\succsim'_i = \succsim_i$ for all i other than m and $\mu^M[\succsim'](m)$ is the first choice for m in $\succsim'_m$, and $\mu^M[\succsim'](m) \succsim_m \mu^M[\succsim](m)$, then for each m_j in M we have $\mu^M[\succsim'](m_j) \succsim_{m_j} \mu^M[\succsim](m_j)$.

Proof. (1) Let $M^* = \{m_j \mid \mu^M[\succsim](m_j) \succ_{m_j} \mu^M[\succsim'](m_j)\}$. Suppose $M^* \neq \emptyset$.

(2) It is clear that all m_j in M^* are matched under $\mu^M[\succsim]$.

(3) Since every individual other than m reports the same preferences under $\succsim$ and $\succsim'$ and $m \notin M^*$, it must be that all m_j in M^* are rejected by their mates under $\succsim^M[\succsim]$ at some step of the deferred acceptance algorithm in $\langle M, W, \succsim'\rangle$.

(4) Let s be the first step of the algorithm in $\langle M, W, \succsim'\rangle$ at which some m_j in M^* is rejected by $w \triangleq \mu^M[\succsim](m_j)$.

(5) Since m_j and w are mutually acceptable, this implies that w must receive a proposal at Step s of the algorithm for $\langle M, W, \succsim'\rangle$ from some m_k who did not propose to her under $\succsim$ and whom she likes more than m_j.

(6) The fact that m_k did not propose to w under $\succsim$ means that $\mu^M[\succsim](m_k) \succ_{m_k} w$.

(7) Then $m_k \in M^*$; otherwise we have the contradiction

$$w \succsim_{m_k} \mu^M[\succsim'](m_k) \succsim_{m_k} \mu^M[\succsim](m_k) \succ_{m_k} w,$$

where the first relation holds because in deferred acceptance algorithm for $\langle M, W, \succsim' \rangle$, m_k is on the waiting list of w at Step s.

(8) So $m_k \neq m$ and $\succsim_{m_k} = \succsim'_{m_k}$ and m_k must have been rejected by $\mu^M[\succsim](m_k)$ in $\langle M, W, \succsim' \rangle$ prior to Step s, which contradicts the choice of s as the first such period.

(9) Consequently, $M^* = \emptyset$ and $\mu^M[\succsim'](m_j) \succsim_{m_j} \mu^M[\succsim](m_j)$ for all m_j in M.

□

2.66 Remark: Lemma shows that if a simple misreport by m leaves m at least as well off as at $\mu^M[\succsim]$, then no man will suffer; that is, every man likes the matching $\mu^M[\succsim']$ resulting from the misreport at least as well as the matching $\mu^M[\succsim]$. This illustrates another way in which the men have common rather than conflicting interests.

2.67 Theorem (Theorem 17 in Dubins and Freedman (1981)): Let $\succsim$ be the true preferences of the agents, and let $\succsim'$ differ from $\succsim$ in that some coalition $\bar{M}$ of the men misreport their preferences. Then there is no matching μ, stable for $\succsim'$, which is preferred to $\mu^M[\succsim]$ by all members of $\bar{M}$.

Proof. It is a corollary of theorem of limits on successful manipulation (Theorem 2.70). □

2.68 Remark: Theorem 2.67 implies that if the man-optimal stable mechanism is used, then no man or coalition of men can improve the outcome for all its members by misreporting preferences.

2.69 For an agent i with true preference $\succsim_i$, the strict preference $\succsim_i^+$ corresponds to $\succsim_i$ if the true preference can be obtained from $\succsim_i^+$ without changing the order of any alternatives, simply by indicating which alternatives are tied.

2.70 Theorem of limits on successful manipulation (Theorem in Demange, Gale and Sotomayor (1987)): Let $\succsim$ be the true preferences (not necessarily strict) of the agents, and let $\succsim'$ differ from $\succsim$ in that some coalition C of men and women misreport their preferences. Then there is no matching μ, stable for $\succsim'$, which is preferred to every stable matching under the true preference profile $\succsim$ by all members of C.

Proof. (1) Suppose that some non-empty subset $\bar{M} \cup \bar{W}$ of men and women misreport their preferences and are strictly better off under some μ, stable under $\succsim'$, than under any stable matching under $\succsim$.

(2) If μ is not individually rational under $\succsim$, then someone, say a man, is matched under μ with a woman not on his true list of acceptable women, so he is surely a liar and is in $\bar{M}$, which is a contradiction.

(3) Assume μ is individually rational under $\succsim$.

(4) Clearly μ is not stable under $\succsim$, since every member in the coalition prefers μ to any stable matching.

(5) Construct a corresponding preference profile $\succsim^+$, with strict preferences, so that, if any agent i is indifferent under $\succsim$ between $\mu(i)$ and some other alternative, then under $\succsim^+$ i prefers $\mu(i)$.

(6) Then (m, w) blocks μ under $\succsim^+$ only if (m, w) blocks μ under $\succsim$.

(7) Since every stable matching under $\succsim^+$ is also stable under $\succsim$,

$$\mu(m) \succ_m \mu^M[\succsim^+](m) \text{ for every } m \text{ in } \bar{M}, \text{ and } \mu(w) \succ_w \mu^W[\succsim^+](w) \text{ for every } w \text{ in } \bar{W}.$$

(8) If $\bar{M}$ is not empty, we can apply the blocking lemma (Lemma 2.36) to the marriage problem $\langle M, W, \succsim^+ \rangle$: there is a pair (m, w) that blocks μ under $\succsim^+$ and so under $\succsim$, such that

$$\mu^M[\succsim^+](m) \succsim_m \mu(m) \text{ and } \mu^M[\succsim^+](w) \succsim_w \mu(w).$$

(9) Clearly m and w are not in $\bar{M} \cup \bar{W}$ and therefore are not misreporting their preferences, so they will also block μ under $\succsim'$, contradicting that μ is stable under $\succsim'$.

(10) If $\bar{M}$ is empty, $\bar{W}$ is not empty and the symmetrical argument applies.

□

2.71 Remark: Theorem 2.70 implies that no matter which stable matching under $\succsim'$ is chosen, at least one of the liars is not better off than he would be at the man-optimal matching under $\succsim$.

2.8 Nonbossiness

☞ 2.72 Definition: A mechanism φ is said to be nonbossy[2] if, for any preference profile $\succ = (\succ_i)_{i \in M \cup W}$ of men and women, any man or woman $i \in M \cup W$ and $\succ'_i$, $\varphi[\succ'_i, \succ_{-i}](i) = \varphi[\succ](i)$ implies $\varphi[\succ'_i, \succ_{-i}] = \varphi[\succ]$.

2.73 Example: DA is not nonbossy.

Let $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2\}$, and preferences given by

m_1	m_2	m_3	w_1	w_2
w_1	w_1	w_2	m_3	m_1
w_2		w_1	m_2	m_3
			m_1	

Table 2.13

The wen-proposing DA outcome is

$$\begin{bmatrix} m_1 & m_2 & m_3 \\ w_2 & \emptyset & w_1 \end{bmatrix}$$

Consider a preference for m_2, $\succ'_{m_2}: \emptyset$. Then the men-proposing DA outcome under this modified preference is

$$\begin{bmatrix} m_1 & m_2 & m_3 \\ w_1 & \emptyset & w_2 \end{bmatrix}$$

So we have just shown that the men-proposing DA is not nonbossy.

✌ 2.74 Theorem (Theorem 1 in Kojima (2010)): There exists no stable mechanism that is nonbossy for marriage markets.

Proof. (1) Consider a problem where $W = \{w_1, w_2, w_3\}$ and $M = \{m_1, m_2, m_3\}$, and preferences are given by

m_1	m_2	m_3	w_1	w_2	w_3
w_3	w_3	w_1	m_1	$\emptyset$	m_3
w_2	w_2	w_2	m_2		m_2
w_1	w_1	w_3	m_3		m_1

Table 2.14

[2]The concept of nonbossiness is due to Satterthwaite and Sonnenschein (1981).

(2) There exists a unique stable matching

$$\varphi[\succ] = \begin{bmatrix} w_1 & w_2 & w_3 & \emptyset \\ m_1 & \emptyset & m_3 & m_2 \end{bmatrix}.$$

(3) Consider $\succ'_{m_2}$ given by

$$\succ'_{m_2} : \emptyset.$$

(4) Now there are two stable matchings, μ and μ', given by

$$\mu = \begin{bmatrix} w_1 & w_2 & w_3 & \emptyset \\ m_3 & \emptyset & m_1 & m_2 \end{bmatrix}, \quad \mu' = \begin{bmatrix} w_1 & w_2 & w_3 & \emptyset \\ m_1 & \emptyset & m_3 & m_2 \end{bmatrix}.$$

(5) Case 1: $\varphi[\succ'_{m_2}, \succ_{-m_2}] = \mu$. Then $\varphi[\succ'_{m_2}, \succ_{-m_2}](m_2) = \varphi[\succ](m_2)$ and $\varphi[\succ'_{m_2}, \succ_{-m_2}] \neq \varphi[\succ]$. Thus, φ is not nonbossy.

(6) Case 2: $\varphi[\succ'_{m_2}, \succ_{-m_2}] = \mu'$.

(i) Consider $\succ'_{w_2}$ given by

$$\succ'_{w_2} : m_1, m_2, m_3.$$

(ii) Then $\varphi[\succ'_{w_2}, \succ'_{w_2}, \succ_{-w_2-m_2}]$ is given by

$$\varphi[\succ'_{w_2}, \succ'_{m_2}, \succ_{-w_2-m_2}] = \begin{bmatrix} w_1 & w_2 & w_3 & \emptyset \\ m_3 & \emptyset & m_1 & m_2 \end{bmatrix}.$$

(iii) Therefore, we have that

$$\varphi[\succ'_{w_2}, \succ'_{m_2}, \succ_{-w_2-m_2}](w_2) = \varphi[\succ'_{m_2}, \succ_{-m_2}](w_2), \text{ and } \varphi[\succ'_{w_2}, \succ'_{m_2}, \succ_{-w_2-m_2}] \neq \varphi[\succ'_{m_2}, \succ_{-m_2}],$$

so φ is not nonbossy.

□

2.75 A rough idea is to note that the men-proposing DA is not nonbossy, but then when preference of a man (say m_2) changes, there are two stable matchings and one of them, which is the woman-optimal stable matching, does not contradict nonbossiness (yet). But then, we can add one more agent, w_2, to make the situation much like the original situation, but the roles of men and women are switched.

Chapter 3

College admissions

Contents

3.1 The formal model

3.1 In this model, there exist two sides of agents referred to as colleges and students. Each student would like to attend a college and has preferences over colleges and the option of remaining unmatched. Each college would like to recruit a maximum number of students determined by their exogenously given capacity. They have preferences over individual students, which translate into preferences over groups of students under a responsiveness assumption.

☞ 3.2 Definition: A college admissions problem $\Gamma = \langle S, C, q, \succ \rangle$ consists of:

- a finite set of students S,
- a finite set of colleges C,
- a quota vector $q = (q_c)_{c \in C}$ such that $q_c \in \mathbb{Z}_+$ is the quota of college c,
- a preference profile for students $\succ_S = (\succ_s)_{s \in S}$ such that $\succ_s$ is a strict preference over colleges and remaining unmatched, denoting the strict preference of student s,
- a preference profile for colleges $\succ_C = (\succ_c)_{c \in C}$ such that $\succ_c$ is a strict preference over students and remaining unmatched, denoting the strict preference of college c.

In this chapter, we will use $\emptyset$ to denote "unmatched".

☞ 3.3 Definition: A matching is the outcome of a problem, and is defined by a function $\mu\colon C \cup S \to 2^S \cup 2^C$ such that

- for each student $s \in S$, $\mu(s) \in 2^C$ with $|\mu(s)| \leq 1$,
- for each college $c \in C$, $\mu(c) \in 2^S$ with $|\mu(c)| \leq q_c$,
- $\mu(s) = c$ if and only if $s \in \mu(c)$.

Alternatively, a matching is a function $\mu\colon S \to C \cup \{\emptyset\}$ such that for each college c, $|\mu^{-1}(c)| \leq q_c$.

3.4 Even though we have described colleges' preferences over students, each college with a quota greater than one must be able to compare groups of students in order to compare alternative matchings, and we have yet to describe the preferences of colleges over groups of students.

Let $\succ_c^{\#}$ denote the preference of college c over all assignments $\mu(c)$ it could receive at some matching μ of the college admissions problem.

Definition: The preference $\succ_c^{\#}$ over sets of students is responsive (to the preferences over individual students) if,[1]

- whenever $s_i, s_j \in S$ and $S' \subseteq S \setminus \{s_i, s_j\}$, $s_i \cup S' \succ_c^{\#} s_j \cup S'$ if and only if $s_i \succ_c s_j$.
- whenever $s \in S$ and $S' \subseteq S \setminus s$, $s \cup S' \succ_c^{\#} S'$ if and only if $s \succ_c \emptyset$, which denotes the remaining unmatched option for a college (and for a student).

3.5 Remark: A college c's preferences $\succ_c^{\#}$ will be called responsive to its preferences over individual students if, for any two assignments that differ in only one student, it prefers the assignment containing the more preferred student (and is indifferent between them if it is indifferent between the students).

3.6 We will henceforth assume that colleges have preferences over groups of students that are responsive to their preferences over individual students.

3.2 Stability

3.7 Definition: A matching μ is blocked by a college $c \in C$ if there exists $s \in \mu(c)$ such that $\emptyset \succ_c s$.

A matching μ is blocked by a student $s \in S$ if $\emptyset \succ_s \mu(s)$.

A matching is individually rational if it is not blocked by any college or student.

3.8 Definition: A matching μ is blocked by a pair $(c, s) \in C \times S$ if

- $c \succ_s \mu(s)$, and
- - either there exists $s' \in \mu(c)$ such that $s \succ_c s'$, or
  - $|\mu(c)| < q_c$ and $s \succ_c \emptyset$.

☞ 3.9 Definition: A matching is stable if it is not blocked by any agent or pair.

3.10 Example: If colleges do not have responsive preferences, the set of stable matchings might be empty.

Consider two colleges and three students with the following preferences, and each college can admit as many as students as it wishes.

It is clear that c_1's preference is not responsive.

The only individually rational matchings without unemployment are

$$\mu_1 = \begin{bmatrix} c_1 & c_2 \\ s_1, s_3 & s_2 \end{bmatrix}, \text{ which is blocked by } (c_2, s_1)$$

[1] By an abuse of notation, we will denote a singleton without $\{\}$.

c_1	c_2	s_1	s_2	s_3
$\{s_1, s_3\}$	$\{s_1, s_3\}$	c_2	c_2	c_1
$\{s_1, s_2\}$	$\{s_2, s_3\}$	c_1	c_1	c_2
$\{s_2, s_3\}$	$\{s_1, s_2\}$			
s_1	s_3			
s_2	s_1			
	s_2			

Table 3.1

$$\mu_2 = \begin{bmatrix} c_1 & c_2 \\ s_1, s_2 & s_3 \end{bmatrix}, \text{which is blocked by } (c_2, \{s_1, s_3\})$$

$$\mu_3 = \begin{bmatrix} c_1 & c_2 \\ s_2, s_3 & s_1 \end{bmatrix}, \text{which is blocked by } (c_2, \{s_1, s_2\})$$

$$\mu_4 = \begin{bmatrix} c_1 & c_2 \\ s_2 & s_1, s_3 \end{bmatrix}, \text{which is blocked by } (c_1, \{s_2, s_3\})$$

$$\mu_5 = \begin{bmatrix} c_1 & c_2 \\ s_1 & s_2, s_3 \end{bmatrix}, \text{which is blocked by } (c_1, \{s_1, s_3\})$$

Now observe that any matching that leaves s_1 unmatched is blocked either by (c_1, s_1) or by (c_2, s_1); any matching that leaves s_2 unmatched is blocked either by (c_1, s_2), (c_2, s_2) or $(c_2, \{s_2, s_3\})$. Finally, any matching that leaves s_3 unmatched is blocked by $(c_2, \{s_1, s_3\})$.

3.11 Definition: A matching μ is group unstable, or it is blocked by a coalition, if there exists another matching μ' and a coalition A, which might consist of multiple students and/or colleges, such that for all students s in A, and for all colleges c in A,

(1) $\mu'(s) \in A$, *i.e.*, every student in A who is matched by μ' is matched to a college in A;

(2) $\mu'(s) \succ_s \mu(s)$, *i.e.*, every student in A prefers his/her new match to his/her old one;

(3) $s' \in \mu'(c)$ implies $s' \in A \cup \mu(c)$, *i.e.*, every college in A is matched at μ' to new students only from A, although it may continue to be matched with some of its old students from $\mu(c)$;

(4) $\mu'(c) \succ_c \mu(c)$, *i.e.*, every college in A prefers its new set of students to its old one.

A matching is group stable if it is not blocked by any coalition.

3.12 Proposition: In college admissions model, a matching is group stable if and only if stable.

Proof. (1) If μ is blocked via coalition A and matching μ', let $c \in A$.

(2) Then the fact that $\mu'(c) \succ_c \mu(c)$ implies that there exists a student s in $\mu'(c) \setminus \mu(c)$ and a $s' \in \mu(c) \setminus \mu'(c)$ such that $s \succ_c s'$.

(3) So $s \in A$, and hence $\mu'(s) \succ_s \mu(s)$.

(4) So s prefers $c = \mu'(s)$ to $\mu(s)$, so μ is blocked by the pair (s, c).

□

3.3 The connection between the college admissions model and the marriage model

3.13 The importance of Proposition 3.12 for the college admissions model goes beyond the fact that it allows us to concentrate on small coalitions. It says that stable and group stable matchings can be identified using only the preferences $\succ$ over individuals—that is, without knowing the preferences $\succ_c^\#$ that each college has over groups of students.

3.14 Consider a particular college admissions problem. We can consider a related marriage market, in which each college c with quota q_c is broken into q_c "pieces" of itself, so that in the related market, the agents will be students and college positions, each having a quota of one.

3.15 Given a college admissions problem $\langle S, C, q, \succ \rangle$, the related marriage problem is constructed as follows:

- "Divide" each college c_ℓ into q_{c_ℓ} separate pieces $c_\ell^1, c_\ell^2, \ldots, c_\ell^{q_{c_\ell}}$, where each piece has a capacity of one; and let each piece have the same preferences over S as college c has. (Since college preferences are responsive, $\succ_c$ is consistent with a unique ranking of students.)

  C^*: The resulting set of college "pieces" (or seats).
- For any student s, extend her preference to C^* by replacing each college c_ℓ in her original preference $\succ_s$ with the block $c_\ell^1, c_\ell^2, \ldots, c_\ell^{q_{c_\ell}}$ in that order.

3.16 Given a matching for a college admissions problem, it is straightforward to define a corresponding matching for its related marriage problem: Given any college c, assign the students who were assigned to c in the original problem one at a time to pieces of c starting with lower index pieces.

3.17 Lemma (Lemma 1 in Roth and Sotomayor (1989)): A matching of a college admissions problem is stable if and only if the corresponding matching of its related marriage problem is stable.

3.18 Remark: Theorem on stability (Theorem 2.35) that says that the set of stable matchings is non-empty for every marriage market will immediately generalize to the case of the college admissions model, via Lemma 3.17.

3.4 Deferred acceptance algorithm and properties of stable matchings

3.19 College-proposing deferred acceptance algorithm.

Step 1: Each college c proposes to its top choice q_c students (if it has fewer individually rational choices than q_c, then it proposes to all its individually rational students). Each student rejects any individually irrational proposal and, if more than one individually rational proposal is received, "holds" the most preferred. Any college c that is rejected will remove the students who have rejected it.

Step k: Any college c that was rejected at the previous step by ℓ students makes a new proposal to its most preferred ℓ students who haven't yet rejected it (if there are fewer than ℓ individually rational students, it proposes to all of them). Each student "holds" her most preferred individually rational offer to date and rejects the rest. Any college c that is rejected will remove the students who have rejected it.

End: The algorithm terminates after a step where no rejections are made by matching each student to the college (if any) whose proposal she is "holding."

3.20 Student-proposing deferred acceptance algorithm.

Step 1: Each student proposes to her top-choice individually rational college (if she has one). Each college c rejects any individually irrational proposal and, if more than q_c individually rational proposals are received, "holds" the most preferred q_c of them and rejects the rest.

Step k: Any student who was rejected at the previous step makes a new proposal to her most preferred individually rational college that hasn't yet rejected her (if there is one). Each college c "holds" at most q_c best student proposals to date, and rejects the rest.

End: The algorithm terminates after a step where no rejections are made by matching each college to the students (if any) whose proposals it is "holding."

✌ 3.21 Theorem on stability (Theorem 1 in Gale and Shapley (1962)): The student- and college-proposing deferred acceptance algorithm give stable matchings for each college admissions model.

Proof. It is a consequence of theorem on stability in marriage market (Theorem 2.35) and Lemma 3.17. □

3.22 In a college admissions model, college c and student s are "achievable" for one another if there is some stable matching at which they are matched.

For each c_ℓ with quota q_ℓ, let a_ℓ be the number of achievable students, and define $k_\ell = \min\{q_\ell, a_\ell\}$.

✌ 3.23 Theorem: The college-proposing deferred acceptance algorithm produces a matching that gives each college c_ℓ its k_ℓ highest ranked achievable students.

Proof. We can prove it by induction.

(1) Suppose that, up to Step r of the algorithm, no student has been removed from the list of a college for whom he or she is achievable, and that at Step $(r+1)$ student s_j holds college c_i, and has been removed from the list of c_k.

(2) Then any matching that matches s_j with c_k, and matches achievable students to c_i, is unstable since s_j ranks c_i higher than c_k and c_i ranks s_j higher than one of its assignees. (This follows since s_j is top-ranked by c_i at the end of Step r, when no achievable students had yet been removed from c_i's list.)

(3) So s_j is not achievable for s_k.

□

3.24 Corollary: There exists a college-optimal stable matching that every college likes as well as any other stable matching, and a student-optimal stable matching that every student likes as well as any other stable matching.

✌ 3.25 Theorem: The student-optimal stable matching is weakly Pareto efficient for the students.

Proof. It follows from Theorem 2.31 and Lemma 3.17. □

3.26 Example: The college-optimal stable matching need not be even weakly Pareto optimal for the colleges.

Proof. (1) Consider the problem consisting of three colleges $\{c_1, c_2, c_3\}$ and four students $\{s_1, s_2, s_3, s_4\}$. College c_1 has a quota of $q_1 = 2$, and both other colleges have a quota of one. The preferences are given by

s_1	s_2	s_3	s_4	c_1	c_2	c_3
c_3	c_2	c_1	c_1	s_1	s_1	s_3
c_1	c_1	c_3	c_2	s_2	s_2	s_1
c_2	c_3	c_2	c_3	s_3	s_3	s_2
				s_4	s_4	s_4

Table 3.2

(2) It is straightforward to see that the college-optimal stable matching is

$$\mu^C = \begin{bmatrix} c_1 & c_2 & c_3 \\ s_3, s_4 & s_2 & s_1 \end{bmatrix}.$$

(3) Consider the matching

$$\mu' = \begin{bmatrix} c_1 & c_2 & c_3 \\ s_2, s_4 & s_1 & s_3 \end{bmatrix}.$$

(4) The matching μ' gives colleges c_2 and c_3 each their first choice student, so they both prefers μ' to μ^C.

(5) Since c_1 has responsive preference, it strictly prefers μ' to μ^C.

(6) Thus every college prefers μ' to μ^C.

□

3.27 Example: The college-optimal stable matching need not be even weakly Pareto optimal for the colleges.

Proof. (1) Consider the problem consisting of two colleges $\{c_1, c_2\}$ with $q_{c_1} = 2$, $q_{c_2} = 1$, and two students $\{s_1, s_2\}$. The preferences are given by

s_1	s_2	c_1	c_2
c_1	c_2	$\{s_1, s_2\}$	s_1
c_2	c_1	s_2	s_2
		s_1	

Table 3.3

(2) It is straightforward to see that the college-optimal stable matching is

$$\mu^C = \begin{bmatrix} c_1 & c_2 \\ s_1 & s_2 \end{bmatrix}.$$

(3) Consider the matching

$$\mu' = \begin{bmatrix} c_1 & c_2 \\ s_2 & s_1 \end{bmatrix}.$$

(4) Both college strictly prefers μ' to μ^C.

□

3.28 Remark: in the marriage problem related to a college admissions problem, it is the college seats that play the role of the agents on the college side of the market. So Theorem 2.31 and Lemma 3.17 tell us that there exists no matching that gives every college a more preferred student in every seat than it gets at the college-optimal stable matching. But of course, as we have just seen, this does not imply that the colleges do not all prefer some other matching.

3.29 Theorem: The set of students admitted and seats filled is the same at every stable matching.

Proof. The proof is immediate via Theorem 2.26 and Lemma 3.17. □

3.30 Lemma (Lemma 3 in Roth and Sotomayor (1989)): Suppose colleges and students have strict individual preferences, and let μ and μ' be stable matchings for $\langle S, C, q, \succ\rangle$, such that $\mu(c) \neq \mu'(c)$ for some c. Let $\bar{\mu}$ and $\bar{\mu}'$ be the stable matchings corresponding to μ and μ' in the related marriage problem. If $\bar{\mu}(c^i) \succ_c \bar{\mu}'(c^i)$ for some seat c^i of c, then $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ for all seats c^j of c.

Proof. (1) It suffices to show that $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ for all $j > i$. To see this, if there exists $j < i$, such that $\bar{\mu}'(c^j) \succ_c \bar{\mu}(c^j)$, then by this claim we have $\bar{\mu}'(c^i) \succ_c \bar{\mu}(c^i)$, which contradicts the fact $\bar{\mu}(c^i) \succ_c \bar{\mu}'(c^i)$.

(2) Suppose that this claim is false. Then there exists an index j such that

$$\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j) \text{ and } \bar{\mu}'(c^{j+1}) \succsim_c \bar{\mu}(c^{j+1}).$$

(3) It is clear that $\bar{\mu}(c^j) \in S$. Then by Theorem 3.29, we know $\bar{\mu}'(c^j)$ is also in S, so denote it by s'.

(4) By decomposition lemma, $c^j = \bar{\mu}'(s') \succ_{s'} \bar{\mu}(s')$.

(5) Since $\bar{\mu}'(c^j) \succ_c \bar{\mu}'(c^{j+1})$, we have $s' = \bar{\mu}'(c^j) \succ_c \bar{\mu}'(c^{j+1}) \succsim_c \bar{\mu}(c^{j+1})$, and hence $s' \neq \bar{\mu}(c^{j+1})$.

(6) Since $\bar{\mu}(c^j) \succ_c s'$, $\bar{\mu}(c^{j+1}) \neq s'$, and c^{j+1} comes right after c^j in the preference of s' in the related marriage problem, we have $\bar{\mu}(c^{j+1}) \succ_c s'$.

(7) So $\bar{\mu}$ is blocked by the pair (s', c^{j+1}), contradicting the stability of μ.

□

3.31 Remark: The proof of Lemma 3.30 actually shows that if $\bar{\mu}(c^i) \succ_c \bar{\mu}'(c^i)$ for some position c^i of c then $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ for all $j > i$.

3.32 Remark: Consider a college c with $q_c = 2$ and preferences $s_1 \succ_c s_2 \succ_c s_3 \succ_c s_4$. Consider two matchings μ and ν such that $\mu(c) = \{s_1, s_4\}$ and $\nu(c) = \{s_2, s_3\}$. Then without knowing anything about the preferences of students and other colleges, we can conclude that μ and ν can not both be stable by Lemma 3.30.

3.33 Theorem (Theorem 1 in Roth (1986)): Any college that does not fill its quota at some stable matching is assigned precisely the same set of students at every stable matching.

Proof. (1) Recall that if a college c has any unfilled positions, these will be the highest numbered c^j at any stable matching of the corresponding marriage problem.

(2) By Theorem 3.29 these positions will be unfilled at any stable matching, that is, $\bar{\mu}(c^j) = \bar{\mu}'(c^j)$ for all such j.

(3) $\bar{\mu}(c^j) = \bar{\mu}'(c^j)$ for all j, since the proof of Lemma 3.30 shows that if $\bar{\mu}(c^i) \succ_c \bar{\mu}'(c^i)$ for some position c^i of c, then $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ for all $j > i$.

□

3.34 Theorem (Theorem 3 in Roth and Sotomayor (1989)): If colleges and students have strict preferences over individuals, then colleges have strict preferences over those groups of students that they may be assigned at stable matchings. That is, if μ and μ' are stable matchings, then a college c is indifferent between $\mu(c)$ and $\mu'(c)$ only if $\mu(c) = \mu'(c)$.

Proof. (1) If $\mu(c) \neq \mu'(c)$, then without loss of generality $\bar{\mu}(c^i) \succ_c \bar{\mu}'(c^i)$ for some position c^i of c, where $\bar{\mu}$ and $\bar{\mu}'$ are the matchings in the related marriage problem corresponding to μ and μ'.

(2) By Lemma 3.30, $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ for all positions c^j of c.

(3) So $\mu(c) \succ_c \mu'(c)$, by repeated application of the fact that c's preferences are responsive and transitive:

$$\begin{aligned}\mu(c) = \{\bar{\mu}(c^1), \bar{\mu}(c^2), \dots, \bar{\mu}(c^{q_\ell})\} &\succ_c \{\bar{\mu}'(c^1), \bar{\mu}(c^2), \dots, \bar{\mu}(c^{q_\ell})\} \\ &\succ_c \{\bar{\mu}'(c^1), \bar{\mu}'(c^2), \dots, \bar{\mu}(c^{q_\ell})\} \succ_c \cdots \succ_c \{\bar{\mu}'(c^1), \bar{\mu}'(c^2), \dots, \bar{\mu}'(c^{q_\ell})\} = \mu'(c).\end{aligned}$$

□

3.35 Theorem (Theorem 4 in Roth and Sotomayor (1989)): Let preferences over individuals be strict, and let μ and μ' be stable matchings for $\langle S, C, \succ, q \rangle$. If $\mu(c) \succ_c \mu'(c)$ for some college c, then $s \succ_c s'$ for all $s \in \mu(c)$ and $s' \in \mu'(c) \setminus \mu(c)$. That is, c prefers every student in its entering class at μ to every student who is in its entering class at μ' but not at μ.

Proof. (1) Consider the related marriage problem $\langle S, C', \succ \rangle$ and the stable matchings $\bar{\mu}$ and $\bar{\mu}'$ corresponding to μ and μ'.

(2) Observe that c fills its quota under μ and μ', since if not, Theorem 3.33 would imply that $\mu(c) = \mu'(c)$.

(3) So $\mu'(c) \setminus \mu(c)$ is a non-empty subset of S.

(4) Let $s' \in \mu'(c) \setminus \mu(c)$, then $s' = \bar{\mu}'(c^j)$ for some position c^j and $s' \notin \mu(c)$, and hence $\bar{\mu}(c^j) \neq \bar{\mu}'(c^j)$.

(5) By Lemma 3.30 $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j) = s'$; otherwise $\mu'(c) \succ_c \mu(c)$, which contradicts the fact $\mu(c) \succ_c \mu'(c)$.

(6) The decomposition lemma (Lemma 2.41) implies $c^j = \bar{\mu}'(s') \succ_{s'} \bar{\mu}(s')$.

(7) So the construction of the related marriage problem implies $c \succ_{s'} \mu(s')$, since $\mu(s') \neq c$.

(8) Thus $s \succ_c s'$ for all $s \in \mu(c)$ by the stability of μ.

□

3.36 Corollary: Let μ and μ' be two stable matchings. For any college c,

- either $i \succ_c j$ for all $i \in \mu(c) \setminus \mu'(c)$ and $j \in \mu'(c) \setminus \mu(c)$,
- or $j \succ_c i$ for all $i \in \mu(c) \setminus \mu'(c)$ and $j \in \mu'(c) \setminus \mu(c)$.

3.37 Remark: Consider again a college c with $q_c = 2$ and preferences $s_1 \succ_c s_2 \succ_c s_3 \succ_c s_4$. Consider two matchings μ and ν such that $\mu(c) = \{s_1, s_3\}$ and $\nu(c) = \{s_2, s_4\}$. Then the theorem says that if μ is stable, ν is not, and vice versa. (Since c's preference is responsive, $\mu(c) \succ_c \mu'(c)$.)

3.38 Corollary (Corollary 1 in Roth and Sotomayor (1989)): Consider a college c with preferences $\succ_c$ over individual students, and let $\succ_c^{\#}$ and $\succ_c^{*}$ be preferences over groups of students that are responsive to $\succ_c$, (but are otherwise arbitrary). Then for every pair of stable matchings μ and μ', $\mu(c)$ is preferred to $\mu'(c)$ under the preferences $\succ_c^{\#}$ if and only if $\mu(c)$ is preferred to $\mu'(c)$ under $\succ_c^{*}$.

Proof. It follows immediately from the theorem and the definition of responsive preferences. □

3.39 Example: Let the preferences over individuals be given by

and let the quotas be $q_{c_1} = 3$, $q_{c_j} = 1$ for $j = 2, \dots, 5$. Then the set of stable outcomes is $\{\mu_1, \mu_2, \mu_3, \mu_4\}$, where

$$\mu_1 = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ s_1, s_3, s_4 & s_5 & s_6 & s_7 & s_2 \end{bmatrix}$$

s_1	s_2	s_3	s_4	s_5	s_6	s_7	c_1	c_2	c_3	c_4	c_5
c_5	c_2	c_3	c_4	c_1	c_1	c_1	s_1	s_5	s_6	s_7	s_2
c_1	c_5	c_1	c_1	c_2	c_3	c_3	s_2	s_2	s_7	s_4	s_1
	c_1					c_4	s_3		s_3		
							s_4				
							s_5				
							s_6				
							s_7				

Table 3.4

$$\mu_2 = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ s_3, s_4, s_5 & s_2 & s_6 & s_7 & s_1 \end{bmatrix}$$
$$\mu_3 = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ s_3, s_5, s_6 & s_2 & s_7 & s_4 & s_1 \end{bmatrix}$$
$$\mu_4 = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ s_5, s_6, s_7 & s_2 & s_3 & s_4 & s_1 \end{bmatrix}$$

Note that these are the only stable matchings, and

$$\mu_1(c_1) \succ^{\#}_{c_1} \mu_2(c_1) \succ^{\#}_{c_1} \mu_3(c_1) \succ^{\#}_{c_1} \mu_4(c_1),$$

for any responsive preferences $\succ^{\#}_{c_1}$.

3.5 Further results for the college admissions model

3.40 Theorem: If μ and μ' are stable matchings for $\langle S, C, \succ, q\rangle$ then $\mu \succ_C \mu'$ if and only if $\mu' \succ_S \mu$. Here $\mu \succ_C \mu'$ means $\mu(c) \succsim_c \mu'(c)$ for all $c \in C$ and $\mu(c) \succ_c \mu'(c)$ for some $c \in C$.

Proof. (1) Suppose that $\mu(c) \succsim_c \mu'(c)$ for all $c \in C$ and $\mu(c) \succ_c \mu'(c)$ for some $c \in C$.

(2) Using Lemma 3.30 in one direction and the responsiveness of the colleges' preferences in the other direction, we can see that this is equivalent to $\bar{\mu}(c') \succsim'_{c'} \bar{\mu}'(c')$ for all $c' \in C'$ and $\bar{\mu}(c') \succ'_{c'} \bar{\mu}'$ for some $c' \in C'$, where $\bar{\mu}$ and $\bar{\mu}'$ are the stable matchings corresponding to μ and μ' for the related marriage problem $\langle S, C', \succ'\rangle$

(3) This in turn is satisfied if and only if $\bar{\mu} \succ_{C'} \bar{\mu}'$ and hence, if and only if $\bar{\mu}' \succ_S \bar{\mu}$ by Theorem 2.24, which implies $\mu' \succ_S \mu$.

□

3.41 Corollary: The optimal stable matching on one side of the problem $\langle S, C, \succ, q\rangle$ is the worst stable matching for the other side.

3.42 In $\langle S, C, \succ, q\rangle$, for any two matchings μ and μ', define the following function on $S \cup C$:

$$\mu \vee_C \mu'(c) = \begin{cases} \mu(c), & \text{if } \mu(c) \succ_c \mu'(c) \\ \mu'(c), & \text{otherwise} \end{cases}, \quad \mu \vee_C \mu'(s) = \begin{cases} \mu(s), & \text{if } \mu'(s) \succ_s \mu(s) \\ \mu'(s), & \text{otherwise} \end{cases}.$$

Similarly, we can define the function $\mu \wedge_C \mu'$.

3.43 Theorem: Let μ and μ' be stable matchings for $\langle S, C, \succ, q\rangle$. Then $\mu \vee_C \mu'$ and $\mu \wedge_C \mu'$ are stable matchings.

Proof. (1) Consider the marriage problem $\langle S, C', \succ'\rangle$ related to $\langle S, C, \succ, q\rangle$ and the stable matchings $\bar{\mu}$ and $\bar{\mu}'$ corresponding to μ and μ'.

(2) We know that $\bar{\lambda} \triangleq \bar{\mu} \vee_{C'} \bar{\mu}'$ is a stable matching for $\langle S, C', \succ'\rangle$.

(3) If $\mu \vee_C \mu'(c) = \mu(c)$, then $\mu(c) \succsim_c \mu'(c)$, and hence $\bar{\mu}(c^i) \succsim_{c^i} \bar{\mu}'(c^i)$ for all positions c^i of c by Lemma 3.30.

(4) Then $\bar{\mu} \vee_{C'} \bar{\mu}'(c^i) = \bar{\mu}(c^i)$ for all positions c^i of c.

(5) If s is in $\mu(c)$, there is some position c^i of c such that $s = \bar{\lambda}(c)$.

(6) (i) To see that $\mu \vee_C \mu'$ is a matching, suppose by the way of contradiction that there are some s in S and c and c' in C with $c \neq c'$ and such that s is contained in both $\mu \vee_C \mu'(c)$ and $\mu \vee_C \mu(c')$.

(ii) Then there exists some position c^i of c, and some position c^j of c', such that $\bar{\lambda}(c^i) = s = \bar{\lambda}(c^j)$, which contradicts the fact that $\bar{\lambda}$ is a matching.

(7) The matching $\mu \vee_C \mu'$ is stable: if $s \succ_c s' \in \mu \vee_C \mu'(c)$, so there is some position c^i of c such that $s' = \bar{\lambda}(c^i)$ and $s \succ_{c^i} \bar{\lambda}(c^i)$. Then by stability of $\bar{\lambda}$, $\bar{\lambda}(s) \succ_s c^i$, which implies that $\mu \vee_C \mu'(s) \succ_s c$ and (c, s) does not block $\mu \vee_C \mu'$.

□

3.44 Corollary: The set of stable matchings forms a lattice under the partial orders $\succ_C$ or $\succ_S$ with the lattice under the first partial order being the dual to the lattice under the second partial order.

3.45 Theorem: If μ and μ' are two stable matchings for $\langle S, C, \succ, q\rangle$ and $c = \mu(s)$ or $c = \mu'(s)$, with $c \in C$ and $s \in S$, then if $\mu(c) \succ_c \mu'(c)$ then $\mu'(s) \succsim_s \mu(s)$; and if $\mu'(s) \succ_s \mu(s)$ then $\mu(c) \succsim_c \mu'(c)$.

Proof. (1) Consider the related marriage problem $\langle S, C', \succ'\rangle$ and the corresponding stable matchings $\bar{\mu}$ and $\bar{\mu}'$.

(2) Define

$$S(\bar{\mu}') = \{s \in S \mid \bar{\mu}'(s) \succ_s \bar{\mu}(s)\}, \text{ and } C'(\bar{\mu}) = \{c^i \in C' \mid \bar{\mu}(c^i) \succ_{c^i} \bar{\mu}'(c^i)\}.$$

Similarly define $S(\bar{\mu})$ and $C'(\bar{\mu}')$.

(3) By decomposition lemma (Lemma 2.41) $\bar{\mu}$ and $\bar{\mu}'$ map $S(\bar{\mu}')$ onto $C'(\bar{\mu})$ and $S(\bar{\mu})$ onto $C'(\bar{\mu}')$.

(4) If $\mu(c) \succ_c \mu'(c)$, Lemma 3.30 implies that $\bar{\mu}(c^i) \succsim_{c^i} \bar{\mu}'(c^i)$ for all position c^i of c.

(5) Then $c^i \notin C'(\bar{\mu}')$ for all positions c^i of c.

(6) Then $\bar{\mu}(c^i)$ and $\bar{\mu}'(c^i)$ are in $S(\bar{\mu}')$ or $\bar{\mu}(c^i) = \bar{\mu}'(c^i)$, for all positions c^i of c.

(7) Since s is matched to some position of c under $\bar{\mu}$ or $\bar{\mu}'$, we have $\mu'(s) \succsim_s \mu(s)$.

□

3.46 Theorem: Suppose $\succ' \ \triangleright_C \ \succ$ and let $\mu^C[\succ']$, $\mu^C[\succ]$, $\mu^S[\succ']$, and $\mu^S[\succ]$ be the corresponding optimal stable matchings. Then

$$\mu^C[\succ] \succsim_C \mu^C[\succ'],\ \mu^C[\succ'] \succsim_S \mu^C[\succ],\ \mu^S[\succ'] \succsim_S \mu^S[\succ] \text{ and } \mu^S[\succ] \succsim_C \mu^S[\succ'].$$

Symmetrical result are obtained if $\succ' \triangleright_S \succ$.

Proof. (1) Suppose $\succ' \triangleright_C \succ$.

(2) Consider the marriage problems $\langle S, \bar{C}, \bar{\succ}\rangle$ and $\langle S, \bar{C}, \bar{\succ}'\rangle$ related to $\langle S, C, \succ, q\rangle$ and $\langle S, C, \succ', q\rangle$ respectively, where $\bar{\succ}(s) = \bar{\succ}'(s)$ for all s in S.

(3) Then $\bar{\succ}' \rhd_{\bar{C}} \bar{\succ}$.

(4) Now apply Theorem 2.44.

□

3.47 Theorem: Suppose C is contained in C' and $\mu^S[\Gamma]$ is the student-optimal matching for $\Gamma = \langle S, C, \succ, q\rangle$ and $\mu^S[\Gamma']$ is the student-optimal matching for $\Gamma' = \langle S, C', \succ', q'\rangle$, where $\succ'$ agrees with $\succ$ on C. Then

$$\mu^S[\Gamma'] \succsim_S' \mu^S[\Gamma] \text{ and } \mu^S[\Gamma] \succsim_C \mu^S[\Gamma'].$$

Symmetrical results are obtained if S is contained in S'.

Proof. (1) Suppose C is contained in C'.

(2) Consider the marriage problem $\langle S, \bar{C}, \bar{\succ}\rangle$ and $\langle S, \bar{C}', \bar{\succ}'\rangle$ related to $\langle S, C, \succ, q\rangle$ and $\langle S, C', \succ', q'\rangle$ respectively, where $\bar{\succ}'$ agrees with $\bar{\succ}$ on $\bar{C}$.

(3) Now apply Theorem 2.47.

□

3.48 Definition: A matching μ' dominates another matching μ via a coalition A contained in $C \cup S$ if for all students s and colleges c in A,

$$\mu'(s) \in A,\ \mu'(c) \in A,\ \mu'(s) \succ_s \mu(s), \text{ and } \mu'(c) \succ_c \mu(c).$$

The core defined via strict domination is the set of matchings that are not dominated by any other matching.

3.49 Definition: A matching μ' weakly dominates μ via a coalition A contained in $C \cup S$ if for all students s and colleges c in A,

$$\mu'(s) \in A,\ \mu'(c) \in A,\ \mu'(s) \succsim_s \mu(s), \text{ and } \mu'(c) \succsim_c \mu(c),$$

and

$$\mu'(s) \succ_s \mu(s) \text{ for some } s \text{ in } A, \text{ or } \mu'(c) \succ_c \mu(c) \text{ for some } c \text{ in } A.$$

The core (defined by weak domination), $\mathcal{C}(\succ)$, is the set of matchings that are not weakly dominated by any other matching.

3.50 Proposition (Theorem A2.2 in Roth (1985b)): When preferences over individuals are strict, the set of stable matchings is $\mathcal{C}(\succ)$.

Proof. Step 1: Every core matching is stable.

(1) If μ is not stable, then μ is unstable via some student s and college c with $s \succ_c s'$ for some s' in $\mu(c)$.

(2) Then μ is weakly dominated via the coalition $c \cup \mu(c) \cup s \setminus s'$ by any matching μ' with $\mu'(s) = c$ and $\mu'(c) = \mu(c) \cup s \setminus s'$.

Step 2: Every stable matching is in the core.

(3) If μ is not in $\mathcal{C}(\succ)$, then μ is weakly dominated by some matching μ' via a coalition A, so some student or college in A prefers μ' to μ.

(4) Suppose some c prefers μ' to μ. Then there must be some student s in $\mu'(c) \setminus \mu(c)$ and some s' in $\mu(c) \setminus \mu'(c)$ such that $s \succ_c s'$. If not, then $s' \succ_c s$ for all s in $\mu'(c) \setminus \mu(c)$ and s' in $\mu(c) \setminus \mu'(c)$, which would imply $\mu(c) \succsim_c \mu'(c)$, since c has responsive preferences. So μ is unstable, since it is blocked by the pair (s, c).

(5) Suppose some student s in A with $\mu'(s) = c$ prefers μ' to μ. Then the fact that $\mu'(c) \succsim_c \mu(c)$ similarly implies that there is a student s' (possibly different from s) in $\mu'(c) \setminus \mu(c)$ and a s'' in $\mu(c) \setminus \mu'(c)$ such that $s' \succ_c s''$. Then μ is blocked by the pair (s', c).

□

3.6 Incentive compatibility

3.51 Throughout this section we fix $S = \{s_1, \ldots, s_p\}$, and $C = \{c_1, \ldots, c_r\}$, so each pair of preference profile and quota profile defines a college admissions problem.

3.52 Let $\mathcal{P}_s$ and $\mathcal{P}_c$ denote the set of all preferences for student s and college c, $\mathcal{P} = (\mathcal{P}_s)^p \times (\mathcal{P}_c)^r$ denote the set of all preference profiles, and $\mathcal{P}_{-i}$ denote the set of all preference profiles for all individuals except i.

Let $\mathcal{Q}_c$ denote the set of all quotas for college c, $\mathcal{Q} = \mathcal{Q}_{c_1} \times \cdots \times \mathcal{Q}_{c_r}$ denote the set of all quota profiles, and $\mathcal{Q}_{-c}$ denote the set of all quota profiles for all individuals except c.

Let $\mathscr{E} = \mathcal{P} \times \mathcal{Q}$, and let $\mathcal{M}$ denote the set of all matchings.

☞ 3.53 A (direct) mechanism is a systematic procedure that determines a matching for each college admissions problem. Formally, it is a function $\varphi\colon \mathscr{E} \to \mathcal{M}$.

☞ 3.54 A mechanism φ is stable if $\varphi[\succsim, q]$ is stable for any $(\succsim, q) \in \mathscr{E}$.

A mechanism φ is Pareto efficient if it is always selects a Pareto efficient matching.

A mechanism φ is individually rational if it is always selects an individually rational matching.

3.55 Let φ^S (or SOSM) and φ^C be the student-optimal and college-optimal stable mechanisms that selects the student-optimal and college-optimal stable matchings for each problem respectively.

3.6.1 Preference manipulation

☞ 3.56 A mechanism φ is strategy-proof if for each $i \in S \cup C$, for each $\succsim_i, \succsim'_i \in \mathcal{P}_i$, for each $\succsim_{-i} \in \mathcal{P}_{-i}$,

$$\varphi[\succsim_{-i}, \succsim_i, q](i) \succsim_i \varphi[\succsim_{-i}, \succsim'_i, q](i).$$

✌ 3.57 Theorem (Theorem 3 in Roth (1982b)): There exists no mechanism that is stable and strategy-proof.

Proof. It follows immediately from Theorem 2.58. □

✌ 3.58 Theorem (Proposition 1 in Alcalde and Barberà (1994)): There exists no mechanism that is Pareto efficient, individually rational, and strategy-proof.

Proof. It follows immediately from Theorem 2.61. □

✌ 3.59 Theorem (Theorem 5 in Roth (1982b)): Truth-telling is a weakly dominant strategy for all students under the student-optimal stable mechanism.

Proof. It follows immediately from Theorem 2.62. □

3.60 Theorem (Proposition 2 in Roth (1985a)): There exists no stable mechanism where truth-telling is a weakly dominant strategy for all colleges.

Proof. (1) Consider the problem consisting of two colleges $\{c_1, c_2\}$ with $q_{c_1} = 2$, $q_{c_2} = 1$, and two students $\{s_1, s_2\}$. The preferences are given by

s_1	s_2	c_1	c_2
c_1	c_2	$\{s_1, s_2\}$	s_1
c_2	c_1	s_2	s_2
		s_1	

Table 3.5

(2) It is straightforward to see that the college-optimal stable matching is

$$\mu^C[\succ_{c_1}, \succ_{c_2}] = \begin{bmatrix} c_1 & c_2 \\ s_1 & s_2 \end{bmatrix}.$$

(3) Now suppose college c_1 reports the manipulated preferences $\succ'_{c_1}$ where only s_2 is acceptable. For this new college admissions problem, the only stable matching is

$$\mu^C[\succ'_{c_1}, \succ_{c_2}] = \begin{bmatrix} c_1 & c_2 \\ s_2 & s_1 \end{bmatrix}.$$

(4) Hence college c_1 benefits by manipulating its preferences under any stable mechanism (including the college-optimal stable mechanism).

□

3.61 Corollary: In the college admissions model, a coalition of agents (in fact, even a single agent) may be able to misreport its preferences so that it does better than at any stable matching.

3.6.2 Capacity manipulation

3.62 A college c manipulates mechanism φ via capacities at problem $\langle \succ, q \rangle$ if

$$\varphi[\succ, q_{-c}, q'_c](c) \succ_c \varphi[\succ, q](c) \text{ for some } q'_c < q_c.$$

A mechanism is immune to manipulation via capacities if it can never be manipulated via capacities.

3.63 Example: The college-optimal stable mechanism is manipulated via capacities:

Proof. (1) Consider the problem consisting of two colleges $\{c_1, c_2\}$ with $q_{c_1} = 2$, $q_{c_2} = 1$, and two students $\{s_1, s_2\}$. The preferences are as follows:

(2) It is straightforward to see that the college-optimal stable matching is

$$\mu^C[\succ, q] = \begin{bmatrix} c_1 & c_2 \\ s_1 & s_2 \end{bmatrix}.$$

s_1	s_2	c_1	c_2
c_1	c_2	$\{s_1, s_2\}$	s_1
c_2	c_1	s_2	s_2
		s_1	

Table 3.6

(3) Let $q'_{c_1} = 1$ be a potential capacity manipulation by college c_1. For this new college admissions problem, the only stable matching is

$$\mu^C[\succ, q'_{c_1}, q_{c_2}] = \begin{bmatrix} c_1 & c_2 \\ s_2 & s_1 \end{bmatrix}.$$

(4) Hence college c_1 benefits by reducing the number of its positions under the college-optimal stable mechanism.

$\square$

3.64 Theorem (Theorem 1 in Sönmez (1997)): Suppose there are at least two colleges and three students. Then there exists no stable mechanism that is immune to manipulation via capacities.

Proof. (1) We first prove the theorem for two colleges and three students.

(2) Let $\phi\colon \mathscr{E} \to \mathcal{M}$ be stable, $C = \{c_1, c_2\}$ and $S = \{s_1, s_2, s_3\}$, $q_{c_1} = q_{c_2} = 2$ and $q'_{c_1} = q'_{c_2} = 1$.

s_1	s_2	s_3	c_1	c_2
c_2	c_1	c_1	$\{s_1, s_2, s_3\}$	$\{s_1, s_2, s_3\}$
c_1	c_2	c_2	$\{s_1, s_2\}$	$\{s_2, s_3\}$
			$\{s_1, s_3\}$	$\{s_1, s_3\}$
			s_1	s_3
			$\{s_2, s_3\}$	$\{s_1, s_2\}$
			s_2	s_2
			s_3	s_1

Table 3.7

(3) The only stable matching for $\langle \succ, q_{c_1}, q_{c_2} \rangle$ is

$$\mu_1 = \begin{bmatrix} c_1 & c_2 \\ s_2, s_3 & s_1 \end{bmatrix}.$$

(4) The only two stable matchings for $\langle \succ, q_{c_1}, q'_{c_2} \rangle$ are μ_1 and

$$\mu_2 = \begin{bmatrix} c_1 & c_2 \\ s_1, s_2 & s_3 \end{bmatrix}.$$

(5) The only stable matching for $\langle \succ, q'_{c_1}, q'_{c_2} \rangle$ is

$$\mu_3 = \begin{bmatrix} c_1 & c_2 \\ s_1 & s_3 \end{bmatrix}.$$

(6) Therefore $\phi[\succ, q_{c_1}, q_{c_2}] = \mu_1$, $\phi[\succ, q'_{c_1}, q'_{c_2}] = \mu_3$, and $\phi[\succ, q_{c_1}, q'_{c_2}] \in \{\mu_1, \mu_2\}$.

(7) If $\phi[\succ, q_{c_1}, q'_{c_2}] = \mu_1$, then $\phi[\succ, q'_{c_1}, q'_{c_2}](c_1) = \mu_3(c_1) = \{s_1\}$ and $\phi[\succ, q_{c_1}, q'_{c_2}](c_1) = \mu_1(c_1) = \{s_2, s_3\}$ and hence

$$\phi[\succ, q'_{c_1}, q'_{c_2}](c_1) \succ_{c_1} \phi[\succ, q_{c_1}, q'_{c_2}](c_1),$$

which implies college c_1 can manipulate ϕ via capacities when its capacity is $q_{c_1} = 2$ and college c_2's capacity is $q'_{c_2} = 1$ by underreporting its capacity as $q'_{c_1} = 1$.

(8) Otherwise $\phi[\succ, q_{c_1}, q'_{c_2}] = \mu_2$ and therefore $\phi[\succ, q_{c_1}, q'_{c_2}](c_2) = \mu_2(c_2) = \{s_3\}$, $\phi[\succ, q_{c_1}, q_{c_2}](c_2) = \mu_1(c_2) = \{s_1\}$. Hence

$$\phi[\succ, q_{c_1}, q'_{c_2}](c_2) \succ_{c_2} \phi[\succ, q_{c_1}, q_{c_2}](c_2)$$

which implies college c_2 can manipulate ϕ via capacities when its capacity is $q_{c_2} = 2$ and college c_1's capacity is $q_{c_1} = 2$ by underreporting its capacity as $q'_{c_2} = 1$.

(9) Hence ϕ is manipulable via capacities completing the proof for the case of two colleges and three students.

(10) Finally we can include colleges whose top choice is keeping all its positions vacant and students whose top choice is staying unmatched to generalize this proof to situations with more than three students and two colleges.

□

3.65 Definition: College preferences are strongly monotonic if for every $c \in C$, for every $T, T' \subset S$,

$$|T'| < |T| \leq q_c \Rightarrow T \succ_c T'.$$

3.66 Theorem (Theorem 5 in Konishi and Ünver (2006)): Suppose college preferences are strongly monotonic. Then the student-optimal stable mechanism is immune to manipulation via capacities.

Proof. Omitted. □

3.67 Remark: Example 3.63 shows that the college-optimal stable mechanism is manipulable via capacities even under strongly monotonic preferences.

3.68 Definition: For each $s \in S$, let $\underline{q}_s$ denote the minimum capacity imposed on school s.

3.69 Theorem (Theorem 1 in Kesten (2012)): DA is immune to capacity manipulation for all school preferences if and only if the priority structure $(\succ, \underline{q})$ is acyclic.

Proof. Omitted. □

Part II

One-sided matching

Chapter 4

Housing market

Contents

4.1 The former model

4.1 Housing market model is introduced by Shapley and Scarf (1974). Each agent owns a house, and a housing market is an exchange (with indivisible objects) where agents have the opinion to trade their house in order to get a better one.

☞ 4.2 Definition: Formally, a housing market is a triple $\langle A, H, \succ, e\rangle$ such that

- $A = \{a_1, a_2, \ldots, a_n\}$ is a set of agents,
- H is a set of houses such that $|A| = |H|$,
- $\succ = (\succ_a)_{a\in A}$ is a strict preference profile such that for each agent $a \in A$, $\succ_a$ is a strict preference over houses. Let $\mathcal{P}_a$ be the set of preferences of agent a. The induced weak preference of agent a is denoted by $\succsim_a$ and for any $h, g \in H$, $h \succsim_a g$ if and only if $h \succ_a g$ or $h = g$.
- $e\colon A \to H$ is an initial endowment matching, that is, $h_i \triangleq h_{a_i} \triangleq e(a_i)$ is the initial endowment of agent i.

☞ 4.3 Definition: In a housing market $\langle A, H, \succ, e\rangle$, a matching (allocation) is a bijection $\mu\colon A \to H$. Here $\mu(a)$ is the assigned house of agent a under matching μ. Let $\mathcal{M}$ be the set of matchings.

☞ 4.4 Definition: A (deterministic direct) mechanism is a procedure that assigns a matching for each housing market $\langle A, H, \succ, e\rangle$.

For the fixed sets of agents A and houses H, a mechanism becomes a function

$$\varphi\colon \times_{a\in A} \mathcal{P}_a \to \mathcal{M}.$$

☞ 4.5 Definition: A matching μ is individually rational if for each agent $a \in A$,

$$\mu(a) \succsim_a h_a = e(a),$$

that is, each agent is assigned a house at least as good as her own occupied house.

A mechanism is individually rational if it always selects an individually rational matching for each housing market.

☞ 4.6 Definition: A matching μ is Pareto efficient if there is no other matching ν such that

- $\nu(a) \succsim_a \mu(a)$ for all $a \in A$, and
- $\nu(a_0) \succ_{a_0} \mu(a_0)$ for some $a_0 \in A$.

A mechanism is Pareto efficient if it always selects a Pareto efficient matching for each housing market.

☞ 4.7 Definition: Define a vector price as a positive real vector assigning a price for each house, *i.e.*,

$$\boldsymbol{p} = (p_h)_{h \in H} \in \mathbb{R}^n_{++}$$

such that p_h is the price of house h.

A matching-price vector pair $(\mu, \boldsymbol{p}) \in \mathcal{M} \times \mathbb{R}^n_{++}$ finds a competitive equilibrium if for each agent $a \in A$,

- $p_{\mu(a)} \leq p_{h_a}$ (budget constraint), and
- $\mu(a) \succsim_a h$ for all $h \in H$ such that $p_h \leq p_{h_a}$ (utility maximization).

A matching is called a competitive equilibrium matching if there exists a price vector which supports the matching to be a competitive equilibrium.

A mechanism is called a competitive equilibrium mechanism if it always selects a competitive equilibrium matching for each housing market, denoted by φ^{eq}.

4.8 Definition: Given a market $\langle A, H, \succ, e\rangle$ and a coalition $B \subseteq A$, a matching μ is a B-matching if for all $a \in B$, $\mu(a) = h_b$ for some $b \in B$.

☞ 4.9 Definition: A matching μ is in the core if there exists no coalition of agents $B \subseteq A$ such that some B-matching $\nu \in \mathcal{M}$ weakly dominates μ, that is,

- $\nu(a) \succsim_a \mu(a)$ for all $a \in B$, and
- $\nu(a_0) \succ_{a_0} \mu(a_0)$ for some $a_0 \in B$.

That is, the core is the collection of matchings such that no coalition could improve their assigned houses even if they traded their initially occupied houses only among each other.

We shall use $\mathcal{C}(\succ)$ or $\mathcal{C}$ to denote the core.

A matching is in the core is called a core matching.

A mechanism is called a core mechanism if it always selects a core matching for each housing market, denoted by φ^{core}.

4.10 Remark: It is clear that the core matching is Pareto efficient and individually rational.

4.11 Definition: A matching μ is in the core defined via strong domination if there exists no coalition of agents $B \subseteq A$ such that some B-matching $\nu \in \mathcal{M}$ strongly dominates μ, that is,

- $\nu(a) \succ_a \mu(a)$ for all $a \in B$.

It is clear that the core is a subset of the core defined via strong domination.

4.2 Top trading cycles algorithm

4.12 Theorem (Theorem in Shapley and Scarf (1974)): The core of a housing market is non-empty and there exists a core matching that can be sustained as part of a competitive equilibrium.

Actually, this theorem is originally stated as follows: The core defined via strong domination is always non-empty, where agents' preferences are allowed to be not strict. Its initial proof makes use of Bondareva-Shapley Theorem.

As an alternative proof, Shapley and Scarf (1974) introduced an iterative algorithm that is a core and competitive equilibrium matching. They attribute this algorithm to David Gale.

4.13 Top trading cycles algorithm.

Step 1: Each agent points to the owner of his favorite house.

Due to the finiteness of agents, there exists at least one cycle (including self-cycles). Moreover, cycles do not intersect.

Each agent in a cycle is assigned the house of the agent he points to and removed from the market.

If there is at least one remaining agent, proceed with the next step.

Step k: Each remaining agent points to the owner of his favorite house among the remaining houses.

Each agent in a cycle is assigned the house of the agent he points to and removed from the market.

If there is at least one remaining agent, proceed with the next step.

End: No agents remain. It is clear that the algorithm will terminate within finite steps. Let Step t denote the last step.

The mechanism determined by top trading cycles algorithm is denoted by φ^{TTC}.

4.14 Notation: In the top trading cycles algorithm, given $\succ$ and e:

- A^k or $A^k[\succ]$ or $A^k[e]$ or $A^k[\succ, e]$: the agents removed at Step k in $\langle A, H, \succ, e\rangle$. If Step t is the last step, then

$$A = A^1 \cup A^2 \cup \cdots \cup A^t.$$

  We refer to $\tilde{A} = \{A^1, A^2, \ldots, A^t\}$ as the cycle structure.

- B^k or $B^k[\succ]$ or $B^k[e]$ or $B^k[\succ, e]$: the remaining agents after Step $k-1$ in $\langle A, H, \succ, e\rangle$. So

$$B^k = A \setminus (A^1 \cup A^2 \cup \cdots \cup A^{k-1}) = A^k \cup A^{k+1} \cup \cdots \cup A^t.$$

- H^k or $H^k[\succ]$ or $H^k[e]$ or $H^k[\succ, e]$: the set of houses that are owned by agents in A^k:

$$H^k = \{h \in H \mid h = e(a) \text{ for some } a \in A^k\}.$$

  Let $H^0 = \emptyset$.

  If Step p is the last step, then

$$H = H^0 \cup H^1 \cup H^2 \cup \cdots \cup H^t.$$

- $G' = \langle B, \succ \rangle$: a directed sub-graph determined by agents $B \subseteq A$ and preference profile $\succ$.
- G^k or $G^k[\succ]$ or $G^k[e]$ or $G^k[\succ, e]$: the directed sub-graph after Step $k-1$ in $\langle A, H, \succ, e\rangle$.
- $\text{Br}_a(H')$ where $a \in A$ and $H' \subseteq H$: agent a's favorite house among H'. Then for each $a \in A^k$, we have

$$\text{Br}_a\left(H \setminus \cup_{\ell=1}^{k-1} H^\ell\right) = \varphi^{\text{TTC}}(a).$$

- $a \xrightarrow{G'} b$ where $G' = \langle B, \succ \rangle$ and $B \subseteq A$: the house of agent b is agent a's favorite house in $\{h_a \mid a \in B\}$ under the preference $\succ_a$.
- $C = (a_{n_1}, a_{n_2}, \ldots, a_{n_m})$ is a chain in the directed sub-graph $G' = \langle B, \succ \rangle$ where $B \subseteq A$: $a_{n_j} \in B$ for $j = 1, 2, \ldots, m$, and

$$a_{n_1} \xrightarrow{G'} a_{n_2} \xrightarrow{G'} \cdots \xrightarrow{G'} a_{n_{m-1}} \xrightarrow{G'} a_{n_m}.$$

  Note that a cycle is a special chain.

4.15 *Proof of "core is non-empty".*

(1) Let B be any coalition. Consider the first j such that $B \cap A^j \neq \emptyset$.

(2) Then we have

$$B \subseteq A^j \cup A^{j+1} \cup \cdots \cup A^t = A \setminus (A^1 \cup A^2 \cup \cdots \cup A^{j-1}).$$

(3) Let $a \in B \cap A^j$. Then a is already getting the favorite possible house available to him in B.

(4) No improvement is possible for her, unless she deals outside of B.

(5) By induction, no agent in B can not strictly improve, and it follows that the outcome produced by top trading cycles algorithm is in the core.

□

4.16 *Proof of "being a competitive equilibrium matching".*

(1) Price vector $\boldsymbol{p}$ is defined as follows:
   - for any $a, b \in A^k$ for some k, set $p_{h_a} = p_{h_b}$;
   - if $a \in A^k$ and $b \in A^\ell$ with $k < \ell$, then set $p_{h_a} > p_{h_b}$.

(2) That is,
   - the prices of the occupied houses whose owners are removed at the same step are set equal to each other;
   - the prices of those whose owners are removed at different steps are set such that the price of a house that leaves earlier is higher than the price of a house that leaves later.

(3) It is easy to check that this price vector $\boldsymbol{p}$ supports the outcome produced by top trading cycles algorithm as a competitive equilibrium.

□

4.17 Remark: It is straightforward to show that any competitive equilibrium matching can be thought of as resulting from top trading cycles algorithm.

4.18 Example of top trading cycles algorithm:

Let

$$A = \{a_1, a_2, \ldots, a_{16}\}.$$

Here h_i is the occupied house of agent a_i. Let the preference profile $\succ$ be given as:

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
h_{15}	h_3	h_1	h_2	h_9	h_6	h_7	h_6	h_{11}	h_7	h_2	h_4	h_6	h_8	h_1	h_5
	h_4	h_3					h_{12}		h_3	h_4	h_{14}	h_{13}			
									h_{12}	h_{16}					
									h_{10}						

Table 4.1

Step 1:

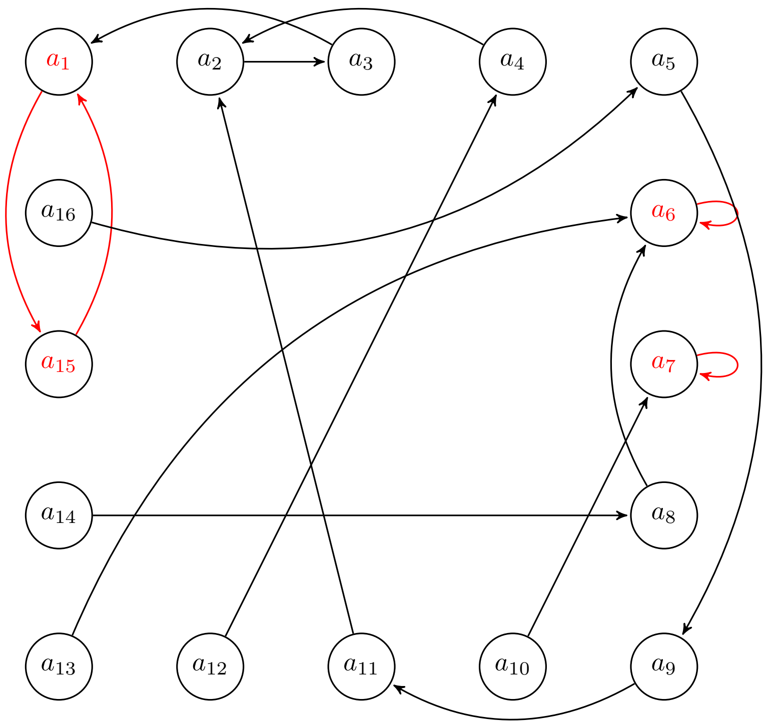


Figure 4.1: Step 1

$A^1 = \{a_1, a_6, a_7, a_{15}\}$.

Step 2: The reduced preferences are as follows:

a_2	a_3	a_4	a_5	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{16}
h_3	h_3	h_2	h_9	h_{12}	h_{11}	h_3	h_2	h_4	h_{13}	h_8	h_5
h_4						h_{12}	h_4	h_{14}			
						h_{10}	h_{16}				

Table 4.2

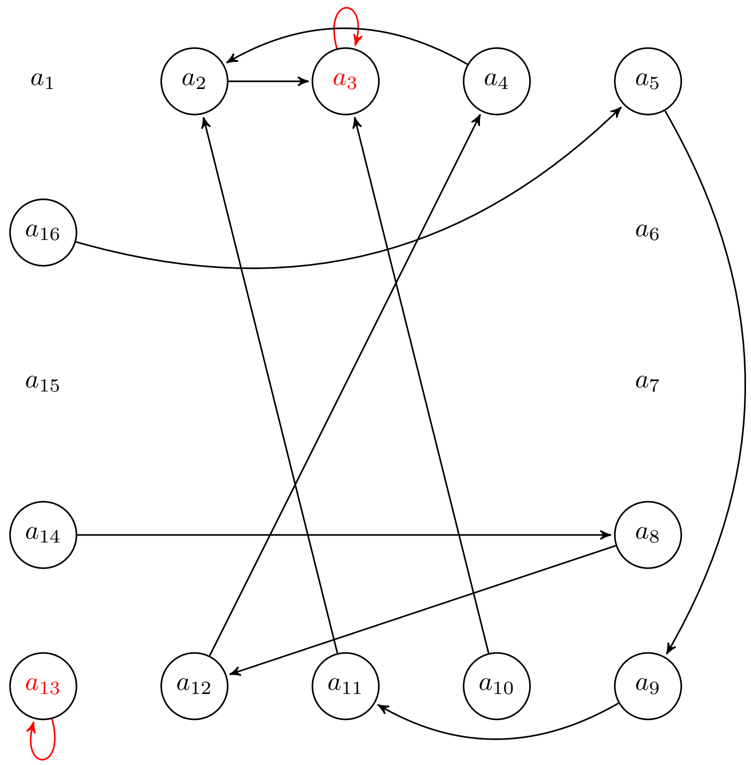


Figure 4.2: Step 2

$A^2 = \{a_3, a_{13}\}$.

Step 3: The reduced preferences are as follows:

a_2	a_4	a_5	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{14}	a_{16}
h_4	h_2	h_9	h_{12}	h_{11}	h_{12}	h_2	h_4	h_8	h_5
					h_{10}	h_4	h_{14}		
						h_{16}			

Table 4.3

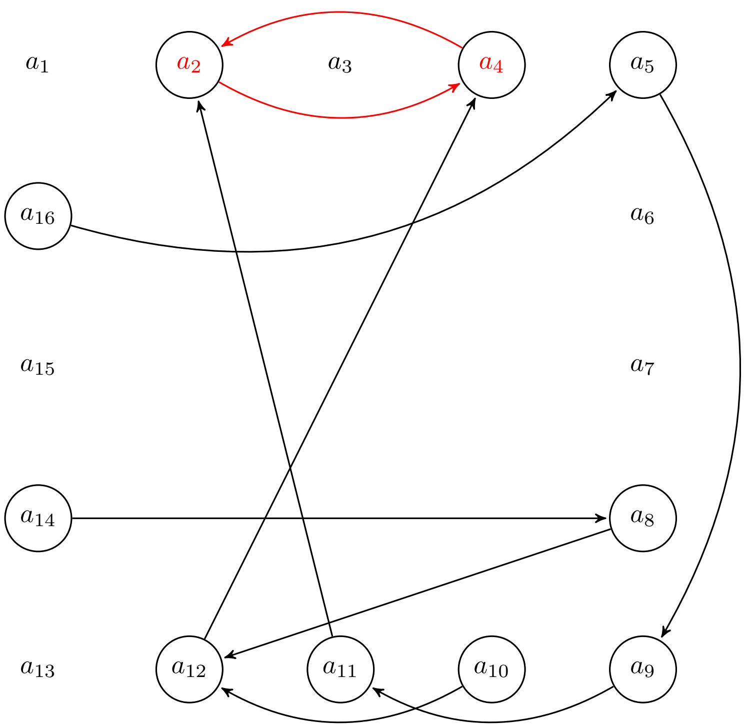


Figure 4.3: Step 3

$A^3 = \{a_2, a_4\}$.

Step 4: The reduced preferences are as follows:

a_5	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{14}	a_{16}
h_9	h_{12}	h_{11}	h_{12}	h_{16}	h_{14}	h_8	h_5
			h_{10}				

Table 4.4

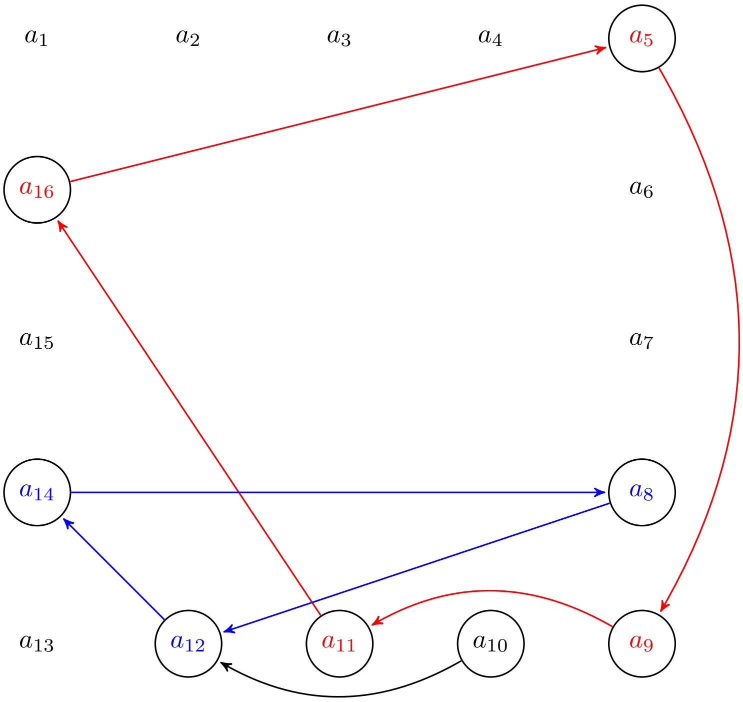


Figure 4.4: Step 4

$A^4 = \{a_5, a_8, a_9, a_{12}, a_{14}, a_{16}\}$.

Step 5: The reduced preferences are as follows:

a_{10}
h_{10}

Table 4.5

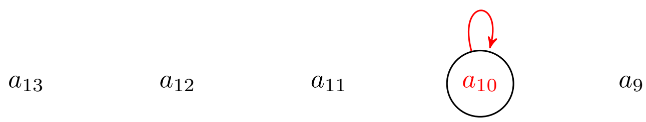


Figure 4.5: Step 5

$A^5 = \{a_{10}\}$.

The outcome is

$$\mu = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ h_{14} & h_4 & h_3 & h_2 & h_9 & h_6 & h_7 & h_{12} & h_{11} & h_{10} & h_{16} & h_{14} & h_{13} & h_8 & h_1 & h_5 \end{bmatrix}.$$

4.19 Lemma (Lemma 1 in Roth and Postlewaite (1977)): If the preference of each agent is strict, then a competitive equilibrium matching (or core matching) weakly dominates any other matching.

Proof. (1) If μ is any competitive equilibrium matching, we can think of μ as being arrived at via trading among top trading cycles $A^1, A^2, \ldots, A^t$.

(2) Let ν be any matching.

(3) If $\mu(a) \neq \nu(a)$ for some $a \in A^1$, μ weakly dominates ν via the coalition A^1 since μ gives each agent of A^1 her most preferred house.

(4) If $\mu(a) = \nu(a)$ for all $a \in A^1$ and $\mu(a) \neq \nu(a)$ for some $a \in A^2$, μ weakly dominates ν via the coalition $A^1 \cup A^2$ since μ gives each agent of A^1 her most preferred house, and each agent of A^2 her most preferred of what was left.

(5) Proceeding in this manner we see that μ weakly dominates all other matchings.

□

4.20 Theorem (Theorem 2 in Roth and Postlewaite (1977)): If the preference of each agent is strict, the core of a housing market has exactly one matching which is also the unique matching that can be sustained at a competitive equilibrium.

Proof. Theorem 4.12 implies that no matching weakly dominates a competitive equilibrium matching (or core matching). Then apply Lemma 4.19. □

4.21 Remark: In a housing market $\langle A, H, \succ, e\rangle$ (with strict preference profile), we have

$$\varphi^{\text{TTC}} = \varphi^{\text{core}} = \varphi^{\text{eq}}.$$

4.22 Chain structure of top trading cycles algorithm.

(1) Consider any agent in A^k at Step $(k-1)$. This agent will take part in a cycle only in the next step. Therefore her favorite house among those left at Step $(k-1)$ is either in H^{k-1} or in H^k.

(2) Note that these should be at least one agent in A^k whose favorite house among those left at Step $(k-1)$ is in H^{k-1}; otherwise agents in A^k would form one or several cycles and trade at Step $(k-1)$. Therefore we have

$$\text{Br}_a(H) \in H^1 \text{ for all } a \in A^1, \text{and } \text{Br}_a(H \setminus \cup_{\ell=1}^{k-2} H^\ell) \in H^{k-1} \cup H^k \text{ for all } a \in A \setminus A^1.$$

(3) Based on this observation, for all $k \geq 2$, we partition the set A^k into the sets of satisfied agents S^k and unsatisfied agents U^k where

$$S^k = S^k[\succ, e] = \left\{a \in A^k \mid \text{Br}_a(H \setminus \cup_{\ell=1}^{k-2} H^\ell) \in H^k\right\},$$
$$U^k = U^k[\succ, e] = \left\{a \in A^k \mid \text{Br}_a(H \setminus \cup_{\ell=1}^{k-2} H^\ell) \in H^{k-1}\right\}.$$

Note that $U^k \neq \emptyset$, $k \geq 2$.

(4) At Step $k-1$, agents in S^k point to an agent in A^k whereas agents in U^k point to an agent in A^{k-1}. The agents in the latter group only in the next step point to an agent in A^k and this follows agents in A^k form one or several cycles.

(5) At Step $k-1$, agents in A^k form one or several chains each of which is tailed by an agent in U^k who possibly follows agents in S^k. Formally the chain structure of A^k is a partition $\{C_1^k, C_2^k, \ldots, C_{r_k}^k\}$ where each chain $C_i^k = (a_{i1}^k, a_{i2}^k, \ldots, a_{in_i}^k)$ is such that

$$\underbrace{a_{i1}^k \xrightarrow{G^{k-1}} a_{i2}^k \xrightarrow{G^{k-1}} \cdots \xrightarrow{G^{k-1}} a_{i(n_i-1)}^k}_{S^k} \xrightarrow{G^{k-1}} \underbrace{a_{in_i}^k}_{U^k} \text{ and } \text{Br}_{a_{in_i}^k}(H \setminus \cup_{\ell=1}^{k-2} H^\ell) \in H^{k-1}.$$

(6) We refer to agent a_{i1}^k as the head and agent $a_{in_i}^k$ as the tail of the chain C_i^k. Let $T^k[\mu] = \{a_{i1}^k \mid i = 1, 2, \ldots, r_k\}$.

(7) At Step k (agents in A^{k-1} with the set of houses H^{k-1} have already been removed), each agent in U^k points to one of these heads (and each of them points to a different one), which in turn converts these chains into one or several cycles.

4.3 Incentive compatibility

4.23 Definition: A mechanism φ is strategy-proof if for each housing market $\langle A, H, \succ, e\rangle$, for each $a \in A$, and for each

$\succ'_a$, we have

$$\varphi[\succ](a) \succsim_a \varphi[\succ_{-a}, \succ'_a](a).$$

4.24 Theorem (Theorem in Roth (1982a)): The core mechanism φ^{TTC} is strategy-proof.

For the proof, we need the following three lemmas.

4.25 Lemma (Lemma 1 in Roth (1982a)): In the top trading cycles algorithm, given $\succ$, if

$$C = (a_{n_1}, a_{n_2}, \dots, a_{n_m})$$

is a chain in $G^k[\succ]$ and $r > k$, then C is a chain in $G^r[\succ]$ if and only if $a_{n_m} \in B^r[\succ]$ (*e.g.*, a_{n_m} has not been removed before Step r).

Proof. (1) If $a_{n_{m-1}} \xrightarrow{G^k[\succ]} a_{n_m}$, then $a_{n_{m-1}} \xrightarrow{G^r[\succ]} a_{n_m}$ if and only if $a_{n_m} \in B^r[\succ]$, due to the top trading cycles algorithm.

(2) By induction, $a_{n_{m-2}} \xrightarrow{G^r[\succ]} a_{n_{m-1}}$ if and only if $a_{n_{m-1}} \in B^r[\succ]$, and so on.

□

4.26 Lemma (Lemma 2 in Roth (1982a)): Let $\succ$ be a strict preference profile, and $\succ'$ be another strict preference profile which differs from $\succ$ only in the preference of agent a_i. Let k and k' be the steps at which agent a_i is removed from the housing market in $\langle A, H, \succ, e\rangle$ and $\langle A, H, \succ', e\rangle$, respectively. Then $B^\ell[\succ]$ and $B^\ell[\succ']$ are same for $1 \leq \ell \leq \min\{k, k'\}$, and have the same cycles for $1 \leq \ell \leq \min\{k, k'\} - 1$.

Proof. (1) It is clear that $B^1[\succ] = B^1[\succ']$.

(2) Since the graphs in $B^1[\succ]$ and $B^1[\succ']$ differs only in the edge emanating from agent a_i, they have the same cycles if $\min\{k, k'\} > 1$, and hence the agents removed at Step 1 from $\succ$ and $\succ'$ are same.

(3) In this case, $B^2[\succ] = B^2[\succ']$, and this lemma follows by induction.

□

4.27 Lemma (Lemma 3 in Roth (1982a)): Let $\succ''$ be a preference profile which differs from $\succ'$ only in the preference of agent a_i, where $\varphi^{\text{TTC}}[\succ'](a_i) \succ''_{a_i} \varphi^{\text{TTC}}[\succ'](a_j)$ for all $j \neq i$. Then we have

$$\varphi^{\text{TTC}}[\succ''](a_i) = \varphi^{\text{TTC}}[\succ'](a_i).$$

Proof. (1) Let k' be the step at which agent a_i with house $h_j \triangleq \varphi^{\text{TTC}}[\succ'](a_i)$ is removed from the market $\langle A, H, \succ', e\rangle$. That is, $a_i, a_j \in B^{k'}[\succ']$.

(2) Let $\varphi^{\text{TTC}}[\succ'](a_i)$ be the initial house of agent a_j.

(3) Let k'' be the step at which agent a_i with house $\varphi^{\text{TTC}}[\succ''](a_i)$ is removed from the market $\langle A, H, \succ'', e\rangle$.

(4) Case 1: $k'' \geq k'$.

(i) That is, agent a_i is still in the market $\langle A, H, \succ'', e\rangle$ at Step k'.

(ii) Then Lemma 4.26 implies that $B^{k'}[\succ'] = B^{k'}[\succ'']$. Hence, $a_i, a_j \in B^{k'}[\succ'] = B^{k'}[\succ'']$.

(iii) Since h_j is top-ranked for agent a_i under $\succ''_{a_i}$, we have $a_i \xrightarrow{G^{k'}[\succ'']} a_j$ and hence

$$G^{k'}[\succ'] = G^{k'}[\succ''].$$

(iv) By the top trading cycles algorithm, a_i with h_j is also removed at Step k' in the market $\langle A, H, \succ'', e\rangle$, that is $\varphi^{\mathrm{TTC}}[\succ''](a_i) = h_j = \varphi^{\mathrm{TTC}}[\succ'](a_i)$ and $k'' = k'$.

(5) Case 2: $k'' < k'$.

(i) That is, agent a_i is removed at Step k'' in the market $\langle A, H, \succ'', e\rangle$.

(ii) Lemma 4.26 implies that at Step $k'' = \min\{k', k''\}$, $B^{k''}[\succ'] = B^{k''}[\succ'']$.

(iii) Since $a_j \in B^{k''}[\succ']$, we have $a_j \in B^{k''}[\succ'']$.

(iv) Therefore, $a_i \xrightarrow{G^{k''}[\succ'']} a_j$, since h_j is top-ranked for agent a_j in $\langle A, H, \succ'', e\rangle$.

(v) Hence h_j is exactly the house which is removed with agent a_i at Step k'' in the market $\langle A, H, \succ'', e\rangle$, that is, $\varphi^{\mathrm{TTC}}[\succ''](a_i) = h_j = \varphi^{\mathrm{TTC}}[\succ'](a_i)$.

□

4.28 *Proof of Theorem 4.24.* Let k and k' be the steps of $\langle A, H, \succ, e\rangle$ and $\langle A, H, \succ', e\rangle$, respectively, at which agent i is removed from the market. Let $h_j = \varphi^{\mathrm{TTC}}[\succ](a_i)$ and $h_{j'} = \varphi^{\mathrm{TTC}}[\succ'](a_i)$. We will see that $h_{j'} \succ_{a_i} h_j$ is impossible.

Lemma 4.27 implies that it is sufficient to consider a preference $\succ'_{a_i}$ that ranks $h_{j'}$ first.

Case 1: $k' \geq k$.

(1) Lemma 4.26 implies that $B^\ell[\succ] = B^\ell[\succ']$ for $1 \leq \ell \leq k$.

(2) It is clear $a_{j'} \in B^{k'}[\succ']$, since agent a_i with house $h_{j'}$ is removed at Step k'.

(3) So $a_{j'} \in B^k[\succ'] = B^k[\succ]$.

(4) If $h_{j'} \succ_{a_i} h_j$, then at Step k, we have $a_i \xrightarrow{G^k[\succ]} a_{j'}$ not $a_i \xrightarrow{G^k[\succ]} a_j$ in the market $\langle A, H, \succ, e\rangle$, which contradicts the fact that a_i is removed with h_j.

Case 2: $k' \leq k$.

(1) Lemma 4.26 implies that $B^\ell[\succ] = B^\ell[\succ']$ for $1 \leq \ell \leq k'$.

(2) Let the chain $C = (a_{j'} \triangleq a_{n_1}, a_{n_2}, \ldots, a_{n_m} \triangleq a_i)$ be the cycle that forms at Step k' in the market $\langle A, H, \succ'$ $, e\rangle$.

(3) Since $\succ$ and $\succ'$ differ only in the a_i's preference, we have

$$a_{j'} = a_{n_1} \xrightarrow{G^{k'}[\succ]} a_{n_2} \xrightarrow{G^{k'}[\succ]} \cdots \xrightarrow{G^{k'}[\succ]} a_{n_m} = a_i,$$

and hence C forms a chain in $G^{k'}[\succ]$.

(4) Since $a_{n_m} = i$ is not removed st Step k in the market $\langle A, H, \succ, e\rangle$, Lemma 4.25 implies that C is a chain in $G^k[\succ]$.

(5) If $h_{j'} \succ_{a_i} h_j$, then at Step k, we have $a_i \xrightarrow{G^k[\succ]} a_{j'}$ not $a_i \xrightarrow{G^k[\succ]} a_j$ in the market $\langle A, H, \succ, e\rangle$, which contradicts the fact that a_i is removed with h_j.

□

4.29 Definition: A mechanism φ is group strategy-proof, if for each housing market $\langle A, H, \succ, e\rangle$, for each non-empty coalition $B \subseteq A$, for each $(\succ'_a)_{a \in B}$, we have for each $a \in B$,

$$\varphi[\succ_{-B}, \succ_B](a) \succsim_a \varphi[\succ_{-B}, \succ'_B](a).$$

4.30 Lemma (Lemma 1 in Bird (1984)): Consider two preference profiles $\succ$ and $\succ'$. If there is an agent $a_i \in A^k[\succ]$ such that $\varphi^{\mathrm{TTC}}[\succ](a_i) \succ_{a_i} \varphi^{\mathrm{TTC}}[\succ'](a_i)$, then there exist agents $a_j \in A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$ and agent $a_\ell \in A^k[\succ] \cup \cdots \cup A^t[\succ]$ such that

$$h_\ell \succ'_{a_j} \varphi^{\mathrm{TTC}}[\succ](a_j).$$

Proof. (1) Assume the contrary. Then

$$\varphi^{\mathrm{TTC}}[\succ](a_j) \succsim'_{a_j} h_\ell,$$

for all $a_j \in A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$ and $a_\ell \in A^k[\succ] \cup \cdots \cup A^t[\succ]$.

(2) It is clear that the equalities above can not hold; otherwise $\varphi^{\mathrm{TTC}}[\succ](a_j) = h_\ell$ due to the strictness of preferences.

(3) Since for each $\varphi^{\mathrm{TTC}}[\succ](a_j) = h_m$ for some $a_m \in A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$, it follows from the top trading cycle algorithm that

$$A^1[\succ'] \cup \cdots \cup A^{k'-1}[\succ'] = A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$$

for some k'.

(4) If $\varphi^{\mathrm{TTC}}[\succ'](a_i) \succ_{a_i} \varphi^{\mathrm{TTC}}[\succ](a_i)$, then $\varphi^{\mathrm{TTC}}[\succ'](a_i)$ must have been taken in an earlier trading cycle under $\succ$.

(5) Thus, $\varphi^{\mathrm{TTC}}[\succ'](a_i) = h_j$ for some $a_j \in A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$.

(6) For preference profile $\succ'$, a_i and a_j are in the same cycle , thus a_i is in $A^1[\succ'] \cup \cdots \cup A^{k'-1}[\succ']$.

(7) But $A^1[\succ] \cup \cdots \cup A^{k-1}[\succ] = A^1[\succ'] \cup \cdots \cup A^{k'-1}[\succ']$ and a_i is not in $A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$. A contradiction.

□

4.31 Remark: This lemma shows that if any agent wants to get a more preferred house, she needs to get an agent in an earlier cycle to change her preference to a house that went in a later trading cycle.

4.32 Theorem (Theorem in Bird (1984)): φ^{TTC} is group strategy-proof.

Proof. (1) Assume each agent a in a subset $B \subseteq A$ reports a preference $\succ'_a$ instead of her true preference $\succ$.

(2) Let a_i be the first agent in B to enter a trading cycle under $\succ$. We will show that a_i can not improve.

(3) Let a_i be in $A^k[\succ]$.

(4) If $\varphi^{\mathrm{TTC}}[\succ'](a_i) \succ_{a_i} \varphi^{\mathrm{TTC}}[\succ](a_i)$, from the lemma there is an agent $a_j \in A^1[\succ] \cup \cdots \cup A^{k-1}[\succ]$ reporting a preference for a house that was assigned in a cycle $q > k-1$ under $\succ$.

(5) Thus, a_j's reported preference $\succ'$ is not same as her true preference $\succ$.

(6) Thus, $a_j \in B$ and a_i can not be the first agent in B.

(7) By induction, every agent in B can not improve her assignment.

□

4.4 Axiomatic characterization of top trading cycles algorithm

4.33 Theorem (Theorem 1 in Ma (1994)): The core mechanism φ^{TTC} is the only mechanism that is individually rational, Pareto efficient, and strategy-proof.

The original proof needs the following five lemmas.

4.34 Given a housing market $\langle A, H, \succ, e\rangle$ and two matchings μ and ν, define

$$J(\mu,\nu,\succ) = \{a \in A \mid \mu(a) \succ_a \nu(a)\}.$$

Similarly we can define $J(\nu,\mu,\succ)$.

Lemma (Lemma 1 in Ma (1994)): Given a housing market $\langle A, H, \succ, e\rangle$, let μ and ν be two Pareto efficient matchings, and $\mu \neq \nu$. Then

$$J(\mu,\nu,\succ) \neq \emptyset.$$

Proof. (1) Suppose $J(\mu,\nu,\succ) = \emptyset$.

(2) Then we have

(i) either $\nu(a) \succsim_a \mu(a)$ for all $a \in A$ and $\nu(a_0) \succ_{a_0} \mu(a_0)$ for some $a_0 \in A$, or

(ii) $\mu(a) = \nu(a)$ for all $a \in A$.

(3) (i) implies that μ is not Pareto efficient, and (ii) implies that $\mu = \nu$.

(4) Both cases leads to a contradiction.

□

Similarly we have $J(\nu,\mu,\succ) \neq \emptyset$.

4.35 Lemma (Lemma 2 in Ma (1994)): Given a housing market $\langle A, H, \succ, e\rangle$, let μ be a core matching, and ν an individually rational and Pareto efficient matching with $\mu \neq \nu$. Then there exists $a \in J(\mu,\nu,\succ)$, such that

$$\mu(a) \succ_a \nu(a) \succ_a e(a).$$

Actually the first relation automatically holds.

Proof. (1) It is clear that μ is Pareto efficient.

(2) By Lemma 4.34, $J(\mu,\nu,\succ) \neq \emptyset$.

(3) Since $\mu(a) \succ_a \nu(a)$ holds for all $a \in J(\mu,\nu,\succ)$, it suffices to show that there exists $a \in J(\mu,\nu,\succ)$ such that $\nu(a) \succ_a e(a)$.

(4) Suppose $e(a) \succsim_a \nu(a)$ for all $a \in J(\mu,\nu,\succ)$.

(5) Since ν is individually rational, $\nu(a) \succsim_a e(a)$ for all $a \in A$, and hence $\nu(a) = e(a)$ for all $a \in J(\mu,\nu,\succ)$.

(6) Let $B = A \setminus J(\mu,\nu,\succ) = J(\nu,\mu,\succ) \cup \big(A \setminus \big(J(\mu,\nu,\succ) \cup J(\nu,\mu,\succ)\big)\big)$.

(7) Since $\nu(a) = e(a)$ for each $a \notin B$, we have $\nu(B) = e(B)$. That is, the restriction of ν on the coalition B is a B-matching.

(8) By Lemma 4.34, we have $J(\nu,\mu,\succ) \neq \emptyset$, and hence ν weakly dominates μ via the coalition B, which contradicts that μ is a core matching.

□

4.36 Given a housing market $\langle A, H, \succ, e\rangle$, define

$$B_\succ = \left\{a \in A \mid \text{there exists a house } h \in H, \text{ such that } \varphi^{\mathrm{TTC}}[\succ](a) \succ_a h \succ_a e(a)\right\}.$$

Define a new preference profile $\succ'=(\succ'_a)_{a\in A}$ as follows:

$$\succ'_a=\begin{cases}\underbrace{\cdots\cdots\cdots}_{\text{truncation of }\succ_a},\varphi^{\mathrm{TTC}}[\succ](a),e(a),\ \underbrace{\cdots\cdots\cdots}_{\text{ranked according to }\succ_a}, & \text{if } a\in B_{\succ};\\ \succ_a, & \text{if } a\in A\setminus B_{\succ}.\end{cases}$$

Lemma (Lemma 3 in Ma (1994)): $\varphi^{\mathrm{TTC}}[\succ]=\varphi^{\mathrm{TTC}}[\succ']=\varphi^{\mathrm{TTC}}[\succ'_{-B},\succ_B]$ for all subsets $B\subseteq A$.

Proof. Obvious. □

4.37 Lemma (Lemma 4 in Ma (1994)): Let φ be an individually rational and Pareto efficient mechanism, then $\varphi^{\mathrm{TTC}}[\succ'
]=\varphi[\succ']$.

Proof. (1) Suppose that $\varphi^{\mathrm{TTC}}[\succ']\neq\varphi[\succ']$.

(2) By Lemma 4.35, there exists $a\in J(\varphi^{\mathrm{TTC}}[\succ'],\varphi[\succ'],\succ')$ such that

$$\varphi^{\mathrm{TTC}}[\succ'](a)\succ'_a\varphi[\succ'](a)\succ'_a e(a). \tag{4.1}$$

(3) Case 1:

(i) For agent $a\in B_{\succ}$, $e(a)$ follows $\varphi^{\mathrm{TTC}}[\succ](a)$ under $\succ'$ immediately.

(ii) Lemma 4.36 implies that

$$\varphi^{\mathrm{TTC}}[\succ](a)=\varphi^{\mathrm{TTC}}[\succ'](a),$$

and hence $e(a)$ follows $\varphi^{\mathrm{TTC}}[\succ'](a)$ under $\succ'$ immediately.

(iii) It contradicts Equation (4.1).

(4) Case 2:

(i) For agent $a\notin B_{\succ}$ and $e(a)$ follows $\varphi^{\mathrm{TTC}}[\succ](a)$ under $\succ$ immediately.

(ii) Lemma 4.36 implies that

$$\varphi^{\mathrm{TTC}}[\succ](a)=\varphi^{\mathrm{TTC}}[\succ'](a),$$

and hence $e(a)$ follows $\varphi^{\mathrm{TTC}}[\succ'](a)$ under $\succ'$ immediately.

(iii) It contradicts Equation (4.1).

(5) Case 3:

(i) For agent $a\notin B_{\succ}$ and $e(a)=\varphi^{\mathrm{TTC}}[\succ](a)$.

(ii) Lemma 4.36 implies that

$$e(a)=\varphi^{\mathrm{TTC}}[\succ](a)=\varphi^{\mathrm{TTC}}[\succ'](a).$$

(iii) It contradicts Equation (4.1).

□

4.38 Lemma (Lemma 5 in Ma (1994)): Let φ be an individually rational, Pareto efficient and strategy-proof mechanism, then $\varphi^{\mathrm{TTC}}[\succ'_{-B},\succ_B]=\varphi[\succ'_{-B},\succ_B]$ for any subset $B\subseteq A$.

Proof. (1) Case 1:

(i) If $B_\succ \cap B = \emptyset$, then

$$[\succ'_{-B}, \succ_B] = [\succ'_{-B\cap B_\succ}, \succ'_{-B\cap -B_\succ}, \succ_B] = [\succ'_{B_\succ}, \succ_{-B_\succ}] = \succ' .$$

(ii) Theorem holds in this case due to Lemma 4.38.

(2) Case 2: If $B \cap B_\succ \neq \emptyset$. Then

$$[\succ'_{-B}, \succ_B] = [\succ'_{-B\cap B_\succ}, \succ'_{-B\cap -B_\succ}, \succ_B] = [\succ'_{-(B\cup -B_\succ)}, \succ_{-B\cap -B_\succ}, \succ_B],$$

and hence it suffices to prove this lemma for all subsets $B \subseteq B_\succ$.

(3) We will prove this lemma by induction. When $|B| = 0$, Lemma 4.37 gives us the desired conclusion.

(4) Assume $\varphi^{\text{TTC}}[\succ'_{-B}, \succ_B] = \varphi[\succ'_{-B}, \succ_B]$ for any B with $|B| = k$.

(5) Suppose $\varphi^{\text{TTC}}[\succ'_{-B}, \succ_B] \neq \varphi[\succ'_{-B}, \succ_B]$ for some B with $|B| = k+1$. For convenience, denote $\succ'' = [\succ'_{-B}, \succ_B]$.

(6) Then by Lemma 4.35, there exists $a \in J(\varphi^{\text{TTC}}[\succ''], \varphi[\succ''], \succ'')$ such that

$$\varphi^{\text{TTC}}[\succ''](a) \succ''_a \varphi[\succ''](a) \succ''_a e(a). \tag{4.2}$$

(7) If $a \in -B$, then by Lemma 4.36 we get from (4.2):

$$\varphi^{\text{TTC}}[\succ](a) \succ'_a \varphi[\succ''](a) \succ'_a e(a),$$

which is impossible by the construction of the $\succ'_a$ (since either $e(a)$ follows $\varphi^{\text{TTC}}[\succ](a)$ immediately or $e(a) = \varphi^{\text{TTC}}[\succ](a)$ under $\succ'_a$ for all $a \in -B$).

(8) If $a \in B$, then by Lemma 4.36 we have:

$$\varphi^{\text{TTC}}[\succ''](a) = \varphi^{\text{TTC}}[\succ'_{-B}, \succ_a, \succ_{B\setminus\{a\}}](a) = \varphi^{\text{TTC}}[\succ'_{-B}, \succ'_a, \succ_{B\setminus\{a\}}](a) = \varphi^{\text{TTC}}[\succ''_{-a}, \succ'_a](a).$$

(9) By induction hypothesis, we have

$$\varphi^{\text{TTC}}[\succ''_{-a}, \succ'_a] = \varphi[\succ''_{-a}, \succ'_a].$$

(10) From Equation (4.2), we have

$$\varphi[\succ''_{-a}, \succ'_a](a) = \varphi^{\text{TTC}}[\succ''_{-a}, \succ'_a](a) = \varphi^{\text{TTC}}[\succ''](a) \succ''_a \varphi[\succ''](a).$$

(11) Substituting for $\succ''$, we get

$$\varphi[\succ'_{-B}, \succ_{B\setminus\{a\}}, \succ'_a] \succ_a \varphi[\succ'_{-B}, \succ_{B\setminus\{a\}}, \succ_a],$$

which contradicts that the mechanism φ is strategy-proof.

□

4.39 *Proof of Theorem 4.33.* It follows Lemma 4.38 immediately. □

4.40 *Alternative proof of Theorem 4.33.* (1) Suppose that there is another mechanism φ satisfying the three conditions.

(2) Fix a preference profile $\succ$.

(3) Let A_1 be the set of agents matched in Step 1 of TTC. We first show that for any agent $a \in A_1$, $\varphi[\succ](a) = \varphi^{\text{TTC}}[\succ](a)$.

(4) Suppose not, then $\varphi[\succ](a)$ is worse.

(5) If $\varphi^{\text{TTC}}[\succ](a) = h_a$, we have a contradiction with individual rationality of φ.

(6) Thus, a trades with others under TTC. Assume that the trading cycle is $a \to k \to \cdots \to 1 \to a$.

(7) Consider a new preference $\succ'_a : h_k, h_a, \emptyset$.

(8) Then $\varphi^{\text{TTC}}[\succ] = \varphi^{\text{TTC}}[\succ'_a, \succ_{-a}]$.

(9) Since φ is individual rational, a must be assigned h_k or h_a under $\varphi[\succ'_a, \succ_{-a}]$.

(10) If she is assigned h_k, then under φ, when her preference is $\succ_a$, she will profitably misreport $\succ'_a$, violating the strategy-proofness of φ.

(11) Thus, $\varphi[\succ'_a, \succ_{-a}](a) = h_a \neq h_k = \varphi^{\text{TTC}}[\succ'_a, \succ_{-a}](a)$.

(12) Since $\varphi[\succ'_a, \succ_{-a}](a) = h_a$, we have $\varphi[\succ'_a, \succ_{-a}](1) \neq h_a = \varphi^{\text{TTC}}[\succ'_a, \succ_{-a}](1)$.

(13) Consider a new preference $\succ'_1 : h_a, h_1, \emptyset$.

(14) Similarly, at $[\succ'_a, \succ'_1, \succ_{-a-1}]$, agent 1 is assigned h_a under φ^{TTC}, but is assigned h_1 under φ.

(15) By induction, at $\succ' = [\succ'_a, \succ'_1, \ldots, \succ'_k]$, $\varphi^{\text{TTC}}[\succ'] = \varphi^{\text{TTC}}[\succ']$, but $\varphi[\succ'](i) = h_i$ for each $i \in \{a, 1, \ldots, k\}$, violating the Pareto efficiency of φ.

(16) By induction on the steps of cycles, we complete the proof.

□

4.41 Theorem 4.33 is "robust" via the following three examples.

4.42 Example 1: a mechanism is individually rational and Pareto efficient, but not strategy-proof.

$A = \{a_1, a_2, a_3\}$, the preference profile $\succ$ is as follows:

a_1	a_2	a_3
h_2	h_1	h_1
h_3	h_3	h_3
h_1	h_2	

Table 4.6

Then both

$$\varphi^{\text{TTC}}[\succ] = \begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_1 & h_3 \end{bmatrix} \text{ and } \mu = \begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_2 & h_1 \end{bmatrix}$$

are individually rational, and Pareto efficient under $\succ$.

Define a mechanism for this market

$$\varphi[\succ'] = \begin{cases} \mu, & \text{if } \succ' = \succ; \\ \varphi^{\text{TTC}}[\succ'], & \text{otherwise.} \end{cases}$$

Now φ is not strategy-proof.

4.43 Example 2: The mechanism in which each agent is assigned her initial house. Clearly this mechanism is individually rational and strategy-proof, but not Pareto efficient.

4.44 Example 3: A mechanism is Pareto efficient and strategy-proof, but not individually rational.

$A = \{a_1, a_2\}$, the mechanism φ in which agent 1 is always assigned the house she likes most. This mechanism is Pareto efficient and strategy-proof.

But under the following preference profile $\succ$

a_1	a_2
h_2	h_2
h_1	

Table 4.7

$$\varphi[\succ] = \begin{bmatrix} a_1 & a_2 \\ h_2 & h_1 \end{bmatrix} \neq \begin{bmatrix} a_1 & a_2 \\ h_1 & h_2 \end{bmatrix} = \varphi^{\text{TTC}}[\succ],$$

and is not individually rational.

4.5 Weak preference

Jaramillo and Manjunath (2012) propose a strategy-proof and Pareto efficient mechanism:

- Pick a tie-breaker and run TTC, but removes a traded cycle only when the agents are not indifferent to object owned by unsatisfied agents.
- Instead, traded agents remain in the problem and their tie-breakers are updated to favor unsatisfied agents

Chapter 5

House allocation

Contents

5.1 The former model

5.1 The house allocation problem is introduced by Hylland and Zeckhauser (1979). In this problem, there is a group of agents and houses. Each agent shall be allocated a house by a central planner using her preferences over the houses.

☞ 5.2 Definition: A house allocation problem is a triple $\langle A, H, \succ\rangle$ such that

- $A = \{a_1, a_2, \ldots, a_n\}$ is a set of agents,
- $H = \{h_1, h_2, \ldots, h_n\}$ is a set of houses,
- $\succ = (\succ_a)_{a\in A}$ is a strict preference profile such that for each agent $a \in A$, $\succ_a$ is a strict preference over houses. Let $\mathcal{P}_a$ be the set of preferences of agent a. The induced weak preference of agent a is denoted by $\succsim_a$ and for any $h, g \in H$, $h \succsim_a g$ if and only if $h \succ_a g$ or $h = g$.

☞ 5.3 Definition: In a house allocation problem $\langle A, H, \succ\rangle$, a matching (allocation) is a bijection $\mu\colon A \to H$. Here $\mu(a)$ is the assigned house of agent a under matching μ. Let $\mathcal{M}$ be the set of matchings.

☞ 5.4 Definition: A (deterministic direct) mechanism is a procedure that assigns a matching for each house allocation problem $\langle A, H, \succ\rangle$.

For the fixed sets of agents A and houses H, a mechanism becomes a function

$$\varphi\colon \times_{a\in A} \mathcal{P}_a \to \mathcal{M}.$$

☞ 5.5 Definition: A matching μ is Pareto efficient if there is no other matching ν such that

- $\nu(a) \succsim_a \mu(a)$ for all $a \in A$, and
- $\nu(a_0) \succ_{a_0} \mu(a_0)$ for some $a_0 \in A$.

Let $\mathscr{E}$ denote the set of all Pareto efficient matchings.

A mechanism is Pareto efficient if it always selects a Pareto efficient matching for each house allocation.

5.2 Simple serial dictatorship and core from assigned endowments

5.6 An ordering $f\colon \{1, 2, \ldots, n\} \to A$ is a one-to-one and onto function. Each ordering induces the following simple mechanism, which is especially plausible if there is a natural hierarchy of agents. Let $\mathcal{F}$ be the set of all orderings.

Simple serial dictatorship induced by an ordering f, denoted by φ^f.

Step 1**:** The highest priority agent $f(1)$ is assigned her top choice house under $\succ_{f(1)}$.

Step k**:** The k-th highest priority agent $f(k)$ is assigned her top choice house under $\succ_{f(k)}$ among the remaining houses.

5.7 Proposition: Simple serial dictatorship induced by an ordering f, φ^f, is Pareto efficient.

Proof. (1) Suppose there is a matching ν that Pareto dominates $\varphi^f[\succ]$.

(2) Consider the agent $a = f(i)$ with the highest priority who obtains a strictly better house in ν than in $\varphi^f[\succ]$.

(3) Then $\nu(a) = \varphi^f[\succ](b)$ for some agent $b = f(j)$ with $j < i$.

(4) By assumption, a is the agent with highest priority such that $\nu(a) \succ_a \varphi^f[\succ](a)$, so $\nu(b) \succ_b \varphi^f[\succ](b)$ is impossible.

(5) Since ν Pareto dominates $\varphi^f[\succ]$, $\nu(b) \succsim_b \varphi^f[\succ](b)$.

(6) Therefore, $\nu(b) = \varphi^f[\succ](b)$, which leads to a contradiction.

□

5.8 Core from assigned endowments μ, denoted by φ^μ: For any house allocation problem $\langle A, H, \succ\rangle$, select the unique element of the core of the housing market $\langle A, H, \succ, \mu\rangle$ where each agent a's initial house is $\mu(a)$. That is,

$$\varphi^\mu = \varphi^{\text{TTC}}[\mu].$$

5.9 Theorem (Lemma 1 in Abdulkadiroğlu and Sönmez (1998)): For any ordering f and any matching μ, the simple serial dictatorship induced by f and the core from assigned endowments μ both yield Pareto efficient matchings. Moreover, for any Pareto efficient matching ν, there is a simple serial dictatorship and a core from assigned endowments that yield it.

Let $\varphi^{\mathcal{F}} = \{\nu \in \mathcal{M} \mid \varphi^f = \nu \text{ for some } f \in \mathcal{F}\}$, and $\varphi^{\mathcal{M}} = \{\nu \in \mathcal{M} \mid \varphi^\mu = \nu \text{ for some } \mu \in \mathcal{M}\}$. Then it suffices to show

$$\varphi^{\mathcal{M}} = \varphi^{\mathcal{F}} = \mathscr{E}.$$

5.10 *Proof of Theorem 5.9, Step 1: "$\varphi^{\mathcal{M}} \subseteq \varphi^{\mathcal{F}}$".*

(1) Let $\nu \in \varphi^{\mathcal{M}}$. Then there exists $\mu \in \mathcal{M}$ with $\nu = \varphi^\mu = \varphi^{\text{TTC}}[\mu]$.

(2) Let Step p be the last step of top trading cycles algorithm. Then let $\{A^1, A^2, \ldots, A^t\}$ be the cycle structure.

(3) By top trading cycles algorithm, for each $k = 1, 2, \ldots, t$ and each $a \in A^k$, we have

$$\text{Br}_a(H \setminus \cup_{\ell=0}^{k-1} H^\ell) = \varphi^{\text{TTC}}[\mu](a) = \varphi^\mu(a) = \nu(a).$$

(4) Let $f\colon \{1, 2, \ldots, n\} \to A$ be the ordering such that for each $k, k' \in \{1, 2, \ldots, t\}$, for each $a \in A^k$, for each $a' \in A^{k'}$, we have

$$k < k' \Rightarrow f^{-1}(a) < f^{-1}(a').$$

That is, f orders agents in A^1 before agents in A^2; agents in A^2 before agents in A^3 and so on.

(5) We will show by induction on i that for all $i \in \{1, 2, \ldots, n\}$ we have $\varphi^f(f(i)) = \nu(f(i))$.

(6) By top trading cycles algorithm and the construction of f, we have

$$\varphi^f(f(1)) = \text{Br}_{f(1)}(H) = \varphi^{\text{TTC}}[\mu](f(1)) = \nu(f(1)).$$

(7) Suppose that $\varphi^f(f(j)) = \nu(f(j))$ for all $j = 1, 2, \ldots, i-1$ where $2 \le i \le n$.

(8) Let $f(i) \in A^k$. We have the following:

- By top trading cycles algorithm, we have

$$\text{Br}_{f(i)}(H \setminus \cup_{\ell=0}^{k-1} H^\ell) = \nu(f(i)).$$

- By the construction of f, we have

$$\cup_{\ell=0}^{k-1} H^\ell \subseteq \cup_{j=1}^{i-1} \nu(f(j)),$$

  and hence

$$H \setminus \cup_{j=1}^{i-1} \nu(f(j)) \subseteq H \setminus \cup_{\ell=0}^{k-1} H^\ell.$$

- $\nu(f(i)) \in H \setminus \cup_{j=1}^{i-1} \nu(f(j))$.

(9) Therefore,

$$\nu(f(i)) = \text{Br}_{f(i)}(H \setminus \cup_{\ell=0}^{k-1} H^\ell) \succsim_{f(i)} \text{Br}_{f(i)}\left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right) \succsim_{f(i)} \nu(f(i)),$$

and hence

$$\nu(f(i)) = \text{Br}_{f(i)}\left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right).$$

(10) It follows that

$$\nu(f(i)) = \text{Br}_{f(i)}\left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right) = \text{Br}_{f(i)}\left(H \setminus \cup_{j=1}^{i-1} \varphi^f(f(j))\right) = \varphi^f(f(i)).$$

□

5.11 *Proof of Theorem 5.9, Step 2: "$\varphi^{\mathcal{F}} \subseteq \mathscr{E}$".*

(1) Let $\mu \in \varphi^{\mathcal{F}}$. Then there exists $f \in \mathcal{F}$ such that $\varphi^f = \mu$.

(2) Let $\nu \in \mathcal{M}$ be such that $\nu(a) \succsim_a \mu(a)$ for all $a \in A$.

(3) We will show that for all $i \in \{1, 2, \ldots, n\}$, $\mu(f(i)) = \nu(f(i))$ by induction on i which in turn will prove that μ can not be Pareto dominated.

(4) Consider agent $f(1)$. We have

$$\varphi^f(f(1)) = \mu(f(1)) = \text{Br}_{f(1)}(H),$$

and therefore $\mu(f(1)) \succsim_{f(1)} \nu(f(1))$.

This, together with $\nu(f(1)) \succsim_{f(1)} \mu(f(1))$ and preference profile being strict, imply that $\mu(f(1)) = \nu(f(1))$.

(5) Suppose that $\nu(f(j)) = \mu(f(j))$ for all $j = 1, 2, \ldots, i-1$ where $2 \leq i \leq n$. We want to show that $\nu(f(i)) = \mu(f(i))$.

(6) We have

$$\begin{aligned}\mu(f(i)) = \varphi^f(f(i)) &= \text{Br}_{f(i)}\left(H \setminus \cup_{\ell=1}^{i-1}\varphi^f(f(\ell))\right) \\ &= \text{Br}_{f(i)}\left(H \setminus \cup_{\ell=1}^{i-1}\mu(f(\ell))\right) = \text{Br}_{f(i)}\left(H \setminus \cup_{\ell=1}^{i-1}\nu(f(\ell))\right),\end{aligned}$$

as well as $\nu(f(i)) \in H \setminus \cup_{\ell=1}^{i-1}\nu(f(\ell))$ and therefore $\mu(f(i)) \succsim_{f(i)} \nu(f(i))$. This, together with $\nu(f(i)) \succsim_{f(i)} \mu(f(i))$ and preference profile being strict imply that $\mu(f(i)) = \nu(f(i))$.

□

5.12 *Proof of Theorem 5.9, Step 3: "$\mathscr{E} \subseteq \varphi^{\mathcal{M}}$".*

(1) Let $\mu \in \mathscr{E}$. Consider the mechanism φ^μ.

(2) Since $\varphi^\mu = \varphi^{\text{TTC}}[\mu]$, φ^μ is individually rational. That is, for all $a \in A$, $\varphi^\mu(a) \succsim_a \mu(a)$.

(3) Since μ is Pareto efficient and the preference profile is strict, we have $\varphi^\mu = \mu$, which in turn implies $\mu \in \varphi^{\mathcal{M}}$, completing the proof of "$\mathscr{E} \subseteq \varphi^{\mathcal{M}}$."

□

✌ 5.13 Theorem (Theorem 1 in Abdulkadiroğlu and Sönmez (1998)): For any house allocation problem, the number of simple serial dictatorships selecting a Pareto efficient matching μ is the same as the number of cores from assigned endowments selecting μ. That is, for all $\nu \in \mathscr{E}$, we have $|\mathcal{M}^\nu| = |\mathcal{F}^\nu|$, where $\mathcal{M}^\nu = \{\mu \in \mathcal{M} \mid \varphi^\mu = \nu\}$ and $\mathcal{F}^\nu = \{f \in \mathcal{F} \mid \varphi^f = \nu\}$.

5.14 *Proof of Theorem 5.13, Step 1: Define "f on $\mathcal{M}^\nu$".*

Let $\nu \in \mathscr{E}$. For any $\mu \in \mathcal{M}$, define $f(\mu)$ as follows:

(1) Apply top trading cycles algorithm to find the cycle structure $\left\{A^1[\mu], A^2[\mu], \ldots, A^{t_\mu}[\mu]\right\}$ for the housing market $\langle A, H, \succ, \mu\rangle$.

(2) For all $t = 2, 3, \ldots, t_\mu$, partition $A^k[\mu]$ into its chains.

(3) Order the agents in $A^1[\mu]$ based on the index of their endowments, starting with the agent whose house has the smallest index. (Recall that the endowment of agent a is $\mu(a)$.)

(4) Order the agents in $A^k[\mu]$, $k = 2, 3, \ldots, t_\mu$ as follows:

   (i) Order the agents in the same chain subsequently, based on their order in the chain, starting with the tail.

   (ii) Order the chains based on the index of the endowments of the heads of the chains (starting the chain whose head has the house with the smallest index).

(5) Order the agents in $A^k[\mu]$ before the agents in $A^{k+1}[\mu]$, $k = 1, 2, \ldots, t_\mu - 1$.

□

5.15 *Proof of Theorem 5.13, Step 2: "f's range is $\mathcal{F}^\nu$".*

(1) Let $\mu \in \mathcal{M}^\nu$. We have $\varphi^\mu = \nu$.

(2) By top trading cycles algorithm, for each $k = 1, 2, \ldots, t_\mu$, for each $a \in A^t[\mu]$, we have

$$\mathrm{Br}_a \left(H \setminus \cup_{\ell=0}^{k-1} H^\ell\right) = \varphi^\mu(a) = \nu(a).$$

(3) By construction, $f(\mu)$ orders agents in $A^1[\mu]$ before the agents in $A^2[\mu]$, agents in $A^2[\mu]$ before the agents in $A^3[\mu]$, and so on.

(4) By the similar method applied in the proof of 5.11, we have the simple serial dictatorship induced by $f(\mu)$, namely $\varphi^{f(\mu)}$, assigns each agent $a \in A$ the house $\nu(a)$.

□

5.16 *Proof of Theorem 5.13, Step 3: "f is one-to-one".*

Claim 1: For any $\mu, \mu' \in \mathcal{M}^\nu$,

$$f(\mu) = f(\mu') \Rightarrow \tilde{A}[\mu] = \tilde{A}[\mu'].$$

(1) Without loss of generality assume $f = f(\mu) = f(\mu')$ orders the agents as $a_1, a_2, \ldots, a_n$.

(2) Let

$$\tilde{A}[\mu] = \Big\{ \underbrace{\{a_1, \ldots, a_{m_1}\}}_{A^1[\mu]}, \underbrace{\{a_{m_1+1}, \ldots, a_{m_2}\}}_{A^2[\mu]}, \ldots, \underbrace{\{a_{m_{k-1}+1}, \ldots, a_{m_k}\}}_{A^k[\mu]}, \ldots, \underbrace{\{a_{m_t-1}, \ldots, a_n\}}_{A^t[\mu]} \Big\},$$

$$\tilde{A}[\mu'] = \Big\{ \underbrace{\{a_1, \ldots, a_{m_1'}\}}_{A^1[\mu']}, \underbrace{\{a_{m_1'+1}, \ldots, a_{m_2'}\}}_{A^2[\mu']}, \ldots, \underbrace{\{a_{m_{k-1}'+1}, \ldots, a_{m_k'}\}}_{A^k[\mu']}, \ldots, \underbrace{\{a_{m_{t'}'-1}, \ldots, a_n\}}_{A^{t'}[\mu']} \Big\}.$$

We want to show that $t = t'$ and $A^k[\mu] = A^k[\mu']$ for all $k = 1, 2, \ldots, t$. We proceed by induction.

(3) Suppose $A^1[\mu] \neq A^1[\mu']$. Without loss of generality suppose that $m_1' < m_1$.

(4) We have agent $a_{m_1'+1} \in A^1[\mu]$, and $\mu \in \mathcal{M}^\nu$, so

$$\mathrm{Br}_{a_{m_1'+1}}(H) = \varphi^\mu(a_{m_1'+1}) = \nu(a_{m_1'+1}).$$

(5) Since $a_{m_1'+1}$ is ordered first in $A^2[\mu']$, she is also ordered first among the agents in her chain.

(6) Then agent $a_{m_1'+1}$ is the tail of her chain, and hence $a_{m_1'+1} \in U^2[\mu']$.

(7) Therefore

$$\mathrm{Br}_{a_{m_1'+1}}(H) \neq \mathrm{Br}_{a_{m_1'+1}}(H \setminus H^1[\mu']) = \varphi^{\mu'}(a_{m_1'+1}) = \nu(a_{m_1'+1}),$$

which leads to a contradiction.

(8) Therefore $A^1[\mu] = A^1[\mu']$.

(9) Suppose $A^\ell[\mu] = A^\ell[\mu']$ for all $\ell = 1, 2, \ldots, k-1$ where $2 \leq k \leq \min\{t, t'\}$.

(10) Then we have $m_{k-1}' = m_{k-1}$. We want to show $A^k[\mu] = A^k[\mu']$.

(11) Suppose, without loss of generality, $m_k' < m_k$.

(12) Then we have $a_{m_k'+1} \in A^k[\mu]$.

(13) Since $\mu \in \mathcal{M}^{\nu}$, we have

$$\mathrm{Br}_{a_{m'_k+1}}(H \setminus \cup_{\ell=0}^{k-1} H^{\ell}[\mu]) = \varphi^{\mu}(a_{m'_k+1}) = \nu(a_{m'_k+1}).$$

(14) Since $a_{m'_k+1}$ is ordered first in $A^{k+1}[\mu']$, she is also ordered first among those agents in her chain.

(15) Then $a_{m'_k+1}$ is the tail of her chain, and hence $a_{m'_k+1} \in U^{k+1}[\mu']$.

(16) Therefore,
$$\mathrm{Br}_{a_{m'_k+1}}(H \setminus \cup_{\ell=0}^{k-1} H^{\ell}[\mu]) = \mathrm{Br}_{a_{m'_k+1}}(H \setminus \cup_{\ell=0}^{k-1} H^{\ell}[\mu']) \in H^{k}[\mu'].$$

(17) Since $a_{m'_k+1} \in A^{k+1}[\mu']$ and $\mu' = \mathcal{M}^{\nu}$, we have

$$\nu(a_{m'_k+1}) = \varphi^{\mu'}[a_{m'_k+1}] \in H^{k+1}[\mu'],$$

and hence $\mathrm{Br}_{a_{m'_k+1}}(H \setminus \cup_{\ell=0}^{k-1} H^{\ell}[\mu]) \neq \nu(a_{m'_k+1})$, which leads to a contradiction.

(18) Therefore $A^k[\mu] = A^k[\mu']$. This also proves that $t = t'$ and hence $\tilde{A}[\mu] = \tilde{A}[\mu']$ by induction.

Claim 2: Suppose $\mu, \mu' \in \mathcal{M}^{\nu}$ are such that $\tilde{A}[\mu] = \tilde{A}[\mu']$. Then

$$f(\mu) = f(\mu') \Rightarrow \mu = \mu'.$$

(19) Let $\mu, \mu' \in \mathcal{M}^{\nu}$ be such that $\tilde{A}[\mu] = \tilde{A}[\mu'] = \{A^1, A^2, \ldots, A^t\}$.

(20) Then we have $H^k[\mu] = H^k[\mu']$ for all $k = 1, 2, \ldots, t$.

(21) Suppose $f(\mu) = f(\mu') = f$. For each $k = 1, 2, \ldots, t$, for each $a \in A^k$, we will show $\mu(a) = \mu'(a)$.

(22) Consider agents in A^1. We have $H^1[\mu] = H^1[\mu']$.

(23) By construction, f orders agents in A^1 based on the index of their endowments. Therefore $f(\mu) = f(\mu')$ implies that $\mu'(a) = \mu(a)$ for all $a \in A^1$.

(24) Consider agents in A^k where $k = 2, 3, \ldots, t$.

(25) Since $H^k[\mu] = H^k[\mu']$ for all $k = 1, 2, \ldots, t$, we have

$$\begin{aligned} U^k[\mu'] &= \left\{a \in A^k \mid \mathrm{Br}_a(H \setminus \cup_{\ell=0}^{k-2} H^{\ell}[\mu']) \in H^{k-1}[\mu']\right\} \\ &= \left\{a \in A^k \mid \mathrm{Br}_a(H \setminus \cup_{\ell=0}^{k-2} H^{\ell}[\mu]) \in H^{k-1}[\mu]\right\} = U^k[\mu], \\ S^k[\mu'] &= A^k \setminus U^k[\mu'] = A^k \setminus U^k[\mu] = S^t[\mu]. \end{aligned}$$

(26) These relations together with $f(\mu) = f(\mu')$ and the construction of f imply that we have the same chain structure for μ and μ'. (Recall that f orders agents in a chain subsequently based on their order in the chain, starting with the tail of the chain who is the only member of chain that is an element of U^k. Therefore for a given ordering f, the set of agents in U^t uniquely determines the chain structure for A^k.)

(27) Let this common chain structure be $\{C_1^k, C_2^k, \ldots, C_{r_k}^k\}$, where for all $i = 1, 2, \ldots, r_k$, we have $C_i^k = (a_{i1}^k, a_{i2}^k, \ldots, a_{in_i}^k)$ with $a_{in_i}^k \in U^k$ and $a_{ij}^k \in S^k$ for all $j = 1, 2, \ldots, n_i - 1$.

(28) By the definition of a chain, for all $i \in 1, 2, \ldots, r_k$ and all $j = 1, 2, \ldots, n_i - 1$, we have

$$\mu\big(a_{i(j+1)}^k\big) = \mathrm{Br}_{a_{ij}^k}(H \setminus \cup_{\ell=0}^{k-2} H^{\ell}[\mu]) = \mathrm{Br}_{a_{ij}^k}(H \setminus \cup_{\ell=0}^{k-2} H^{\ell}[\mu']) = \mu'\big(a_{i(j+1)}^k\big).$$

(29) Since the chain structure is the same for endowments μ and μ', the set of heads is also the same for both endowments. That is $T^k[\mu] = T^k[\mu'] \triangleq T$.

(30) Therefore we have $\mu(a) = \mu'(a)$ for all $a \in A^k \setminus T^k$.

(31) We also have

$$\begin{aligned}\{h \in H \mid \mu'(a) = h \text{ for some } a \in T^k\} &= H^k \setminus \{h \in H \mid \mu'(a) = h \text{ for some } a \in A^k \setminus T^k\} \\ &= H^k \setminus \{h \in H \mid \mu(a) = h \text{ for some } a \in A^k \setminus T^k\} \\ &= \{h \in H \mid \mu(a) = h \text{ for some } a \in T^k\}.\end{aligned}$$

That is, the set of agents T^k collectively own the same set of houses under endowments μ and μ'.

(32) By the construction of f, heads of chains are ordered based on their endowments, $f(\mu) = f(\mu')$ implies $\mu(a) = \mu'(a)$ for all $a \in T^k$, and hence $\mu(a) = \mu'(a)$ for all $a \in A^k$.

□

5.17 *Proof of Theorem 5.13, Step 4: "f is onto".*

(1) By Step 2 and Step 3 we have

$$|\mathcal{F}^\nu| \geq |\mathcal{M}^\nu| \text{ for all } \nu \in \mathscr{E}.$$

(2) Therefore

$$\sum_{\nu \in \mathscr{E}} |\mathcal{F}^\nu| \geq \sum_{\nu \in \mathscr{E}} |\mathcal{M}^\nu|.$$

(3) By Theorem 5.9,

$$\sum_{\nu \in \varphi^{\mathcal{F}}} |\mathcal{F}^\nu| \geq \sum_{\nu \in \varphi^{\mathcal{M}}} |\mathcal{M}^\nu|.$$

(4) Both the left-hand side and the right-hand side of the inequality are equal to the number of orderings, $n!$.

(5) Hence $|\mathcal{M}^\nu| = |\mathcal{F}^\nu|$ for all $\nu \in \mathscr{E}$.

□

5.3 Incentive compatibility

☞ 5.18 Definition: A mechanism φ is strategy-proof if for each house allocation problem $\langle A, H, \succ\rangle$, for each $a \in A$, and for each $\succ'_a$, we have

$$\varphi[\succ](a) \succsim_a \varphi[\succ_{-a}, \succ'_a](a).$$

✌ 5.19 Theorem: The simple serial dictatorship induced by an ordering f is strategy-proof.

Proof. (1) Let f be an ordering.

(2) The first agent $f(1)$ of the ordering obtains the favorite house for her when she tells the truth, so she has no incentives to lie.

(3) The second agent $f(2)$ of the ordering gets her favorite house among the remaining houses, so she has no incentives to lie.

(4) And so on.

□

5.20 Definition: A mechanism φ is group strategy-proof if for each house allocation problem $\langle A, H, \succ\rangle$, there is no group of agents $B \subseteq A$ and preferences $\succ'_B$ such that

- $\varphi[\succ'_B, \succ_{-B}](a) \succsim_a \varphi[\succ_B, \succ_{-B}](a)$ for all $a \in B$ and
- $\varphi[\succ'_B, \succ_{-B}](a_0) \succ_{a_0} \varphi[\succ_B, \succ_{-B}](a_0)$ for some $a_0 \in B$.

In words, a mechanism is group strategy-proof if no group of agents can jointly misreport preferences in such a way to make some member strictly better off while no one in the group is made worse off.

5.21 Theorem: The simple serial dictatorship induced by an ordering f is group strategy-proof.

Proof. An intuition is that the mechanism only uses preference information of an agent when it is her turn to choose, so the best she can do is to report her true favorite remaining good as her favorite choice. Whenever she does so, the subsequent part of the mechanism proceeds exactly as when she reports true preferences. □

5.4 Axiomatic characterization of simple serial dictatorship

5.22 Let σ be a permutation (relabeling) of houses. Let $\succ^\sigma$ be the preference profile where each house h is renamed to $\sigma(h)$. That is, $g \succ^\sigma_a h$ if and only if $\sigma^{-1}(g) \succ_a \sigma^{-1}(h)$.

☞ Definition: A mechanism φ is neutral if, for any house allocation problem and permutation σ,

$$\varphi[\succ^\sigma](a) = \sigma\big(\varphi[\succ](a)\big) \text{ for all } a \in A.$$

$$\begin{tikzcd} \mathcal{P} \arrow[r, "\sigma"] \arrow[d, "\varphi"'] & \mathcal{P} \arrow[d, "\varphi"] \\ \mathcal{M} \arrow[r, "\sigma"'] & \mathcal{M} \end{tikzcd}$$

Figure 5.1

5.23 Example (Example in Svensson (1999)): Let $A = \{1, 2, 3\}$ and $H = \{a, b, c\}$. Let φ be a mechanism defined so that if a is the best element in H according to preference $\succ_2$, then $\varphi[\succ](1)$ is the best element in $\{b, c\}$ according to preference $\succ_1$, $\varphi[\succ](2) = a$ and $\varphi[\succ](3)$ is the remaining element. If all other cases, $\varphi[\succ](1)$ is the best element in H according to preference $\succ_1$, $\varphi[\succ](2)$ is the best element in $H \setminus \{a\}$ according to preference $\succ_2$ and $\varphi[\succ](3)$ is the remaining element.

Hence, the mechanism φ is serially dictatorial for all preference profiles except for those where individual 2 has a as the best element.

This mechanism is obviously not neutral—the element a has a special position.

☞ 5.24 Definition: A mechanism φ is nonbossy if for any $\succ$, $a \in A$ and $\succ'_a$,

$$\varphi[\succ](a) = \varphi[\succ'_a, \succ_{-a}](a) \text{ implies } \varphi[\succ] = \varphi[\succ'_a, \succ_{-a}].$$

5.25 Lemma (Lemma 1 in Svensson (1999)): Let φ be a strategy-proof and nonbossy mechanism, $\succ$ and $\succ'$ two preference profiles such that for $h \in H$ and $a \in A$, $\varphi[\succ](a) \succ'_a h$ if $\varphi[\succ](a) \succ_a h$. Then $\varphi[\succ] = \varphi[\succ']$.

Proof. Step 1: To prove $\varphi[\succ] = \varphi[\succ'_a, \succ_{-a}]$.

(1) From strategy-proofness, it follows that

$$\varphi[\succ](a) \succsim_a \varphi[\succ'_a, \succ_{-a}](a).$$

(2) By the assumption of the lemma,

$$\varphi[\succ](a) \succsim'_a \varphi[\succ'_a, \succ_{-a}](a).$$

(3) Strategy-proofness also implies that

$$\varphi[\succ'_a, \succ_{-a}](a) \succsim'_a \varphi[\succ](a).$$

(4) Hence

$$\varphi[\succ](a) = \varphi[\succ'_a, \succ_{-a}](a).$$

(5) Finally nonbossiness implies

$$\varphi[\succ] = \varphi[\succ'_a, \succ_{-a}].$$

Step 2:

(6) For $\succ$ and $\succ'$, let $\succ^r = (\succ'_1, \succ'_2, \ldots, \succ'_{r-1}, \succ_r, \ldots, \succ_n)$ a preference profile for each $r = 1, 2, \ldots, n+1$.

(7) Then it follows that

$$\varphi[\succ^r] = \varphi[\succ_r, \succ^r_{-r}] = \varphi[\succ'_r, \succ^r_{-r}] = \varphi[\succ^{r+1}].$$

(8) Since $\varphi[\succ] = \varphi[\succ^1]$ and $\varphi[\succ'] = \varphi[\succ^{n+1}]$, they are same.

□

5.26 Theorem (Theorem 1 in Svensson (1999)): A mechanism φ is strategy-proof, nonbossy and neutral mechanism if and only if it is a simple serial dictatorship.

Proof. It suffices to prove the "only if" direction.

Step 1: Consider the preference profile $\succ$ where all agents share the common preference and $h_1 \succ_a h_2 \succ_a \cdots \succ_a h_n$ for all $a \in A$.

(1) Let $f\colon \{1, 2, \ldots, n\} \to A$ be an ordering given by

$$f(j) = (\varphi[\succ])^{-1}(h_j).$$

(2) Clearly, $\varphi[\succ](f(j))$ is the best element in

$$H \setminus \{\varphi[\succ](f(1)), \varphi[\succ](f(2)), \ldots, \varphi[\succ](f(j-1))\},$$

according to the common preference.

(3) Then it is obvious that φ and φ^f coincide on the set of such preference profiles.

Step 2: Consider the preference profile $\succ'$ where all agents share the common preference and $h_{i_1} \succ'_a h_{i_2} \succ'_a \cdots \succ'_a h_{i_n}$ for all $a \in A$.

(4) Define a permutation σ on H as follows: $\sigma(h_j) = h_{i_j}$ for all h_j.

(5) Then $\succ'=\succ^\sigma$.

(6) Neutrality implies $\varphi[\succ'](a) = \varphi[\succ^\sigma](a) = \sigma\big(\varphi[\succ](a)\big)$ for all $a \in A$.

(7) Therefore,

$$\varphi[\succ'](a) = h_{i_j} \iff \sigma\big(\varphi[\succ](a)\big) = h_{i_j} \iff \varphi[\succ](a) = \sigma^{-1}(h_{i_j}) = h_j \iff a = f(j),$$

that is, agent a gets the j-th favorite house under $\varphi[\succ']$ if and only if she is the j-th turn to choose in the procedure φ^f.

(8) Thus, $\varphi[\succ'](a) = h_{i_j} \iff \varphi^f[\succ'] = h_{i_j}$.

(9) Hence, $\varphi = \varphi^f$ coincide on the set of such preference profiles.

Step 3: Consider a general preference profile $\succ'$.

(10) Define $\{h_{i_j}\}_{j=1}^n$ according to:

$$h_{i_j} \text{ is the best element in } H \setminus \{h_{i_1}, h_{i_2}, \ldots, h_{i_{j-1}}\} \text{ according to } \succ'_{f(j)} .$$

(11) Let $\succ''$ be a preference profile where all agents share the common preference, and satisfy:

$$h_{i_1} \succ''_a h_{i_2} \succ''_a \cdots \succ''_a h_{i_n}.$$

(12) From Step 2, $\varphi[\succ''] = \varphi^f[\succ'']$.

(13) Clearly, $\varphi^f[\succ''](f(j)) = h_{i_j} = \varphi^f[\succ'](f(j))$ for each $j = 1, 2, \ldots, n$. Thus, $\varphi^f[\succ''] = \varphi^f[\succ']$.

(14) It remains to show that $\varphi[\succ''] = \varphi[\succ']$.

(15) Let $h \in H$ and $h_{i_j} = \varphi^f[\succ''](f(j)) = \varphi[\succ''](f(j)) \succsim''_{f(j)} h$.

(16) Then $h \in H \setminus \{h_{i_1}, h_{i_2}, \ldots, h_{i_{j-1}}\}$.

(17) By the definition of $\{h_{i_j}\}$, we have

$$\varphi[\succ''](f(j)) = h_{i_j} \succsim'_{f(j)} h.$$

(18) By Lemma 5.25, we have $\varphi[\succ''] = \varphi[\succ']$.

□

5.27 Corollary: A mechanism φ is group strategy-proof and neutral mechanism if and only if it is a simple serial dictatorship.

Proof. It follows immediately from Theorem 9.16 and Theorem 5.26. □

5.5 Consistency

5.28 For any problem $\Gamma = \langle A, H, \succ\rangle$, any nonempty subset A' of A, and any allocation μ, the reduced problem of Γ with respect to A' under μ is

$$r^\mu_{A'}(\Gamma) = \langle A', \mu(A'), (\succ_i|_{\mu(A')})_{i\in A'}\rangle,$$

where $\mu(A')$ is the remaining houses after the agents in $A \setminus A'$ have left with their assigned houses, and $\succ_i|_{\mu(A')}$ is the restriction of agent i's preference to the remaining houses.

☞ 5.29 Definition: A mechanism φ is consistent if for any problem $\Gamma = \langle A, H, \succ \rangle$, any nonempty subset A' of A, and any allocation μ, one has

$$\varphi[\Gamma](a) = \varphi[r^{\mu}_{A'}(\Gamma)](a) \text{ for each } a \in A'.$$

A mechanism φ is pairwise consistent if for any problem $\Gamma = \langle A, H, \succ \rangle$, any nonempty subset A' of A with even cardinality, and any allocation μ, one has

$$\varphi[\Gamma](a) = \varphi[r^{\mu}_{A'}(\Gamma)](a) \text{ for each } a \in A'.$$

☞ 5.30 Definition: In the problem $\Gamma = \langle A, H, \succ \rangle$, the allocation μ' strongly Pareto dominates μ if every agent in A is strictly better off under μ' than under μ.

A mechanism is weakly Pareto optimal if it never chooses allocations that are strongly Pareto dominated.

✌ 5.31 Theorem (Corollary 1 in Ergin (2000)): If a mechanism is weakly Pareto optimal, pairwise consistent, and pairwise neutral, then it is a simple serial dictatorship.

Proof. Omitted. □

5.6 Random house allocation

5.32 A lottery p is a probability distribution over matchings,

$$p = (p_1, p_2, \ldots, p_{n!}),$$

with $\sum_k p_k = 1$ and $p_k \geq 0$ for all k.

We denote the lottery that assigns probability 1 to matching μ by p^{μ}. Let $\Delta(\mathcal{M})$ be the set of all lotteries.

5.33 Random serial dictatorship, φ^{rsd}, (or random priority, φ^{rp}) is defined as

$$\varphi^{\text{rsd}} = \text{RSD} = \varphi^{\text{rp}} = \text{RP} = \sum_{f \in \mathcal{F}} \frac{1}{n!} p^{\varphi^f}.$$

5.34 Core from random endowments, φ^{cre}, is defined as

$$\varphi^{\text{cre}} = \sum_{\mu \in \mathcal{M}} \frac{1}{n!} p^{\varphi^{\mu}}.$$

5.35 Theorem (Theorem 2 in Abdulkadiroğlu and Sönmez (1998)):

$$\varphi^{\text{rsd}} = \varphi^{\text{cre}}.$$

Proof. We have $n!$ simple serial dictatorships and $n!$ cores from assigned endowments. By Theorem 5.9 the members of both classes select Pareto efficient matchings and by Theorem 5.13 the number of simple serial dictatorships selecting a particular Pareto efficient matching ν is the same as the number of cores from assigned endowments selecting ν. Therefore random serial dictatorship which randomly selects a simple serial dictatorship with uniform distribution leads to the same lottery as the core from random endowments which randomly selects a core from assigned endowment with uniform distribution. □

Chapter 6

House allocation with existing tenants

Contents

6.1 The former model

6.1 Motivated by real-life on-campus housing practices, Abdulkadiroğlu and Sönmez (1999) introduced a house allocation problem with existing tenants: A set of houses shall be allocated to a set of agents by a centralized clearing house. Some of the agents are existing tenants, each of whom already occupies a house, referred to as an occupied house, and the rest of the agents are newcomers. Each agent has strict preferences over houses. In addition to occupied houses, there are vacant houses. Existing tenants are entitled not only to keep their current houses but also to apply for other houses.

☞ 6.2 Definition: A house allocation problem with existing tenants, denoted by $\langle A_E, A_N, H_O, H_V, \succ\rangle$, consists of

- a finite set of existing tenants A_E,
- a finite set of new applicants A_N,
- a finite set of occupied houses $H_O = \{h_i \colon a_i \in A_E\}$,
- a finite set of vacant houses H_V, and
- a strict preference profile $\succ = (\succ_i)_{i\in A_E\cup A_N}$.

Let $A = A_E \cup A_N$ denote the set of all agents and $H = H_O \cup H_V \cup \{h_0\}$ denote the set of all houses plus the null house.

Agent i's strict preference $\succ_i$ is on H. Let $\mathcal{P}$ be the set of all strict preferences on H. Let $\succsim_i$ be agent i's induced weak preference. We assume that the null house h_0 is the last choice for each agent.

☞ 6.3 Definition: A matching $\mu\colon A \to H$ is an assignment of houses to agents such that

- every agent is assigned one house, and
- only the null house h_0 can be assigned to more than one agent.

For any agent $a \in A$, we refer to $\mu(a)$ as the assignment of agent i under μ. Let $\mathcal{M}$ be the set of all matchings.

☞ 6.4 Definition: A direct mechanism is a procedure that assigns a matching for each house allocation problem with existing tenants $\langle A_E, A_N, H_O, H_V, \succ\rangle$.

☞ 6.5 Definition: A matching is Pareto efficient if there is no other matching that makes all agents weakly better off and at least one agent strictly better off.

A mechanism is Pareto efficient if it always selects a Pareto efficient matching for each house allocation problem with existing tenants.

☞ 6.6 Definition: A matching is individually rational if no existing tenant strictly prefers his endowment to his assignment.

A mechanism is individually rational if it always selects an individually rational matching for each house allocation problem with existing tenants.

☞ 6.7 Definition: A mechanism φ is strategy-proof if for each house allocation problem with existing tenants $\langle A_E, A_N, H_O, H_V, \succ\rangle$, for each $a \in A$, for each $\succ'_a$, we have

$$\varphi[\succ](a) \succsim_a \varphi[\succ'_a, \succ_{-a}](a).$$

6.2 Real-lief mechanisms

6.8 Given a group $B \subseteq A$ of agents, an ordering of these agents is a one-to-one function $f\colon \{1, 2, \ldots, |B|\} \to B$.

Given a group $B \subseteq A$ of agents and a set $G \subseteq H$ of houses, the serial dictatorship induced by ordering f is defined as follows: The agent who is ordered first under f gets her top choice from G, the next agent gets her top choice among remaining houses, and so on.

6.2.1 Random serial dictatorship with squatting rights

6.9 Random serial dictatorship with squatting rights:

Phase 1: Every existing tenant $a \in A_E$ reports whether she is "In" or "Out" and a strict preference $\succ_a$.

Every new applicant $a \in A_N$ reports a strict preference $\succ_a$.

Phase 2: Every existing tenant $a \in A_E$ who reports "Out" is assigned her current house.

Phase 3: Let $B = A_N \cup \{a \in A_E \mid a \text{ chooses "In"}\}$ and $G = H_V \cup \{h_i \in H_O \mid a_i \text{ chooses "In"}\}$.

(1) An ordering f of agents in B is randomly chosen from a given distribution of orderings.

(2) Houses in G are assigned to these agents based on the serial dictatorship induced by f under the reported preference profile.

6.10 Random serial dictatorship with squatting rights suffers a major deficiency: since it does not guarantee each existing tenant a house that is at least as good as her own, some of them may choose to stay "Out" (*i.e.*, use their squatting rights), and this may result in the loss of potentially large gains from trade.

6.2.2 Random serial dictatorship with waiting list

6.11 Random serial dictatorship with waiting list, induced by a given ordering f of agents:

Start: Define the set of available houses for Step 1 to be the set of vacant houses.

Define the set of acceptable houses for agent a to be

- the set of all houses in case agent a is a new applicant, and
- the set of all houses better than her current house h_a in case she is an existing tenant.

Step 1**:** The agent with the highest priority among those who have at least one acceptable available house is assigned her top available house and removed from the process.

Her assignment is deleted from the set of available houses for Step 2. In case she is an existing tenant, her current house becomes available for Step 2.

Step k**:** The set of available houses for Step k is defined at the end of Step $k-1$.

The agent with the highest priority among all remaining agents who has at least one acceptable available house is assigned her top available house and removed from the process.

Her assignment is deleted from the set of available houses for Step $k+1$. In case she is an existing tenant, her current house becomes available for Step $k+1$.

End: If there is at least one remaining agent and one available house that is acceptable to at least one of them, then the process continues.

When the process terminates, those existing tenants who are not re-assigned keep their current houses.

6.12 Example: Let $A_E = \{a_1, a_2, a_3\}$, $A_N = \emptyset$, $H_O = \{h_1, h_2, h_3\}$, and $H_V = \{h_4\}$. Here the existing tenant a_i occupies the house h_i for $i = 1, 2, 3$.

Let the agents be ordered as a_1-a_2-a_3 and let the preferences be as follows:

a_1	a_2	a_3
h_2	h_3	h_1
h_3	h_1	h_4
h_1	h_2	h_3
h_4	h_4	h_2
h_0	h_0	h_0

Table 6.1

Start: The set of available houses is $\{h_4\}$. The sets of acceptable houses for agents a_1 and a_2 both are $\emptyset$. The set of acceptable houses for agent a_3 is $\{h_4\}$.

Step 1: h_4 is acceptable to only a_3. So, a_3 is assigned h_4. The set of available houses becomes $\{h_3\}$.

Step 2: h_3 is acceptable to both a_1 and a_2. Since a_1 has the higher priority, a_1 is assigned h_3. The set of available houses becomes $\{h_1\}$.

Step 3: h_1 is acceptable to a_2, then a_2 is assigned h_1.

End: Since there are no remaining agents at the end of Step 3, the process terminates and the final matching is

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ h_3 & h_1 & h_4 \end{bmatrix}.$$

6.13 Random serial dictatorship with waiting list is inefficient.

Consider the example in the previous item. The outcome is Pareto dominated by

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_3 & h_1 \end{bmatrix}.$$

6.2.3 MIT-NH4 mechanism

6.14 The following mechanism is used at the residence NH4 of MIT.

6.15 MIT-NH4 mechanism, given an ordering f, works as follows:

Phase 1: The first agent is tentatively assigned his or her top choice among all houses, the next agent is tentatively assigned his top choice among the remaining houses, and so on, until a squatting conflict occurs.

Phase 2: A squatting conflict occurs if it is the turn of an existing tenant but every remaining house is worse than his or her current house. That means someone else, the conflicting agent, is tentatively assigned the existing tenant's current house.

When this happens

(1) the existing tenant is assigned his or her current house and removed from the process, and

(2) all tentative assignments starting with the conflicting agent and up to the existing tenant are erased.

At this point the squatting conflict is resolved and the process starts over again with the conflicting agent. Every squatting conflict that occurs afterwards is resolved in a similar way.

End: The process is over when there are no houses or agents left. At this point all tentative assignments are finalized.

6.16 Example: Let $A_E = \{a_1, a_2, a_3, a_4\}$, $A_N = \{a_5\}$, $H_O = \{h_1, h_2, h_3, h_4\}$ and $H_V = \{h_5\}$. Here the existing tenant a_k occupies the house h_k for $k = 1, 2, 3, 4$. Let the ordering f order the agents as a_1-a_2-a_3-a_4-a_5 and let the preferences be as follows:

a_1	a_2	a_3	a_4	a_5
h_3	h_4	h_5	h_3	h_4
h_4	h_5	h_3	h_5	h_5
h_5	h_2	h_4	h_4	h_3
h_1	h_3	h_2	h_2	h_1
h_2	h_1	h_1	h_1	h_2
h_0	h_0	h_0	h_0	h_0

Table 6.2

Step 1: First agent a_1 is tentatively assigned h_3, next agent a_2 is tentatively assigned h_4, next agent a_3 is tentatively assigned h_5, and next its agent a_4's turn and a squatting conflict occurs. The conflicting agent is agent a_2 who was tentatively assigned h_4. Agent a_2's tentative assignment, as well as that of agent a_3, is erased. Agent a_4 is assigned his or her current house h_4 and removed from the process. This resolves the squatting conflict.

Step 2: The process starts over with the conflicting agent a_2. Agent a_2 is tentatively assigned h_5 and next it is agent a_3's turn and another squatting conflict occurs. The conflicting agent is agent a_1 who was tentatively assigned h_3. His tentative assignment, as well as that of agent a_2 are erased. Agent a_3 is assigned his current house h_3 and removed from the process. This resolves the second squatting conflict.

Step 3: The process starts over with the conflicting agent a_1. He is tentatively assigned h_5, next agent a_2 is tentatively assigned h_2 and finally agent a_5 is tentatively assigned h_1. At this point all tentative assignments are finalized.

Therefore the final matching is

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_5 & h_2 & h_3 & h_4 & h_1 \end{bmatrix}.$$

6.17 While it is innovative, the MIT-NH4 mechanism does not resolve the inefficiency problem.

Consider the example in the previous item, the outcome is Pareto dominated by both

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_3 & h_2 & h_5 & h_4 & h_1 \end{bmatrix} \text{ and } \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_4 & h_2 & h_5 & h_3 & h_1 \end{bmatrix}.$$

6.3 Top trading cycles algorithm

6.18 Top trading cycles algorithm, induced by a given ordering f of agents.

Step 1**:** Define the set of available houses for this step to be the set of vacant houses.

- Each agent a points to her favorite house under her reported preference.
- Each occupied house points to its occupant.
- Each available house points to the agent with highest priority (*i.e.*, $f(1)$).

Since the numbers of agents and houses are finite, there is at least one cycle, here a cycle is an ordered list of agents and houses $(j_1, j_2, \ldots, j_k)$ where j_1 points to j_2, j_2 points to j_3, ..., j_k points to j_1.

Every agent who participates in a cycle is assigned the house that she points to, and removed with her assignment.

Whenever there is an available house in a cycle, the agent with the highest priority, $f(1)$, is also in the same cycle. If this agent is an existing tenant, then her house $h_{f(1)}$ can not be in any cycle and it becomes available for Step 2.

All available houses that are not removed remain available.

Step k**:** The set of available houses for Step k is defined at the end of Step $k-1$.

- Each remaining agent a points to her favorite house among the remaining houses under her reported preference.
- Each remaining occupied house points to its occupant.
- Each available house points to the agent with highest priority among the remaining agents.

There is at least one cycle. Every agent in a cycle is assigned the house that she points to and removed with her assignment.

If there is an available house in a cycle then the agent with the highest priority in this step is also in the same cycle. If this agent is an existing tenant, then her house can not be in any cycle and it becomes available for Step $k+1$.

All available houses that are not removed remain available.

End: If there is at least one remaining agent and one remaining house, then the process continues.

We use φ^f to denote the top trading cycles mechanism induced by the ordering f.

6.19 Example: Let $A_E = \{a_1, a_2, a_3, a_4\}$, $A_N = \{a_5\}$, $H_O = \{h_1, h_2, h_3, h_4\}$ and $H_V = \{h_5, h_6, h_7\}$. Here the existing tenant a_i occupies the house h_i for $i = 1, 2, 3, 4$. Let the ordering f order the agents as a_1-a_2-a_3-a_4-a_5 and let the preferences be as follows:

a_1	a_2	a_3	a_4	a_5
h_2	h_7	h_2	h_2	h_4
h_6	h_1	h_1	h_4	h_3
h_5	h_6	h_4	h_3	h_7
h_1	h_5	h_7	h_6	h_1
h_4	h_4	h_3	h_1	h_2
h_3	h_3	h_6	h_7	h_5
h_7	h_2	h_5	h_5	h_6
h_0	h_0	h_0	h_0	h_0

Table 6.3

Step 1:

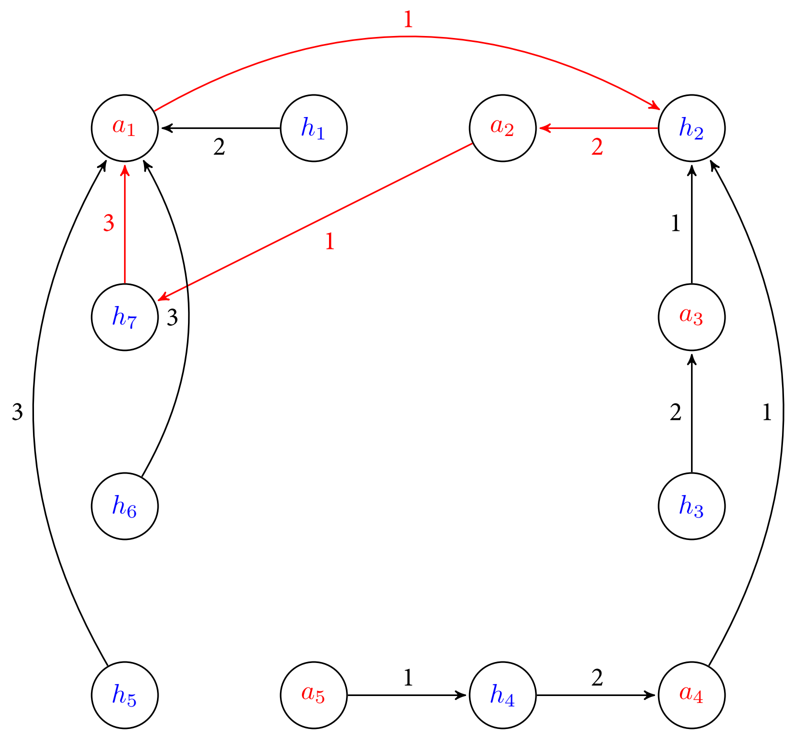


Figure 6.1: Step 1

The set of available houses in Step 1 in $H_V = \{h_5, h_6, h_7\}$. The only cycle that is formed at this step is

$$(a_1, h_2, a_2, h_7).$$

Therefore a_1 is assigned h_2 and a_2 is assigned h_7.

Step 2: The reduced preferences are as follows:

a_3	a_4	a_5
h_1	h_4	h_4
h_4	h_3	h_3
h_3	h_6	h_1
h_6	h_1	h_5
h_5	h_5	h_6
h_0	h_0	h_0

Table 6.4

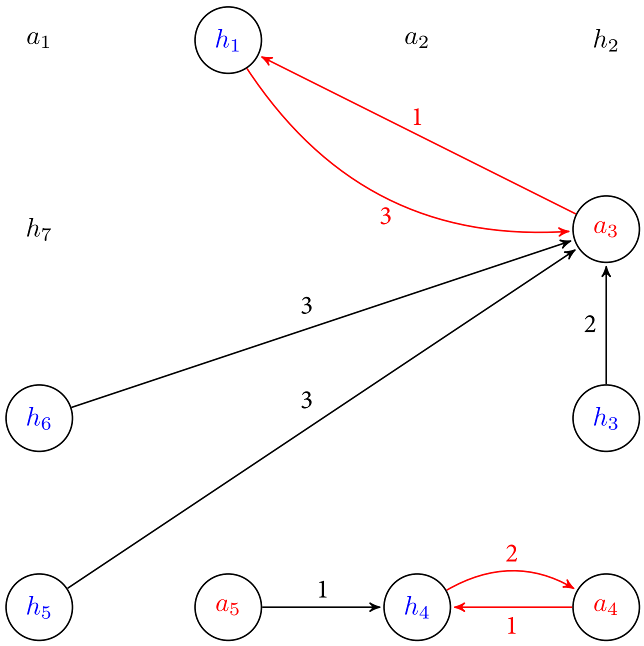


Figure 6.2: Step 2

Since a_1 leaves in Step 1, house h_1 becomes available in Step 2. Therefore the set of available houses for Step 2 is $\{h_1, h_5, h_6\}$. The available houses h_1, h_5 and h_6 all point to agent a_3, now the highest ranking agent. There are two cycles (a_3, h_1) and (a_4, h_4). Therefore a_3 is assigned h_1 and a_4 is assigned her own house h_4.

Step 3: The reduced preferences are as follows:

a_5
h_3
h_5
h_6
h_0

Table 6.5

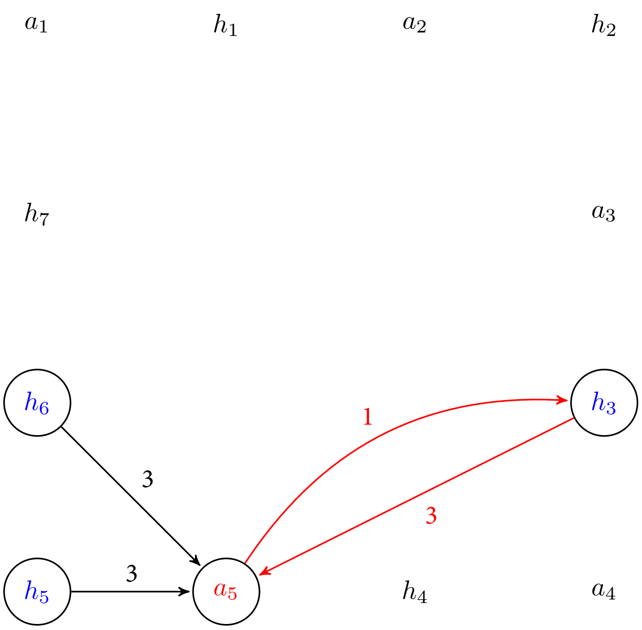


Figure 6.3: Step 3

Since a_3 leaves in Step 2, house h_3 becomes available for Step 3. Therefore the set of available houses for Step 3 is $\{h_3, h_5, h_6\}$. The available houses h_3, h_5, and h_6 all point to the only remaining agent a_5. The only cycle is (a_5, h_3). Therefore a_5 is assigned h_3.

There are no remaining agents so the algorithm terminates and the matching it induces is:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_2 & h_7 & h_1 & h_4 & h_3 \end{bmatrix}$$

6.20 Theorem (Proposition 1 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism φ^f is Pareto efficient.

Proof. (1) Consider the top trading cycles algorithm. Any agent who leaves at Step 1 is assigned his or her top choice and cannot be made better off.

(2) Any agent who leaves at Step 2 is assigned his or her top choice among those houses remaining at Step 2 and since the preferences are strict he or she cannot be made better off without hurting someone who left at Step 1.

(3) Proceeding in a similar way, no agent can be made better off without hurting someone who left at an earlier step. Therefore the mechanism φ^f is Pareto efficient.

□

6.21 Theorem (Proposition 2 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism φ^f is individually rational.

Proof. (1) Consider the top trading cycles algorithm. For any existing tenant $a \in A_E$, his or her house h_a points to him or her until he or she leaves.

(2) Therefore the assignment of a cannot be worse than his endowment h_a.

□

6.22 Theorem (Theorem 1 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism φ^f is strategy-proof.

Proof. The proof is analogous to the proof of Theorem 4.24. □

6.23 Theorem (Theorem 2 in Abdulkadiroğlu and Sönmez (1999)): Let f be an ordering, and φ a mechanism that is Pareto efficient, individually rational, and strategy-proof. Then for all $t = 1, 2, \ldots, |A|$, we have

$$\begin{aligned} &\big[\exists \succ \in \mathcal{P} \text{ such that } \varphi[\succ](f(t)) \succ_{f(t)} \varphi^f[\succ](f(t))\big] \\ \Rightarrow &\big[\exists \tilde{\succ} \in \mathcal{P} \text{ and } s < t \text{ such that } \varphi[\tilde{\succ}](f(s)) \succ_{f(s)} \varphi^f[\tilde{\succ}](f(s))\big]. \end{aligned} \tag{6.1}$$

Without loss of generality, we label agents so that $a_t = f(t)$ for all t.

6.24 Remark: There are many applications where agents are naturally ordered based on their seniority. Let f denote this ordering. Then Theorem 6.23 shows that there is no Pareto efficient, individually rational and strategy-proof mechanism which always better respects the seniority of the agents.

6.25 *Proof of Theorem 6.23, Step 1.* "For all $\succ \in \mathcal{P}$, $\varphi^f[\succ](a_1) \succsim_{a_1} \varphi[\succ](a_1)$".

(1) Suppose that there exists $\succ \in \mathcal{P}$ such that $\varphi[\succ](a_1) \succ_{a_1} \varphi^f[\succ](a_1)$.

(2) Denote $\mu \triangleq \varphi^f[\succ]$ and $h^* \triangleq \varphi[\succ](a_1)$. Then we have $h^* \succ_{a_1} \mu(a_1)$.

(3) Let B be the set of agents who leave the top trading cycles algorithm before a_1 under $\succ$. Similarly let G be the set of houses that leave the algorithm before agent a_1.

(4) Since all vacant houses point to agent a_1 as long as she is in the market, we have $B \subseteq A_E$.

(5) Moreover, since the only way agents in B leave the market before agent a_1 is by trading among themselves, we have

$$G = \{g \in H \mid g = h_j \text{ for some } a_j \in B\} = \{g \in H \mid g = \varphi^f[\succ](a_j) = \mu(a_j) \text{ for some } a_j \in B\}. \tag{6.2}$$

(6) Furthermore, for each $a_j \in B$, and for each $h \in H \setminus G$,

$$\mu(a_j) \succ_{a_j} h,$$

otherwise agent a_j would not leave the market before house h does.

(7) Due to the top trading cycles algorithm, it is only houses in G that can be better for agent a_1 than her assignment $\mu(a_1)$, and hence

$$h^* = \varphi[\succ](a_1) \in G.$$

(8) For all $a_j \in B$, let $\succ'_{a_j}$ be such that

$$\mu(a_j) \succsim'_{a_j} h_j \succ'_{a_j} h \text{ for all } h \in H \setminus \{\mu(a_j), h_j\}.$$

(9) Since φ is individually rational and $B \subseteq I_E$, we have

$$\varphi[\succ_{-B}, \succ'_B](a_j) \in \{\mu(a_j), h_j\}$$

for all $a_j \in B$, and therefore

$$\{\varphi[\succ_{-B}, \succ'_B](a_j) \mid a_j \in B\} = \{\mu(a_j), h_j \mid a_j \in B\} = G$$

be Relation (6.2).

(10) This, together with Pareto efficiency of φ, implies that

$$\varphi[\succ_{-B}, \succ'_B](a_j) = \mu(a_j) \text{ for all } a_j \in B.$$

(11) Let $\Omega = \{\sigma_t\}_{t=1}^{|B|!}$ be the set of all orderings of B where $\sigma_t(s)$ is the agent who is s-th in ordering σ_t.

(12) For any $\sigma \in \Omega$, construct the following sequence of preference profiles:

$$\begin{aligned}
\succ^0 (\sigma) &= \big(\succ_{-B}, \succ'_{\sigma(1)}, \succ'_{\sigma(2)}, \dots, \succ'_{\sigma(|B|)} \big) = (\succ_{-B}, \succ'_B) \\
\succ^1 (\sigma) &= \big(\succ_{-B}, \succ_{\sigma(1)}, \succ'_{\sigma(2)}, \dots, \succ'_{\sigma(|B|)} \big) \\
&\vdots \\
\succ^t (\sigma) &= \big(\succ_{-B}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \dots, \succ_{\sigma(t)}, \succ'_{\sigma(t+1)}, \dots, \succ'_{\sigma(|B|)} \big) \\
&\vdots \\
\succ^{|B|-1} (\sigma) &= \big(\succ_{-B}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \dots, \succ_{\sigma(|B|-1)}, \succ'_{\sigma(|B|)} \big) \\
\succ^{|B|} (\sigma) &= \big(\succ_{-B}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \dots, \succ_{\sigma(|B|-1)}, \succ_{\sigma(|B|)} \big) = \succ .
\end{aligned}$$

(13) For any $\sigma \in \Omega$, we have

$$\{\varphi[\succ^0 (\sigma)](a_j) \mid a_j \in B\} = \{\varphi[\succ_{-B}, \succ'_B](a_j) \mid a_j \in B\} = G,$$

that is, all houses in G are assigned to agents in B by the mechanism φ under $\succ^0 (\sigma)$.

(14) Recall that $a_1 \notin B$, $h^* \in G$ and $h^* = \varphi[\succ](a_1) = \varphi[\succ^{|B|} (\sigma)](a_1)$. That is, some agent in B is assigned a house that is not in G by the mechanism φ under $\succ^{|B|} (\sigma) = \succ$.

(15) For any $\sigma \in \Omega$, let $\succ^{n(\sigma)} (\sigma)$ be the first member of the sequence $\{\succ^0 (\sigma), \succ^1 (\sigma), \dots, \succ^{|B|} (\sigma)\}$ under which an agent in B is assigned a house that is not in G by the mechanism φ.

(16) Pick an ordering $\sigma^* \in \Omega$ with

$$n(\sigma^*) = \min_{\sigma \in \Omega} n(\sigma).$$

(17) By definition of $n(\sigma^*)$, there exist $a_k \in B$ and $h \in H \setminus G$, such that

$$\varphi[\succ^{n(\sigma^*)} (\sigma^*)](a_k) = h.$$

(18) Note that $\succ_{a_k}^{n(\sigma^*)}(\sigma^*)$ is either $\succ_{a_k}$ or $\succ'_{a_k}$. Note that the individual rationality guarantees that the mechanism φ will not assign a_k a house which is worse then h_k. Hence $\succ_{a_k}^{n(\sigma^*)}(\sigma^*) = \succ'_{a_k}$ contradicts Relation in the previous item.

(19) Consider the preference profile

$$\succ^{n(\sigma^*)-1}(\sigma^*) = (\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ'_k).$$

(20) Note that the number of agents in B revealing the preference $\succ_{a_j}$ is $n(\sigma^*) - 1$ under this profile.

(21) Therefore the minimality of $n(\sigma^*)$ ensures

$$\{\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)] \mid a_j \in B\} = G,$$

and this together with Pareto efficiency of φ ensures

$$\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_j) = \mu(a_j) \text{ for all } a_j \in B.$$

(22) In particular, this is true for agent $a_k \in B$:

$$\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_k) = \mu(a_k).$$

(23) Therefore,

$$\begin{aligned}\varphi[\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ'_{-k}](a_k) &= \varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_k) = \mu(a_k)\\ &\succ_{a_k} h = \varphi[\succ^{n(\sigma^*)}(\sigma^*)](a_k) = \varphi[\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ_{-k}](a_k),\end{aligned}$$

contradicting the strategy-proofness of φ.

□

6.26 *Proof of Theorem 6.23, Step 2.* Let $\ell \leq |A|$ and suppose Relation (6.1) holds for all $t = 1, 2, \ldots, \ell - 1$. Then Relation (6.1) holds for $t = \ell$ as well.

(1) Suppose Relation 6.1 holds for all $t = 1, 2, \ldots, \ell - 1$ but not for $t = \ell$.

(2) Then there exists a preference profile $\succ \in \mathcal{P}$ such that

$$\varphi[\succ](a_\ell) \succ_{a_\ell} \varphi^f[\succ](a_\ell),$$

and for all $\tilde{\succ} \in \mathcal{P}$ and $t = 1, 2, \ldots, \ell - 1$,

$$\varphi[\tilde{\succ}](a_t) \tilde{\succsim}_{a_t} \varphi^f[\tilde{\succ}](a_t).$$

(3) If there exists $\tilde{\succ} \in \mathcal{P}$ and $t \in \{1, 2, \ldots, \ell - 1\}$ such that

$$\varphi[\tilde{\succ}](a_t) \tilde{\succ}_{a_t} \varphi^f[\tilde{\succ}](a_t),$$

then by the induction hypothesis there exists $\succ^* \in \mathcal{P}$ and $s < t \leq \ell - 1$ such that

$$\varphi^f[\succ^*](a_s) \succ^*_{a_s} \varphi[\succ^*](a_s)$$

contradicting Relation in the previous item.

(4) Therefore, for each $\tilde{\succ} \in \mathcal{P}$ and $t = 1, 2, \ldots, \ell - 1$,

$$\varphi[\tilde{\succ}](a_t) = \varphi^f[\tilde{\succ}](a_t).$$

(5) Denote $\mu = \varphi^f[\succ]$ and $h^* = \varphi[\succ](a_\ell)$.

(6) Let B be the set of agents who leave the top trading cycles algorithm before agent a_ℓ under $\succ$. Similarly, let G be the set of houses that leave the algorithm before agent a_ℓ. Define $B_1 = B \setminus \{a_1, a_2, \ldots, a_{\ell-1}\}$.

(7) Note that no agent in A_N, unless she is a member of $\{a_1, a_2, \ldots, a_{\ell-1}\}$, can leave the market before agent a_ℓ. Therefore $B_1 \subseteq A_E$. (All vacant houses point to a_ℓ after $a_1, a_2, \ldots, a_\ell$ being removed)

(8) We also have for each $a_j \in B_1$, $h_j \in G$, for otherwise agent a_j could not form a cycle before agent a_ℓ left the market.

(9) For agent a_ℓ, it is only houses in G that can be better than her assignment $\mu(a_\ell)$. Therefore $h^* \in G$.

(10) For all $a_j \in B_1$, let $\succ'_{a_j} \in \mathcal{P}$ be such that

$$\mu(j) \succsim'_{a_j} h_j \succ'_{a_j} h \text{ for all } h \in H \setminus \{\mu(a_j), h_j\}.$$

(11) Clearly $\varphi^f[\succ_{-B_1}, \succ'_{B_1}] = \varphi^f[\succ] = \mu$, and hence for each $i = 1, 2, \ldots, \ell$,

$$\varphi[\succ_{-B_1}, \succ'_{B_1}](a_i) = \varphi^f[\succ_{-B_1}, \succ'_{B_1}](a_i) = \varphi^f[\succ](a_i) = \mu(a_i),$$

which in turn implies

$$\varphi[\succ_{-B_1}, \succ'_{B_1}](a_j) = \mu(a_j) \text{ for all } a_j \in B \setminus B_1$$

for $B \setminus B_1 \subseteq \{a_1, a_2, \ldots, a_{\ell-1}\}$.

(12) Individual rationality of φ implies

$$\varphi[\succ_{-B_1}, \succ'_{B_1}](a_j) \in \{\mu(a_j), h_j\}.$$

(13) Since $\mu(a_j) \in G$ for all $a_j \in B$, we have

$$\{\varphi[\succ_{-B_1}, \succ'_{B_1}](a_j) \mid a_j \in B\} \subseteq G.$$

(14) Pareto efficiency of φ implies that

$$\varphi[\succ_{-B_1}, \succ'_{B_1}](a_j) = \mu(a_j) \text{ for all } a_j \in B.$$

(15) Let $\Omega = \{\sigma_t\}_{t=1}^{|B_1|!}$ be the set of all orderings of B_1 where $\sigma_t(s)$ is the agent who is s-th in ordering σ_t. For any $\sigma \in \Omega$, construct the following sequence of preference profiles:

$$\begin{aligned}
\succ^0(\sigma) &= \big(\succ_{-B_1}, \succ'_{\sigma(1)}, \succ'_{\sigma(2)}, \ldots, \succ'_{\sigma(|B_1|)}\big) = (\succ_{-B_1}, \succ'_{B_1}) \\
\succ^1(\sigma) &= \big(\succ_{-B_1}, \succ_{\sigma(1)}, \succ'_{\sigma(2)}, \ldots, \succ'_{\sigma(|B_1|)}\big) \\
&\vdots \\
\succ^t(\sigma) &= \big(\succ_{-B_1}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \ldots, \succ_{\sigma(t)}, \succ'_{\sigma(t+1)}, \ldots, \succ'_{\sigma(|B_1|)}\big)
\end{aligned}$$

$$\vdots$$

$$\succ^{|B_1|-1}(\sigma) = \big(\succ_{-B_1}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \ldots, \succ_{\sigma(|B_1|-1)}, \succ'_{\sigma(|B_1|)} \big)$$

$$\succ^{|B_1|}(\sigma) = \big(\succ_{-B_1}, \succ_{\sigma(1)}, \succ_{\sigma(2)}, \ldots, \succ_{\sigma(|B_1|-1)}, \succ_{\sigma(|B_1|)} \big) = \succ .$$

(16) Since

$$\{\varphi[\succ^0(\sigma)](a_j) \mid a_j \in B\} = \{\varphi[\succ_{-B_1}, \succ'_{B_1}](a_j) \mid a_j \in B\} = G,$$

all houses in G are assigned to agents in B by mechanism φ under $\succ^0[\sigma]$.

(17) Note that $\varphi^f[\succ^k(\sigma)] = \varphi^f[\succ] = \mu$ for all $k \in \{0, 1, \ldots, |B_1|\}$.

(18) Therefore for each $a_j \in B \setminus B_1$, for each $k \in \{0, 1, \ldots, |B_1|\}$, we have

$$\varphi[\succ^k(\sigma)](a_j) = \varphi^f[\succ^k(\sigma)](a_j) = \varphi^f[\succ](a_j) = \mu(a_j) \in G.$$

(19) Recall that $a_\ell \notin B$, $h^* \in G$ and $h^* = \varphi[\succ](a_\ell) = \varphi[\succ^{|B_1|}(\sigma)](a_\ell)$. That is, some agent in B is assigned a house which is not in G by the mechanism φ under $\succ^{|B_1|}(\sigma) = \succ$.

(20) For each $\sigma \in \Omega$, let $\succ^{n(\sigma)}(\sigma)$ be the first member of the sequence $\{\succ^0(\sigma), \succ^1(\sigma), \ldots, \succ^{|B_1|}(\sigma)\}$ under which an agent in B is assigned a house that is not in G by the mechanism φ.

(21) Pick an ordering $\sigma^* \in \Omega$ with

$$n(\sigma^*) = \min_{\sigma \in \Omega} n(\sigma).$$

(22) By definition of $n(\sigma^*)$, there exist $a_k \in B$ and $h \in H \setminus G$, such that

$$\varphi[\succ^{n(\sigma^*)}(\sigma^*)](a_k) = h.$$

(23) We have already shown that $a_k \notin B_1$.

(24) Note that $\succ_{a_k}^{n(\sigma^*)}(\sigma^*)$ is either $\succ_{a_k}$ or $\succ'_{a_k}$. Note that the individual rationality guarantees that the mechanism φ will not assign a_k a house which is worse then h_k. Hence $\succ_{a_k}^{n(\sigma^*)}(\sigma^*) = \succ'_{a_k}$ contradicts Relation in the previous item.

(25) Consider the preference profile

$$\succ^{n(\sigma^*)-1}(\sigma^*) = (\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ'_k).$$

(26) Note that the number of agents in B_1 revealing the preference $\succ_{a_j}$ is $n(\sigma^*) - 1$ under this profile.

(27) Therefore the minimality of $n(\sigma^*)$ ensures

$$\{\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)] \mid a_j \in B\} = G,$$

and this together with Pareto efficiency of φ ensures

$$\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_j) = \mu(a_j) \text{ for all } a_j \in B.$$

(28) In particular, this is true for the agent $a_k \in B$:

$$\varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_k) = \mu(a_k).$$

(29) Therefore,

$$\varphi[\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ'_{-k}](a_k) = \varphi[\succ^{n(\sigma^*)-1}(\sigma^*)](a_k) = \mu(a_k)$$
$$\succ_{a_k} h = \varphi[\succ^{n(\sigma^*)}(\sigma^*)](a_k) = \varphi[\succ_{-k}^{n(\sigma^*)}(\sigma^*), \succ_{-k}](a_k),$$

contradicting the strategy-proofness of φ.

□

6.4 You request my house—I get your turn algorithm

6.27 You request my house—I get your turn (YRMH-IGYT) algorithm, induced by a given ordering f:

Phase 1: Assign the first agent her top choice, the second agent her top choice among the remaining houses, and so on, until someone demands the house of an existing tenant.

Phase 2: If at that point the existing tenant whose house is requested is already assigned another house, then do not disturb the procedure.

Otherwise, modify the remainder of the ordering by inserting this existing tenant before the requestor at the priority order and proceed with the Phase 1 through this existing tenant.

Similarly, insert any existing tenant who is not already served just before the requestor in the priority order once her house is requested by an agent.

Phase 3: If at any point a cycle forms, it is formed by exclusively existing tenants and each of them requests the house of the tenant who is next in the cycle. A cycle is an ordered list $(h_1, a_1, \ldots, h_k, a_k)$ of occupied houses and existing tenants where agent a_1 demands the house a_2, h_2, agent a_2 demands the house of agent a_3, h_3, . . . , agent a_k demands the house of a_1, h_1.

In such case, remove all agents in the cycle by assigning them the house they demand and proceed similarly.

6.28 Example.

- $A_E = \{a_1, a_2, \ldots, a_9\}$ is the set of existing tenants,
- $A_N = \{a_{10}, a_{11}, \ldots, a_{16}\}$ is the set of new applicants, and
- $H_V = \{h_{10}, h_{11}, \ldots, h_{16}\}$ is the set of vacant houses.

Suppose that each existing tenant a_k occupies h_k for each $k = 1, 2, \ldots, 9$. Let the preference profile $\succ$ be given as:

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
h_{15}	h_3	h_1	h_2	h_9	h_6	h_7	h_6	h_{11}	h_7	h_2	h_4	h_6	h_8	h_1	h_5
	h_4	h_3					h_{12}		h_3	h_4	h_{14}	h_{13}			
									h_{12}	h_{16}					
									h_{10}						

Table 6.6

Let $f = (a_{13}, a_{15}, a_{11}, a_{14}, a_{12}, a_{16}, a_{10}, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9)$ be the ordering of the agents.

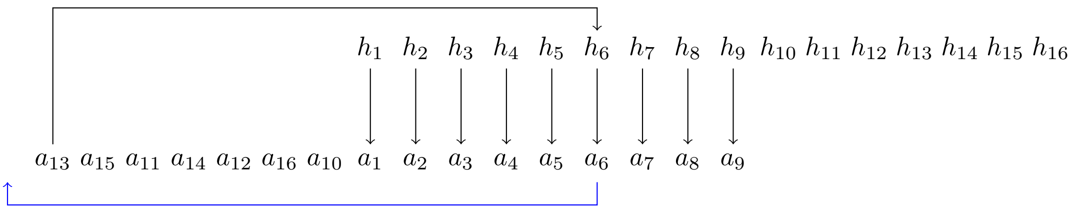


Figure 6.4: Step 1

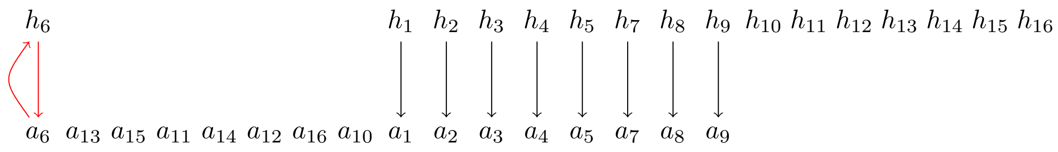


Figure 6.5: Step 2

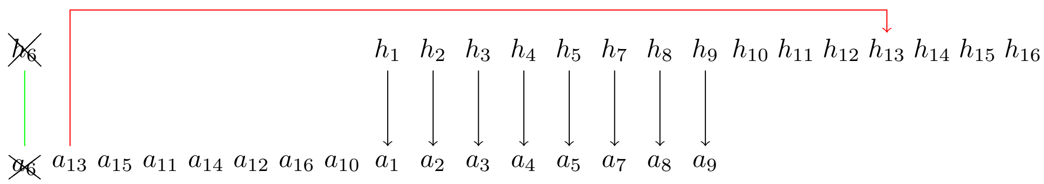


Figure 6.6: Step 3

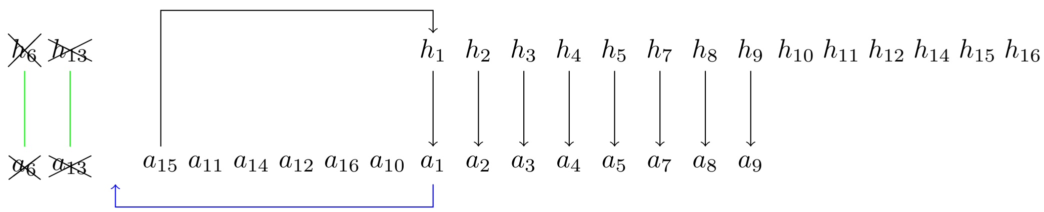


Figure 6.7: Step 14

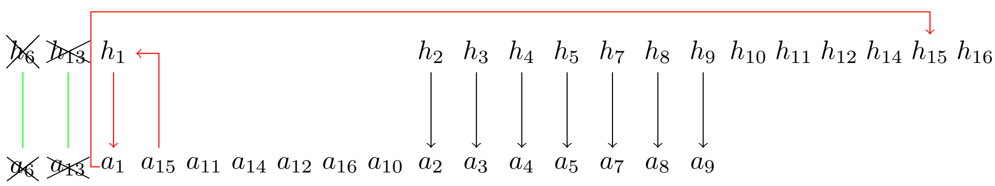


Figure 6.8: Step 5

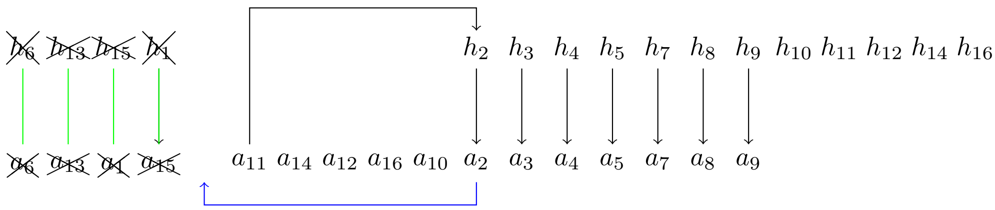


Figure 6.9: Step 6

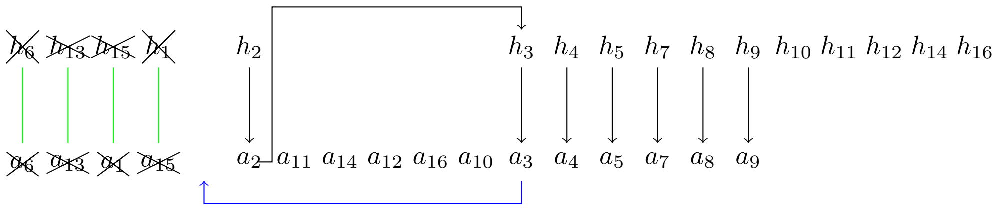


Figure 6.10: Step 7

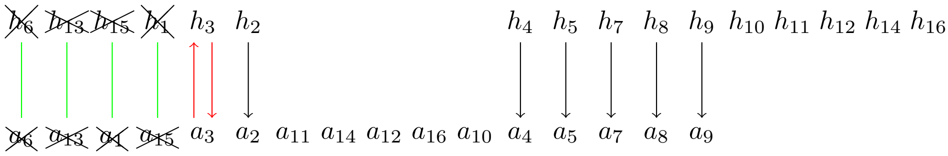


Figure 6.11: Step 8

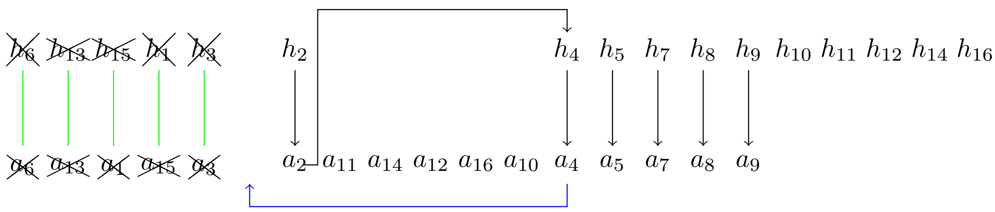


Figure 6.12: Step 9

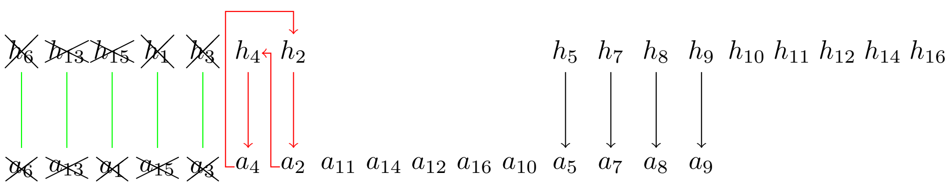


Figure 6.13: Step 10

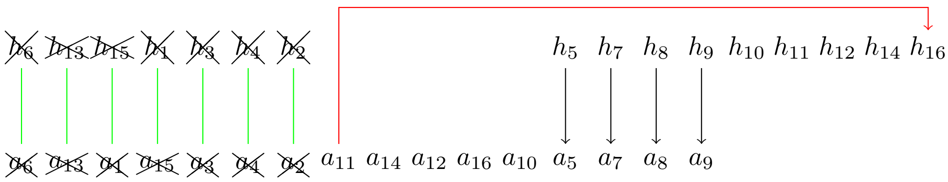


Figure 6.14: Step 11

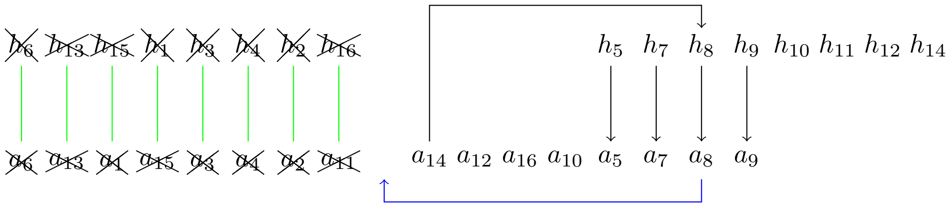


Figure 6.15: Step 12

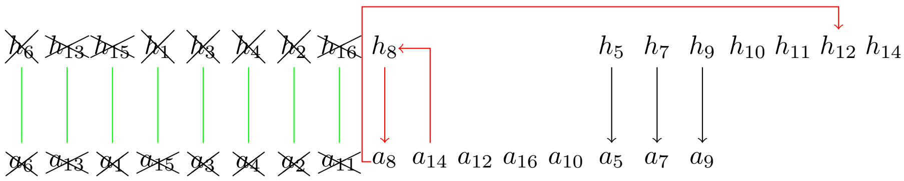


Figure 6.16: Step 13

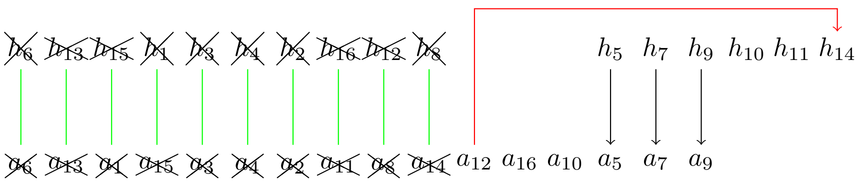


Figure 6.17: Step 14

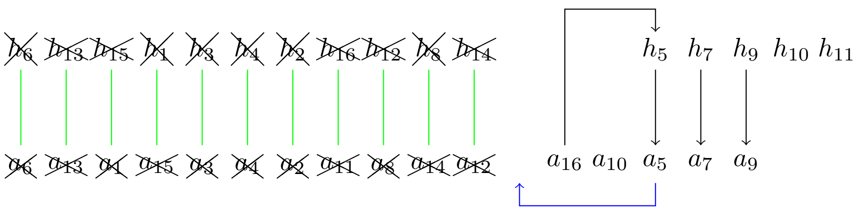


Figure 6.18: Step 15

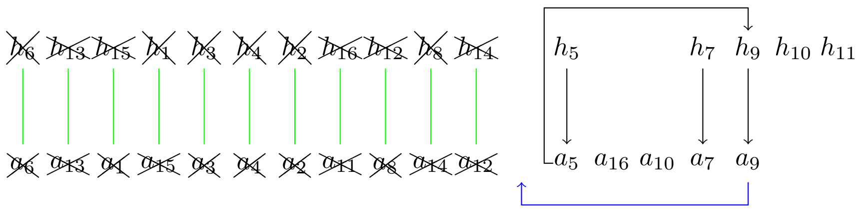


Figure 6.19: Step 16

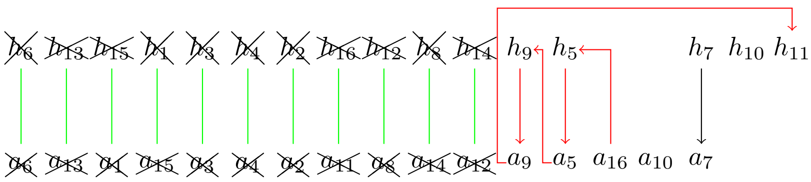


Figure 6.20: Step 17

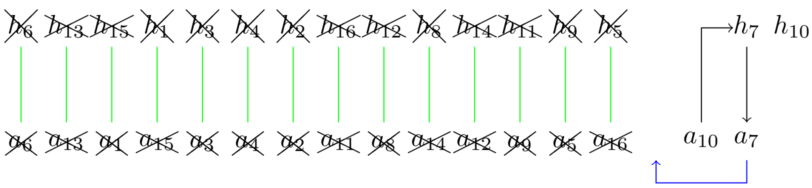


Figure 6.21: Step 18

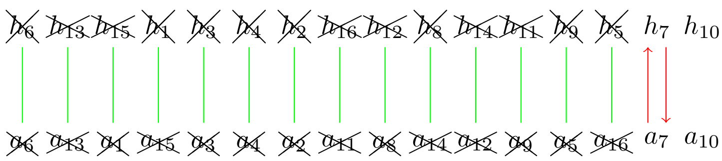


Figure 6.22: Step 19

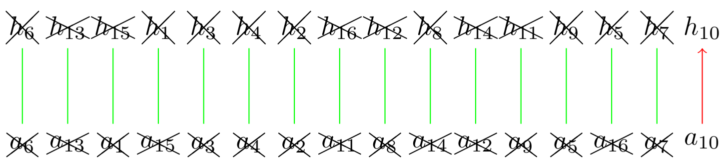


Figure 6.23: Step 20

The outcome of the algorithm is

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ h_{15} & h_4 & h_3 & h_2 & h_9 & h_6 & h_7 & h_{12} & h_{11} & h_{10} & h_{16} & h_{14} & h_{13} & h_8 & h_1 & h_5 \end{bmatrix}.$$

✌ 6.29 Theorem (Theorem 3 in Abdulkadiroğlu and Sönmez (1999)): For a given ordering f, the YRMH-IGYT algorithm yields the same outcome as the top trading cycles algorithm.

Proof. For any set B of agents and set G of houses remaining in the algorithm, YRMH-IGYT algorithm assigns the next series of houses in one of two possible ways.

- Case 1: There is a sequence of agents $a_1, a_2, \ldots, a_k$ (which may consist of a single agent) where agent a_1 has the highest priority in B and demands house of a_2, agent a_2 demands house of a_3, ..., agent a_{k-1} demands house of a_k, and a_k demands an available house h. At this point agent a_k is assigned house h, the next agent a_{k-1} is assigned house h_k (which just became available), ..., and finally agent a_1 is assigned house h_2. Note that the ordered list $(h, a_1, h_2, a_2, \ldots, h_k, a_k)$ is a (top trading) cycle for the pair (B, G).
- Case 2: There is a loop $(a_1, a_2, \ldots, a_k)$ of agents. When that happens agent a_1 is assigned the house of a_2, agent a_2 is assigned house of a_3, ..., agent a_k is assigned house of a_1. In this case $(h_1, a_1, h_2, a_2, \ldots, h_k, a_k)$ is a (top trading) cycle for the pair (B, G).

Hence the YRMH-IGYT algorithm locates a cycle and implements the associated trades for any sets of remaining agents and houses. □

6.5 Axiomatic characterization of YRMH-IGYT

6.30 Let $\sigma\colon H \to H$ be a permutation for vacant houses. That is, σ is a bijection such that $\sigma(h) = h$ for any $h \in H_O \cup \{h_0\}$.

Given a preference profile $\succ$, let $\succ^\sigma$ be a preference profile where σ is a permutation for vacant houses. That is, $g \succ^\sigma_a h$ if and only if $\sigma^{-1}(g) \succ_a \sigma^{-1}(h)$.

☞ 6.31 Definition: A mechanism is weakly neutral if labeling of vacant houses has no effect on the outcome of the mechanism.

Formally, a mechanism φ is weakly neutral if for any house allocation problem with existing tenants and any permutation for vacant houses, we have

$$\varphi[\succ^\sigma](a) = \sigma\big(\varphi[\succ](a)\big) \text{ for any } a \in A.$$

6.32 For any problem $\Gamma = \langle A_E, A_N, H_O, H_V, \succ\rangle$, any $A' \subseteq A$, any $H' \subseteq H$, and any matching μ, the reduced problem of Γ with respect to A' and H' under μ is

$$r^\mu_{A',H'}[\Gamma] = \langle A'_E, A'_N, H'_O, H'_V, (\succ_a|_{H'})_{a\in A'}\rangle$$

when $\big(\mu(A\setminus A')\cup(H\setminus H')\big)\cap\{h_a\}_{a\in A'_E} = \emptyset$, where $A'_E = A'\cap A_E$, $A'_N = A'\cap A_N$, $H'_O = \big(H'\setminus\mu(A\setminus A')\big)\cap H_O$, $H'_V = \big(H'\setminus\mu(A\setminus A')\big)\cap H_V$, and $\succ_a|_{H'}$ is the restriction of agent i's preference to the remaining houses.

☞ 6.33 Definition: A mechanism φ is consistent if for any problem $\Gamma = \langle A_E, A_N, H_O, H_V, \succ\rangle$, any $A' \subseteq A$, any $H' \subseteq H$, and any matching μ, one has

$$\varphi[\Gamma](a) = \varphi\left[r^{\varphi[\Gamma]}_{A',H'}(\Gamma)\right](a) \text{ for each } a \in A'.$$

✌ 6.34 Theorem (Theorem 1 in Sönmez and Ünver (2010)): A mechanism is Pareto efficient, individually rational, strategy-proof, weakly neutral, and consistent if and only if it is a YRMH-IGYT mechanism.

Proof. Omitted. □

6.6 Random house allocation with existing tenants

6.35 Here we assume that $|A_E| = n$ and $|A_N| = |H_V| = m$.

6.36 Let $\mathcal{M}^* = \{\mu \in \mathcal{M} \mid \mu(a) = h_a \text{ for all } a \in A_E\}$ be the set of matchings which assign each existing tenant her current house. For given A, H and for any $\mu \in \mathcal{M}^*$, define a mechanism φ^μ as follows: For any preference profile mechanism φ^μ interprets μ as the initial allocation and chooses the core of the induced housing market.

Since the preferences are fixed throughout the paper, we denote the outcome of mechanism φ^μ also with φ^μ dropping the argument in $\varphi^\mu[\succ]$.

Core from random endowments, φ^{cre}, is defined as

$$\varphi^{\text{cre}} = \sum_{\mu \in \mathcal{M}^*} \frac{1}{m!} \varphi^\mu.$$

6.37 Let $\mathcal{F}^* = \{f \text{ is an ordering of agents} \mid f^{-1}(a) < f^{-1}(a') \text{ for all } a \in A_N \text{ and } a' \in A_E\}$.

YRMH–IGYT (or TTC) from random priority, φ^{rp}, is defined as

$$\varphi^{\text{rp}} = \sum_{f \in \mathcal{F}^*} \frac{1}{n!m!} \varphi^f.$$

6.38 Theorem (Theorem 1 in Sönmez and Ünver (2005)): φ^{cre} and φ^{rp} are equivalent.

Proof. Omitted. □

Chapter 7

Hierarchical exchange rule

Pápai (2000), Pycia and Ünver (2015)

Chapter 8

Lottery mechanism

Contents

8.1 The former model

☞ 8.1 A random assignment problem, denoted by $\Gamma = \langle N, O, \succ \rangle$, consists of

- N is a finite set of agents,
- O is a finite set of indivisible objects, where $|N| = |O| = n$, and
- $\succ = (\succ_i)_{i \in N}$, where $\succ_i$ is agent i's strict preference. We write $a \succsim_i b$ if and only if $a \succ_i b$ or $a = b$.

☞ 8.2 A deterministic assignment (or simply assignment) is a one-to-one mapping from N to O; it will be uniquely represented as a permutation matrix $X = (X_{io})$ (an $n \times n$ matrix with entries 0 or 1 and exactly one non-zero entry per row and one per column).

We identify rows with agents and columns with objects.

$$X_{io} = \begin{cases} 1, & \text{if agent } i \text{ receives object } o \text{ under the assignment } X; \\ 0, & \text{if agent } i \text{ does not receive object } o \text{ under the assignment } X. \end{cases}$$

Let $\mathcal{D}$ denote the set of permutation matrices/determnistic assignments.

8.3 Example for deterministic assignment.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

☞ 8.4 A random assignment is a bistochastic matrix $P = (P_{io})_{i\in N, o\in O}$ (a matrix with non-negative entries, with each row and column summing to 1). The value P_{io} describes the probability that the agent i receives the object o.

☞ 8.5 A lottery assignment is a probability distribution p over the set of deterministic assignments, where $p(X)$ denotes the probability of the deterministic assignment X.

8.6 We associate to each lottery assignment p a random assignment P is the following way:

$$P = \sum_{X\in\mathcal{D}} p(X)\cdot X.$$

By the classical Birkhoff-von Neumann theorem (see Pulleyblank (1995), page 187–188), any bistochastic matrix can be written (not necessarily uniquely) as a convex combination of permutation matrices.

Henceforth, we identify lottery assignments with the corresponding random assignments and use these terms interchangeably. Let $\mathcal{R}$ denote the set of random assignments.

☞ 8.7 A (lottery) mechanism is a procedure the assigns a random assignment P for each random assignment problem $\langle N, O, \succ\rangle$.

8.8 A von Neumann-Morgenstern utility function u_i is a real-valued mapping from O to $\mathbb{R}$.

We extend the domain of u_i to the set of random assignments as follows. Agent i's expected utility for the random assignment P is

$$u_i(P) = \sum_{o\in O} P_{io}\cdot u_i(o).$$

We say u_i is consistent/compatible with $\succ_i$ when $u_i(a) > u_i(b)$ if and only if $a \succ_i b$.

8.2 Random priority mechanism

8.9 An ordering $f\colon \{1, 2, \ldots, n\} \to N$ is a one-to-one and onto function.

Let $\mathcal{F}$ be the set of orderings. Given an ordering f and preference profile $\succ$, the corresponding simple serial dictatorship assignment is denoted by $\varphi^f[\succ]$, defined as usual.

8.10 Random priority, RP, (or random serial dictatorship, RSD) is defined as

$$\mathrm{RP} = \varphi^{\mathrm{rp}} = \mathrm{RSD} = \varphi^{\mathrm{rsd}} = \sum_{f \in \mathcal{F}} \frac{1}{n!} \varphi^f.$$

Note that, in this chapter, $\varphi^f[\succ]$, RP$[\succ]$ and RSD$[\succ]$ will be treated as matrices.

8.11 Core from random endowments, φ^{cre}, is defined as

$$\varphi^{\mathrm{cre}} = \sum_{\mu \in \mathcal{M}} \frac{1}{n!} p^{\varphi^\mu}.$$

8.12 Theorem (Theorem 2 in Abdulkadiroğlu and Sönmez (1998)):

$$\varphi^{\mathrm{rsd}} = \varphi^{\mathrm{cre}}.$$

8.13 Pathak and Sethuraman (2011)

8.3 Simultaneous eating algorithm and probabilistic serial mechanism

8.14 Let $\omega_i \colon [0,1] \to \mathbb{R}_+$ be agent i's eating speed function, that is, $\omega_i(t)$ is the speed at which agent i is allowed to eat at time t.

The speed $\omega_i(t)$ is non-negative and the total amount that agent i will eat between $t = 0$ and $t = 1$ (the end time of the algorithm) is one:

$$\int_0^1 \omega_i(t)\,\mathrm{d}t = 1.$$

Let W denote the set of eating speed functions:

$$W = \left\{ \omega_i \colon [0,1] \to \mathbb{R}_+ \;\middle|\; \omega_i \text{ is measurable and } \int_0^1 \omega_i(t)\,\mathrm{d}t = 1 \right\}.$$

8.15 Simultaneous eating algorithm. Given the profile of eating speeds $\omega = (\omega_i)_{i \in N}$ and the preference profile $\succ$, the algorithm lets each agent i eat her best available good at the prespecified speeds.

8.16 For any $o \in O' \subseteq O$, let $N(o, O') = \{i \in N \mid o \succ_i b \text{ for all } b \in O', b \neq o\}$.

Given the profile of eating speeds $\omega = (\omega_i)_{i \in N}$ and the preference profile $\succ$, the outcome of simultaneous eating algorithm is defined by the following recursive procedure.

Step 0 Let $t^0 = 0$, $O^0 = O$, $P^0 = (0)$, the $n \times n$ matrix of zeros.

Step k Suppose that $t^0, O^0, P^0, \ldots, t^{k-1}, O^{k-1}, P^{k-1}$ are already defined.

- For any $o \in O^{k-1}$, define

$$t^k(o) = \begin{cases} \min \left\{ t \;\middle|\; \displaystyle\sum_{i \in N(o, O^{k-1})} \int_{t^{k-1}}^{t} \omega_i(s)\,\mathrm{d}s + \sum_{i \in N} P_{io}^{k-1} = 1 \right\}, & \text{if } N(o, O^{k-1}) \neq \emptyset, \\ +\infty, & \text{if } N(o, O^{k-1}) = \emptyset. \end{cases}$$

Each agent in $N(o, O^{k-1})$ will eat the object o immediately after time $t = t^{k-1}$, and $t^k(o)$ specifies the time when the object o will be eaten away given that no new agent enter.

- Define
$$t^k = \min_{o \in O^{k-1}} t^k(o).$$
From t^{k-1} onwards, once an object is eaten away, then this instant is denoted as t^k. Note that, at the instant t^k, there could be more than one objects which are eaten away.
- Define
$$O^k = O^{k-1} \setminus \{o \mid t^k(o) = t^k\}.$$
The set $\{o \mid t^k(o) = t^k\}$ is exactly the set of objects which are eaten away at instant t^k, and the set O^k denotes the set of objects which remain after t^k.
- Define $P^k = (P^k_{io})$:
$$P^k_{io} = \begin{cases} P^{k-1}_{io} + \int_{t^{k-1}}^{t^k} \omega_i(s)\, \mathrm{d}s, & \text{if } i \in N(o, O^{k-1}), \\ P^{k-1}_{io}, & \text{otherwise.} \end{cases}$$
Between t^{k-1} and t^k, if agent i eats object o (no matter whether o is eaten away at instant t^k), then she will obtain a quantity $\int_{t^{k-1}}^{t^k} \omega_i(s)\, \mathrm{d}s$ of object o.
The relation $\int_0^1 \omega_i(s)\, \mathrm{d}s$ guarantees that $P^k_{io} \leq 1$.

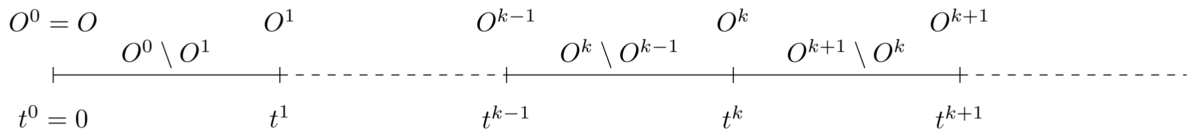


Figure 8.1

8.17 By the construction, $O^k \subsetneqq O^{k-1}$ for all k, and hence $O^n = \emptyset$ and $P^n = P^{n+1} = \cdots$.

The matrix P^n is the random assignment corresponding to the profile of eating speed functions $\omega = (\omega_i)_{i \in N}$ and the preference profile $\succ$: $P_\omega[\succ] = P^n$.

8.18 The probabilistic serial assignment at a given preference profile $\succ$ is the random assignment corresponding to the profile of uniform eating speeds $\omega_i(t) = 1$ for all $i \in N$, all $t \in [0, 1]$. It is denoted by PS$[\succ]$ or $\varphi^{\mathrm{ps}}[\succ]$.

8.4 Efficiency

8.4.1 Basics

☞ 8.19 Given a preference profile $\succ$. An assignment X Pareto dominates an assignment Y at $\succ$ if
$$X \neq Y, \text{ and } X_i \succsim_i Y_i \text{ for all } i \in N,$$
where X_i denotes the object agent i receives under X.

An assignment X is Pareto optimal at $\succ$ if there is no assignment that Pareto dominates it at $\succ$.

☞ 8.20 Given a preference profile $\succ$ and a profile of von Neumann-Morgenstern utilities u.

- A random assignment P is ex ante efficient at u, if P is Pareto optimal in $\mathcal{R}$ at u.
- A lottery assignment p is ex post efficient at $\succ$, if all assignments in its support are Pareto optimal at $\succ$.

  A random assignment P is ex post efficient at $\succ$, if it admits a lottery assignment that is ex post efficient at $\succ$.

By Theorem 5.9, P is ex post efficient at $\succ$ if and only if it takes the form

$$P = \sum_{f\in\mathcal{F}} \mu_f \cdot \varphi^f[\succ] \text{ for some convex system of weights } \mu_f,$$

where φ^f is the simple serial dictatorship induced by the ordering f.

8.4.2 Ordinal efficiency

☞ 8.21 Given a preference $\succ_i$. A lottery P_i first-order stochastically dominates another lottery Q_i with respect to $\succ_i$, denoted by $P_i \operatorname{sd}(\succ_i) Q_i$, if

$$\sum_{k\colon o_k \succsim_i o_j} P_{ik} \geq \sum_{k\colon o_k \succsim_i o_j} Q_{ik} \text{ for all } j.$$

That is, P_i first-order stochastically dominates Q_i if and only if

- the probability of receiving the first choice is at least as much in P_i as in Q_i, and in general,
- for any k, the probability of receiving one of top k choices is at least as much in P_i as in Q_i.

☞ 8.22 Given a preference profile $\succ$. A random assignment P ordinally dominates (or stochastically dominates) another random assignment Q at $\succ$ if $P \neq Q$ and for each agent i, the lottery P_i first-order stochastically dominates the lottery Q_i with respect to $\succ_i$, where P_i is the i-th row of the matrix P which represents the lottery allocation of agent i.

8.23 Proposition: P_i first-order stochastically dominates Q_i with respect to $\succ_i$ if and only if $u_i \cdot P_i \geq u_i \cdot Q_i$ for any von Neumann-Morgenstern utility function u_i consistent with $\succ_i$. Here u_i is regarded as a utility vector $\big(u_i(o_1), u_i(o_2), \ldots, u_i(o_n)\big)$.

Moreover, $P_i \neq Q_i$ implies that the corresponding inequality is strict.

"$\Rightarrow$". (1) Without loss of generality, let us assume $o_1 \succ_i o_2 \succ_i \cdots \succ_i o_n$.

(2) Then we have

$$\sum_{k=1}^{t} P_{ik} \geq \sum_{k=1}^{t} Q_{ik} \text{ for all } t = 1, 2, \ldots, n.$$

(3) For any von Neumann-Morgenstern utility function u_i which is consistent with $\succ_i$, we have $u_i(o_j) - u_i(o_{j+1}) \geq 0$ for all $j = 1, \ldots, n-1$, and hence

$$\begin{aligned} u_i \cdot P_i &= \sum_{k=1}^{n} u_i(o_k) P_{ik} \\ &= u_i(o_n) \sum_{k=1}^{n} P_{ik} + [u_i(o_{n-1}) - u_i(o_n)] \sum_{k=1}^{n-1} P_{ik} + [u_i(o_{n-2}) - u_i(o_{n-1})] \sum_{k=1}^{n-2} P_{ik} + \cdots \end{aligned}$$

$$+ [u_i(o_j) - u_i(o_{j+1})] \sum_{k=1}^{j} P_{ik} + \cdots + [u_i(o_1) - u_i(o_2)] \sum_{k=1}^{1} P_{ik}$$
$$\geq u_i(o_n) \sum_{k=1}^{n} Q_{ik} + [u_i(o_{n-1}) - u_i(o_n)] \sum_{k=1}^{n-1} Q_{ik} + \cdots + [u_i(o_1) - u_i(o_2)] \sum_{k=1}^{1} Q_{ik}$$
$$= u_i \cdot Q_i$$

□

"$\Leftarrow$". (1) Without loss of generality, let us assume $o_1 \succ_i o_2 \succ_i \cdots \succ_i o_n$. Then it suffices to show that

$$\sum_{k=1}^{t} P_{ik} \geq \sum_{k=1}^{t} Q_{ik} \text{ for all } t = 1, 2, \ldots, n.$$

(2) Assume that $1 \leq s \leq n$ is the first number such that $\sum_{k=1}^{s} P_{ik} < \sum_{k=1}^{s} Q_{ik}$.

(3) Take $\varepsilon > 0$, and construct a von Neumann-Morgenstern utility function u_i such that

$$0 < u_i(o_j) - u_i(o_{j+1}) \begin{cases} < \varepsilon, & \text{if } j \neq s \\ > \frac{n-1}{\sum_{k=1}^{s} Q_{ik} - P_{ik}} \varepsilon & \text{if } j = s \end{cases}.$$

(4) Therefore, we have

$$u_i \cdot P_i - u_i \cdot Q_i < \varepsilon \cdot \sum_{t \neq s,n} \sum_{k=1}^{t} [P_{ik} - Q_{ik}] - (n-1)\varepsilon < 0,$$

which contradicts the hypothesis.

□

8.24 The random assignment P is ordinally efficient at $\succ$ if it is not ordinally dominated at $\succ$ by any other random assignment.

☞ 8.25 Proposition (Lemma 2 in Bogomolnaia and Moulin (2001)): Given a random assignment P, a preference profile $\succ$, and a profile u of von Neumann-Morgenstern utilities consistent with $\succ$.

(i) If P is ex ante efficient at u, then it is ordinally efficient at $\succ$; the converse statement holds for $n = 2$ but fail for $n \geq 3$.

(ii) If P is ordinally efficient at $\succ$, then it is ex post efficient at $\succ$; the converse statement holds for $n \leq 3$ but fail for $n \geq 4$.

8.26 *Proof of Proposition 8.25, Statement (i).*

If P is ex ante efficient at u, then it is ordinally efficient at $\succ$.

(1) Suppose that P is not ordinally efficient at $\succ$.

(2) Then there exists another random assignment Q which ordinally dominates P at $\succ$.

(3) Then by Proposition 8.23, we have $u_i \cdot P_i \geq u_i \cdot Q_i$ for all i.

(4) Moreover, $P_i \neq Q_i$ for some i, and hence the corresponding inequality is strict so that P is ex ante Pareto inferior to Q.

P may not be ex ante efficient at u, if it is ordinally efficient at $\succ$ when $n \geq 3$.

(1) Consider the following example: there are three agents $\{1, 2, 3\}$, three objects $\{a, b, c\}$, unanimous ordinal preferences $a \succ_i b \succ_i c$ and the consistent von Neumann-Morgenstern utilities:

$$u_1(x) = \begin{cases} 1, & \text{if } x = a \\ 0.8, & \text{if } x = b \\ 0, & \text{if } x = c \end{cases}, \quad u_2(x) = u_3(x) = \begin{cases} 1, & \text{if } x = a \\ 0.2, & \text{if } x = b \\ 0, & \text{if } x = c \end{cases}.$$

(2) It is clear that the random assignment $P = (P_{ik})$ is not ex ante efficient, where $P_{ik} = \frac{1}{3}$. P leads to a utility profile $(0.6, 0.4, 0.4)$, and the random assignment $Q = (Q_{ik})$ yields to a utility profile $(0.8, 0.5, 0.5)$, where $Q_{1b} = 1$, $Q_{2a} = Q_{2c} = Q_{3a} = Q_{3c} = \frac{1}{2}$.

(3) Every random assignment is ordinally efficient. Otherwise, assume P is not ordinally efficient, and is stochastically dominated by Q.

(4) Then $P \neq Q$, and $\sum_{k\colon o_k \succsim_i o_j} Q_{ik} \geq \sum_{k\colon o_k \succsim_i o_j} P_{ik}$ for all i and j.

(5) Then,

$$\sum_i \sum_{k\colon o_k \succsim_i o_j} Q_{ik} \geq \sum_i \sum_{k\colon o_k \succsim_i o_j} P_{ik} \text{ for all } j.$$

(6) Since three agents have the same ordinal preference, we have

$$\sum_i \sum_{k\colon o_k \succsim_i o_j} Q_{ik} = \sum_{k\colon o_k \succsim_i o_j} \sum_i Q_{ik} = \sum_{k\colon o_k \succsim_i o_j} 1 = \sum_{k\colon o_k \succsim_i o_j} P_{ik} = \sum_i \sum_{k\colon o_k \succsim_i o_j} P_{ik} \text{ for all } j,$$

which leads to a contradiction.

□

8.27 *Proof of Proposition 8.25, Statement (ii).*

If P is ordinally efficient at $\succ$, then it is ex post efficient at $\succ$.

(1) Suppose P is not ex post efficient at $\succ$.

(2) Consider a decomposition of P as a convex combination of deterministic assignments:

$$P = \sum_X p(X) \cdot X.$$

(3) Then there is an element X that is Pareto inferior at $\succ$ and such that $p(X) > 0$.

(4) Let Y be a deterministic assignment Pareto superior to X.

(5) Upon replacing X by Y in the summation, we obtain a random assignment that stochastically dominates P (note that stochastic dominance is preserved by convex combinations).

When $n \geq 3$, P may not be ex post efficient at $\succ$, if it is ordinally efficient at $\succ$.

(1) Consider the following example: there are four agents $\{1, 2, 3, 4\}$, four objects $\{a, b, c, d\}$. The preferences are as follows:

1 and 2	3 and 4
a	b
b	a
c	d
d	c

Table 8.1

(2) Consider the following two random assignments

$$P=\begin{pmatrix}\frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12}\\ \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12}\\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12}\\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12}\end{pmatrix} \text{ and } Q=\begin{pmatrix}\frac{1}{2} & 0 & \frac{1}{2} & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & \frac{1}{2} & 0 & \frac{1}{2}\end{pmatrix}.$$

(3) Every agent gets her first choice with probability $\frac{1}{2}$ under Q, and first choice with $\frac{5}{12}$ and second choice with $\frac{1}{12}$ under P.

Every agent gets her third choice with probability $\frac{1}{2}$ under Q, and third choice with $\frac{5}{12}$ and fourth choice with $\frac{1}{12}$ under P.

Therefore, P is stochastically dominated by Q, and hence not ordinally efficient.

(4) It is straightforward to check that $P=\sum_{f\in\mathcal{F}}\frac{1}{n!}\varphi^f[\succ]$, so P is ex post efficient.

When $n\leq 3$, if P is ex post efficient at $\succ$, then it is ordinally efficient at $\succ$.

(1)

□

8.28 Proposition (Theorem 1 in McLennan (2002)): If P is ordinally efficient at $\succ$, then there is a profile u of von Neumann-Morgenstern utilities which is consistent with $\succ$, such that P is ex ante efficient at u.

8.4.3 Efficiency of random priority

8.29 Recall Theorem 5.9 (Lemma 1 in Abdulkadiroğlu and Sönmez (1998)): Simple serial dictatorship is Pareto efficient.

8.30 Example (Section 2 in Bogomolnaia and Moulin (2001)): Random serial dictatorship is not ordinally efficient.

Consider the following example: there are four agents $\{1,2,3,4\}$, four objects $\{a,b,c,d\}$. The preferences are as follows:

1 and 2	3 and 4
a	b
b	a
c	d
d	c

Table 8.2

The outcome of random serial dictatorship is

$$P = \begin{pmatrix} \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \end{pmatrix},$$

which is ordinally dominated by

$$Q = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Every agent gets her first choice with probability $\frac{1}{2}$ under Q, and first choice with $\frac{5}{12}$ and second choice with $\frac{1}{12}$ under P.

Every agent gets her third choice with probability $\frac{1}{2}$ under Q, and third choice with $\frac{5}{12}$ and fourth choice with $\frac{1}{12}$ under P.

8.4.4 Efficiency of probabilistic serial

8.31 Given a preference profile $\succ$ and a random assignment P, we define a binary relation $\tau(P, \succ)$ on O as follows:

$$a\tau(P, \succ)b \Leftrightarrow \text{there exists } i \in N, \text{ such that } a \succ_i b \text{ and } P_{ib} > 0.$$

8.32 Proposition (Lemma 3 in Bogomolnaia and Moulin (2001)): The random assignment P is ordinally efficient at profile $\succ$ if and only if the relation $\tau(P, \succ)$ is acyclic.

"$\Rightarrow$". (1) Assume P is ordinally efficient.

(2) Suppose the relation $\tau(P, \succ)$, denoted τ for simplicity, has a cycle:

$$o_K \tau o_{K-1} \tau \cdots \tau o_2 \tau o_1 = o_K.$$

(3) Without loss of generality, we assume that the objects o_k, $k = 1, 2, \dots, K-1$ are all different.

(4) By definition of τ, we can construct a sequence $i_1, i_2, \dots, i_{K-1}$ in N such that:

$$P_{i_k o_k} > 0 \text{ and } o_{k+1} \succ_{i_k} o_k \text{ for all } k = 1, 2, \dots, K-1.$$

(5) Choose $\delta > 0$ such that

$$\delta \leq P_{i_k o_k} \text{ for all } k = 1, 2, \dots, K-1.$$

(6) Define a matrix $\Delta = (\delta_{io})$ as follows:

$$\begin{cases} \delta_{i_k o_k} = -\delta, & \text{for } k = 1, 2, \dots, K-1, \\ \delta_{i_k o_{k+1}} = \delta, & \text{for } k = 1, 2, \dots, K-1, \\ 0, & \text{otherwise.} \end{cases}$$

(7) Define a matrix $Q = P + \Delta$.

(8) By construction, Q is a bistochastic matrix and hence a random assignment.

(9) Moreover, Q stochastically dominates P, because one goes from P_{i_k} to Q_{i_k} by shifting some probability from object o_k to the preferred object o_{k+1}.

□

"⇐". (1) Suppose P is stochastically dominated at $\succ$ by Q.

(2) Let i_1 be an agent such that $P_{i_1} \neq Q_{i_1}$.

(3) Since Q_{i_1} first-order stochastically dominates P_{i_1}, there exist two objects o_1 and o_2 such that

$$o_2 \succ_{i_1} o_1, P_{i_1 o_1} > Q_{i_1 o_1} \geq 0, \text{ and } P_{i_1 o_2} < Q_{i_1 o_2}.$$

(4) In particular, $o_2 \tau(P, \succ) o_1$.

(5) By feasibility of Q, there exists an agent i_2 such that $P_{i_2 o_2} > Q_{i_2 o_2} \geq 0$.

(6) Since P is stochastically dominated at $\succ$ by Q, there exists o_3, such that

$$o_3 \succ_{i_2} o_2 \text{ and } P_{i_2 o_3} < Q_{i_2 o_3}.$$

(7) Hence, $o_3 \tau o_2$, and so on, until by finiteness of N and O we find a cycle of the relation τ.

□

✌ 8.33 Theorem (Theorem 1 in Bogomolnaia and Moulin (2001)): Fix a preference profile $\succ$. For every profile of eating speed functions $\omega = (\omega_i)_{i \in N}$, the random assignment $P_\omega[\succ]$ is ordinally efficient.

Conversely, for every ordinally efficient random assignment P at $\succ$, there exists a profile $\omega = (\omega_i)_{i \in N}$ such that $P = P_\omega[\succ]$.

8.34 *Proof.* (1) Suppose that for some ω, $P_\omega[\succ]$ is not ordinally efficient.

(2) By Proposition 8.32, we can find a cycle in the relation τ:

$$o_0 \tau o_1 \tau \cdots \tau o_{k-1} \tau o_k \tau \cdots \tau o_K \tau o_0.$$

(3) For each k, let i_k be an agent such that

$$o_{k-1} \succ_{i_k} o_k \text{ and } P_{i_k o_k} > 0.$$

(4) Let t^k be the first step t in simultaneous eating algorithm when the agent i_k starts to acquire good o_k, *i.e.*, the least t for which $P^t_{i_r o_r} \neq 0$.

(5) For agent i_k, since $o_{k-1} \succ_{i_k} o_k$, at instant t^k, the object o_{k-1} has already been fully distributed, *i.e.*, $o_{k-1} \notin O^{t^k}$.

(6) Thus $t^{k-1} < t^k$ for all $k = 1, 2, \ldots, K+1$, which is a contradiction since $o_0 = o_{K+1}$.

□

8.35 *Proof.* (1) Fix an ordinally efficient assignment P.

(2) Let

$$\bar{O}^0 = O \text{ and } B^1 = \{o \in \bar{O}^0 \mid \nexists b \in \bar{O}^0 \text{ such that } b\tau o\},$$

that is, B^1 is the set of maximal elements of $\bar{O}^0$ under τ.

(3) Let

$$B^k = \{o \in \bar{O}^{k-1} \mid \nexists b \in \bar{O}^{k-1} \text{ such that } b\tau o\} \text{ and } \bar{O}^k = \bar{O}^{k-1} \setminus B^k.$$

It is clear that this sequence will stop in finite steps. Assume this sequence stops at a Step K, for which $B^K = \bar{O}^{K-1}$.

Note that $\{B^1, B^2, \ldots, B^K\}$ forms a partition of O.

(4) For all $k = 1, 2, \ldots, K$, when $\frac{k-1}{K} \le t \le \frac{k}{K}$,

$$\omega_i(t) \triangleq \begin{cases} K \cdot P_{io}, & \text{if } o \in B^k \text{ and } i \in N(o, \bar{O}^{k-1}), \\ 0, & \text{otherwise.} \end{cases}$$

We will check that P is the result of the simultaneous eating algorithm with eating speeds ω and that $\bar{O}^0, \ldots, \bar{O}^K$ coincide with $O^0, \ldots, O^K$ from this algorithm.

(5) For any $o \in B^k$, for any $b \in \bar{O}^{k-1}$, for any $i \in N$, either $o \succsim_i b$ or $P_{io} = 0$.

Thus for the agent i with $P_{io} > 0$, we have $o \succsim_i b$ for any $b \in \bar{O}^{k-1}$, and hence $i \in N(o, \bar{O}^{k-1})$.

If $N(o, \bar{O}^{k-1}) = \emptyset$, then for every agent i, $P_{io} = 0$, and hence $\sum_i P_{io} = 0$, which is a contradiction.

(6) Therefore, from $\frac{k-1}{K}$ to $\frac{k}{K}$, only objects in the set B^k will be eat in the simultaneous eating algorithm.

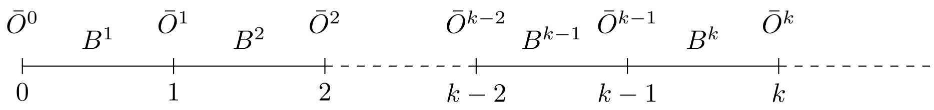


Figure 8.2

(7) From 0 to $\frac{1}{K}$, for each object $o \in B^1$,

- every agent i with $P_{io} > 0$ will eat object o with the speed $S \cdot P_{io}$, and
- every agent i with $P_{io} = 0$ will not eat object o.

At the instant $\frac{1}{K}$, every object o in B^1 will be eaten away since $\sum_i S \cdot P_{io} \cdot \frac{1}{S} = 1$.

(8) Hence $t^1 = \frac{1}{K}$, $O^1 = \bar{O}^1$, and P^1 is as follows:

$$P^1_{io} = \begin{cases} P_{io}, & \text{if } o \in B^1, \\ 0, & \text{if } o \notin B^1. \end{cases}$$

(9) We proceed by induction. Suppose that

$$t^{k-1} = \frac{k-1}{K}, \; O^{k-1} = \bar{O}^{k-1}, \text{ and } P^{k-1}_{io} = \begin{cases} P_{io}, & \text{if } o \in B^1 \cup B^2 \cup \cdots \cup B^{k-1}, \\ 0, & \text{otherwise.} \end{cases}$$

(10) For any $o \in \bar{O}^{k-1}$, we have $o \notin B^1 \cup B^2 \cup \cdots \cup B^{k-1}$, and hence $P^{k-1}_{io} = 0$.

(11) Therefore, we have

$$\sum_{i\in N(o,\bar{O}^{k-1})} \int_{\frac{k-1}{K}}^{t} \omega_i(s)\,\mathrm{d}s + \underbrace{\sum_{i\in N} P_{io}^{k-1}}_{=0}$$

$$= \begin{cases} \displaystyle\sum_{i\in N(o,\bar{O}^{k-1})} \int_{\frac{k-1}{K}}^{t} S\cdot P_{io}\,\mathrm{d}s = [Kt-(k-1)]\cdot \underbrace{\sum_{i\in N(o,\bar{O}^{k-1})} P_{io}}_{=1} = Kt-(k-1), & \text{if } o\in B^s, \\ 0, & \text{if } o\notin B^s. \end{cases}$$

(12) So,

$$t^k(o) = \begin{cases} \frac{k}{K}, & \text{if } o\in B^k, \\ +\infty, & \text{otherwise.} \end{cases}$$

(13) Thus, $t^k = \frac{k}{K}$, $O^k = \bar{O}^k$ and

$$P_{io}^k = \begin{cases} P_{io}, & \text{if } o\in B^1\cup B^2\cup\cdots\cup B^k, \\ 0, & \text{otherwise.} \end{cases}$$

□

8.5 Fairness

8.5.1 Anonymity

8.36 A mechanism P is anonymous if the mapping $\succ\mapsto P[\succ]$ is symmetric from the n preferences $\succ_i$ to the n assignments P_i.

8.37 Remark: In view of Theorem 8.33, the PS assignment is the simplest fair (anonymous) selection from the set of ordinally efficient assignments at a given preference profile.

The following result shows that whenever we use a simultaneous eating algorithm to construct an anonymous assignment rule, we must end up with the PS mechanism.

8.38 Proposition (Lemma 4 in Bogomolnaia and Moulin (2001)): Fix at profile of eating speeds ω. Let P be the mechanism derived from ω at all profiles. P is anonymous if and only if it coincides with PS.

8.39 *Proof.* We only prove "only if" direction.

(1) Fix ω and P as in the statement of the proposition and assume P is anonymous. We fix a preference profile $\succ$.

(2) The partial assignment obtained under PS at any moment $t\in[0,1]$ is anonymous, so under $\succ=(\succ_i)$ or its permutations, objects $o_1, o_2, \ldots, o_k, \ldots, o_n$ are eaten away in the same order and at the same instants $0 = x_0 < x_1 \leq x_2 \leq \cdots \leq x_k \leq \cdots \leq x_n = 1$.

(3) Under PS, an agent can change the good she eats only at one of the instants x_k, and the set of agents who eat a given good can only expand with time.

(4) Let $N(o_k)$ be the set of agents who eat good o_k in $[x_{k-1}, x_k]$. If $|N(o_k)| = 1$, then o_k is entirely assigned to one agent and $x_k = 1 = x_n$. Thus, $|N(o_k)| \geq 2$ whenever $x_k < x_n$.

(5) Claim: Suppose there exists instants $0 = y_0 < y_1 \leq y_2 \leq \cdots \leq y_k \leq \cdots \leq y_n = 1$ such that at y_k all agents get under P exactly the x_k fraction of their unit share of goods, *i.e.*, $\int_0^{y_k} \omega_i(t)\,\mathrm{d}t = x_k$ for all i and k. Then P coincides with PS.

(i) Suppose that assignments are the same at $x_1, \ldots, x_{k-1}$ under PS and at $y_1, \ldots, y_{k-1}$ under P.

(ii) Under PS during $[x_{k-1}, x_k]$ each agent eats her best among the objects still available $o_k, \ldots, o_n$, and the fraction $x_k - x_{k-1}$ eaten by everyone will not exhaust any object before x_k.

(iii) Since $x_k - x_{k-1}$ is exactly the fraction each agent eats during the interval $[y_{k-1}, y_k]$ under P, the set of objects which are eat during $[x_{k-1}, x_k]$ under PS is the same as that during $[y_{k-1}, y_k]$ under P, and hence they will end up at y_k with the same partial assignment as at x_k under PS.

(6) In the following, we will check that such $y_1, \ldots, y_n$ exist.

(7) Define

$$\bar{t}_i(k) = \max\left\{ t \,\middle|\, \int_0^t \omega_i(s)\,\mathrm{d}s \geq x_k \right\}, \qquad \underline{t}_i(k) = \min\left\{ t \,\middle|\, \int_0^t \omega_i(s)\,\mathrm{d}s \geq x_k \right\}$$
$$\bar{t}(k) = \min_i \bar{t}_i(k), \qquad \underline{t}(k) = \max \underline{t}_i(k),$$

that is, $[\underline{t}_i(k), \bar{t}_i(k)]$ is the largest interval during which the total fraction of objects eaten by an agent i stays equal to x_k.

(8) Proceed by induction on k. Suppose that under P all agents are able to eat exactly the fractions $x_1, \ldots, x_{k-1}$ by the instants $y_1, \ldots, y_{k-1}$ respectively. If $\underline{t}(k) \leq \bar{t}(k)$ then choose any $y_k \in [\underline{t}(k), \bar{t}(k)]$.

(9) In the following, we will show that $\underline{t}(k) > \bar{t}(k)$ is impossible by contradiction.

(10) Since $|N(o_k)| \geq 2$ whenever $x_k < x_n$, we focus on the case such that $|N(o_k)| \geq 2$.

(11) Consider the permutations $\succ^1$ and $\succ^2$ of $\succ$, such that agents 1 and 2 are in $N(o_k)$,

$$\bar{t}(k) = \bar{t}_1(k) \text{ and } \underline{t}(k) = \underline{t}_2(k) \text{ under } \succ^1,$$

and $\succ^2$ is obtained from $\succ^1$ by exchanging agents 1 and 2.

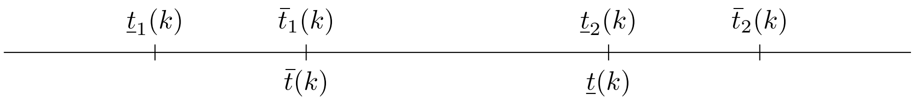


Figure 8.3

(12) We have

$$\begin{aligned}
&\sum_{i \in N(o_k)} \int_{y_{k-1}}^{\bar{t}(k)} \omega_i(s)\,\mathrm{d}s && \text{at } \bar{t}(k) < \underline{t}(k) = \underline{t}_2(k) \text{, agent 2 is still eating } o_k \\
&< |N(o_k)| \cdot (x_k - x_{k-1}) && \text{amount of } o_k \text{ left after } y_{k-1} \\
&< \sum_{i \in N(o_k)} \int_{y_{k-1}}^{\underline{t}(k)} \omega_i(s)\,\mathrm{d}s && \text{at } \underline{t}(k) > \underline{t}_1(k) \text{, agent 1 starts to eat another object}
\end{aligned}$$

(13) For any object o_j with $j > k$, we have

$$\sum_{i\in N(o_k)} \int_{y_{k-1}}^{\bar{t}(k)} \omega_i(s)\,\mathrm{d}s \leq |N(o_j)| \cdot (x_k - x_{k-1}) \leq \text{amount of } o_j \text{ left after } y_{k-1}.$$

Moreover, the equality is possible only if $x_j = x_k$.

(14) Thus under $\succ^1$ and $\succ^2$, no object among $o_k, \ldots, o_n$ is eaten away before $\bar{t}(k)$, and o_k will be exhausted at some instants s^1 and s^2 respectively, where $s^1, s^2 \in \big(\bar{t}(k), \underline{t}(k)\big)$.

(15) For any $s \in \big(\bar{t}(k), \underline{t}(k)\big)$,

- under $\succ^1$, the fraction of objects agent 1 gets by time s is larger than x_k, while the fraction of objects agent 2 gets by the time s is smaller than x_k, and
- under $\succ^2$, the fraction of objects agent 2 gets by time s is larger than x_k, while the fraction of objects agent 1 gets by the time s is smaller than x_k.

(16) By induction hypothesis, all agents get exactly the same partial assignment at x_{k-1} under PS and at y_{k-1} under P.

(17) As a result,

- agent 1 will get more and agent 2 less than x_k of objects under $\succ^1$, and
- agent 2 will get more and agent 1 less than x_k of objects under $\succ^2$.

This contradicts the anonymity of P.

□

8.5.2 Envy-freeness

8.40 A random assignment P is envy-free at a profile $\succ$ if $P_i \,\mathrm{sd}(\succ_i) P_j$ for all $i, j \in N$.

A random assignment P is weakly envy-free at a profile $\succ$ if for all $i, j \in N$,

$$P_j \,\mathrm{sd}(\succ_i) P_i \Rightarrow P_i = P_j.$$

8.41 Proposition (Proposition 1 in Bogomolnaia and Moulin (2001)): For any preference profile $\succ$,

(i) the assignment PS[$\succ$] is envy-free;

(ii) the assignment RP[$\succ$] is weakly envy-free;

(iii) the assignment RP[$\succ$] is envy-free for $n = 2$;

(iv) the assignment RP[$\succ$] may not be envy-free for $n \geq 3$.

8.42 *Proof.* (1) Fix a preference profile $\succ$ and an agent i, and label the objects in such a way that

$$o_1 \succ_i o_2 \succ_i \cdots \succ_i o_n.$$

Let $P = \mathrm{PS}[\succ]$.

(2) It suffices to show

$$\sum_{k=1}^{t} P_{io_k} \geq \sum_{k=1}^{t} P_{jo_k} \text{ for all } j \in N \text{ and } t = 1, 2, \ldots, n.$$

(3) Keep in mind that $\omega_i(t) = 1$ for all $i \in N$ and $t \in [0, 1]$.

(4) Let k_1 be the step at which o_1 is fully allocated, namely

$$a \in O^{k_1 - 1} \setminus O^{k_1}.$$

(i) Because o_1 is top-ranked in i's preference list, we have $i \in N(o_1, O^k)$ for all $k \leq k_1 - 1$.

(ii) Since from $t = 0$ to $t = t^{k_1}$, agent i eats object o_1, we have

$$P_{io_1}^{k_1} = t^{k_1} \geq P_{jo_1}^{k_1} \text{ for all } j \in N.$$

(iii) Since o_1 is fully allocated at instant k_1, we have

$$P_{io_1} = P_{io_1}^{k_1} \geq P_{jo_1}^{k_1} = P_{jo_1} \text{ for all } j \in N.$$

(5) Let k_2 be the step at which $\{a, b\}$ is fully allocated, that is,

$$\{a, b\} \cap O^{k_2 - 1} \neq \emptyset, \text{ and } \{a, b\} \cap O^{k_2} = \emptyset.$$

(i) Note that $k_1 \leq k_2$, and that $i \in N(o_1, O^k) \cup N(o_2, O^k)$ for all $k \leq k_2 - 1$.

(ii) Hence we have

$$P_{io_1} + P_{io_2} = P_{io_1}^{k_2} + P_{io_2}^{k_2} = t^{k_2} \geq P_{jo_1}^{k_2} + P_{jo_2}^{k_2} = P_{jo_1} + P_{jo_2} \text{ for all } j \in N.$$

(6) Repeating this argument we find that P_i first-order stochastically dominates P_j at $\succ_i$, as desired.

□

8.43 *Proof.* (1) Let $\succ$ be a preference profile at which $P_2 \,\mathrm{sd}(\succ_1) P_1$, we will show that $P_2 = P_1$, where $P = \mathrm{RP}$.

(2) Label the objects as follows

$$o_1 \succ_1 o_2 \succ_1 \cdots \succ_1 o_n.$$

(3) For any ordering f where 1 precedes 2, let $\bar{f}$ be the ordering obtained from f by permuting 1 and 2. Clearly $\big\{\{f, \bar{f}\} \mid f \in \mathcal{F}\big\}$ forms a partition of $\mathcal{F}$.

(4) Since $\succ$ is fixed, we omit it in $\varphi[\succ]$.

(5) Consider o_1:

(i) If 2 gets o_1 in $\varphi^{\bar{f}}$, so does 1 in φ^f.
In φ^f, 2 can not get o_1 since 1 would get o_1 before 2 anyway.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & \quad & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_1 & & \cancel{o_1} \\ & \nwarrow & \\ o_1? & & o_1 \end{pmatrix} \end{array}$$

(ii) Therefore in the random assignment $Q = (Q_{io}) \triangleq \frac{\varphi^f + \varphi^{\bar{f}}}{2}$, we have $Q_{2o_1} \leq Q_{1o_1}$.

(iii) Since $P = RP$ is a convex combination of such assignments, $P_{2o_1} \leq P_{1o_1}$.

(iv) From assumption $P_2 \,\mathrm{sd}(\succ_1) P_1$, we have $Q_{2o_1} = Q_{1o_1}$ for all pairs f and $\bar{f}$, and hence for such pair

- either 1 gets o_1 in φ^f and 2 gets o_1 in $\varphi^{\bar{f}}$,

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_1 & \cancel{o_1} \\ \cancel{o_1} & o_1 \end{pmatrix} \end{array} \tag{8.1}$$

- or none of 1, 2 gets o_1 in any of φ^f or $\varphi^{\bar{f}}$.

$$\begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} \cancel{o_1} & \cancel{o_1} \\ \cancel{o_1} & \cancel{o_1} \end{pmatrix} \end{array} \tag{8.2}$$

(6) Consider o_2:

(i) If 2 gets o_2 in $\varphi^{\bar{f}}$, then by Equation (8.2), 1 can not get o_1 in φ^f, and hence 1 gets o_2 in φ^f.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_2 & & \\ & \nwarrow & \\ & & o_2 \end{pmatrix} \end{array}$$

(ii) If 2 gets o_2 in φ^f, then 1 gets o_1 in φ^f since in φ^f 1 precedes 2.
By Equation (8.1), 2 gets o_1 in $\varphi^{\bar{f}}$.
In φ^f, 2 gets o_2, so in $\varphi^{\bar{f}}$, when 2 has already got o_1, 1 should get o_2.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_1 & \leftarrow & o_2 \\ & \searrow & \\ o_2 & \leftarrow & o_1 \end{pmatrix} \end{array}$$

(iii) Therefore $Q_{2o_2} \leq Q_{1o_2}$ in Q, and hence $P_{2o_2} \leq P_{1o_2}$.

(iv) By the assumption $P_{2o_1} + P_{2o_2} \geq P_{1o_1} + P_{1o_2}$ and the fact $P_{2o_1} = P_{1o_1}$, we have $P_{2o_2} = P_{1o_2}$ and $Q_{2o_2} = Q_{1o_2}$ for all pairs f and $\bar{f}$.

(v) Therefore, for any pair f and $\bar{f}$, the allocations of $o_1, o_2, O \setminus \{o_1, o_2\}$ are "symmetric" between f and $\bar{f}$, that is, if φ^f has $1 \to x$ and $2 \to y$ where x and y are o_1, o_2 or $O \setminus \{o_1, o_2\}$, then $\varphi^{\bar{f}}$ has $1 \to y$ and $2 \to x$. Here x is $O \setminus \{o_1, o_2\}$ means that x is some element of $O \setminus \{o_1, o_2\}$.

(7) We proceed by induction. Let $P_{1o_i} = P_{2o_i}$ for all $i = 1, 2, \ldots, k-1$. Suppose also that for any $x, y \in \{o_1, o_2, \ldots, o_{k-1}, O \setminus \{o_1, o_2, \ldots, o_{k-1}\}\}$, whenever 1 receives x and 2 receives y in φ^f, 1 receives y and 2 receives x in $\varphi^{\bar{f}}$.

(8) If 2 gets o_k in $\varphi^{\bar{f}}$, then by induction hypothesis 1 gets an object from $O \setminus \{o_1, o_2, \ldots, o_{k-1}\}$ in φ^f. Since o_k is the best for her in this set and it is available, 1 gets o_k in φ^f.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_k & & \\ & \nwarrow & \\ & & o_k \end{pmatrix} \end{array}$$

(9) If 2 gets o_k in φ^f, then 1 gets o_ℓ with $\ell < k$ in φ^f. Then by induction hypothesis, 2 gets o_ℓ in $\varphi^{\bar{f}}$. Hence o_k is available for 1 in $\varphi^{\bar{f}}$. But by induction hypothesis, 1 has to get some object from $O \setminus \{o_1, o_2, \ldots, o_{k-1}\}$ in $\varphi^{\bar{f}}$, so she gets o_k.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_\ell & \leftarrow & o_k \\ & \searrow & \\ o_k & \leftarrow & o_\ell \end{pmatrix} \end{array}$$

(10) It follows that $Q_{2o_k} \leq Q_{1o_k}$, and hence $P_{2o_k} \leq P_{1o_k}$.

(11) Since $\sum_{i=1}^k P_{2o_i} \geq \sum_{i=1}^k P_{1o_i}$ by assumption and $P_{1o_i} = P_{2o_i}$ ($i = 1, 2, \ldots, k-1$) by the induction hypothesis, we deduce as above $P_{2o_k} = P_{1o_k}$.

□

8.44 *Proof.* Consider the case where $|N| = |O| = 2$.

(1) If agents' top choices are difference, then RP $=$ PS.

(2) If agents' top choices are same, then it is easy to show that RP $=$ PS.

(3) Therefore, RP is envy-free in this case.

□

8.45 *Proof.* (1) Consider the example with three agents $1, 2, 3$ and three objects a, b, c, and the preferences are as follows:

1	2	3
a	a	b
b	c	a
c	b	c

Table 8.3

(2) It is clear that

$$\text{RP} = \text{RP}[\succ] = \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{5}{6} & \frac{1}{6} \end{pmatrix}$$

(3) Consider the following consistent von Neumann-Morgenstern utility $u_1(a) = 10$, $u_1(b) = 9$ and $u_1(c) = 0$, then we have

$$u_1 \cdot \text{RP}_3 = \frac{5}{6}u_1(b) + \frac{1}{6}u_1(c) = 7.5 > 6.5 = \frac{1}{2}u_1(a) + \frac{1}{6}u_1(b) + \frac{1}{3}u_1(c) = u_1 \cdot \text{RP}_1 .$$

That is, in the RP assignment, agent 1 envy the allocation of agent 3.

(4) By Proposition 8.23, RP_1 sd$(\succ_1)$ RP_3 does not hold. Hence, RP is not envy-free.

□

8.5.3 Equal treatment of equals

8.46 Definition: A mechanism $P\colon \succ \mapsto P[\succ]$ has the property "equal treatment of equals" if

$$\succ_i = \succ_j \Rightarrow P_i[\succ] = P_j[\succ].$$

8.47 Proposition: PS and RP have the property "equal treatment of equals."

8.6 Incentive compatibility

8.48 Definition: A mechanism $P\colon \succ \mapsto P[\succ]$ is strategy-proof if for all $i \in N$, $\succ$ and $\succ_i^*$, we have

$$P_i[\succ] \operatorname{sd}(\succ_i) P_i[\succ_{-i}, \succ_i^*].$$

8.49 Definition: A mechanism $P\colon \succ \mapsto P[\succ]$ is weakly strategy-proof if for all $i \in N$, $\succ$ and $\succ_i^*$, we have

$$P_i[\succ_{-i}, \succ_i^*] \operatorname{sd}(\succ_i) P_i[\succ] \Rightarrow P_i[\succ_{-i}, \succ_i] = P_i[\succ].$$

8.50 Proposition (Proposition 1 in Bogomolnaia and Moulin (2001)):

(i) RP is strategy-proof;

(ii) PS is weakly strategy-proof;

(iii) PS is strategy-proof for $n = 2$;

(iv) PS is not strategy-proof for $n \geq 3$.

8.51 *Proof.* For any ordering f, the priority mechanism $\succ \mapsto \varphi^f[\succ]$ is obviously strategy-proof. This property is preserved by convex combinations.

(1)

□

8.52 *Proof.* (1) Let $N(o,t)$ be the (possibly empty) set of agents who eat object o at time t. Thus, if t is such that $t^{s-1} \leq t < t^s$ for some $k = 1, 2, \ldots, n$, then

$$N(o,t) = \begin{cases} O(o, O^{k-1}), & \text{if } o \in O^{k-1}, \\ \emptyset, & \text{if } o \notin O^{k-1}. \end{cases}$$

(2) Let $n(o,t) = |N(o,t)|$, and

$$t(o) = \sup\{t \mid n(o,t) \geq 1\},$$

that is, $t(o)$ is the time at which o is eaten away.

(3) Note that $n(o,t)$ is non-decreasing in t on $[0, t(o))$, because once agent i joins $N(o,t)$, she keeps eating object o until its exhaustion.

(4) Moreover,

$$\int_0^{t(o)} n(o,t) \, \mathrm{d}t = 1,$$

because one unit of object o is allocated during the entire algorithm.

(5) Fix $\succ$, and agent denoted as agent 1, and a misreport $\succ_1^*$ by this agent.

(6) Let $P = \mathrm{PS}[\succ]$ and $P^* = \mathrm{PS}[\succ_{-1}, \succ_1^*]$, and similarly $N(o,t)$, $N^*(o,t)$, and so on.

(7) Label the objects so that $a \succ_1 b \succ_1 c \succ \cdots$.

(8) Assume $P_1^* \operatorname{sd}(\succ_1) P_1$ and show $P_1^* = P_1$.

(9) If $P_{1a} = 1$, it is trivial that $P_1^* = P_1$. So we assume $P_{1a} < 1$ from now on.

(10) At profile $\succ$, agent 1 is eating o during the whole interval $[0, t(a))$, and hence $t(a) = P_{1a}$.

(11) At profile $\succ^*$, agent 1 eats o on a subset of $[0, t^*(a))$, and hence $t(a) = P_{1a} \leq P_{1a}^* \leq t^*(a)$.

(12) Claim: for all $t \in [0, t(a))$ and all agents $i \neq 1$, we have

$$i \in N(a,t) \Rightarrow i \in N^*(a,t).$$

(13) Thus we have $N(a,t) \setminus \{1\} \subseteq N^*(a,t) \setminus \{1\}$, and hence

$$\begin{aligned} &\int_0^{t(a)} |N(a,t) \setminus \{1\}| \,\mathrm{d}t + P_{1a} = \int_0^{t(a)} n(a,t) \,\mathrm{d}t = 1 \\ = &\int_0^{t^*(a)} n^*(a,t) \,\mathrm{d}t = \int_0^{t^*(a)} |N^*(a,t) \setminus \{1\}| \,\mathrm{d}t + P_{1a}^*. \end{aligned}$$

(14) Therefore, $t(a) = t^*(a)$ and $N(a,t) = N^*(a,t)$ for all $t \in [0, t(a))$

(15) Thus, $P_{1a} = P_{1a}^*$ and the PS algorithms under $\succ$ and $\succ^*$ coincide on the interval $[0, t(a))$.

(16) It should be clear that the above argument can be repeated: the assumption $P_1^* \operatorname{sd}(\succ_1) P_1$ gives $P_{1b}^* \geq P_{1b}$ and we show successively $t(b) \geq t^*(b)$, then $N(b,t) \setminus \{1\} \subseteq N^*b, t \setminus \{1\}$ on the interval $[0, t(b))$, implying $t(b) = t^*(b)$ and so on.

□

8.53 *Proof of Claim.*

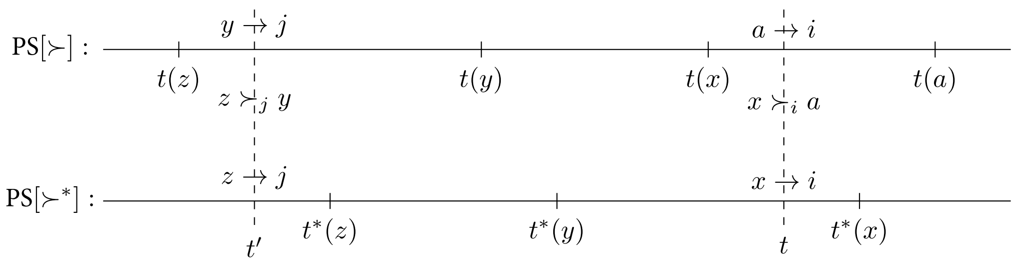


Figure 8.4

(1) Suppose there is an agent $i \neq 1$ and a time $t \in [0, t(a))$ such that

$$i \in N(a,t) \text{ and } i \in N^*(x,t) \text{ for some object } x \neq a.$$

(2) Under $\succ^*$, since $t < t(a) \leq t^*(a)$, the object a is available, and hence $x \succ_i^* a$.

(3) Since $\succ_i^* = \succ_i$, we have x has been eaten away at t under $\succ$, and hence $t(x) \leq t < t^*(x)$.

(4) Let B be the set of objects x such that $x \neq a$ and $t(x) < t^*(x)$. By the argument above, $B \neq \emptyset$.

(5) Take $y \in B$, such that $t(y)$ is minimal. Note that $t(y) \leq t(x) \leq t < t(a)$.

(6) Since $t(y) < t^*(y)$, we have at some time $t' < t(y)$, there is an agent j such that

$$j \in N(y,t) \text{ and } j \notin N^*(j,t).$$

Otherwise, $N(y,t) \subseteq N^*(y,t)$ for some $t \in [0, t(y))$. Combined with

$$\int_0^{t(y)} n(y,t)\,\mathrm{d}t = 1 = \int_0^{t^*(y)} n^*(y,t)\,\mathrm{d}t,$$

and the fact that $n^*(y,t)$ is non-decreasing in t, we have $t(y) = t^*(y)$, which contradicts the definition of B: $t(y) < t^*(y)$.

(7) Since $t' < t(y) < t(a)$ and agent 1 eats a over the whole interval $[0, t(a))$ under $\succ$, we have agent j can not be agent 1.

(8) Let z be the object that agent j eats at t' under $\succ^*$: $j \in N^*(z, t')$.

(9) Since $t' < t(y) < t^*(y)$, y is available at t' under $\succ^*$, and hence $z \succ_j y$.

(10) Since j eats y at t' under $\succ$ and $\succ_j^* = \succ_j$, z is no longer available at t' under $\succ$. Hence $t(z) \leq t' < t^*(z)$.

(11) Since $t(z) \leq t' < t(y) \leq t(x) < t(a)$, we have $z = a$, and hence $z \in B$, which contradicts the definition of y.

□

8.7 RP vs PS

8.54 Comparison of RP and PS:

	RP = RSD	PS
Ordinal efficiency	×	√
Ex post efficiency	√	√
Envy-freeness	×	√
Weak envy-freeness	√	√
Equal treatment of equals	√	√
Strategy-proofness	√	×
Weak strategy-proofness	√	√

Table 8.4

8.55 Proposition (Proposition 2 in Bogomolnaia and Moulin (2001)): Fix $n = 3$.

- The random priority mechanism is characterized by the combination of three axioms: ordinal efficiency, equal treatment of equals, and strategy-proofness
- The probabilistic serial mechanism is characterized by the combination of three axioms: ordinal efficiency, envy-freeness, and weak strategy-proofness.

8.8 Impossibility results

8.56 Theorem (Zhou (1990)): Incompatibility of ex ante efficiency, equal treatment of equals, and strategy-proofness.

8.57 Theorem (Theorem 2 in Bogomolnaia and Moulin (2001)): Fix $n \geq 4$. Then there is no mechanism meeting the following three requirements: ordinal efficiency, equal treatment of equals, and strategy-proofness.

8.9 Large markets

Kojima and Manea (2010) show that for any given utility functions of the agents, when there are sufficiently many copies of each object, PS will be strategy-proof.

Che and Kojima (2010) show that PS and RSD are asymptotically equivalent, as the size of the market increases.

Part III

School choice

Chapter 9

Introduction to school choice

Contents

9.1 The former model

☞ 9.1 A school choice problem is a five-tuple $\langle I, S, q, P, \succsim\rangle$, where

- $I = \{i_1, i_2, \ldots, i_n\}$ is a finite set of students,
- $S = \{s_1, s_2, \ldots, s_m\}$ is a finite set of schools,
- $q \triangleq (q_s)_{s\in S}$ is a quota profiles for schools where $q_s \in \mathbb{Z}_+$ is the quota of school s,
- $P \triangleq (P_i)_{i\in I}$ is a strong preference profile for students where P_i is a strict preference relation over $S \cup \{\emptyset\}$, denoting the strict preference relation of student i,
- $\succsim \triangleq (\succsim_s)_{s\in S}$ is a weak priority profile for schools where $\succsim_s$ is a weak priority relation over $I \cup \{\emptyset\}$, denoting the weak priority of school s.

Here $\emptyset$ represents remaining unmatched. For each $i \in I$, let R_i be the symmetric extension of P_i, that is, sR_is' if and only if sP_is' or $s = s'$.

9.2 In school choice problem, the priorities of schools are exogenous, that is, students are strategic agents but schools are simply objects to be consumed. So a school choice problem is a one-sided matching problem. It is one difference between the school choice problem and the college admission problem.

If each school has a strong priority relation $\succ_s$, then it is clear that a school choice problem naturally associates with an isomorphic college admission problem by letting each school s's preference relation be its priority relation $\succ_s$.

☞ 9.3 In a school choice problem $\langle I, S, q, P, \succsim \rangle$, a matching is a function $\mu\colon I \to S \cup \{\emptyset\}$ such that for each school s, $|\mu^{-1}(s)| \leq q_s$.

Let $\mathcal{M}$ denote the set of all matchings.

9.4 In a school choice problem $\langle I, S, q, P, \succsim \rangle$, let $\mathcal{P}$ denote the sets of all the possible preferences for students. We allow only students to report preferences, and schools' priorities are exogenously given and publicly known.[1]

Then a mechanism $\varphi^{\succsim}$ or simply φ selects a matching $\varphi[P]$ for every $P \in \mathcal{P}^n$. Formally, φ is a function

$$\varphi\colon \mathcal{P}^n \to \mathcal{M}.$$

9.5 Typical goals of school authorities are:

- efficient placement,
- fairness of outcomes,
- easy for participants to understand and use, *etc.*

☞ 9.6 A matching μ' (Pareto) dominates μ if for all $i \in I$, $\mu'(i) R_i \mu(i)$, and for some $i' \in I$, $\mu'(i) P_i \mu(i)$.

A matching is Pareto efficient if it is not dominated.

A mechanism φ is Pareto efficient if $\varphi[P]$ is Pareto efficient for all $P \in \mathcal{P}^n$.

A mechanism φ dominates ψ if

- for all P, $\varphi[P](i) R_i \psi[P](i)$ for all i
- for some P, $\varphi[P](i) P_i \psi[P](i)$ for some i

☞ 9.7 A matching μ is individually rational if no student prefers being unmatched to her assignment.

A mechanism φ is individually rational if $\varphi[P]$ is individually rational for all $P \in \mathcal{P}^n$.

☞ 9.8 A matching μ is non-wasteful if no student prefers a school with one or more empty seats to her assignment. That is, μ is non-wasteful if, whenever i prefers s to her assignment $\mu(i)$, $|\mu^{-1}(s)| = q_s$.

A mechanism φ is non-wasteful if $\varphi[P]$ is non-wasteful for all $P \in \mathcal{P}^n$.

☞ 9.9 We say that student i desires school s at μ if $s P_i \mu(i)$.

A matching μ eliminates justified envy if no student i prefers the assignment of another student j while at the same time having higher priority at school $\mu(j)$.

A mechanism φ eliminates justified envy if $\varphi[P]$ eliminates justified envy for all $P \in \mathcal{P}^n$.

9.10 Lemma (Lemma 2 in Balinski and Sönmez (1999)): Assume that each school has a strict priority relation. A matching is individually rational, non-wasteful, and eliminates justified envy if and only if it is stable for its associated college admissions problem.

9.11 Remark: In school choice, stability can be understood as a fairness criterion.

9.12 In a school choice problem $\langle I, S, q, P, \succsim \rangle$, a matching μ is constrained efficient if it is stable and is not Pareto dominated by any other stable matching.

9.13 A mechanism φ is strategy-proof if no student can benefit from misreporting, *i.e.*, truth-telling is a weakly dominant strategy for all students under the mechanism φ. Formally,

$$\varphi[P_i, P_{-i}](i) R_i \varphi[P_i', P_{-i}](i), \text{ for all } i, P_i', P.$$

9.14 A mechanism φ is group strategy-proof if for any P, there do not exist $J \subseteq I$, and $P_J' = (P_i')_{i \in J}$ such that

$$\varphi[P_J', P_{-J}](i) R_i \varphi[P](i) \text{ for all } i \in J, \text{ and } \varphi[P_J', P_{-J}](j) P_j \varphi[P](j) \text{ for some } j \in J.$$

9.15 A mechanism φ is nonbossy if for any P, $i \in I$ and P_i',

$$\varphi[P](i) = \varphi[P_i', P_{-i}](i) \text{ implies } \varphi[P] = \varphi[P_i', P_{-i}].$$

Nonbossiness ensures that students can not be bossy, that is, change the matching for others, by reporting different preferences, without changing their own.

9.16 Theorem (Lemma 1 in Pápai (2000)): A mechanism φ is group strategy-proof if and only if it is strategy-proof and nonbossy.

Proof. It is obvious that group strategy-proofness implies strategy-proofness and nonbossiness. So it suffices to show the other direction.

(1) Suppose the mechanism φ is strategy-proof and nonbossy.

(2) Let $J \subseteq I$, $P \in \mathcal{P}^n$, and P_J' be such that for all $i \in J$,

$$\varphi[P_J', P_{-J}](i) R_i \varphi[P](i).$$

We will show that $\varphi[P_J', P_{-J}] = \varphi[P]$.

(3) Without loss of generality, let $J = \{1, 2, \ldots, k\}$.

(4) For all $i \in J$, let P_i'' preserve the order P_i, except, let top-ranked school be $\varphi[P_J', P_{-J}](i)$.

(5) Strategy-proofness implies that $\varphi[P](1) R_1 \varphi[P_1'', P_{-1}](1)$.

- If $\varphi[P_J', P_{-J}](1) P_1 \varphi[P](1)$.
  (i) Then $\varphi[P_J', P_{-J}](1) \notin o_1(P_{-1})$, where the student i's option set at P_{-i} is defined by

  $$o_i(P_{-i}) \triangleq \{s \in S \colon \text{there exists } P_i''' \text{ such that } \varphi[P_i''', P_{-i}](i) = s\}.$$

  Otherwise, $\varphi[P_1''', P_{-1}](1) = \varphi[P_J', P_{-J}](1) P_1 \varphi[P](1)$ for some P_1''', which violates the strategy-proofness.

  (ii) Hence, given P_{-1}, student 1 can not get $\varphi[P_J', P_{-J}](1)$.

  (iii) That is, the top-ranked object of P_1'' can not be obtained.

  (iv) Therefore, by comparing P_1 and P_1'', we have $\varphi[P_1'', P_{-1}](1) = \varphi[P](1)$.
- If $\varphi[P_J', P_{-J}](1) = \varphi[P](1)$.
  (i) By definition of P_1'', $\varphi[P](1)$ is student 1's top-ranked school.

  (ii) Therefore $\varphi[P_1'', P_{-1}](1) = \varphi[P](1)$.

[1]In many school districts, schools are not allowed to submit their own preferences; Instead, school priorities are given by law.

(6) By nonbossiness, we have $\varphi[P_1'', P_{-1}] = \varphi[P]$.

(7) Repeating the same argument for students $2, 3, \ldots, k$, we get $\varphi[P_J'', P_{-J}] = \varphi[P]$:

$$\varphi[P_1'', P_2'', P_{-\{1,2\}}](2) = \varphi[P_1'', P_2, P_{-\{1,2\}}](2).$$

(8) Under the preference P_i'', $\varphi[P_J', P_{-J}](i)$ is student i's top-ranked school, so no school is ranked above it.

(9) Therefore, for all $i \in J$ and $s \in S$, $sR_i''\varphi[P_J', P_{-J}](i)$ implies $sR_i'\varphi[P_J', P_{-J}](i)$.

(10) By strategy-proofness, we have $\varphi[P_i'', P_{J\setminus\{i\}}', P_{-J}](i) R_i'' \varphi[P_i', P_{J\setminus\{i\}}', P_{-J}](i)$, and hence

$$\varphi[P_i'', P_{J\setminus\{i\}}', P_{-J}](i) R_i' \varphi[P_i', P_{J\setminus\{i\}}' P_{-J}](i).$$

(11) By strategy-proofness again, we have $\varphi[P_i'', P_{J\setminus\{i\}}', P_{-J}](i) = \varphi[P_J', P_{-J}](i)$.

(12) By nonbossiness, $\varphi[P_i'', P_{J\setminus\{i\}}', P_{-J}] = \varphi[P_J', P_{-J}]$.

(13) By the similar method above, we have $\varphi(P_J'', P_{-J}) = \varphi(P_J', P_{-J})$.

(14) Therefore we have

$$\varphi[P_J', P_{-J}] = \varphi[P],$$

which implies that φ is group strategy-proof.

□

9.17 Remark: Theorem 9.16 is a general result for one-sided matchings.

9.2 Boston school choice mechanism

9.18 The most commonly used school choice mechanism is that used by the Boston Public School until 2005.

9.19 The Boston mechanism.[2]

1 For each school a priority ordering is exogenously determined. (In case of Boston, priorities depend on home address, whether the student has a sibling already attending a school, and a lottery number to break ties.)

2 Each student submits a preference ranking of the schools.

3 The final phase is the student assignment based on preferences and priorities:

Step 1: In Step 1 only the top choices of the students are considered. For each school, consider the students who have listed it as their top choice and assign seats of the school to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her top choice.

Step k: Consider the remaining students. In Step k only the k-th choices of these students are considered. For each school still with available seats, consider the students who have listed it as their k-th choice and assign the remaining seats to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her k-th choice.

End: The algorithm terminates when no more students are assigned. At each step, every assignment is final.

9.20 Example: There are three students $\{i_1, i_2, i_3\}$ and three schools $\{s_1, s_2, s_3\}$ each with one seat, with the following preferences and priorities:

[2] This name came from the fact that it was in use for school choice in Boston Public Schools before it was replaced by the student-proposing DA.

i_1	i_2	i_3	s_1	s_2	s_3
s_2	s_1	s_1	i_1	i_2	i_2
s_1	s_2	s_2	i_3	i_1	i_1
s_3	s_3	s_3	i_2	i_3	i_3

Table 9.1

Step 1: Student i_1 is on the list of school s_2, and students i_2 and i_3 are on the list of school s_1 where i_3 has higher priority. So i_1 is assigned to s_2, i_3 is assigned to s_1, and i_2 remains unmatched.

Step 2: Student i_2 is on the list of school s_3, so she is assigned to s_3.

The matching is

$$\mu = \begin{bmatrix} i_1 & i_2 & i_3 \\ s_2 & s_3 & s_1 \end{bmatrix}.$$

9.21 The Boston mechanism is not necessarily stable.

Consider Example 9.20. The matching μ is blocked by the pair (i_2, s_2).

9.22 The Boston mechanism assigns as many students as possible to their first choices based on their submitted preferences; next, as many students as possible to their second choices; and so on.

The major drawback of this widely used mechanism is its lack of strategy-proofness.

Consider Example 9.20, if i_2 reports her preference as $s_2 P_2' s_1 P_2' s_3$ instead, the Boston mechanism produces the following matching The matching is

$$\mu' = \begin{bmatrix} i_1 & i_2 & i_3 \\ s_3 & s_2 & s_1 \end{bmatrix},$$

and the student i_2 benefits from submitting a false preference.

9.23 As seen in this example, a student who ranks a school as her second choice loses her priority to students who rank it as their first choice, so that it is risky for the student to use her first choice at a highly sought-after school if she has relatively low priority there. So the Boston mechanism gives students incentive to misrepresent their preferences by improving the ranking of schools in their choice lists for which they have high priority.

9.24 Worries in Boston mechanism is real.

St. Petersburg Times (14 September 2003):

> Make a realistic, informed selection on the school you list as your first choice. It's the cleanest shot you will get at a school, but if you aim too high you might miss. Here's why: If the random computer selection rejects your first choice, your chances of getting your second choice school are greatly diminished. That's because you then fall in line behind everyone who wanted your second choice school as their first choice. You can fall even farther back in line as you get bumped down to your third, fourth and fifth choices.

The 2004–2005 BPS School Guide:

> For a better choice of your 'first choice' school . . . consider choosing less popular schools.

9.3 Deferred acceptance algorithm and student-optimal stable mechanism

9.25 DA was implemented in Boston in 2006 and is in use.

9.26 In a school choice problem $\langle I, S, q, P, \succ\rangle$ with given strict priorities $\succ$, let DA$^{\succ}$ (or SOSM$^{\succ}$) or DA (resp. or SOSM) denote the student-optimal stable mechanism, which is produced by Gale and Shapley's student-proposing deferred acceptance algorithm.

9.27 Theorem: For any given $(P, \succ)$, DA produces a matching that is stable at $(P, \succ)$, which is also at least as good for every student as any other stable matching at $(P, \succ)$.

Proof. Recall Theorem 3.21 and Corollary 3.24. □

9.28 Theorem: Given fixed priorities $\succ$, DA is strategy-proof (for students).

Proof. Recall Theorem 3.59. □

✌ 9.29 Theorem (Theorem 3 in Alcalde and Barberà (1994)): DA is the unique stable and strategy-proof mechanism in school choice problems.

Proof. We will show that any stable mechanism φ which does not always choose the matching resulting from the student-proposing DA will be manipulable.

(1) Suppose $\varphi[P]$ was not the the student optimal matching at P. There will then be some student $i \in I$ who is not assigned to her optimal school DA$[P](i)$.

(2) It is clear that DA$[P](i) P_i \varphi[P](i)$.

(3) Consider a new preference P_i' of i: P_i' keeps the same ranking among schools and sets the schools behind DA$[P](i)$ unacceptable.

(4) Clearly, DA$[P]$ is stable under $[P_i', P_{-i}]$.

(5) By Theorem 3.29, we know that the set of students remaining unassigned is the same at all stable matchings for the given preference profile $[P_i', P_{-i}]$.

(6) Since $\varphi[P_i', P_{-i}]$ is another stable matching, $\varphi[P_i', P_{-i}](i) P_i \emptyset$, and hence $\varphi[P_i', P_{-i}](i) R_i\, \mathrm{DA}[P](i)$.

(7) Since DA$[P](i) P_i \varphi[P](i)$, we have

$$\varphi[P_i', P_{-i}](i) R_i\, \mathrm{DA}[P](i) P_i \varphi[P](i),$$

that is, i can manipulate φ at P via P_i'.

□

9.30 The major drawback of DA is its lack of efficiency.

Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, where $I = \{i, j, k\}$, $S = \{s_1, s_2\}$, $q_{s_1} = q_{s_2} = 1$, and

i	j	k	s_1	s_2
s_2	s_1	s_1	i	k
s_1		s_2	j	i
			k	

Table 9.2

The matching produced by DA is

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix},$$

Step	1	2	3	End
s_1	$j, \cancel{k}$	j	$\cancel{j}, i$	i
s_2	i	$\cancel{i}, k$	k	k
$\emptyset$	k	i	j	j

Table 9.3

and the procedure is

It is clear that μ is dominated by the matching

$$\mu' = \begin{bmatrix} i & j & k \\ s_2 & \emptyset & s_1 \end{bmatrix}.$$

The efficiency of DA will be detailedly discussed in Chapters 10.

9.31 Remark: DA is strategy-proof and stable, but not efficient. Are there mechanisms that improve the efficiency of students without sacrificing the other two properties?

- Stability will be lost for sure, since DA produces the student-optimal stable matching.
- Strategy-proofness will also be lost, due to the following impossibility result.

9.32 Theorem (Proposition 1 in Kesten (2010), Theorem 1 in Abdulkadiroğlu *et al.* (2009), Proposition 1 in Erdil (2014)): If φ is a strategy-proof and non-wasteful mechanism, then there is no strategy-proof mechanism that Pareto dominates φ.

Proof. (1) Suppose the matching ν weakly Pareto dominates $\mu \triangleq \varphi(P)$. We will show that the same set of students is matched under μ and ν.

(2) If i is matched under μ, *i.e.*, $\mu(i) P_i \emptyset$. Since $\nu(i) R_i \mu(i)$, i is also matched under ν.

(3) If i is unmatched under μ, since μ is non-wasteful, i does not prefer any not fully assigned school under μ to $\emptyset$.

(4) If i is matched under ν, then i must be matched to some school s that is fully assigned at μ.

(5) Consequently, some student matched to fully assigned schools under μ must be matched with some seat that was vacant under μ and is worse off. Contradiction to ν dominates μ.

(6) Suppose there exists ψ that dominates φ. Then there exists P such that $\psi[P](i) R_i \varphi[P](i)$ for all i and $\psi[P](j) P_j \varphi[P](j)$ for some j.

(7) Let $s = \psi[P]_j$. Consider P'_j: $s, \emptyset$.

(8) At (P'_j, P_{-j}), since φ is strategy-proof, φ can only assign j with $\emptyset$, because assigning her s will lead j to misreport P'_j when her true preference is P_j.

(9) Since $\varphi[P'_j, P_{-j}]$ is weakly dominated by $\psi[P'_j, P_{-j}]$, the same set of students is matched; hence $\psi[P'_j, P_{-j}](j) = \emptyset$.

(10) However, under the mechanism ψ, j will have incentive to report P_j when her true preference is P'_j, when others have preferences P_{-j}:

$$\psi[P_j, P_{-j}](j) = s P'_j \emptyset = \psi[P'_j, P_{-j}](j).$$

This violates the strategy-proofness of ψ.

$\square$

9.33 Corollary: Given strict school priorities, no Pareto efficient and strategy-proof mechanism Pareto dominates DA.

9.34 It has been empirically documented that the efficiency loss of DA can be significant in practice; see Abdulkadiroğlu *et al.* (2009). This creates a trade-off between efficiency and strategy-proofness.

9.35 The efficiency improvement of DA will be detailedly discussed in Chapters 11.

9.4 Top trading cycles mechanism

9.36 For school choice problems, TTC and DA can be viewed as two competing mechanisms. However the school system finally chose DA: the story says the idea of "trading priorities" in TTC did not appeal to policy makers.

9.37 Top trading cycles mechanism.

Assign a counter for each school which keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools.

Step 1: Each student points to her favorite school under her announced preferences. Each school points to the student who has the highest priority for the school.

Since the number of students and schools are finite, there is at least one cycle. (A cycle is an ordered list of distinct schools and distinct students $(s_1, i_1, s_2, i_2, \ldots, s_k, i_k)$ where s_1 points to i_1, i_l points to $s_2, \ldots, s_k$ points to i_k, i_k points to s_l.) Moreover, each school can be part of at most one cycle. Similarly, each student can be part of at most one cycle. Every student in a cycle is assigned a seat at the school she points to and is removed.

The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. Counters of all other schools stay put.

Step k: Each remaining student points to her favorite school among the remaining schools and each remaining school points to the student with highest priority among the remaining students.

There is at least one cycle. Every student in a cycle is assigned a seat at the school that she points to and is removed.

The counter of each school in a cycle is reduced by one and if it reduces to zero the school is also removed. Counters of all other schools stay put

End: The algorithm terminates when no more students are assigned. At each step, every assignment is final.

9.38 The intuition for this mechanism is that it starts with students who have the highest priorities, and allows them to trade the schools for which they have the highest priorities in case a Pareto improvement is possible.

9.39 Theorem (Proposition 3 in Abdulkadiroğlu and Sönmez (2003)): The top trading cycles mechanism is Pareto efficient.

Proof. Recall Theorem 4.12. □

9.40 Theorem (Proposition 4 in Abdulkadiroğlu and Sönmez (2003)): The top trading cycles mechanism is strategy-proof.

Proof. Recall Theorem 4.24. □

i	j	k	s_1	s_2	s_3
s_2	s_1	s_1	i	j	j
s_1	s_2	s_2	k	i	i
s_3	s_3	s_3	j	k	k

Table 9.4

9.41 Remark: The top trading cycles mechanism does not completely eliminate justified envy.

Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, where $I = \{i, j, k\}$, $S = \{s_1, s_2, s_3\}$, $q_{s_1} = q_{s_2} = q_{s_3} = 1$, and

The matching produced by TTC is

$$\mu = \begin{bmatrix} i & j & k \\ s_2 & s_1 & s_3 \end{bmatrix}.$$

It is clear that j violates k's priority at school s_1, since $k \succ_{s_1} j$ and $s_1 P_3 s_3 = \mu(3)$.

9.42 Remark: Although TTC is Pareto efficient and DA is not, the two are not Pareto ranked in general.

Consider Example 9.41, the outcome of DA and TTC are

$$\mu^{\text{DA}} = \begin{bmatrix} i & j & k \\ s_2 & s_3 & s_1 \end{bmatrix} \text{ and } \mu^{\text{TTC}} = \begin{bmatrix} i & j & k \\ s_2 & s_1 & s_3 \end{bmatrix},$$

where neither matching Pareto dominates the other one.

9.5 Equitable top trading cycles mechanism

Hakimov and Kesten (2014)

9.6 Nash equilibrium outcome

Theorem (Ergin and Sönmez (2006)): When priorities are strict, the set of Nash equilibrium outcomes of the preference revelation game induced by the Boston mechanism is equal to the set of stable matchings of the associated college admissions game under true preferences.

Corollary 1 When priorities are strict, the dominant-strategy equilibrium outcome of the Gale-Shapley student-optimal stable mechanism either Pareto-dominates or is equal to the Nash equilibrium outcomes of the Boston mechanism.

Theorem 31 (Haeringer and Klijn 2007) When priorities are strict and students can reveal only a limited number of schools in their preference lists, the Gale-Shapley student-optimal stable mechanism may have Nash equilibria in undominated strategies that induce justified envy.

Marilda Sotomayor (2008 IJGT)

Chapter 10

Acyclicity

Contents

10.1 Cycles and efficiency of deferred acceptance algorithm

☞ 10.1 Definition (Definition 1 in Ergin (2002)): Given a priority structure $\succ$ and quota profile q, a cycle is $a, b \in S$, $i, j, k \in I$ such that the following are satisfied:

(C) Cycle condition: $i \succ_a j \succ_a k \succ_b i$.

(S) Scarcity condition: There exist disjoint sets of students $I_a, I_b \subseteq I \backslash \{i, j, k\}$ such that $|I_a| = q_a - 1$, $|I_b| = q_b - 1$, $i' \succ_a j$ for every $i' \in I_a$, and $i'' \succ_b i$ for every $i'' \in I_b$.

A priority structure $\succ$ (or $(\succ, q)$) is acyclic if there exists no cycle.

10.2 Consider the school choice problem $\langle I, S, q, P, \succ\rangle$ in Example 9.30, where $I = \{i, j, k\}$, $S = \{a, b\}$, $q_a = q_b = 1$, and

i	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 10.1

The matching produced by DA is

$$\mu = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

A mutually beneficial agreement between i and k would be to get schools a and b respectively by exercising their priority rights, and then to make an exchange so that finally i gets b and k gets a.

However the final matching would violate the priority of j for a, contradicting the allocation on the basis of specified priorities.

Here the priority structure is not acyclic, since j may block a potential matching between i and k without affecting his own position, that is

$$i \succ_a j \succ_a k \succ_b i.$$

10.3 Remark: The scarcity condition requires that there are enough people with higher priority for a and b such that there may be instants when i, j, and k would compete for admission in either a or b.

10.4 For any problem $\Gamma = \langle I, S, q, P, \succ\rangle$, any $I' \subseteq I$, any $q' = (q'_s)_{s\in S}$ with $q'_s \leq q_s$ for each $s \in S$, there is a unique student-optimal stable mechanism outcome for $\langle I', S, q', P_{I'}, \succ|_{I'}\rangle$.

The reduced problem $r^{\mu}_{I'}(\Gamma) = \langle I', S, q', P_{I'}, \succ|_{I'}\rangle$ is the smaller problem consisting of students I' and remaining positions after students $I \setminus I'$ have left with their matchings under the matching μ, where $q'_s = q_s - |\mu^{-1}(s) \setminus I'|$.

☞ 10.5 Definition: The student-optimal stable mechanism is consistent if for any problem $\Gamma = \langle I, S, q, P, \succ\rangle$ and any nonempty subset $I' \subseteq I$,

$$\mathrm{DA}[\Gamma](i) = \mathrm{DA}\left[r_{I'}^{\mathrm{DA}[\Gamma]}(\Gamma)\right](i) \text{ for each } i \in I'.$$

10.6 Remark: Consistency requires that once a matching is determined and a group of students receive their colleges before the others, the rule should not change the matching of the remaining students in the reduced problem involving the remaining students and colleges.

✌ 10.7 Theorem (Theorem 1 in Ergin (2002)): For any $\succ$ and q, the following are equivalent:

(i) $\succ$ is acyclic.

(ii) $\mathrm{DA}^{\succ}$ is Pareto efficient.

(iii) $\mathrm{DA}^{\succ}$ is consistent.

(iv) $\mathrm{DA}^{\succ}$ is group strategy-proof.

10.8 Definition: Given a priority structure $\succ$, a generalized cycle is constituted of distinct $s_0, s_1, \ldots, s_{n-1} \in S$ and $i', i_0, i_1, \ldots, i_{n-1} \in I$ with $n \geq 2$ such that the following are satisfied:

(C') $i_0 \succ_{s_0} i' \succ_{s_0} i_{n-1} \succ_{s_{n-1}} i_{n-2} \succ_{s_{n-2}} \cdots \succ_{s_3} i_2 \succ_{s_2} i_1 \succ_{s_1} i_0$.

(S') There exist disjoint sets of agents $I_{s_0}, I_{s_1}, \ldots, I_{s_{n-1}} \subseteq I \setminus \{i', i_0, i_1, \ldots, i_{n-1}\}$ such that

$$I_{s_0} \subseteq U_{s_0}(i'),\ I_{s_1} \subseteq U_{s_1}(i_0),\ I_{s_2} \subseteq U_{s_2}(i_1), \ldots, I_{s_{n-2}} \subseteq U_{s_{n-2}}(i_{n-3}),\ I_{s_{n-1}} \subseteq U_{s_{n-1}}(i_{n-2}),$$

and $|I_{s_l}| = q_{s_l} - 1$ for all $l = 0, 1, \ldots, n-1$, where $U_s(i) \triangleq \{j \in I \mid j \succ_s i\}$.

10.9 Lemma: If DA is not Pareto efficient, then $\succ$ has a generalized cycle.

Proof.

Part 1: Suppose that DA is not Pareto efficient, that is, there exist P and μ', such that μ' Pareto dominates $\mu = \mathrm{DA}[P]$. We will show that there exist students

$$i_0, i_1, \ldots, i_{n-1}, i_n = i_0 \in I$$

with $n \geq 2$, such that each student envies the next under μ.

(1) Let $J = \{i \in I \mid \mu'(i) P_i \mu(i)\}$, since μ' Pareto dominates μ, $J \neq \emptyset$.

(2) Moreover, for any student $i \in I \setminus J$, he/she should be indifferent between $\mu(i)$ and $\mu'(i)$, and hence $I \setminus J = \{i \in I \mid \mu'(i) = \mu(i)\}$.

(3) For each $i \in J$, we also have $\mu'(i) \in S$, since $\mu'(i) P_i \mu(i) R_i \emptyset$.

(4) For each $i \in J$, since $\mu'(i) P_i \mu(i)$, i has been rejected by $\mu'(i)$ at a step under μ. So at that step $\mu'(i)$'s waiting list must be full, and therefore at last the school $\mu'(i)$ has full quota, *i.e.*, $|\mu^{-1}(\mu'(i))| = q_{\mu'(i)}$.

(5) Fix $i \in J$. Claim: There is some student in J who was assigned to $\mu'(i)$ under μ.

   (i) Otherwise the set of $q_{\mu'(i)}$ students who were assigned to $\mu'(i)$ under μ would be a subset of $I \setminus J$, and hence they would be assigned to $\mu(i)$ also under μ', since $I \setminus J = \{i \in I \mid \mu'(i) = \mu(i)\}$.

   (ii) Since $i \in J$ is also assigned to $\mu'(i)$ under μ', there are at least $q_{\mu'(i)} + 1$ students assigned to $\mu'(i)$ under μ', which leads to a contradiction.

(6) Define the correspondence $\pi\colon J \twoheadrightarrow J$ by $\pi(i) = \mu^{-1}(\mu'(i)) \cap J$. By the above argument, π is non-empty valued.

(7) We can choose a selection $\bar{\pi}$ of π such that for any $i, j \in J$ with $\mu'(i) = \mu'(j)$, we have $\bar{\pi}(i) = \bar{\pi}(j) \in J$. Hence we have $\mu\bar{\pi} = \mu'$.

(8) For each $i \in J$, since $\mu(i) \neq \mu'(i)$, we have $\bar{\pi}(i) \neq i$. Therefore there is $n \geq 2$ and n distinct students

$$i_1, i_2, \ldots, i_n = i_0 \in J$$

with $i_r = \bar{\pi}(i_{r-1})$ for $r = 1, 2, \ldots, n$.

(9) Set $s_r = \mu(i_r)$ for $r = 1, 2, \ldots, n$. Then $s_r = \mu(i_r) = \mu(\bar{\pi}(i_{r-1})) = \mu'(i_{r-1})$ for $r = 1, 2, \ldots, n$.

(10) Since $i_1, i_2, \ldots, i_n = i_0$ are distinct, $s_1, s_2, \ldots, s_n = s_0$ are also distinct by the particular choice of the selection $\bar{\pi}$.

(11) Now we have showed that $s_r = \mu(i_r) = \mu'(i_{r-1}) P_{i_{r-1}} \mu(i_{r-1})$ for $r = 1, 2, \ldots, n$.

(12) Since μ is stable, we have $i_r \succ_{s_r} i_{r-1}$ for $r = 1, 2, \ldots, n$. Therefore we have

$$i_0 \succ_{s_0} i_{n-1} \succ_{s_{n-1}} i_{n-2} \succ_{s_{n-2}} \cdots \succ_{s_3} i_2 \succ_{s_2} i_1 \succ_{s_1} i_0.$$

Part 2:

(1) Let k be the latest step under μ when someone in $\{i_0, i_1, \ldots, i_{n-1}\}$ applies to (and is accepted) the school to which he is assigned under μ.

(2) Without loss of generality, suppose that i_0 applies to $s_0 = \mu(i_0)$ at this step.

(3) After that step, all students in $\{i_0, i_1, \ldots, i_{n-1}\}$ never get rejected again, since they are in the waiting list of their final allocation.

(4) For $r = 0, 1, \ldots, n-1$, since $s_r P_{i_{r-1}} s_{r-1}$, i_{r-1} was rejected by s_r at an earlier step than when he applied to s_{r-1}, which is earlier than Step k.

(5) Therefore at the end of Step $k-1$, s_r's waiting list must be full, for $r = 0, 1, \ldots, n-1$.

(6) Note that at the end of Step $k-1$, s_0's waiting list does not include any $i_r \in \{i_1, i_2, \ldots, i_{n-1}\}$. Otherwise i_r would apply to s_r at a step later k, a contradiction.

(7) We can find $i' \in I$ distinct from $i_0, i_1, \ldots, i_{n-1}$ such that he is rejected by s_0 at Step k.

(8) Since i' is accepted to the waiting list of s_0 when i_{n-1} is rejected by i_0, we have $i_0 \succ_{s_0} i' \succ_{s_0} i_{n-1}$.

(9) For any $r \in \{0, 1, \ldots, n-1\}$, let I_{s_r} be the set of students in the waiting list of s_r other than i_r at the end of Step k. It is now straightforward to see that condition (S') is also satisfied.

□

10.10 Lemma (Lemma in Narita (2009)): If $\succ$ has a generalized cycle, then $\succ$ has a cycle.

Proof. Suppose that $\succ$ and q have a generalized circle and let the size of the shortest generalized cycle be $n > 2$, that is, $s_0, s_1, \ldots, s_{n-1} \in S$, $i', i_0, i_1, \ldots, i_{n-1} \in I$ and $I_{s_0}, I_{s_1}, \ldots, I_{s_{n-1}} \subseteq I \setminus \{i', i_0, i_1, \ldots, i_{n-1}\}$ constitute the shortest generalized cycle of size $n > 2$.

$$i_0 \succ_{s_0} i' \succ_{s_0} i_{n-1} \succ_{s_{n-1}} i_{n-2} \succ_{s_{n-2}} \cdots \succ_{s_3} i_2 \succ_{s_2} i_1 \succ_{s_1} i_0.$$

Case (1-1): Suppose $i_0 \succ_{s_2} i_2$ and for all $i \in I_{s_2}$, $i \succ_{s_2} i_2$.

(1) We have $i_0 \succ_{s_2} i_2 \succ_{s_2} i_1 \succ_{s_1} i_0$.

(2) $I_{s_1}, I_{s_2} \subseteq I \setminus \{i_0, i_1, i_2\}$ are disjoint sets satisfying

$$I_{s_2} \subseteq U_{s_2}(i_2),\ I_{s_1} \subseteq U_{s_1}(i_0),\ |I_{s_2}| = q_{s_2} - 1,\ |I_{s_1}| = q_{s_1} - 1.$$

(3) Therefore, $s_2, s_1 \in S$, $i_0, i_2, i_1 \in I$ and $I_{s_2}, I_{s_1} \subseteq I \setminus \{i_0, i_2, i_1\}$ constitute a cycle, *i.e.*, a generalized cycle of size 2, which is a contradiction.

Case (1-2): Suppose $i_0 \succ_{s_2} i_2$ and there exists $i \in I_{s_2}$ such that $i_2 \succ_{s_2} i$.

(1) Since $i \in I_{s_2} \subseteq U_{s_2}(i_1)$, we have $i \succ_{s_2} i_1$, and hence $i_2 \succ_{s_2} i \succ_{s_2} i_1$.

(2) Let i' be the minimum element in I_{s_2} with respect to $\succ_{s_2}$, and $I'_{s_2} = I_{s_2} \cup \{i_2\} \setminus \{i'\}$.

(3) Then, $i_0 \succ_{s_2} i_2 \succ_{s_2} i' \succ_{s_2} i_1 \succ_{s_1} i_0$.

(4) $I_{s_1}, I'_{s_2} \subseteq I \setminus \{i_0, i_1, i'\}$ are disjoint sets satisfying

$$I'_{s_2} \subseteq U_{s_2}(i'),\ I_{s_1} \subseteq U_{s_1}(i_0),\ |I_{s_1}| = q_{s_1} - 1,\ |I'_{s_2}| = q_{s_2} - 1.$$

(5) Therefore, $s_2, s_1 \in S$, $i_0, i', i_1 \in I$, and I'_{s_2}, I_{s_1} constitute a cycle, which is a contradiction.

Case (2-1): Suppose $i_2 \succ_{s_2} i_0$, and for all $i \in I_{s_2}$, $i \succ_{s_2} i_0$.

(1) Then we have

$$i_0 \succ_{s_0} i' \succ_{s_0} i_{n-1} \succ_{s_{n-1}} i_{n-2} \succ_{s_{n-2}} \cdots \succ_{s_3} i_2 \succ_{s_2} i_0.$$

(2) $I_{s_0}, I_{s_2}, I_{s_3}, \ldots, I_{s_{n-1}} \subseteq S \setminus \{i', i_0, i_2, i_3, \ldots, i_{n-1}\}$ are disjoint sets satisfying

$$I_{s_0} \subseteq U_{s_0}(i'),\ I_{s_2} \subseteq U_{s_2}(i_0),\ I_{s_3} \subseteq U_{s_3}(i_2), \ldots, I_{s_{n-2}} \subseteq U_{s_{n-2}}(i_{n-3}),\ I_{s_{n-1}} \subseteq U_{s_{n-1}}(i_{n-2}).$$

(3) We also have $|I_{s_r}| = q_{s_r} - 1$ for all $r = 0, 2, 3, \ldots, n-1$.

(4) Therefore, $s_0, s_2, s_3, \ldots, s_{n-1} \in S$, $i', i_0, i_2, i_3, \ldots, i_{n-1} \in I$ and $I_{s_0}, I_{s_2}, I_{s_3}, \ldots, I_{s_{n-1}}$ constitute a generalized cycle of size $n-1$, which is a contradiction.

Case (2-2): Suppose $i_2 \succ_{s_2} i_0$, and there exists $i \in I_{s_2}$ such that $i_0 \succ_{s_2} i$.

(1) Since $i \in I_{s_2} \subseteq U_{s_2}(i_1)$, we have $i \succ_{s_2} i_1$, and hence $i_0 \succ_{s_2} i \succ_{s_2} i_1$.

(2) Let i'' be the minimum element in I_{s_2} with respect to $\succ_{s_2}$, and $I''_{s_2} = I_{s_2} \cup \{i_2\} \setminus \{i''\}$.

(3) Then, $i_0 \succ_{s_2} i'' \succ_{s_2} i_1 \succ_{s_1} i_0$.

(4) $I_{s_1}, I''_{s_2} \subseteq I \setminus \{i_0, i_1, i''\}$ are disjoint sets satisfying

$$I''_{s_2} \subseteq U_{s_2}(i''),\ I_{s_1} \subseteq U_{s_1}(i_0),\ |I_{s_1}| = q_{s_1} - 1,\ |I''_{s_2}| = q_{s_2} - 1.$$

(5) Therefore, $s_2, s_1 \in S$, $i_0, i'', i_1 \in I$, and I''_{s_2}, I_{s_1} constitute a cycle, which is a contradiction.

□

10.11 *Proof of Theorem 10.7, Part 1: "acyclicity implies Pareto efficiency".* It follows immediately from two lemmas above.

□

10.12 *Proof of Theorem 10.7, Part 2: "Pareto efficiency implies consistency".*

(1) Assume DA is not consistent.

(2) Then, there is P, and $\emptyset \neq I' \subsetneqq I$, such that

$$\mu|_{I'} \neq \mu',$$

where $\mu = \mathrm{DA}[I, S, q, P, \succ]$, and $\mu' = \mathrm{DA}[r^{\mu}_{I'}(I, S, q, P, \succ)]$.

(3) Then by Corollary 3.24, μ' Pareto dominates $\mu|_{I'}$ in the reduced problem.

(4) Then the matching ν defined by

$$\nu(i) = \begin{cases} \mu'(i), & \text{if } i \in I', \\ \mu(i), & \text{otherwise.} \end{cases}$$

Pareto dominates μ, contradiction.

□

10.13 *Proof of Theorem 10.7, Part 3: "consistency implies group strategy-proofness".*

(1) By Corollary 3.24, DA is strategy-proof.

(2) By Theorem 9.16, it suffices to show that DA is nonbossy.

(3) Suppose that DA is consistent.

(4) Let i, P and P'_i be given and set

$$\mu = \mathrm{DA}[I, S, q, P, \succ], \text{and } \nu = \mathrm{DA}[I, S, q, P'_i, P_{-i}, \succ].$$

(5) Assume $\mu(i) = \nu(i)$, then two reduced problems $r^{\mu}_{I\setminus\{i\}}(I, S, q, P, \succ)$ and $r^{\nu}_{I\setminus\{i\}}(I, S, q, P'_i, P_{-i}, \succ)$ are same.

(6) By consistency of DA,

$$\begin{aligned} \mu|_{I\setminus\{i\}} &= \mathrm{DA}[I, S, q, P, \succ]|_{I\setminus\{i\}} = \mathrm{DA}[r^{\mu}_{I\setminus\{i\}}(I, S, q, P, \succ)], \\ \nu|_{I\setminus\{i\}} &= \mathrm{DA}[I, S, q, P'_i, P_{-i}, \succ]|_{I\setminus\{i\}} = \mathrm{DA}[r^{\nu}_{I\setminus\{i\}}(I, S, q, P'_i, P_{-i}, \succ)]. \end{aligned}$$

(7) Therefore, $\mu|_{I\setminus\{i\}} = \nu|_{I\setminus\{i\}}$.

(8) Since $\mu(i) = \nu(i)$ and $\mu|_{I\setminus\{i\}} = \nu|_{I\setminus\{i\}}$, we conclude that $\mu = \nu$.

□

10.14 *Proof of Theorem 10.7, Part 4: "group strategy-proofness implies acyclicity".*

(1) Suppose that $\succ$ has a cycle with a, b, i, j, k ($i \succ_a j \succ_a k \succ_b i$), I_a and I_b.

(2) Consider the preference profile P, where

- students in I_a and I_b respectively rank a and b as their top choice,
- the preferences of i, j and k are as follows,

i	j	k
b	a	a
a		b

Table 10.2

- students outside $I_a \cup I_b \cup \{i, j, k\}$ prefer not to be assigned to any school.

(3) Let $I' = \{i, j, k\}$, $P_{-j} = P_{-j}$, and P'_j rank $\emptyset$ at the top.

(4) Then we have

$$\mathrm{DA}[I, S, q, P_{I\setminus I'}, P_{I'}, \succ] = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}, \text{ and } \mathrm{DA}[I, S, q, P_{I\setminus I'}, P'_{I'}, \succ] = \begin{bmatrix} i & j & k \\ b & \emptyset & a \end{bmatrix},$$

which contradicts the group strategy-proofness of DA under the true preferences P.

□

10.15 *Proof of Theorem 10.7, Part 5: "consistency implies acyclicity".*

(1) Suppose that $\succ$ has a cycle with a, b, i, j, k ($i \succ_a j \succ_a k \succ_b i$), I_a and I_b.

(2) Consider the preference profile P, where

- students in I_a and I_b respectively rank a and b as their top choice,
- the preferences of i, j and k are as follows,

i	j	k
b	a	a
a		b

Table 10.3

- students outside $I_a \cup I_b \cup \{i, j, k\}$ prefer not to be assigned to any school.

(3) Then, the student-optimal stable mechanism outcome μ for $\langle I, S, q, P, \succ\rangle$ is

$$\mu = \begin{bmatrix} a & b \\ I_a \cup \{i\} & I_b \cup \{j\} \end{bmatrix}.$$

(4) Consider the reduced problem

$$r^{\mu}_{\{i,k\}}(I, S, q, P, \succ) = \langle \{i, k\}, S, q', P_{\{i,k\}}, \succ|_{\{i,k\}} \rangle,$$

is such that the preferences of i and k are as above, $q'_a = q'_b = 1$, and $q'_s = q_s$ for any $s \in S \setminus \{a, b\}$.

(5) The student-optimal stable mechanism outcome μ' of this reduced problem is

$$\mu' = \begin{bmatrix} a & b \\ k & i \end{bmatrix}.$$

(6) Since $\mu' \neq \mu|_{\{i,k\}}$, DA is not consistent.

□

10.16 Theorem (Theorem 2 in Ergin (2002)): $(\succ, q)$ is cyclical if and only if there exist student i and schools s_1, s_2 such that i's rank is larger than $q_{s_1} + q_{s_2}$ at s_1 or s_2, and $|r_{s_1}(i) - r_{s_2}(i)| > 1$, where $r_s(i)$ is the rank of student i at school s.

Proof. Omitted. □

10.2 Robust stability

☞ 10.17 Definition (Definition 1 in Kojima (2011)): A mechanism φ is robustly stable if the following conditions are satisfied:

(1) φ is stable.

(2) φ is strategy-proof.

(3) There exist no $i \in I$, $s \in S$, $P \in \mathcal{P}^{|I|}$, and $P'_i \in \mathcal{P}$, such that

- $sP_i\varphi[P](i)$, and
- $i \succ_s i'$ for some $i' \in \varphi[P'_i, P_{-i}](s)$ or $|\varphi[P'_i, P_{-i}](s)| < q_s$.

In words, a mechanism is robustly stable if it is stable, strategy-proof, and also immune to a combined manipulation, where a student first misrepresents his or her preferences and then blocks the matching that is produced by the centralized mechanism.

10.18 Theorem (Theorem 1 in Kojima (2011)): There exists a priority structure $\succ$ and a quote profile q for which there is no robustly stable mechanism.

Proof. (1) DA is stable and strategy-proof.

(2) Moreover, it is the unique stable and strategy-proof mechanism (see Theorem 9.29).

(3) It suffices to show that DA is not immune to the combined manipulation.

(4) Consider a problem with $I = \{1, 2, 3\}$, $S = \{a, b\}$, $q_a = q_b = 1$, and

1	2	3	a	b
b	a	a	1	3
a		b	2	1
			3	2

Table 10.4

(5) Under the true preferences (P_2, P_{-2}), the DA produces

$$\begin{bmatrix} 1 & 2 & 3 \\ a & \emptyset & b \end{bmatrix}.$$

(6) Now consider a false preference $P_2' : \emptyset$. Then, under (P_{-2}, P_2'), DA produces

$$\begin{bmatrix} 1 & 2 & 3 \\ b & \emptyset & a \end{bmatrix}.$$

(7) Since $aP_2\emptyset = \text{DA}[P_2, P_{-2}](2)$ and $2 \succ_a 3 \in \text{DA}[P_2', P_{-2}](a)$, DA is not robustly stable. More specifically, student 2 has incentives to first report P_2' and then block $\text{DA}[P_2', P_{-2}]$.

□

✌ 10.19 Theorem (Theorem 2 in Kojima (2011)): Given $\langle I, S, q, P, \succ\rangle$, DA is robustly stable if and only if the priority structure $(\succ, q)$ is acyclic.

10.20 *Proof of Theorem 10.19, Part 1: "robust stability implies acyclicity".* We show the claim by contraposition.

(1) Suppose that the priority structure is not acyclic. Then, by definition, there exist $a, b \in S$, $i, j, k \in I$ such that

- $i \succ_a j \succ_a k \succ_b i$
- there exist disjoint sets $I_a, I_b \subseteq I \setminus \{i, j, k\}$ such that $|I_a| = q_a - 1$, $|I_b| = q_b - 1$, $i' \succ_a j$ for all $i \in I_a$, and $i'' \succ_b i$ for all $i \in I_b$.

(2) Consider the following preference profile P of students:

i	j	k	$i' \in I_a$	$i'' \in I_b$	$i \in I \setminus [\{i, j, k\} \cup I_a \cup I_b]$
b	a	a	a	b	
a		b			

Table 10.5

It is easy to see that $\text{DA}[P](j) = \emptyset$.

(3) Now consider a false preference of student j, $P_j' : \emptyset$.

(4) We have $\text{DA}[P_j', P_{-j}](k) = a$. Since

$$aP_j\emptyset = \text{DA}[P](j) \text{ and } j \succ_a k \in \text{DA}[P_j', P_{-j}](a),$$

DA is not robustly stable.

□

10.21 *Proof of Theorem 10.19, Part 2: "acyclicity implies robust stability".* Prove by contradiction.

(1) Assume DA is not robustly stable. Since DA is stable and strategy-proof, we will have the following condition: Condition A: There exists $s \in S$, $c \in C$, $P \in \mathcal{P}^{|S|}$ and $P'_s \in \mathcal{P}$, such that

- $c P_s \, \mathrm{DA}[P](s)$;
- $s \succ_c s'$ for some $s' \in \mathrm{DA}[P'_s, P_{-s}](c)$ or $|\,\mathrm{DA}[P'_s, P_{-s}](c)| < q_c$.

(2) Let $P' = (P'_s, P_{-s})$.

(3) Case 1: Suppose $\mathrm{DA}[P'](s) = \emptyset$.

(i) Let
$$P''_s : c, \emptyset, \quad P'' = (P''_s, P_{-s}).$$

(ii) If $\mathrm{DA}[P''](s) = c$. Since we have
$$\mathrm{DA}[P''](s) = c P_s \, \mathrm{DA}[P](s),$$
this is a contradiction to strategy-proofness of DA.

(iii) If $\mathrm{DA}[P''](s) = \emptyset$ which equals to $\mathrm{DA}[P'](s)$).
Then, by definition of P''_s, we have
$$c P''_s \emptyset = \mathrm{DA}[P''](s). \tag{10.1}$$
Since $(\succ, q)$ is acyclic, DA is nonbossy, and hence $\mathrm{DA}[P''] = \mathrm{DA}[P']$.
By Condition A, we will have $s \succ_c s'$ for some $s' \in \mathrm{DA}[P'](c) = \mathrm{DA}[P''](c)$, or $|\,\mathrm{DA}[P''](c)| = |\,\mathrm{DA}[P'](c)| < q_c$.
This and relation (10.1) means that $\mathrm{DA}[P'']$ is unstable under P'', contradicting the fact that DA is a stable mechanism.

(4) Case 2: Suppose $\mathrm{DA}[P'](s) \neq \emptyset$. Let
$$P''_s : \emptyset, \quad P'' = (P''_s, P_{-s}).$$
By the comparative statics, $|\,\mathrm{DA}[P'](c)| > |\,\mathrm{DA}[P''](c)|$, and if $|\,\mathrm{DA}[P'](c)| = |\,\mathrm{DA}[P''](c)| = q_c$, then there exists $s'' \in \mathrm{DA}[P''](c)$, such that $s' \succsim_c s''$ for all $s' \in \mathrm{DA}[P'](c)$.
Therefore Condition A is satisfied with respect to s, c and P''_s and, since $\mathrm{DA}[P''](s) = \emptyset$, the analysis reduces to Case 1.

□

10.22 Remark: Given that DA is the unique stable and strategy-proof mechanism (see Theorem 9.29), this theorem implies that, given the market, there exists a robustly stable mechanism if and only if the priority structure is acyclic.

10.3 Group robust stability

Afacan (2012)

10.4 Strong cycles and stability of top trading cycles algorithm

Lemma 1, Proposition 2 in Kesten (2006)

Chapter 11

Efficiency improvement on student-optimal stable mechanism

Contents

11.1 When the priority structure contains cycles, DA is not Pareto efficient as shown in Theorem 10.7. In Remark 9.31 and Corollary 9.33, we also show that we can not improve the efficiency of students without sacrificing the stability and strategy-proofness. In this chapter, we will focus on how to improve the efficiency with minimal hurt on stability and strategy-proofness.

11.1 Efficiency-adjusted deferred acceptance algorithm

11.2 Example: Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, where $I = \{i, j, k\}$, $S = \{s_1, s_2\}$, $q_{s_1} = q_{s_2} = 1$, and

i	j	k	s_1	s_2
s_2	s_1	s_1	i	k
s_1		s_2	j	i
			k	

Table 11.1

The matching produced by DA is

$$\begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix},$$

and the procedure is

Step	1	2	3	End
s_1	$j, \cancel{k}$	j	$\cancel{j}, i$	i
s_2	i	$\cancel{i}, k$	k	k
$\emptyset$	k	i	j	j

Table 11.2

11.3 In Example 11.2, when the DA algorithm is applied to this problem, student j causes student k to be rejected from school s_1 and starts a chain of rejections that ends back at school s_1, forming a full cycle and causing student j himself to be rejected. There such a cycle has resulted in loss of efficiency

By applying to school s_1, student j "interrupts" a desirable settlement between students i and k without affecting her own placement and artificially introduces inefficiency into the outcome. The key idea behind the mechanism produced by Kesten (2010) is based on preventing students such as student j of this example from interrupting settlements among other students.

11.4 Coming back to Example 11.2, suppose student j consents to give up her priority at school s_1, *i.e.*, if she is okay with accepting the the unfairness caused by matching k to s_1. Thus, school s_1 is to be removed from student j's preferences without affecting the relative ranking of the other schools in her preferences.

Note that, when we rerun DA, replacing the preferences of student j with her new preferences, there is no change in the placement of student j. But, because the previously mentioned cycle now disappears, students i and k each move one position up in their preferences. Moreover, the new matching is now Pareto-efficient. To be more detailed, the preference profiles become

i	j	k	s_1	s_2
s_2		s_1	i	k
s_1		s_2	j	i
			k	

Table 11.3

The matching produced by DA is

$$\begin{bmatrix} i & j & k \\ s_2 & \emptyset & s_1 \end{bmatrix},$$

and the procedure is

Step	1	End
s_1	k	k
s_2	i	i
$\emptyset$	j	j

Table 11.4

☞ 11.5 Definition: Given a problem to which DA is applied, let i be a student who is tentatively placed at a school s at some Step t and rejected from it at some later Step t'. If there is at least one other student who is rejected from school s after Step $t-1$ and before Step t', that is, rejected at a Step $l \in \{t, t+1, \ldots, t'-1\}$, then we call student i an interrupter for school s and the pair (i, s) an interrupting pair of Step t'.

11.6 In real-life applications, it is imperative that each student be asked for permission to waive her priority for a critical school in cases similar to Example 11.2. We incorporate this aspect of the problem into the procedure by dividing the set of students into two groups: those students who consent to priority waiving and those who do not.

11.7 Lemma: If the outcome of DA is inefficient for a problem, then there exists one interrupting pair in DA. However, the converse is not necessarily true, *i.e.*, an interrupting pair does not always result in efficiency loss.

Proof. (1) Fix a school choice problem. Let α denote the outcome of DA, which is Pareto dominated by another matching β.

(2) There exists a student i_1 such that $\beta(i_1) \succ_{i_1} \alpha(i_1)$.

(3) Under the matching α, all the seats of school $\beta(i_1)$ are full.

(4) Since β Pareto dominates α, there is a student i_2 who is placed at school $\beta(i_1)$ under α, and who is placed at a better school $\beta(i_2)$ under β.

(5) Under the matching α, all the seats of school $\beta(i_2)$ are full.

(6) Since β Pareto dominates α, there is a student i_3 who is placed at school $\beta(i_2)$ under α, and who is placed at a better school $\beta(i_3)$ under β.

(7) Continuing in a similar way, we conclude that because matching β Pareto dominates matching α, there is a student i_k who is placed at school $\beta(i_{k-1})$ under , and who is placed at the school $\beta(i_1)$ under β, which is better for her.

(8) That is, there is a cycle of students $(i_1, i_2, \ldots, i_k)$ $(k \geq 2)$, such that each student prefers the school the next student in the cycle (for student i_k it is i_1) is placed at under α to the school she is placed at under the same matching.

(9) Let $i \in \{i_1, i_2, \ldots, i_k\}$ be the student in this cycle who is the last (or, one of the last, if there are more than one such students) to apply to the school, say school s, that she is placed at at the end of DA.

(10) Then the student in the above cycle who prefers school s to the school she is placed at under α was rejected from there at an earlier step.

(11) Then, when student i applies to school s, all the seats are already full and because student i is placed at this school at the end of DA, some student i' is rejected.

(12) Thus, student i' is an interrupter for school s.

□

Consider an interrupting pair (i, s): it is possible that student i's rejection from school s (at Step t according to the above definition) could be caused by some student j whose application to school s has not been directly or indirectly triggered by the student that student i displaced from school s when she is tentatively admitted. In such cases as these, the DA outcome does not suffer efficiency loss due to the presence of an interrupter.

11.8 Efficiency-adjusted deferred acceptance mechanism (EADAM):

Round 0: Run DA for $(P, \succ)$.

Round $k \geq 1$: (1) Find the last step of DA in Round $k-1$ in which a consenting interrupter is rejected from the school for which she is an interrupter.

(2) Identify all interrupting pairs of that step each of which contains a consenting interrupter.

(3) For each identified interrupting pair (i, s), remove school s from the preferences of student i without changing the relative order of the remaining schools. Do not make any changes in the preferences of the remaining students.

(4) Rerun DA with the new preference profile.

End: If there are no interrupting pairs, then stop.

When we say student i is an interrupter of Round t, this means that student i is identified as an interrupter during Round $t+1$ in DA that was run at the end of Round t.

11.9 Example (Example 5 in Kesten (2010)): Let $I = \{i_1, i_2, i_3, i_4, i_5, i_6\}$ and $S = \{s_1, s_2, \ldots, s_5\}$, where $q_{s_1} = q_{s_2} = q_{s_3} = q_{s_4} = 1$ and $q_{s_5} = 2$. The priorities for the schools and the preferences of the students are given as follows: Suppose that all students consent.

s_1	s_2	s_3	s_4	s_5	i_1	i_2	i_3	i_4	i_5	i_6
i_2	i_3	i_1	i_4	$\vdots$	s_2	s_3	s_3	s_1	s_1	s_4
i_1	i_6	i_6	i_3		s_1	s_1	s_4	s_2	s_5	s_1
i_5	i_4	i_2	i_6		s_3	s_5	s_2	s_4	$\vdots$	s_3
i_6	i_1	i_3	$\vdots$		$\vdots$	$\vdots$	$\vdots$			s_2
i_4	$\vdots$	$\vdots$								s_5
i_3										

Table 11.5

Round 0:

Step	1	2	3	4	5	6	7	8	9	10	End
s_1	$\cancel{i_4}, i_5$	i_5	$i_1, \cancel{i_5}, \cancel{i_6}$	i_1	$\cancel{i_1}, i_2$	i_2	i_2	i_2	i_2	i_2	i_2
s_2	i_1	$\cancel{i_1}, i_4$	i_4	i_4	i_4	i_4	$\cancel{i_4}, i_6$	i_6	$i_3, \cancel{i_6}$	i_3	i_3
s_3	$i_2, \cancel{i_3}$	i_2	i_2	$\cancel{i_2}, i_6$	i_6	$i_1, \cancel{i_6}$	i_1	i_1	i_1	i_1	i_1
s_4	i_6	$i_3, \cancel{i_6}$	i_3	i_3	i_3	i_3	i_3	$\cancel{i_3}, i_4$	i_4	i_4	i_4
s_5				i_5	i_5	i_5	i_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3, i_4	i_1, i_6	i_5, i_6	i_2	i_1	i_6	i_4	i_3	i_6		

Table 11.6

Round 1: The last step in which an interrupter is rejected from the school she is an interrupter for is Step 9, where the interrupting pair is (i_6, s_2). We remove school s_2 from the preferences of student i_6. We then rerun DA with the new preference profile:

Step	1	2	3	4	5	6	7	End
s_1	$\cancel{i_4}, i_5$	i_5	$i_1, \cancel{i_5}, \cancel{i_6}$	i_1	$\cancel{i_1}, i_2$	i_2	i_2	i_2
s_2	i_1	$\cancel{i_1}, i_4$	i_4	i_4	i_4	i_4	i_4	i_4
s_3	$i_2, \cancel{i_3}$	i_2	i_2	$\cancel{i_2}, i_6$	i_6	$i_1, \cancel{i_6}$	i_1	i_1
s_4	i_6	$i_3, \cancel{i_6}$	i_3	i_3	i_3	i_3	i_3	i_3
s_5				i_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3, i_4	i_1, i_6	i_5, i_6	i_2	i_1	i_6		

Table 11.7

Round 2: The last step in which an interrupter is rejected from the school she is an interrupter for is Step 6, where the interrupting pair is (i_6, s_3). We remove school s_3 from the (updated) preferences of student i_6. We then rerun DA with the new preference profile:

Step	1	2	3	4	End
s_1	$\xcancel{i_4}, i_5$	i_5	$i_1, \xcancel{i_5}, \xcancel{i_6}$	i_1	i_1
s_2	i_1	$\xcancel{i_1}, i_4$	i_4	i_4	i_4
s_3	$i_2, \xcancel{i_3}$	i_2	i_2	i_2	i_2
s_4	i_6	$i_3, \xcancel{i_6}$	i_3	i_3	i_3
s_5				i_5, i_6	i_5, i_6
$\emptyset$	i_3, i_4	i_1, i_6	i_5, i_6		

Table 11.8

Round 3: The last step in which an interrupter is rejected from the school she is an interrupter for is Step 3, where the interrupting pair is (i_5, s_1). We remove school s_1 from the preferences of student i_5 and keep the preferences of the remaining students the same. We then rerun DA with the new preference profile:

Step	1	2	3	4	5	6	End
s_1	i_4	i_4	$\xcancel{i_4}, i_6$	i_6	$i_1, \xcancel{i_6}$	i_1	i_1
s_2	i_1	i_1	i_1	$\xcancel{i_1}, i_4$	i_4	i_4	i_4
s_3	$i_2, \xcancel{i_3}$	i_2	i_2	i_2	i_2	i_2	i_2
s_4	i_6	$i_3, \xcancel{i_6}$	i_3	i_3	i_3	i_3	i_3
s_5	i_5	i_5	i_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3	i_6	i_4	i_1	i_6		

Table 11.9

Round 4: The last step in which an interrupter is rejected from the school she is an interrupter for is Step 5, where the interrupting pair is (i_6, s_1). We remove school s_1 from the (updated) preferences of student i_6. We then rerun DA with the new preference profile:

Step	1	2	3	End
s_1	i_4	i_4	i_4	i_4
s_2	i_1	i_1	i_1	i_1
s_3	$i_2, \xcancel{i_3}$	i_2	i_2	i_2
s_4	i_6	$i_3, \xcancel{i_6}$	i_3	i_3
s_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3	i_6		

Table 11.10

End: There are no interrupting pairs; hence we stop.

11.10 Because the numbers of schools and students are finite, the algorithm eventually terminates in a finite number of steps. Since DA runs in two consecutive rounds of EADAM are identical until the first step a consenting interrupter applies to the school for which she is an interrupter, in practice the EADAM outcome can be computed conveniently by only rerunning the relevant last steps of DA. Note also that each round of EADAM consists of a run of DA that is a polynomial-time procedure (*e.g.*, see Gusfield and Irving (1989)). Then because a student can be identified as an interrupter at most $|S|$ times, these iterations need to be done at most $|I||S|$ times, giving us a computationally simple polynomial-time algorithm.

11.11 Remark: Why shall we start with the last interrupter(s) in the algorithm?

Case 1: Handle all the interrupters simultaneously.

Let $I = \{i_1, i_2, i_3\}$ and $S = \{s_1, s_2, s_3\}$, where each school has only one seat. The priorities for the schools and the preferences of the students are given as follows:

i	j	k	s_1	s_2	s_3
s_1	s_1	s_2	k	i	$\vdots$
s_2	s_2	s_1	i	j	
s_3	s_3	s_3	j	j	

Table 11.11

The procedure of DA is

Step	1	2	3	4	5	End
s_1	$\cancel{i}, j$	i	$\cancel{i}, k$	k	k	k
s_2	k	$j, \cancel{k}$	j	$i, \cancel{j}$	i	i
s_3					j	j
$\emptyset$	i	k	i	j		

Table 11.12

The outcome of DA for this problem is not Pareto efficient. There are two interrupting pairs within the algorithm: (i_1, s_1) and (i_2, s_2).

Now consider the revised problem where we remove school s_1 from student i_1's preferences and school s_2 from those of student i_2. The procedure of DA to the revised problem is as follows:

Step	1	2	3	End
s_1	j	$\cancel{j}, k$	k	k
s_2	$i, \cancel{k}$	i	i	i
s_3			j	j
$\emptyset$	k	j		

Table 11.13

The outcome does not change (*i.e.*, still inefficient) even though there are no interrupters left in the new algorithm.

Case 2: Start with the earliest interrupter.

Consider the example above. Note that student i_1 was identified as an interrupter at Step 3 before student i_2, who was identified at Step 4. Thus, let us then consider the revised problem where we only remove school s_1 from student i_1's preferences. The procedure of DA to the revised problem is as follows:

Step	1	2	3	4	End
s_1	j	$\cancel{j}, k$	k	k	k
s_2	$i, \cancel{k}$	i	$i, \cancel{j}$	i	i
s_3				j	j
$\emptyset$	k	j	j		

Table 11.14

Once again, there is no change in the outcome. Hence, this approach does not work either.

11.12 Theorem (Theorem 1 in Kesten (2010)): The EADAM Pareto dominates the DA as well as any mechanism which eliminates justified envy. If no student consents, the two mechanisms are equivalent. If all students consent, then the EADAM outcome is Pareto-efficient. In the EADAM outcome all nonconsenting students' priorities are respected; however, there may be consenting students whose priorities for some schools are violated with their permission.

11.13 Lemma (Lemma A.1 in Kesten (2010)): Given a problem, the matching obtained at the end of Round r ($r \geq 1$) of EADAM places each student at a school that is at least as good for her as the school she was placed at at the end of Round $r-1$.

Proof. (1) Suppose by contradiction that there are a problem, a Round r ($r \geq 1$), of EADAM, and a student i_1 such that the school student i_1 is placed at in Round r is worse for her than the school s_1^{r-1} she was placed at in Round $r-1$.

(2) This means that when we run DA in Round r, student i_1 is rejected from school s_1^{r-1}.

(3) Then there is a student $i_2 \in I \setminus \{i_1\}$ who is placed at school s_1^{r-1} in Round r and who was placed at a school s_2^{r-1} (in Round $r-1$) that is better for her than school s_1^{r-1}.

(4) This means there is a student $i_3 \in I \setminus \{i_1, i_2\}$ who is placed at school s_2^{r-1} in Round r and who was placed at a school s_3^{r-1} that is better for her than school s_2^{r-1} in Round $r-1$, and so on.

(5) Thus, there must be a student $i_k \in I \setminus \{i_1, \ldots, i_{k-1}\}$ who is the first student to apply to a school s_{k-1}^{r-1} that is worse for her than the school s_k^{r-1} she was placed at in Round $r-1$.

(6) Case 1: Student i_k is not an interrupter of Round $r-1$.

- (i) The preferences of student i_k are the same in Rounds r and $r-1$.
- (ii) Thus, there is a student who is placed at school s_k^{r-1} in Round r and who did not apply to it in Round $r-1$.
- (iii) This contradicts the assumption that student i_k is the first student to apply to a school that is worse for her than the school she was placed at in Round $r-1$.

(7) Case 2: Student i_k is an interrupter of Round $r-1$.

- (i) In Round r, student i_k, instead of applying to the school she is an interrupter for, applied to her next choice, say school s^*.
- (ii) Student i_k also applied to school s^* in Round $r-1$.
- (iii) Thus, there is a student who is placed at school s_k^{r-1} in Round r and who did not apply to it in Round $r-1$.
- (iv) But then, this again contradicts the assumption that student i_k is the first student to apply to a school that is worse for her than the school she was placed at in Round $r-1$.

□

11.14 Corollary (Corollary 1 in Kesten (2010)): If all students consent, then EADAM selects the Pareto efficient matching which eliminates justified envy whenever it exists.

11.2 Simplified efficiency-adjusted deferred acceptance algorithm

☞ 11.15 Definition: A school s is underdemanded at a matching μ if no student prefers s to her assignment under μ.

It is straightforward to see that a school is underdemanded at the DA matching if and only if it never rejects any student throughout the DA procedure.

☞ 11.16 Definition: A school is tier-0 underdemanded at matching μ if it is underdemanded at μ. For any integer $k > 0$, a school is tier-k underdemanded at matching μ if

- it is desired only by students matched with lower-tier underdemanded schools at μ, and
- it is desired by at least one of the students matched with tier-$(k-1)$ underdemanded schools at μ.

School s is essentially underdemanded at matching μ if it is tier-k underdemanded at μ for some integer $k \geq 0$.

11.17 The set of essentially underdemanded schools at the DA matching can also be identified through a recursive process, by reviewing the DA procedure that produces this DA matching. Tier-0 underdemanded schools are the schools that never reject any student throughout the DA procedure. After removing tier-0 underdemanded schools and the students matched with them, tier-1 underdemanded schools are the remaining schools that never reject any remaining students throughout the DA procedure, and so on.

11.18 Example (Example 1 in Tang and Yu (2014)): There are four schools $\{s_1, s_2, s_3, s_4\}$, each with one seat, and four students $\{i_1, i_2, i_3, i_4\}$. Their priorities and preferences are as follows:

s_1	s_2	s_3	s_4	i_1	i_2	i_3	i_4
i_1	i_3	i_2	i_4	s_2	s_1	s_1	s_3
i_2	i_1	i_4	$\vdots$	s_1	s_3	s_2	s_4
i_3	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$							

Table 11.15

The DA procedure is

Step	1	2	3	4	5	End
s_1	$i_2, \cancel{i_3}$	i_2	$i_1, \cancel{i_2}$	i_1	i_1	i_1
s_2	i_1	$\cancel{i_1}, i_3$	i_3	i_3	i_3	i_3
s_3	i_4	i_4	i_4	$\cancel{i_2}, i_4$	i_4	i_4
s_4					i_2	i_2
$\emptyset$	i_3	i_1	i_2	i_2		

Table 11.16

By definition, school s_4 is (tier-0) underdemanded at the DA matching, since it never rejects any student throughout the DA procedure.

Let remove student i_2 who is matched to s_4. Then school s_3 is tier-1 underdemanded at the DA matching.

After removing student i_4, it is clear that s_1 and s_2 are not underdemanded.

11.19 Definition: Student i is not Pareto improvable (or, simply, unimprovable) at DA$[P, \succ]$ if for every matching μ that Pareto dominates DA$[P, \succ]$, $\mu(i) = \text{DA}[P, \succ](i)$.

11.20 Lemma (Lemma 1 in Tang and Yu (2014)): At the DA matching, all students matched with essentially underdemanded schools are not Pareto improvable.

Therefore, the concept of (essentially) underdemanded schools offers us a convenient way to identify a large set of unimprovable students. The lemma above still holds if the DA matching is replaced with any non-wasteful matching.

11.21 Simplified EADAM:

Round 0: Run DA for the problem $\langle P, \succ \rangle$.

Round k: This round consists of three steps:

(1) Identify the schools that are underdemanded at the round-$(k-1)$ DA matching, settle the matching at these schools, and remove these schools and the students matched with them.

(2) For each removed student i who does not consent, each remaining school s that student i desires and each remaining student j such that $i \succ_s j$, remove s from j's preference.

(3) Rerun DA (the round-k DA) for the subproblem that consists of only the remaining schools and students.

End: Stop when all schools are removed.

11.22 Example (Examples 2 and 3 in Tang and Yu (2014)): Let $I = \{i_1, i_2, i_3, i_4, i_5, i_6\}$ and $S = \{s_1, s_2, \ldots, s_5\}$, where $q_{s_1} = q_{s_2} = q_{s_3} = q_{s_4} = 1$ and $q_{s_5} = 2$. The priorities for the schools and the preferences of the students are given as follows:

s_1	s_2	s_3	s_4	s_5	i_1	i_2	i_3	i_4	i_5	i_6
i_2	i_3	i_1	i_4	$\vdots$	s_2	s_3	s_3	s_1	s_1	s_4
i_1	i_6	i_6	i_3		s_1	s_1	s_4	s_2	s_5	s_1
i_5	i_4	i_2	i_6		s_3	s_5	s_2	s_4	$\vdots$	s_3
i_6	i_1	i_3	$\vdots$		$\vdots$	$\vdots$	$\vdots$			s_2
i_4	$\vdots$	$\vdots$								s_5
i_3										

Table 11.17

Suppose that all students consent.

Round 0:

Step	1	2	3	4	5	6	7	8	9	10	End
s_1	$\cancel{i_4}, i_5$	i_5	$i_1, \cancel{i_5}, \cancel{i_6}$	i_1	$\cancel{i_1}, i_2$	i_2	i_2	i_2	i_2	i_2	i_2
s_2	i_1	$\cancel{i_1}, i_4$	i_4	i_4	i_4	i_4	$\cancel{i_4}, i_6$	i_6	$i_3, \cancel{i_6}$	i_3	i_3
s_3	$i_2, \cancel{i_3}$	i_2	i_2	$\cancel{i_2}, i_6$	i_6	$i_1, \cancel{i_6}$	i_1	i_1	i_1	i_1	i_1
s_4	i_6	$i_3, \cancel{i_6}$	i_3	i_3	i_3	i_3	i_3	$\cancel{i_3}, i_4$	i_4	i_4	i_4
s_5				i_5	i_5	i_5	i_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3, i_4	i_1, i_6	i_5, i_6	i_2	i_1	i_6	i_4	i_3	i_6		

Table 11.18

Round 1: At round-0 DA matching, s_5 is the only underdemended school, and students i_5 and i_6 are matched with it. Remove s_5 together with i_5 and i_6, and rerun DA with the rest of the schools and students. The procedure of round-1 DA is illustrated in the following table:

Step	1	2	End
s_1	i_4	i_4	i_4
s_2	i_1	i_1	i_1
s_3	$i_2, \cancel{i_3}$	i_2	i_2
s_4		i_3	i_3
$\emptyset$	i_3		

Table 11.19

Round 2: At the end of round-1, all schools are underdemanded except for s_3. So in round-2, we first remove all other schools and their matched students, and then run DA for s_3 and i_2. The round-2 DA is trivial and the algorithm stops immediately afterward. The final matching is the same as the round-1 DA matching.

11.23 The simplified EADAM preserves the iterative structure of Kesten's EADAM, while taking a new perspective by focusing on unimprovable students instead of (only) interrupters.

The new perspective leads to several differences.

- First, at the end of each round, we remove all students matched with underdemanded schools, and thereby remove all of their desired applications instead of removing only the last interruption.
- Second, after the removal of non-consenting students—since we already know which matchings among the remaining schools and students would violate their priorities—we modify the preferences of the remaining students accordingly to avoid violations of their priorities in future rounds of the algorithm.

✌ 11.24 Lemma (Lemma 2 in Tang and Yu (2014)): For each $k \geq 1$, the round-k DA matching of the simplified EADAM weakly Pareto dominates that of round-$(k-1)$.

11.25 Lemma (Proposition 1 in Tang and Yu (2014)): The simplified EADAM is well-defined and stops within $|S \cup \{\emptyset\}| + 1 = m + 2$ rounds.

✌ 11.26 Theorem (Theorem 1 in Tang and Yu (2014)): The simplified EADAM is Pareto efficient when all students consent and is constrained efficient otherwise.

11.27 Theorem (Theorem 2 in Tang and Yu (2014)): Under the simplified EADAM, the assignment of any student does not change whether she consents or not.

11.28 Lemma (Lemma 3 in Tang and Yu (2014)): The lastly rejected interrupters of the DA procedure are matched with essentially underdemanded schools at the DA matching.

✌ 11.29 Theorem (Theorem 3 in Tang and Yu (2014)): For every school choice problem with consent, the simplified EADAM produces the same matching as Kesten's EADAM does.

11.3 Stable improvement cycle algorithm

11.30 In a school choice problem $\langle I, S, q, P, \succ \rangle$ with a given matching μ, for each school s, let d_s be the highest $\succ_s$-priority student among those who desire s (*i.e.*, who prefer s to her assignment under μ).

☞ 11.31 Definition: A stable improvement cycle consists of distinct students $i_1, i_2, \ldots, i_n = i_0$ ($n \geq 2$) such that for each $\ell = 0, 1, \ldots, n-1$,

(1) i_ℓ is matched to some school under μ;

(2) i_ℓ desires $\mu(i_{\ell+1})$; and

(3) $i_\ell = D_{\mu(i_{\ell+1})}$.

11.32 Given a stable improvement cycle, define a new matching μ' by:

$$\mu'(j) = \begin{cases} \mu(j), & \text{if } j \notin \{i_1, i_2, \ldots, i_n\}; \\ \mu(i_{\ell+1}), & \text{if } j = i_\ell. \end{cases}$$

Note that the matching μ' continues to be stable and it Pareto dominates μ.

✌ 11.33 Theorem (Theorem 1 in Erdil and Ergin (2008)): Fix $\succ$ and P, and let μ be a stable matching. If μ is Pareto dominated by another stable matching ν, then it admits a stable improvement cycle.

11.34 Lemma (Lemma 1 in Erdil and Ergin (2008)): Let $\succ$ and P be given. Suppose that μ is a stable matching that is Pareto dominated by a (not necessarily stable) matching ν. Let I' denote the set of students who are strictly better off under ν and let $S' = \mu(I')$ be the set of schools to which students in I' are assigned under μ. Then we have:

(i) Students who are not in I' have the same match under μ and ν;

(ii) The number of students in I' who are assigned to a school s are the same in μ and ν; in particular, $S' = \nu(I')$;

(iii) Each student in I' is assigned to a school in μ and in ν.

Proof. (i) For each $i \in I \setminus I'$, i is indifferent between $\mu(i)$ and $\nu(i)$. Thus, $\mu(i) = \nu(i)$.

(ii) We first show that $|I' \cap \mu^{-1}(s)| \geq |I' \cap \nu^{-1}(s)|$ for any school s.

(1) Suppose that $|I' \cap \mu^{-1}(s)| < |I' \cap \nu^{-1}(s)|$ for some school s.

(2) Together with (i), this implies that the number of students in I who are assigned to s under μ is less than the number of students who are assigned to s under ν.

(3) Hence, s must have empty seats under μ.

(4) For any $i \in I' \cap \nu^{-1}(s)$, $s = \nu(i) P_i \mu(i)$, that is, i desires s which has empty seats under μ, a contradiction to the stability of μ.

Now suppose the inequality $|I' \cap \mu^{-1}(s)| \geq |I' \cap \nu^{-1}(s)|$ holds strictly for some school s^*.

(5) Summing across all schools we have

$$\sum_{s \in S} |I' \cap \mu^{-1}(s)| > \sum_{s \in S} |I' \cap \nu^{-1}(s)|.$$

(6) Hence, the number of students in I' who are assigned to some school under μ is more than the number of students in I' who are assigned to some school in ν.

(7) There exists a student $i \in I'$ who is assigned to a school under μ, but not under ν.

(8) Since $\emptyset = \nu(i) P_i \mu(i)$, this contradicts the stability of μ.

(iii) (1) From (ii), we have

$$|I'| \geq \sum_{s \in S} |I' \cap \mu^{-1}(s)| = \sum_{s \in S} |I' \cap \nu^{-1}(s)|.$$

(2) It suffices to show that the inequality above cannot hold strictly.

(3) Suppose for a contradiction that

$$|I'| > \sum_{s \in S} |I' \cap \mu^{-1}(s)| = \sum_{s \in S} |I' \cap \nu^{-1}(s)|.$$

(4) Hence, there exists a student $i \in I'$ who is unmatched under ν.

(5) Note that i has to be matched under μ; otherwise, she would be indifferent between μ and ν, a contradiction to her being in I'.

(6) But then $\emptyset = \nu(i) P_i \mu(i)$, a contradiction to the stability of μ.

□

11.35 *Proof of Theorem 11.33.*

(1) Suppose μ and ν are stable matchings and ν Pareto dominates μ.

(2) Let I' denote the set of students who are strictly better off under ν. Let $S' = \mu(I')$ be the set of schools to which students in I' are assigned to under μ.

(3) For any $s \in S'$, since $\mu(I') = \nu(I') = S'$, there exists a student i such that i desires s at μ and is assigned to s under ν.

(4) For any $s \in S'$, let i_s denote the highest $\succ_s$-priority student among those in I' that desire s at μ.

(5) Let school $\mu(i_s)$ point to s.

(6) By Lemma 11.34, $\mu(i_s) \in S'$.

(7) Since i_s desires s at μ, $\mu(i_s) \neq s$.

(8) Thus, we can repeat this for each school $s \in S'$ and find a school $t \in S' \setminus \{s\}$ that points to s.

(9) Since each school in S' is pointed to by a different school in S', there exists a cycle of distinct schools $s_1, s_2, \ldots, s_n = s_0$ $(n \geq 2)$ in S', where s_ℓ points to $s_{\ell+1}$ for $\ell = 0, 1, \ldots, n-1$.

(10) Let $i_\ell = i_{s_{\ell+1}}$ for $\ell = 0, 1, \ldots, n-1$. Then $\mu(i_\ell) = s_\ell$, and i_ℓ desires $s_{\ell+1} = \mu(i_{\ell+1})$ at μ.

(11) Let d_s denote the highest $\succ_s$-priority students among those who desire s at μ. In the following, we will show that $i_\ell = d_{\mu(i_{\ell+1})}$. For simplicity, denote $d_{\mu(i_{\ell+1})}$ by j.

(12) Suppose $i_\ell \neq j$. Thus, $j \notin I'$ and $j \succ_{\mu(i_{\ell+1})} i_\ell$.

(13) Then $\mu(j) = \nu(j)$ by Lemma 11.34.

(14) Since j desires $\mu(i_{\ell+1})$ at μ, j also desires $\mu(i_{\ell+1})$ at ν.

(15) This contradicts the stability of ν, since j has high $\succ_{\mu(i_{\ell+1})}$-priority than i_ℓ, who is matched to $\mu(i_{\ell+1})$ under ν.

□

11.36 Stable improvement cycles algorithm:

Step 0: Run DA algorithm and obtain a temporary matching μ^0.

Step $k \geq 1$: (1) Find a stable improvement cycle for μ^{k-1}: for schools s and t, let $s \to t$ if the student d_t is matched to s under μ^{k-1}.

(2) If there are any cycles, select one. For each $s \to t$ in this cycle, carry out this stable improvement cycle to obtain μ^k.

End: The algorithm stops when there is no cycle.

11.4 Weak priorities

11.37 In the context of school choice, it might be reasonable to assume that the students have strict preferences, but school priority orders are typically determined according to criteria that do not provide a strict ordering of all the students. Instead, school priorities are weak orderings with quite large indifference classes. For instance, in Boston there are mainly four indifference classes for each school in the following order: (a) the students who have siblings at that school (sibling) and are in the reference area of the school (walk zone), (b) sibling, (c) walk zone, and (d) all other students.

Common practice in these cases is to exogenously fix an ordering of the students, chosen randomly, and break all the indifference classes according to this fixed strict ordering. Then, one can apply the DA algorithm to obtain the student optimal stable matching with respect to the strict priority profile derived from the original one. Since the breaking of indifferences does not switch the positions of any two students in any priority order, the outcome would also be stable with respect to the original priority structure.

11.38 Tie-breaking rules do not necessarily brings us the student-optimal stable matching.

Example: Consider the school choice problem $\langle I, S, q, P, \succsim\rangle$, where $I = \{i, j, k\}$, $S = \{s_1, s_2\}$, $q_{s_1} = q_{s_2} = 1$, and

i	j	k	s_1	s_2
s_2	s_1	s_1	i	k
s_1		s_2	j, k	i

Table 11.20

The tie-breaking rule either breaks $\succsim_{s_1}$ as $i \succ_{s_1} j \succ_{s_1} k$ or as $i \succ_{s_1} k \succ_{s_1} j$, and the corresponding DA produces two stable matching, respectively

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix}, \quad \nu = \begin{bmatrix} i & j & k \\ s_2 & \emptyset & s_1 \end{bmatrix}.$$

Clearly, μ is Pareto dominated by ν.

11.39 Tie-breaking rules may lead to a stable matching such that there may be another stable matching that is better off for everyone.

Example: $I = \{i, j, k\}$, $S = \{s_1, s_2, s_3\}$, each school has one seat,

i	j	k	s_1	s_2	s_3
s_2	s_3	s_2	i	j	k
s_1	s_2	s_3	j, k	i, k	i, j
s_3	s_1	s_1			

Table 11.21

Assume ties are broken in the order i, j, k for each school. DA with this tie-breaking finds

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & s_2 & s_3 \end{bmatrix},$$

but everyone prefers

$$\mu' = \begin{bmatrix} i & j & k \\ s_1 & s_3 & s_2 \end{bmatrix},$$

and μ' is stable with respect to the original priority.

11.40 As a consequence of Theorem 9.32, for any tie-breaking rule, there is no strategy-proof mechanism that Pareto dominates the DA outcome with this tie-breaking rule.

11.41 If we need to use DA, what tie-breaking should be used?

- Single tie breaking: Use one lottery to decide order on all students and, whenever two students are in the same priority class, break the tie using the ordering.
- Multiple tie breaking: Draw one lottery for each school, and whenever two students are in the same priority class for a school, break the tie using the ordering for that particular school.

Abdulkadiroğlu *et al.* (2015) show that, when there is no intrinsic priority and the market is large, DA-STB is more efficient than DA-MTB.

Intuition: DA's inefficiency comes from students displacing each other. That is less likely in STB than in MTB.

11.42 In a school choice problem $\langle I, S, q, P, \succ\rangle$ with a given matching μ, for each school s, let D_s be the set of highest $\succ_s$-priority students among those who desire s (*i.e.*, who prefer s to her assignment under μ).

☞ 11.43 Definition: A stable improvement cycle consists of distinct students $i_1, i_2, \ldots, i_n = i_0$ ($n \geq 2$) such that for each $\ell = 0, 1, \ldots, n-1$,

(1) i_ℓ is matched to some school under μ;

(2) i_ℓ desires $\mu(i_{\ell+1})$; and

(3) $i_\ell \in D_{\mu(i_{\ell+1})}$.

11.44 Given a stable improvement cycle, define a new matching μ' by:

$$\mu'(j) = \begin{cases} \mu(j), & \text{if } j \notin \{i_1, i_2, \ldots, i_n\}; \\ \mu(i_{\ell+1}), & \text{if } j = i_\ell. \end{cases}$$

Note that the matching μ' continues to be stable and it Pareto dominates μ.

✌ 11.45 Theorem (Theorem 1 in Erdil and Ergin (2008)): Fix $\succsim$ and P, and let μ be a stable matching. If μ is Pareto dominated by another stable matching ν, then it admits a stable improvement cycle.

Proof.

(1) Suppose μ and ν are stable matchings and ν Pareto dominates μ.

(2) Let I' denote the set of students who are strictly better off under ν. Let $S' = \mu(I')$ be the set of schools to which students in I' are assigned to under μ.

(3) For any $s \in S'$, since $\mu(I') = \nu(I') = S'$, there exists a student i such that i desires s at μ and is assigned to s under ν.

(4) For any $s \in S'$, let i_s denote the highest $\succ_s$-priority student among those in I' that desire s at μ.

(5) Let school $\mu(i_s)$ point to s.

(6) By Lemma 11.34, $\mu(i_s) \in S'$.

(7) Since i_s desires s at μ, $\mu(i_s) \neq s$.

(8) Thus, we can repeat this for each school $s \in S'$ and find a school $t \in S' \setminus \{s\}$ that points to s.

(9) Since each school in S' is pointed to by a different school in S', there exists a cycle of distinct schools $s_1, s_2, \ldots, s_n = s_0$ ($n \geq 2$) in S', where s_ℓ points to $s_{\ell+1}$ for $\ell = 0, 1, \ldots, n-1$.

(10) Let $i_\ell = i_{s_{\ell+1}}$ for $\ell = 0, 1, \ldots, n-1$. Then $\mu(i_\ell) = s_\ell$, and i_ℓ desires $s_{\ell+1} = \mu(i_{\ell+1})$ at μ.

(11) Let D_s denote the set of highest $\succ_s$-priority students among those who desire s at μ. In the following, we will show that $i_\ell \in D_{\mu(i_{\ell+1})}$.

(12) Suppose $i_\ell \notin D_{\mu(i_{\ell+1})}$. Thus, $D_{\mu(i_{\ell+1})}$ has no intersection with I'.

(13) For any $j \in D_{\mu(i_{\ell+1})}$, we have $j \notin I'$ and $j \succ_{\mu(i_{\ell+1})} i_\ell$.

(14) Since $j \notin I'$, $\mu(j) = \nu(j)$ by Lemma 11.34.

(15) Since j desires $\mu(i_{\ell+1})$ at μ, j also desires $\mu(i_{\ell+1})$ at ν.

(16) This contradicts the stability of ν, since j has high $\succ_{\mu(i_{\ell+1})}$-priority than i_ℓ, who is matched to $\mu(i_{\ell+1})$ under ν.

□

11.46 Stable improvement cycles algorithm:

Step 0: Run DA algorithm and obtain a temporary matching μ^0.

Step $k \geq 1$: (1) Find a stable improvement cycle for μ^{k-1}: for schools s and t, let $s \to t$ if some student $i \in D_t$ is matched to s under μ^{k-1}.

(2) If there are any cycles, select one. For each $s \to t$ in this cycle, select a student $i \in D_t$ with $\mu^{k-1}(i) = s$. Carry out this stable improvement cycle to obtain μ^k.

End: The algorithm stops when there is no cycle.

11.47 Starting with an arbitrary stable matching, SIC produces a stable constrained efficient stable matching.

11.48 EADAM and simplified EADAM can also be applied to resolve the efficiency loss resulting from tie-breakings. See Kesten (2010) and Tang and Yu (2014).

11.49 There may not exist a strategy-proof selection of constrained efficient matchings.

Example: Let $I = \{i, j, k\}$, $S = \{a, b, c\}$, each school has one seat,

i	j	k	a	b	c
b	b	a	i	k	k
c	c	b	j	i, j	j
a	a	c	k		i

Table 11.22

The two constrained efficient matchings are

$$\mu = \begin{bmatrix} i & j & k \\ b & c & a \end{bmatrix} \text{ and } \mu' = \begin{bmatrix} i & j & k \\ c & b & a \end{bmatrix}.$$

Let both P'_i and P'_j be b, a, c. At $(P'_a, P_{-a}, \succsim)$, only μ is constrained efficient, and at $(P'_b, P_{-b}, \succsim)$, only μ' is constrained efficient.

If φ is a constrained efficient mechanism, then $\varphi[P'_a, P_{-a}, \succsim]$ has to be μ, and $\varphi[P'_b, P_{-b}, \succsim]$ has to be μ'. So at $(P, \succsim)$, one needs to select one of them. However, whenever φ selects the matching that is more favorable to one of a and b, the other will misreport.

11.50 Theorem (Theorem 1 in Abdulkadiroğlu *et al.* (2009)): For any tie-breaking rule, there is no strategy-proof mechanism that always results in a (Pareto) better matching than the DA with the tie-breaking.

11.51 Remark: Whatever improvement over DA with tie breaking may become non-strategy-proof.

Chapter 12

Affirmative action

Contents

12.1 Affirmative action policies have been widely used in public education although they have also received various criticisms. There are two affirmative action policies:

- Majority quota: the number of majority students matched to school s cannot exceed the majority quota q_s^M.
- Minority reserve: if the number of minority students matched to school s is less than the minority reserve r_s^m, then minority students are always preferred to majority students.

12.2 We are interested in the question whether these affirmative action policies really benefit minority students.

12.1 The formal model

☞ 12.3 A market is tuple $\Gamma = \langle I, S, (q_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$, where

- I is a finite set of students. The set of students are partitioned to two subsets, the set I^M of majority students and I^m of minority students.
- S is a finite set of schools.
- For each $s \in S$, q_s is the total capacity of school s.
- For each school $s \in S$, P_s is a strict priority order over the set of students.
- For each student $i \in I$, $\succ_i$ is a strict preference over S and being unmatched (being unmatched is denoted by $\emptyset$). If $s \succ_i \emptyset$, then school s is said to be acceptable to student i.

☞ 12.4 A matching μ is a mapping from I to $S \cup \{\emptyset\}$ such that $|\mu^{-1}(s)| \leq \mu_s$ for all $s \in S$.

☞ 12.5 A matching μ Pareto dominates matching ν if $\mu(i) \succsim_i \nu(i)$ for all $i \in I$ and $\mu(i) \succ_i \nu(i)$ for at least one $i \in I$. A matching is Pareto efficient if it is not Pareto dominated by another matching.

Affirmative action policies are implemented to improve the matches of minorities, sometimes at the expense of majorities. Therefore, we also need an efficiency concept to analyze the welfare of minority students. A matching μ Pareto dominates matching ν for minorities if $\mu(i) \succsim_i \nu(i)$ for all $i \in I^m$ and $\mu(i) \succ_i \nu(i)$ for at least one $i \in I^m$. A matching is Pareto efficient for minorities if it is not Pareto dominated for minorities by another matching.

☞ 12.6 A mechanism φ is a function that, for each market Γ, associates a matching $\varphi(\Gamma)$.

12.2 Affirmative action policies with majority quotas

12.7 For each $s \in S$, let q_s^M be the type-specific capacity for majority students ($q_s^M \leq q_s$), which is implemented by prohibiting schools to admit more than q_s^M of majority students.

Given $(q_s^M)_{s\in S}$, a matching μ is feasible under majority quotas if $|\mu^{-1}(s) \cap I^M| \leq q_s^M$ for all $s \in S$. This condition requires that the number of majority students matched to each school s is at most its type-specific capacity q_s^M.

☞ 12.8 Definition: Given $(q_s^M)_{s\in S}$, a matching μ is stable under majority quotes if

(1) $\mu(i) \succsim_i \emptyset$ for each $i \in I$, and

(2) if $s \succ_i \mu(i)$, then either

(i) $i \in I^m$, $|\mu^{-1}(s)| = q_s$ and $i' P_s i$ for all $i' \in \mu^{-1}(s)$, or

(ii) $i \in I^M$, $|\mu^{-1}(s) \cap I^M| < q_s^M$, $|\mu^{-1}(s)| = q_s$ and $i' P_s i$ for all $i' \in \mu^{-1}(s)$, or

(iii) $i \in I^M$, $|\mu^{-1}(s) \cap I^M| = q_s^M$, and $i' P_s i$ for all $i' \in \mu^{-1}(s) \cap I^M$.

All conditions except for (2-iii) are standard. Condition (2-iii) describes a case in which a potential blocking is not realized because of a type-specific capacity constraint for the majority students: Student i wants to be matched with school s, but she is a majority student and the seats for majority students are filled by students who have higher priority than i at s.

☞ 12.9 Definition: A mechanism φ is stable under majority quotes if $\varphi(\Gamma)$ is a stable matching under majority quotes in Γ for any given Γ.

 12.10 Deferred acceptance algorithm with majority quotas.

Step 1: Start with a matching in which no student is matched. Each student i applies to her first choice school (call it s). The school s rejects i if

- q_s seats are filled by students who have higher priority than i at s, or
- $i \in I^M$ and q_s^M seats are filled by students in I^M who have higher priority than i at s.

Each school s keeps all other students who applied to s.

Step k: Start with the tentative matching obtained at the end of Step $k-1$. Each student i applies to her first choice school (call it s) among all schools that have not rejected i before. The school s rejects i if

- q_s seats are filled by students who have higher priority than i at s, or
- $i \in I^M$ and q_s^M seats are filled by students in I^M who have higher priority than i at s. Each school s keeps all other students who applied to s.

End: The algorithm terminates at a step in which no rejection occurs, and the tentative matching at that step is finalized.

✌ 12.11 Theorem: Abdulkadiroğlu and Sönmez (2003) show that the outcome of the deferred acceptance algorithm with majority quotas is the student-optimal stable matching, a stable matching that is unanimously most preferred by all students among all stable matchings.

 12.12 Top trading cycles mechanism with majority quotas.

Step 1: Start with a matching in which no student is matched. For each school s, set its total counter at its total capacity q_s and its majority-specific counter at its type-specific capacity q_s^M. Each school points to a student who has the highest priority at that school. Each student i points to her most preferred school that still has a seat for her, that is, a school whose total counter is strictly positive and, if $i \in I^M$, its majority-specific counter is strictly positive. There exists at least one cycle (if a student points to $\emptyset$, it is regarded as a cycle). Every student in a cycle receives the school she is pointing to and is removed. The counter of each school is reduced by one. If the assigned student is in I^M, then the school matched to that student reduces its majority-specific counter by one. If no student remains, terminate. Otherwise, proceed to the next step.

Step k: Start with the matching and counter profile reached at the end of Step $k-1$. Each school points to a student who has the highest priority at that school. Each student i points to her most preferred school that still has a seat for her, that is, a school whose total counter is strictly positive and, if $i \in I^M$, its majority-specific counter is strictly positive. There exists at least one cycle (if a student points to $\emptyset$, it is regarded as a cycle). Every student in a cycle receives the school she is pointing to and is removed. The counter of each school is reduced by one. If the assigned student is in I^M, then the school matched to that student reduces its majority-specific counter by one. If no student remains, terminate. Otherwise, proceed to the next step.

End: If no student remains, terminate.

☞ 12.13 Definition: Market $\tilde{\Gamma} = \langle I, S, (\tilde{\boldsymbol{q}}_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$ is said to have a stronger quota-based affirmative action policy than $\Gamma = \langle I, S, (\boldsymbol{q}_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$ if, for every $s \in S$,

$$q_s = \tilde{q}_s \text{ and } q_s^M \geq \tilde{q}_s^M.$$

The definition requires that the type-specific capacity for the majority be smaller in $\tilde{\Gamma}$ than in Γ, so that a stronger restriction is imposed in the former.

A matching mechanism φ is said to respect the spirit of quota-based affirmative action if there are no markets Γ and $\tilde{\Gamma}$ such that $\tilde{\Gamma}$ has a stronger quota-based affirmative action policy than Γ and $\varphi(\tilde{\Gamma})$ is Pareto dominated by $\varphi(\Gamma)$ for the minority.

✌ 12.14 Theorem (Theorem 1 in Kojima (2012)): There exists no stable mechanism that respects the spirit of quota-based affirmative action. In particular, deferred acceptance algorithm with majority quotas does not respect the spirit of quota-based affirmative action.

Proof. (1) Consider the following market: $I = \{i_1, i_2, i_3\}$ with $I^M = \{i_1, i_2\}$ and $I^m = \{i_3\}$, $S = \{s_1, s_2\}$, $\boldsymbol{q}_{s_1} = (2, 2)$, $\boldsymbol{q}_{s_2} = (1, 1)$, preferences and priorities are as follows:

(2) There exists a unique stable matching

$$\mu = \begin{bmatrix} s_1 & s_2 \\ i_1, i_2 & i_3 \end{bmatrix}.$$

i_1	i_2	i_3	s_1	s_2
s_1	s_1	s_2	i_1	i_2
	s_2	s_1	i_2	i_3
			i_3	i_1

Table 12.1

(3) Consider $\tilde{\boldsymbol{q}} = (\tilde{\boldsymbol{q}}_{s_1}, \boldsymbol{q}_{s_2})$, where $\tilde{\boldsymbol{q}}_{s_1} = (2,1)$. Then market $\tilde{\Gamma} = \langle I, S, (\tilde{\boldsymbol{q}}_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$ has a stronger quota-based affirmative action policy than $\Gamma = \langle I, S, (\boldsymbol{q}_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$.

(4) In market $\tilde{\Gamma}$, there is a unique stable matching

$$\tilde{\mu} = \begin{bmatrix} s_1 & s_2 \\ i_1, i_3 & i_2 \end{bmatrix}.$$

(5) Student i_3 is strictly worse off under $\tilde{\mu}$ than under μ. Therefore $\tilde{\mu}$ is Pareto dominated by μ for the minority. Since μ and $\tilde{\mu}$ are the unique stable matchings of Γ and $\tilde{\Gamma}$, respectively, this completes the proof.

□

12.15 In the example presented in the proof, it is not only the minority student but also the majority students that are weakly worse off in $\tilde{\Gamma}$. In other words, this example shows that a stronger quota-based affirmative action constraint can induce a Pareto inferior matching (for all students).

The reason that a quota for majority students can have adverse effects on minority students is simple. Consider a situation in which a school c is mostly desired by majorities. Then having a majority quota for c decreases the number of majority students who can be assigned to c even if there are empty seats. This, in turn, increases the competition for other schools and thus can even make the minority students worse off.

12.16 The following example illustrate the case where the quota-based affirmative action constraint benefits everyone, including the majority students, under the student-optimal stable mechanism.

Consider the following market: $I = \{i_1, i_2, i_3, i_4\}$ with $I^M = \{i_1, i_2\}$ and $I^m = \{i_3, i_4\}$, $S = \{s_1, s_2\}$, $\boldsymbol{q}_{s_1} = (2,2)$, $\boldsymbol{q}_{s_2} = (1,1)$, preferences and priorities are as follows:

i_1	i_2	i_3	i_4	s_1	s_2
s_1	s_1	s_1	s_2	i_1	i_3
		s_2	s_1	i_4	i_4
				i_2	$\vdots$
				i_3	

Table 12.2

The unique stable matching is

$$\mu = \begin{bmatrix} s_1 & s_2 & \emptyset \\ i_1, i_4 & i_3 & i_2 \end{bmatrix}$$

Suppose that the capacity of s_1 is changed to $\tilde{\boldsymbol{q}}_{s_1} = (2,1)$, so that the new market has a stronger quota-based affirmative action policy than the original one. The student-optimal stable matching of this modified market is

$$\tilde{\mu} = \begin{bmatrix} s_1 & s_2 & \emptyset \\ i_1, i_3 & i_4 & i_2 \end{bmatrix}$$

Every student is weakly better off under $\tilde{\mu}$ than under μ: Students i_1 and i_2 are indifferent, whereas i_3 and i_4 are strictly better off.

12.17 There exist markets Γ and $\tilde{\Gamma}$ such that $\tilde{\Gamma}$ has a stronger quota-based affirmative action policy than Γ and, for any stable mechanism φ, matching $\varphi(\Gamma)$ is Pareto dominated by $\varphi(\tilde{\Gamma})$ for the minority.

Let φ be an arbitrary stable mechanism. Consider the following market: $I = \{i_1, i_2, i_3\}$, $I^M = \{i_1, i_2\}$, $I^m = \{i_3\}$, $S = \{s\}$, $\boldsymbol{q}_s = (q_s, q_s^M) = (2, 2)$, $P_s\colon i_1, i_2, i_3$, $s \succ_i \emptyset$ for any s. In this market there exists a unique stable matching

$$\mu = \begin{bmatrix} s & \emptyset \\ i_1, i_2 & i_3 \end{bmatrix}$$

Consider $\tilde{\boldsymbol{q}}_s = (2, 1)$. The new market imposes quota-based affirmative action on the original market. There is a unique stable matching

$$\tilde{\mu} = \begin{bmatrix} s & \emptyset \\ i_1, i_3 & i_2 \end{bmatrix}$$

Clearly, μ is Pareto dominated by $\tilde{\mu}$ for the minority.

✌ 12.18 Theorem (Theorem 3 in Kojima (2012)): The TTC mechanism does not respect the spirit of quota-based affirmative action.

☞ 12.19 Definition: Market $\tilde{\Gamma} = \langle I, S, (\boldsymbol{q}_s)_{s\in S}, (\tilde{P}_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$ is said to have a stronger priority-based affirmative action policy than $\Gamma = \langle I, S, (\boldsymbol{q}_s)_{s\in S}, (P_s)_{s\in S}, (\succ_i)_{i\in I}\rangle$ if, for every $s \in S$ and $i, i' \in I$,

$$iR_s i' \text{ and } i \in I^m \text{ implies } i\tilde{R}_s i'.$$

This policy is sometimes called preferential treatment and is based on a simple idea: A priority-based affirmative action policy promotes the ranking of a minority student at schools relative to majority students while keeping the relative ranking of each student within her own group fixed.

A matching mechanism φ is said to respect the spirit of priority-based affirmative action if there are no markets Γ and $\tilde{\Gamma}$ such that $\tilde{\Gamma}$ has a stronger quota-based affirmative action policy than Γ and $\varphi(\tilde{\Gamma})$ is Pareto dominated by $\varphi(\Gamma)$ for the minority.

✌ 12.20 Theorem (Theorem 2 in Kojima (2012)): There exists no stable mechanism that respects the spirit of priority-based affirmative action. In particular, deferred acceptance algorithm with majority quotas does not respect the spirit of quota-based affirmative action.

✌ 12.21 Theorem (Theorem 4 in Kojima (2012)): The TTC mechanism does not respect the spirit of priority-based affirmative action.

12.22 Theorem (Theorem 3 in Hafalir *et al.* (2013)): There exists no Pareto efficient and strongly group strategy-proof mechanism that is weakly preferred by all students to the top trading cycles algorithm with majority quotas.

12.3 Affirmative action policies with minority reserves

12.23 For each $s \in S$, let r_s^m be the type-specific capacity for minority students ($r_s^m \leq q_s$), which gives priority to minority students up to the reserve numbers.

Given $(r_s^m)_{s\in S}$, a matching μ is feasible under majority quotas if $|\mu^{-1}(s)\cap I^m|\geq r_s^m$ for all $s\in S$. This condition requires that the number of minority students matched to each school s is at least its type-specific capacity r_s^m.

Whenever we compare the effects of minority reserves $(r_s^m)_{s\in S}$ and majority quotas $(q_s^M)_{s\in S}$, we assume that $r_s^m+q_s^M=q_s$ for each $s\in S$.

☞ 12.24 Definition: Given $(r_s^m)_{s\in S}$, a matching μ is stable under minority reserves if

(1) $\mu(i)\succsim_i\emptyset$ for each $i\in I$, and

(2) if $s\succ_i\mu(i)$, then either

(i) $i\in I^m$, $|\mu^{-1}(s)|=q_s$ and $i'P_si$ for all $i'\in\mu^{-1}(s)$, or

(ii) $i\in I^M$, $|\mu^{-1}(s)\cap I^m|>r_s^m$, $|\mu^{-1}(s)|=q_s$ and $i'P_si$ for all $i'\in\mu^{-1}(s)$, or

(iii) $i\in I^M$, $|\mu^{-1}(s)\cap I^m|\leq r_s^m$, and $i'P_si$ for all $i'\in\mu^{-1}(s)\cap I^M$.

Condition (2-i) describes a situation where (i,s) does not form a blocking pair because i is a minority student and s prefers all students in s to i. In condition (2-ii), whereas blocking does not happen because i is a majority student, the number of minority students in s exceeds minority reserves and s prefers all students in s to i. Finally, in condition (2-iii), (i,s) does not form a blocking pair because i is a majority student, the number of minority students in s does not exceed minority reserves, and s prefers all majority students in s to i.

☞ 12.25 Definition: A mechanism φ is stable under minority reserves if $\varphi(\Gamma)$ is a stable matching under minority reserves in Γ for any given Γ.

12.26 Deferred acceptance algorithm with minority reserves:

Step 1: Start with the matching in which no student is matched. Each student i applies to her first-choice school. Each school s first accepts as many as r_s^m minority applicants with the highest priorities if there are enough minority applicants. Then it accepts applicants with the highest priorities from the remaining applicants until its capacity is filled or the applicants are exhausted. The rest of the applicants, if any remain, are rejected by s.

Step k: Start with the tentative matching obtained at the end of step $k-1$. Each student i who got rejected at step $k-1$ applies to her next-choice school. Each school s considers the new applicants and students admitted tentatively at step $k-1$. Among these students, school s first accepts as many as r_s^m minority students with the highest priorities if there are enough minority students. Then it accepts students with the highest priorities from the remaining students. The rest of the students, if any remain, are rejected by s. If there are no rejections, then stop.

End: The algorithm terminates when no rejection occurs and the tentative matching at that step is finalized.

12.27 Proposition (Proposition 1 in Hafalir *et al.* (2013)): The student-proposing deferred acceptance algorithm with minority reserves produces a stable matching that assigns the best outcome among the set of stable matching outcomes for each student and is weakly group strategy-proof.

Proof idea

✌ 12.28 Theorem (Theorem 1 in Hafalir *et al.* (2013)): Consider majority quotas $(q_s^M)_{s\in S}$ and minority reserves $(r_s^m)_{s\in S}$ such that $r_s^m=q_s-q_s^M$ for each $s\in S$. Take a matching μ that is stable under majority quotas $(q_s^M)_{s\in S}$. Then either μ is stable under minority reserves $(r_s^m)_{s\in S}$ or there exists a matching that is stable under minority reserves $(r_s^m)_{s\in S}$ that Pareto dominates μ.

Improvement algorithm.

12.29 Theorem (Theorem 2 in Hafalir *et al.* (2013)): Consider minority reserves $(r_s^m)_{s\in S}$. Let μ^r and μ be the matchings produced by the student-proposing deferred acceptance algorithm with or without minority reserves $(r_s^m)_{s\in S}$, respectively, for a given preference profile. Then there exists at least one minority student i such that $\mu^r(i) \succsim_i \mu(i)$.

12.30 On very peculiar cases, such as the example below, imposing minority reserves can make some minorities worse off while leaving the rest indifferent.

Example 1 in Hafalir *et al.* (2013).

12.31 Theorem (Proposition 2 in Hafalir *et al.* (2013)):

12.32 Theorem (Proposition 3 in Hafalir *et al.* (2013)):

12.33 Theorem (Proposition 4 in Hafalir *et al.* (2013)):

12.34 Top trading cycles algorithm with minority reserves

12.35 Theorem (Proposition 5 in Hafalir *et al.* (2013)): The top trading cycles algorithm with minority reserves is Pareto efficient and strongly group strategy-proof.

12.36 Theorem (Theorem 4 in Hafalir *et al.* (2013)):

12.4 Controlled school choice

Echenique and Yenmez (2015)

Part IV

Kidney exchange

Chapter 13

Kidney exchange I

Contents

13.1 Background

13.1 Transplant is an important treatment of serious kidney diseases. Over 90,000 patients are on waiting lists for kidney in the US. In 2011, there were

- 11,043 transplants from diseased donors,
- 5,771 transplants from living donors, while
- 4,697 patients died while on the waiting list (and 2,466 others were removed because they were "too sick to transplant").

13.2 Buying and selling kidneys is illegal in the US as well as many other countries.

Section 301 of the National Organ Transplant Act states:

> It shall be unlawful for any person to knowingly acquire, receive or otherwise transfer any human organ for valuable consideration for use in human transplantation.

《人体器官移植条例》第三条：

> 任何组织或者个人不得以任何形式买卖人体器官，不得从事与买卖人体器官有关的活动。

Given that constraint, donation is the most important source of kidneys.

13.3 There are two sources of donation:

- Deceased donors: In the US and Europe a centralized priority mechanism is used for the allocation of deceased donor kidneys. The patients are ordered in a waiting list, and the first available donor kidney is given to the patient who best satisfies a metric based on the quality of the match, waiting time in the queue, age of the patient, and other medical and fairness criteria.

- Living donors: Living donors usually come from friends or relatives of a patient (because the monetary transaction is prohibited).

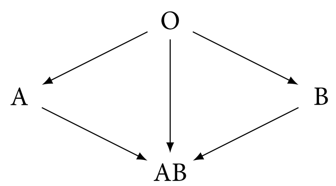
Live donation has been increasing recently.

Donor types	2008	1998	1988
All donors	10,920	9,761	5,693
Deceased donors	5,992	5,339	3,876
Live donors	4,928	4,422	1,817

Table 13.1: Number of donors by donor types. Data obtained at http://www.optn.org/.

13.4 For a successful transplant, the donor kidney needs to be compatible with the patient.

(1) Blood type compatibility: There are four blood types, O, A, B and AB.



- O type patients can receive kidneys from O type donors.
- A type patients can receive kidneys from O or A type donors.
- B type patients can receive kidneys from O or B type donors.
- AB type patients can receive kidneys from donors of any blood type (that is, O, A, B or AB).

(2) There is another compatibility issue around some proteins called HLA Tissue Compatibility.

13.5 A problem with transplant from live donors: transplant is carried out if the donor kidney is compatible with the patient. Otherwise the willing donor goes home and the patient cannot get transplant.

13.6 Question: Is there any way to increase the number and quality of transplant?

13.7 A paired exchange (aka paired conation) involves two incompatible patient-donor pairs such that the patient in each pair feasibly receives a transplant from the donor in the other pair. This pair of patients exchange donated kidneys. The number of pairs in a paired exchange can be larger than two.

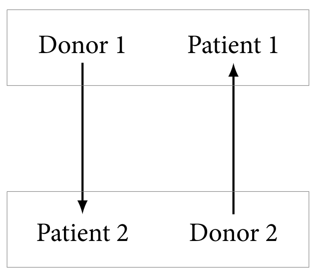


Figure 13.1: A paired exchange.

Take a look at the web page of Alliance for Paired Donation at http://paireddonation.org/.

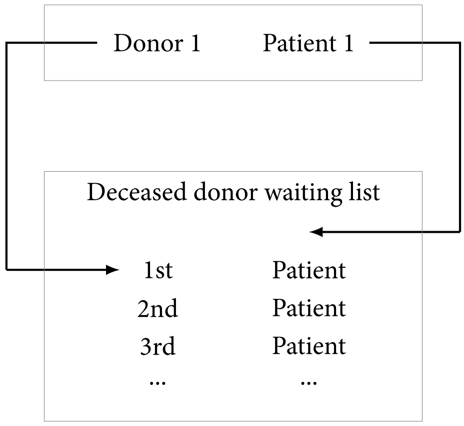


Figure 13.2: A list exchange.

13.8 A list exchange involves an exchange between one incompatible patient-donor pair and the deceased donor waiting list. The patient in the pair becomes the first priority person on the deceased donor waiting list in return for the donation of her donor's kidney to someone on the waiting list.

List exchanges can potentially harm O blood-type patients waiting on the deceased donor waiting list. Since the O blood type is the most common blood type, a patient with an incompatible donor is most likely to have O blood herself and a non-O bloodtype incompatible donor. Thus, after the list exchange, the blood type of the donor sent to the deceased donor waiting list has generally non-O blood, while the patient placed at the top of the list has O blood. Thus, list exchanges are deemed ethically controversial.

13.2 The model

13.9 Definition: A kidney exchange problem consists of:

- a set of donor kidney-transplant patient pairs $\{(k_1, t_1), \dots, (k_n, t_n)\}$,
- a set of compatible kidneys $K_i \subseteq K = \{k_1, \dots, k_n\}$ for each patient t_i, and
- a strict preference relation $\succ_i$ over $K_i \cup \{k_i, w\}$ where w refers to the priority in the waiting list in exchange for kidney k_i.

13.10 A matching is a function that specifies which patient obtains which kidney (or waiting list). We assume that the waiting list can be matched with any number of patients.

A kidney exchange mechanism is a systematic procedure to select a matching for each kidney exchange problem.

13.11 A matching is Pareto-efficient if there is no other matching that makes everybody weakly better off and at least one patient strictly better off.

A mechanism is Pareto-efficient if it always chooses Pareto-efficient matchings.

13.12 A matching is individually rational if each patient is matched with an option that is weakly better than her own paired-donor.

A mechanism is individually rational if it always selects an individually rational matching.

13.13 A mechanism is strategy-proof if it is always the best strategy for each patient to:

- reveal her preferences over other available kidneys truthfully, and

- declare the whole set of her donors (in case she has multiple donors) to the system without hiding any (the model treats each patient as having a single donor, but the extension to multiple donors is straightforward).

13.3 Multi-way kidney exchanges with strict preferences

13.14 In Roth *et al.* (2004)'s design the underlying assumptions are as follows:

- Any number of patient-donor pairs can participate in an exchange, *i.e.*, exchanges are possibly multi-way.
- Patients have heterogeneous preferences over compatible kidneys; in particular, no two kidneys have the same quality, *i.e.*, the preferences of a patient are strict and they linearly order compatible kidneys, the waiting list option, and her own paired-donor.
- List exchanges are allowed.

13.15 Under these assumptions, this model is very similar to the house allocation model with existing tenants. We will consider a class of mechanisms that clear through an iterative algorithm.

13.16 In each step,

- each patient t_i points either toward a kidney in $K_i \cup \{k_i\}$ or toward w, and
- each kidney k_i points to its paired recipient t_i.

13.17 A cycle is an ordered list of kidneys and patients $(k_1, t_1, k_2, t_2, \ldots, k_m, t_m)$ such that kidney k_1 points to a patient t_1, patient t_1 points to kidney k_2, . . . , kidney k_m points to patient t_m, and patient t_m points to kidney k_1.

13.18 Cycles larger than a single pair are associated with direct exchanges, very much like the paired-kidney-exchange programs, but may involve more than two pairs, so that patient t_1 is assigned kidney k_2, patient t_2 is assigned kidney k_3, . . . , patient t_m is assigned kidney k_1.

Note that each kidney or patient can be part of at most one cycle and thus no two cycles intersect.

13.19 A w-chain is an ordered list of kidneys and patients $(k_1, t_1, k_2, t_2, ..., k_m, t_m)$ such that kidney k_1 points to patient t_1, patient t_1 points to kidney k_2, . . . , kidney k_m points to patient t_m, and patient t_m points to w.

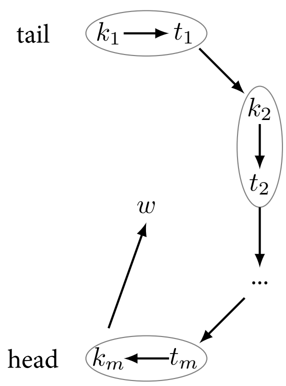


Figure 13.3: A w-chain.

We refer to the pair (k_m, t_m) whose patient receives a cadaver kidney in a w-chain as the head and the pair (k_1, t_1) whose donor donates to someone on the cadaver queue as the tail of the w-chain.

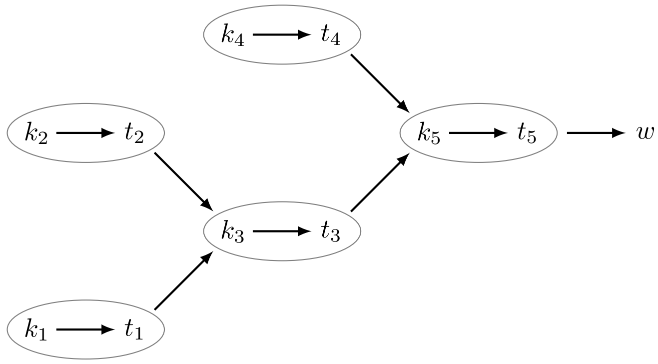


Figure 13.4: Five w-chains.

13.20 w-chains are associated with indirect exchanges but unlike in a cycle, a kidney or a patient can be part of several w-chains.

One practical possibility is choosing among w-chains with a well-defined chain selection rule, very much like the rules that establish priorities on the cadaveric waiting list.

- The current pilot indirect exchange programs in the United States choose the minimal w-chains, consisting of a single donor-recipient pair, but this may not be efficient.
- Selection of longer w-chains will benefit other patients as well, and therefore the choice of a chain selection rule has efficiency implications.
- Chain selection rules may also be used for specific policy objectives such as increasing the inflow of type O living donor kidneys to the cadaveric waiting list.

13.21 Lemma (Lemma 1 in Roth *et al.* (2004)): Consider a graph in which both the patient and the kidney of each pair are distinct nodes as is the wait-list option w. Suppose that each patient points either toward a kidney or w, and each kidney points to its paired recipient. Then either there exists a cycle, or each pair is the tail of some w-chain.

Proof. (1) Consider a graph where each patient points toward either a kidney or w, and each kidney points to its paired recipient.

(2) Suppose that there is no cycle.

(3) Consider an arbitrary pair (k_i, t_i). Start with kidney k_i, and follow the path in the graph.

(4) Since there are no cycles, no kidney or patient can be encountered twice. Hence by the finiteness of pairs, the path will terminate at w. This is the w-chain initiated by pair (k_i, t_i) completing the proof.

□

13.22 Fixed parameters: First, we take the operation of the cadaver queue as fixed. The cadaver queue can be thought of as a stochastic arrival process of cadavers and patients, interacting with a scoring rule that determines which patients are offered which cadaver kidneys.

We also take as fixed how patients whose donors donate a kidney to someone on the queue are given high priority on the queue, *e.g.*, by being given points in the scoring rule.

We also take as given the size of the live kidney exchange; *i.e.*, the set of patient-donor pairs is taken to be fixed.

13.23 For the mechanism defined below, we assume that when one among multiple w-chains must be selected, a fixed chain selection rule is invoked. We will consider a number of such rules, and their implications for incentives, efficiency, and equity.

Below we list a number of plausible chain selection rules:

(a) Choose minimal w-chains, and remove them.

(b) Choose the longest w-chain and remove it. If the longest w-chain is not unique, then use a tiebreaker to choose among them.

(c) Choose the longest w-chain and keep it. If the longest w-chain is not unique, then use a tiebreaker to choose among them.

(d) Prioritize patient-donor pairs in a single list. Choose the w-chain starting with the highest priority pair, and remove it.

(e) Prioritize patient-donor pairs in a single list. Choose the w-chain starting with the highest priority pair, and keep it.

(f) Prioritize the patient-donor pairs so that pairs with type O donor have higher priorities than those who do not. Choose the w-chain starting with the highest priority pair; remove it in case the pair has a type O donor, but keep it otherwise.

13.24 Throughout the procedure kidneys are assigned to patients through a series of exchanges. Some patients and their assigned kidneys will be immediately removed from the procedure, while others will remain with their assignments but they will assume a passive role. So at any point in the procedure, some agents may no longer be participants, some participants will be active, and the others passive.

13.25 For a given kidney exchange problem, the top trading cycles and chains (TTCC) mechanism determines the exchanges as follows.

Step 1: Initially all kidneys are available and all agents are active. At each stage of the procedure

- each remaining active patient t_i points to the best remaining unassigned kidney or to the waiting list option w, whichever is more preferred,
- each remaining passive patient continues to point to her assignment, and
- each remaining kidney k_i points to its paired patient t_i.

Step 2: By Lemma 13.21, there is either a cycle, or a w-chain, or both.

(a) Proceed to Step 3 if there are no cycles. Otherwise, locate each cycle, and carry out the corresponding exchange (*i.e.*, each patient in the cycle is assigned the kidney he is pointing to). Remove all patients in a cycle together with their assignments.

(b) Each remaining patient points to his top choice among remaining kidneys, and each kidney points to its paired recipient. Locate all cycles, carry out the corresponding exchanges, and remove them. Repeat until no cycle exists.

Step 3: If there are no pairs left, we are done. Otherwise, by Lemma 13.21, each remaining pair initiates a w-chain. Select only one of the chains with the chain selection rule. The assignment is final for the patients in the selected w-chain. In addition to selecting a w-chain, the chain selection rule also determines:

(a) whether the selected w-chain is removed, or

(b) the selected w-chain in the procedure although each patient in it is henceforth passive. If the w-chain is removed, then the tail kidney is assigned to a patient in the deceased donor waiting list. Otherwise, the tail kidney remains available in the problem for the remaining steps.

Step 4: Each time a w-chain is selected, a new series of cycles may form. Repeat Steps 2 and 3 with the remaining active patients and unassigned kidneys until no patient is left. If there exist some tail kidneys of w-chains remaining at this point, remove all such kidneys and assign them to the patients in the deceased-donor waiting list.

13.26 Example (Example 1 in Roth *et al.* (2004)): Consider a kidney exchange problem with 12 pairs as follows:

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
k_9	k_{11}	k_2	k_5	k_3	k_3	k_6	k_6	k_3	k_{11}	k_3	k_{11}
k_{10}	k_3	k_4	k_9	k_7	k_5	k_1	k_4	k_{11}	k_1	k_6	k_3
k_1	k_5	k_5	k_1	k_{11}	k_8	k_3	k_{11}	w	k_4	k_5	k_9
	k_6	k_6	k_8	k_4	k_6	k_9	k_2		k_5	k_{11}	k_8
	k_2	k_7	k_{10}	k_5		k_{10}	k_3		k_6		k_{10}
		k_8	k_3			k_1	k_8		k_7		k_{12}
		w	w			w			w		

Table 13.2

Suppose that patients are ordered in a priority-list based on their indices starting with the patient with the smallest index. We use the following chain selection rule: choose the longest w-chain. In case the longest w-chain is not unique, choose the w-chain with the highest priority patient; if the highest priority patient is part of more than one, choose the w-chain with the second highest priority patient, and so on. Keep the selected w-chains until the termination.

Round 1: There is a single cycle $C_1 = (k_{11}, t_{11}, k_3, t_3, k_2, t_2)$. Remove the cycle by assigning k_{11} to t_2, k_3 to t_{11}, and k_2 to t_3.

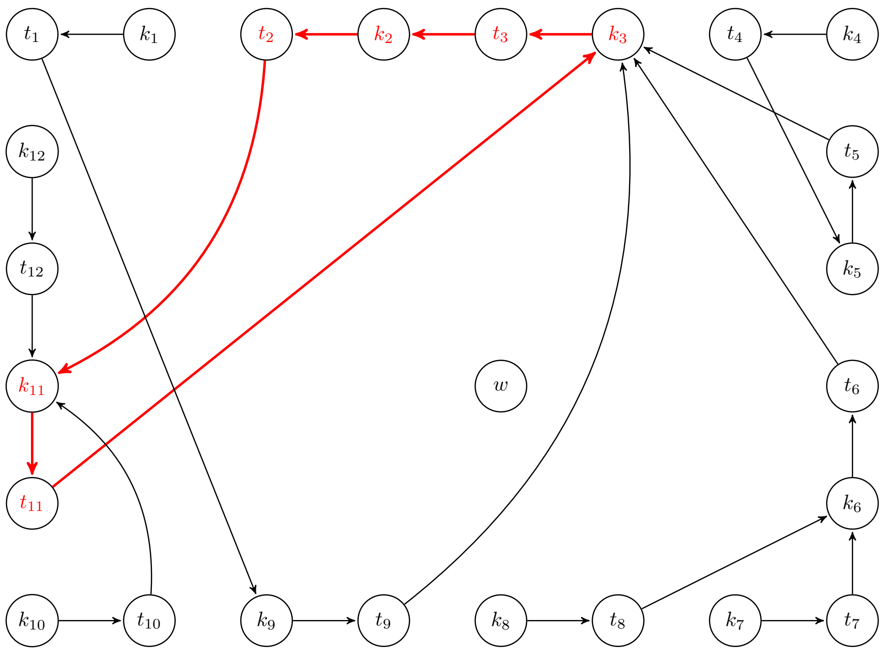


Figure 13.5: Round 1

Round 2: Upon removing cycle C_1, a new cycle $C_2 = (k_7, t_7, k_6, t_6, k_5, t_5)$. Remove it by assigning k_7 to t_5, k_6 to t_7, and k_5 to t_6.

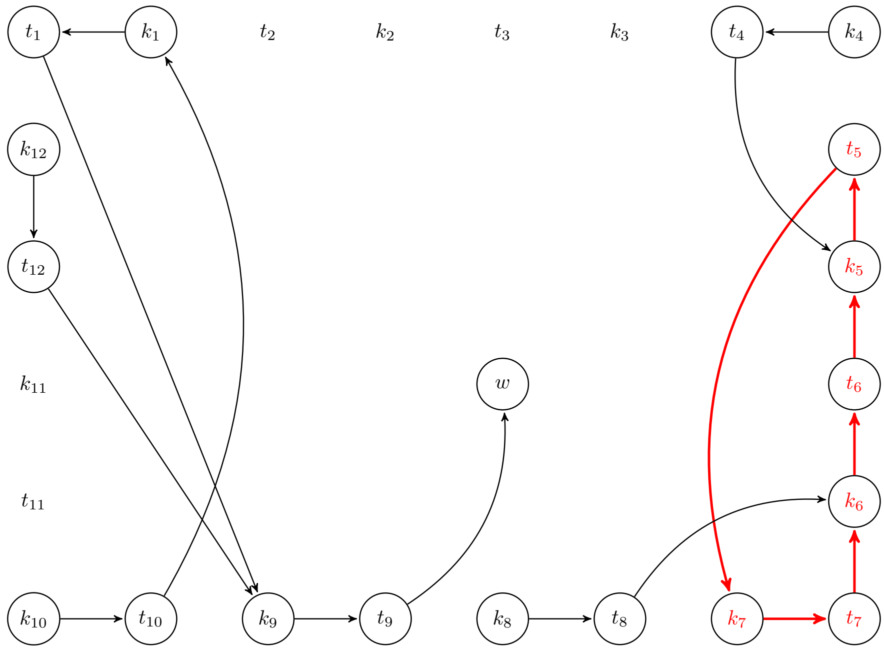


Figure 13.6: Round 2

Round 3: No new cycle forms, and hence each kidney-patient pair starts a w-chain. The longest w-chains are $W_1 = (k_8, t_8, k_4, t_4, k_9, t_9)$ and $W_2 = (k_{10}, t_{10}, k_1, t_1, k_9, t_9)$. Since t_1, the highest priority patient, is in W_2 but not in W_1, choose and fix W_2. Assign w to t_9, k_9 to t_1, and k_1 to t_{10} but do not remove them. Kidney k_{10}, the kidney at the tail of W_2, remains available for the next round.

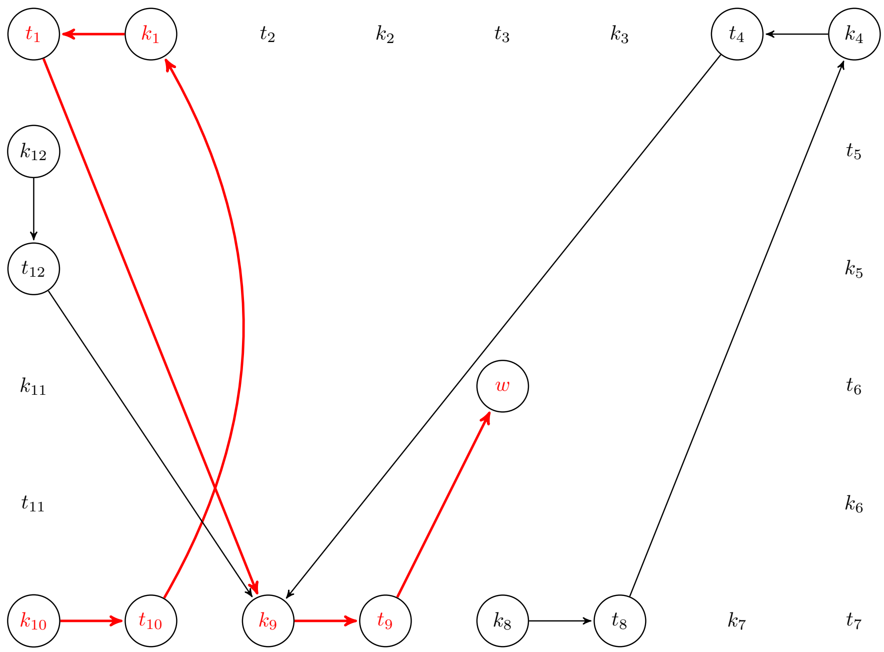


Figure 13.7: Round 3

Round 4: Upon fixing the w-chain W_2, a new cycle $C_3 = (k_4, t_4, k_8, t_8)$ forms. Remove it by assigning k_4 to t_8 and k_8 to t_4.

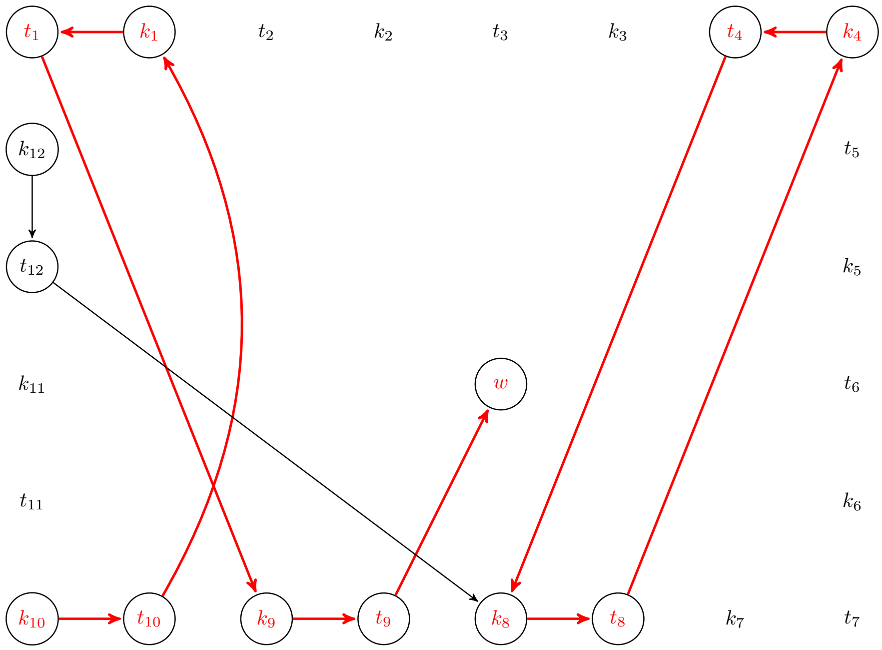


Figure 13.8: Round 4

Round 5: No new cycles form, and the pair (k_{12}, t_{12}) "joins" W_2 from its tail to form the longest w-chain $W_3 =$ ⊠$(k_{12}, t_{12}, k_{10}, t_{10}, k_1, t_1, k_9, t_9)$. Fix W_3, and assign k_{10} to t_{12}. Since no patient is left, w-chain W_3 is removed, and kidney k_{12} at its tail is offered to the highest priority patient at the cadaveric waiting list.

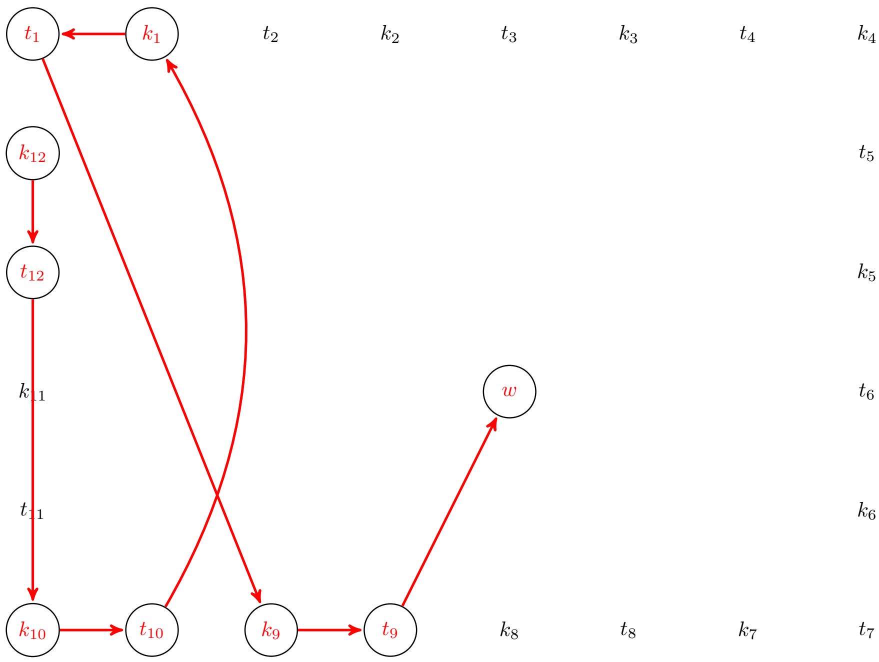


Figure 13.9: Round 5

✌ 13.27 Theorem (Theorem 1 in Roth *et al.* (2004)): Consider a chain selection rule such that any w-chain selected at a nonterminal round remains in the procedure, and thus the kidney at its tail remains available for the next round. The TTCC mechanism, implemented with any such chain selection rule, is efficient.

Proof. (1) Let the TTCC mechanism be implemented with a chain selection rule such that any w-chain selected at a nonterminal round remains in the procedure and the kidney at its tail remains available for the next round.

(2) Any patient whose assignment is finalized in Round 1 has received his top choice and cannot be made better off.

(3) Any patient whose assignment is finalized in Round 2 has received his top choice among the kidneys not already assigned as part of an exchange (since chains are not removed, so the kidney at their tail remains available), and cannot be made better off without hurting a patient whose assignment was finalized in Round 1.

(4) Proceeding in a similar way, no patient can be made better off without hurting a patient whose assignment is finalized in an earlier round.

(5) Therefore, TTCC mechanism selects a Pareto-efficient matching at any given time provided that w-chains are removed at the termination.

□

13.28 Consider a class of priority-based chain selection rules that covers rules (d), (e), and (f): each ordering of patient-donor pairs together with a fixed pair defines a chain selection rule, and it is given as follows:

(1) Order donor-patient pairs in a single priority list, and fix a pair (k_j, t_j).

(2) Whenever a w-chain is to be selected, select the w-chain starting with the highest priority pair (k_i, t_i), and remove the w-chain if the pair (k_i, t_i) has strictly higher priority than the fixed pair (k_j, t_j), and keep it until termination otherwise.

13.29 Lemma (Lemma 2 in Roth *et al.* (2004)): Consider the TTCC mechanism implemented with a priority-based chain selection rule. Fix the stated preferences of all patients except patient t_i at P⊠i. Suppose that in the algorithm the assignment of patient t_i is finalized at Round s under P_i and at Round s' under P_i'. Suppose that $s \leq s'$⊠. Then the remaining active patients and unassigned kidneys at the beginning of Round s are the same, whether patient t_i announces P_i or P_i'.

Proof. (1) Patient t_i fails to participate in a cycle or a selected w-chain prior to Round s under either preference.

(2) Therefore, at any round prior to Round s not only the highest priority active patient is the same, whether patient t_i announces P_i or P_i', but also the same cycles/w-chains form, and in case there are no cycles, the same w-chain is selected, whether patient t_i announces P_i or P_i'. Hence the remaining active patients and unassigned kidneys at the beginning of Round s are the same, whether patient t_i announces P_i or P_i'.

□

13.30 Theorem (Theorem 2 in Roth *et al.* (2004)): Consider the chain selection rules (a), (d), (e), and (f). The TTCC mechanism, implemented with any of these chain selection rules, is strategy-proof.

Among these four chain selection rules, the last two are especially appealing: Rule (e) yields an efficient and strategy-proof mechanism, whereas Rule (f) gives up efficiency in order to increase the inflow of type O kidneys to the cadaveric waiting list.

13.31 *Proof.* We first consider the chain selection rule (a).

(1) Recall that for each patient t_i, the relevant part of preference P_i is the ranking up to k_i or w, whichever is more preferred.

(2) Given the preference profile $(P_i)_{i=1}^n$, construct a new preference profile $(P_i')_{i=1}^n$ as follows:

- for each patient t_i with $k_i P_i w$, let $P'_i = P_i$,
- for each patient t_i with $w P_i k_i$, construct P'_i from P_i by swapping the ranking of k_i and w.

(3) Note that $k_i P'_i w$ for each patient t_i and because the relevant part of preferences are the more preferred of k_i and w, $\langle\{(k_i, t_i)\}i = 1^n, (P'_i)_{i=1}^n\rangle$, is a housing market.

(4) Let μ⊠ denote the outcome of the TTC mechanism for this housing market, and construct matching ν from matching μ as follows: if $P'_i \neq P_i$ and $\mu(t_i) = k_i$, then $\nu(t_i) = w$, otherwise, $\nu(t_i) = \mu(t_i)$.

(5) The key observation is that ν is the outcome of the TTCC mechanism when it is implemented with the minimal w-chain selecting chain selection rule.

(6) Therefore, by Theorem 4.24, a patient can never receive a more preferred kidney by a preference misrepresentation.

(7) He can receive the wait-list option w by a misrepresentation but cannot profit from it. That is because the TTCC mechanism never assigns a patient a kidney that is inferior to w. Hence TTCC is strategy-proof with this choice of chain selection rule.

Next consider any of the priority-based chain selection rules.

(1) Consider a patient t_i with true preferences P_i. Fix an announced preference profile P_{-i} for all other patients.

(2) We want to show that revealing his true preferences P_i is at least as good as announcing any other preferences P'_i under the TTCC mechanism.

(3) Let s and s' be the rounds at which patient t_i leaves the algorithm under P_i and P'_i, respectively.

(4) Case 1: $s < s'$.
   (i) By Lemma 13.29 the same kidneys remain in the algorithm at the beginning of Round s whether patient t_i announces P_i or P'_i.
   (ii) Moreover, patient t_i is assigned his top choice remaining at Round s under P_i.
   (iii) Therefore, his assignment under P_i is at least as good as his assignment under P'_i.

(5) Case 2: $s \geq s'$. After announcing P'_i, the assignment of patient t_i is finalized either by joining a cycle, or by joining a selected w-chain. We will consider the two cases separately.

(6) Case 2a: The assignment of patient t_i is finalized either by joining a cycle under P'_i.
   (i) Let $(k^1, t^1, k^2, \ldots, k^r, t_i)$ be the cycle patient t_i joins, and thus k^1 be the kidney he is assigned under P'_i.
   (ii) Next suppose that he reveals his true preferences P_i.
   (iii) Consider Round s'. By Lemma 13.29, the same active patients and available kidneys remain at the beginning of this round whether patient t_i announces P'_i or P_i.
   (iv) Therefore, at Round s', kidney k^1 points to patient t^1, patient t^1 points to kidney k^2, . . . , kidney k^r points to patient t_i.
   (v) Moreover, they keep on doing so as long as patient t_i remains.
   (vi) Since patient t_i truthfully points to his best remaining choice at each round, he either receives a kidney better than kidney k^1 or eventually points to kidney k^1, completes the formation of cycle $(k^1, t^1, k^2, \ldots, k^r, t_i)$, and gets assigned kidney k^1.

(7) Case 2b: The assignment of patient t_i is finalized by joining a selected w-chain under P'_i.
   (i) Let $(k^1, t^1, k^2, \ldots, k^r, t_i = t^r, k^{r+1}, \ldots, k^{r+m}, t^{r+m})$ be the selected w-chain patient t_i joins, where $r \geq 1$ and $m \geq 0$, under P'_i.
   (ii) Therefore, under P'_i, patient t_i is assigned the kidney k^{r+1} if $m \geq 1$, and the wait-list option w if $m = 0$.

(iii) Also note that, given the considered class of priority-based chain selection rules, pair (k^1, t^1) is the highest priority pair in Round s'.

(iv) Next suppose that patient t_i reveals his true preferences P_i.

(v) Consider Round s'. By Lemma 13.29, the same active patients and available kidneys remain at the beginning of this round whether patient t_i announces P_i' or P_i.

(vi) We will complete the proof by showing that, upon announcing his truthful preferences P_i, the assignment of patient t_i is finalized in Round s' and thus he is assigned his top choice available at the beginning of Round s.

(vii) Recall that for this case there is no cycle in Round s' when patient t_i announces P_i'.

(viii) Therefore, when he announces his true preferences P_i, either there is no cycle in Round s' or there is one cycle that includes him.

(ix) If it is the latter, then his assignment is finalized in Round s', and we are done.

(x) Otherwise, each pair initiates a w-chain by Lemma 13.21, and one of these w-chains has to be selected.

(xi) By the choice of a priority-based chain selection rule, this will be the w-chain that starts with the highest priority pair (k_1, t_1).

(xii) But the path starting with kidney k_1 passes through patient t_i and therefore the selected w-chain includes patient t_i.

(xiii) Hence in this case as well his assignment is finalized in Round s' completing the proof.

□

13.32 Example (Example 2 in Roth *et al.* (2004)): Strategy-proofness of TTCC is lost if one adopts a chain selection rule that chooses among the longest w-chains.

Consider the problem in Example 13.26, but suppose that patient t_4 misrepresents his preferences as P_4': $k_5, k_1, k_9, \ldots$ improving the ranking of kidney k_1. While Round 1 and Round 2 remain as in Example 13.26, Round 3 changes, and this time the longest w-chain at Round 3 is ⊠$W_4 = (k_8, t_8, k_4, t_4, k_1, t_1, k_9, t_9)$. Therefore, patient t_4 is assigned kidney k_1 instead of kidney k_8, making his preference misrepresentation profitable.

13.33 Proposition (Proposition 1 in Krishna and Wang (2007)): The TTCC algorithm induced by chain selection rule (e) is equivalent to the YRMH-IGYT algorithm.

13.34 Recall Theorem 6.34: A mechanism is Pareto efficient, individually rational, strategy-proof, weakly neutral, and consistent if and only if it is a YRMH-IGYT mechanism.

13.35 Simulation result

Chapter 14

Kidney exchange II

Contents

14.1 Two-way kidney exchanges with 0–1 preferences

14.2 Multi-way kidney exchanges with 0–1 preferences

14.3 Dynamic kidney exchange

14.4 Lung exchange

Part V

Matching with transfers

Bibliography

Atila Abdulkadiroğlu and Yeon-Koo Che, The role of priorities in assigning indivisible objects: A characterization of top trading cycles, working paper, 2010.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda, Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered, *American Economic Review* **101** (2011), 1–14.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda, Expanding "choice" in school choice, *American Economic Journal: Microeconomics* **7(1)** (2015), 1–42. 160

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth, Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match, *American Economic Review* **99:5** (2009), 1954–1978. 133, 134, 162

Atila Abdulkadiroğlu and Tayfun Sönmez, Random serial dictatorship and the core from random endowments in house allocation problems, *Econometrica* **66:3** (1998), 689–701. 70, 72, 79, 105, 110

Atila Abdulkadiroğlu and Tayfun Sönmez, House allocation with existing tenants, *Journal of Economic Theory* **88** (1999), 233–260. 81, 88, 89, 98

Atila Abdulkadiroğlu and Tayfun Sönmez, School choice: a mechanism design approach, *American Economic Review* **93:3** (2003), 729–747. 134, 165

Atila Abdulkadiroğlu and Tayfun Sönmez, Ordinal efficiency and dominated sets of assignments, *Journal of Economic Theory* **112:1** (2003), 157–172.

Azar Abizada, Stability and incentives for college admissions with budget constraints, *Theoretical Economics*, forthcoming.

Mustafa Oğuz Afacan, Group robust stability in matching markets, *Games and Economic Behavior* **74** (2012), 394–398. 145

Nikhil Agarwal, An empirical model of the medical match, *American Economic Review* **105:7** (2015), 1939–1978.

Jóse Alcalde and Salvador Barberà, Top dominance and the possibility of strategy-proof stable solutions to matching problems, *Economic Theory* **4** (1994), 417–435. 27, 44, 132

Itai Ashlagi, Mark Braverman and Avinatan Hassidim, Stability in large matching markets with complementarities, *Operation Research* **62:4** (2014), 713–732.

Sophie Bade, Serial dictatorship: the unique optimal allocation rule when information is endogenous, *Theoretical Economics*, forthcoming.

Michel Balinski and Tayfun Sönmez, A tale of two mechanisms: student placement, *Journal of Economic Theory* **84** (1999), 73–94. 128

Charles G. Bird, Group incentive compatibility in a market with indivisible goods, *Economics Letters* **14** (1984), 309–313. 63

Anna Bogomolnaia and Hervé Moulin, A new solution to the random assignment problem, *Journal of Economic Theory* **100** (2001), 295–328. 108, 110, 111, 112, 114, 116, 120, 122

Eric Budish, The combinatorial assignment problem: approximate competitive equilibrium from equal incomes, *Journal of Political Economy* **119** (2011), 1061–1103.

Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom, Designing random allocation mechanisms: Theory and applications, *American Economic Review* **103:2** (2013), 585–623.

Gabriel Carroll, A general equivalence theorem for allocation of indivisible objects, *Journal of Mathematical Economics* **51** (2014), 163–177.

Yeon-Koo Che and Youngwoo Koh, Decentralized college admissions, *Journal of Political Economy*, forthcoming.

Yeon-Koo Che and Fuhito Kojima, Asymptotic equivalence of probabilistic serial and random priority mechanisms, *Econometrica* **78:5** (2010), 1625–1672. 123

Yan Chen and Tayfun Sönmez, Improving efficiency of on-campus housing: An experimental study, *American Economic Review* **92:5** (2002), 1669–1686.

Yan Chen and Onur Kesten, Chinese college admissions and school choice reforms: Theory and experiments, working paper.

Yan Chen and Tayfun Sönmez, School choice: An experimental study, *Journal of Economic Theory* **127** (2006), 202–231.

Gabrielle Demange, David Gale and Marilda Sotomayor, A further note on the stable matching problem, *Discrete Applied Mathematics* **16** (1987), 217–222. 17, 30

Lester E. Dubins and David A. Freedman, Machiavelli and the Gale-Shapley algorithm, *American Mathematical Monthly* **88:7** (1981), 485–494. 29, 30

Umut Mert Dur and M. Utku Ünver, Two-sided matching via balanced exchange: Tuition and worker exchanges, working paper (r&r *Journal of Political Economy*)

Federico Echenique and Juan Sebastiàn Pereyra, Strategic complementarities and unraveling in matching markets, *Theoretical Economics*, forthcoming. 169

Federico Echenique and M. Bumin Yenmez, How to control controlled school choice, *American Economic Review*, **105:8** (2015), 2679–2694. 169

Lars Ehlers and Aytek Erdil, Efficient assignment respecting priorities, *Journal of Economic Theory* **145** (2010), 1269–1282.

Aytek Erdil, Strategy-proof stochastic assignment, *Journal of Economic Theory* **151** (2014), 146–162. 133

Aytek Erdil and Haluk Ergin, What's the matter with tie-breaking? Improving efficiency in school choice, *American Economic Review* **98** (2008), 669–689. 157, 160

Haluk Ergin, Consistency in house allocation problems, *Journal of Mathematical Economics* **34** (2000), 77–97. 79

Haluk Ergin, Efficient resource allocation on the basis of priorities, *Econometrica* **70:6** (2002), 2489–2497. 137, 138, 143

Haluk Ergin and Tayfun Sönmez, Games of school choice under the Boston mechanism, *Journal of Public Economics* **90** (2006), 215–237. 135

Haluk Ergin, Tayfun Sönmez and M. Utku Ünver, Lung exchange, working paper, 2014.

David Gale and Lloyd Shapley, College admissions and the stability of marriage, *American Mathematical Monthly* **69** (1962:1), 9–15. 11, 13, 37

David Gale and Marilda Sotomayor, Some remarks on the stable matching problem, *Discrete Applied Mathematics* **11** (1985), 223–232. 15, 17, 21, 22, 23

Dan Gusfield and Robert W. Irving, *The Stable Marriage Problem: Structure and Algorithms*, Cambridge, MA: MIT Press, 1989. 151

Guillaume Haeringer and Flip Klijn, Constrained school choice, *Journal of Economic Theory* **144** (2009), 1921–1947.

Isa E. Hafalir, M. Bumin Yenmez and Muhammed A. Yildirim, Effective affirmative action in school choice, *Theoretical Economics* **8** (2013), 325–363. 167, 168, 169

Rustamdjan Hakimov and Onur Kesten, The equitable top trading cycles mechanism for school choice, working paper, 2014. 135

Tadashi Hashimoto, Daisuke Hirata, Onur Kesten, Morimitsu Kurino, and M. Utku Ünver, Two axiomatic approaches to the probabilistic serial mechanism, *Theoretical Economics* **9:1** (2014), 253–277.

J. S. Hwang, Modelling on college admissions in terms of stable marriages, working paper, Academia Sinica. 17

Aanund Hylland and Richard Zeckhauser, The efficient allocation of individuals to positions, *Journal of Political Economy* **87:2** (1979), 293–314. 69

Paula Jaramillo and Vikram Manjunath, The difference indifference makes in strategy-proof allocation of objects, *Journal of Economic Theory* **147:5** (2012), 1913–1946. 68

Onur Kesten, On two competing mechanisms for priority-based allocation problems, *Journal of Economic Theory* **127** (2006), 155–171. 145

Onur Kesten, Why do popular mechanisms lack efficiency in random environments? *Journal of Economic Theory* **144** (2009), 2209–2226.

Onur Kesten, School choice with consent, *The Quarterly Journal of Economics* **125** (2010), 1297–1348. 133, 148, 150, 153, 161

Onur Kesten, On two kinds of manipulation for school choice problems, *Economic Theory* **51** (2012), 677–693. 47

Onur Kesten and Morimitsu Kurino, Do outside options matter in school choice? A new perspective on the efficiency vs. strategy-proofness trade-off, working paper, Carnegie Mellon University, 2013.

Bettina Klaus and Flip Klijn, Stable matchings and preferences of couples, *Journal of Economic Theory* **121:1** (2005), 75–106.

Bettina Klaus, Flip Klijn and Toshifumi Nakamura, Corrigendum to "Stable matchings and preferences of couples", *Journal of Economic Theory* **144:5** (2009), 2227–2233.

Fuhito Kojima, Impossibility of stable and nonbossy matching mechanisms, *Economics Letters* **107** (2010), 69–70. 31

Fuhito Kojima, Robust stability in matching markets, *Theoretical Economics* **6** (2011), 257–267. 143, 144

Fuhito Kojima, School choice: Impossibilities for affirmative action, *Games and Economic Behavior* **75** (2012), 685–693. 165, 167

Fuhito Kojima and Mihai Manea, Incentives in the probabilistic serial mechanism, *Journal of Economic Theory* **145** (2010), 106–123.

Fuhito Kojima and Mihai Manea, Axioms for deferred acceptance, *Econometrica* **78** (2010), 633–653. 123

Fuhito Kojima and Parag Pathak, Incentives and stability in large two-sided matching markets, *American Economic Review* **99** (2008), 608–627.

Fuhito Kojima, Parag Pathak and Alvin E. Roth, Matching with couples: Stability and incentives in large markets, *Quarterly Journal of Economics* **128:4** (2013), 1585–1632.

Scott Duke Kominers and Tayfun Sönmez, Matching with slot-specific priorities: theory, *Theoretical Economics*, forthcomeing.

Hideo Konishi and M. Utku Ünver, Games of capacity manipulation in hospital-intern markets, *Social Choice and Welfare* **27** (2006), 3–24. 47

Donald Knuth, Marriages stables. 14

Aradhna Krishna and Yu Wang, The relationship between top trading cycles mechanism and top trading cycles and chains mechanism, *Journal of Economic Theory* **132:1** (2007), 539–547. 184

Jinpeng Ma, Strategy-proofness and the strict core in a market with indivisibilities, *International Journal of Game Theory* **23** (1994), 75–83. 63, 64, 65

, Mihai Manea, Asymptotic ordinal inefficiency of random serial dictatorship, *Theoretical Economics* **4:2** (2009), 165–197.

Andrew McLennan, Ordinal efficiency and the polyhedral separating hyperplane theorem, *Journal of Economic Theory* **105** (2002), 435–449. 110

D. G. McVitie and L.B.Wilson, Stable marriage assignment for unequal sets, *BIT Numerial Mathematics* **10** (1970), 295–309. 15

Eiichi Miyagawa, Strategy-Proofness and the Core in House Allocation Problems, *Games and Economic Behavior* **38:2** (2002), 347–361.

Yusuke Narita, Efficient resource allocation on the basis of priority: a comment, working paper, Massachusetts Institute of Technology, 2009. 140

Szilvia Pápai, Strategyproof assignment by hierarchical exchange, *Econometrica* **68** (2000), 1403–1433. 101, 129

Parag Pathak and Jay Sethuraman, Lotteries in student assignment: An equivalence result, *Theoretical Economics* **6** (2011), 1–17. 105

arag Pathak and Tayfun Sönmez (2008), Leveling the playing field: Sincere and sophisticated players in the Boston mechanism, *American Economic Review* **98:4** (2013), 1636–1652.

arag Pathak and Tayfun Sönmez (2013), School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation, *American Economic Review* **103:1** (2013), 80–106.

William R. Pulleyblank, Matching and extensions, in *Handbook of Combinatorics, Volume 1* (Ronald L. Graham, Martin Grötschel and László Lovász Eds.), 179–232, MIT Press, Cambridge, Massachusetts, 1995. 104

Marek Pycia and M. Utku Ünver, Incentive compatible allocation and exchange of discrete resources, working paper, 2015. 101

Alvin E. Roth, Incentive compatibility in a market with indivisible goods, *Economics Letters* **9** (1982), 127–132. 61

Alvin E. Roth, The economics of matching: stability and incentives, *Mathematics of Operations Research* **7:4** (1982), 671–628. 16, 26, 29, 44

Alvin E. Roth, The college admissions problem is not equivalent to the marriage problem, *Journal of Economic Theory* **36** (1985), 277–288. 45

Alvin E. Roth, Common and conflicting interests in two-sided matching markets, *European Economic Review* **27** (1985), 75–96. 43

Alvin E. Roth, On the allocation of residents to rural hospitals: a general property of two-sided matching markets, *Econometrica* **54:2** (1986), 425–427. 39

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

Alvin E. Roth, What have we learned from market design?, *Economic Journal* **118** (2008), 285–310.

Alvin E. Roth, Deferred acceptance algorithms: history, theory, practice, and open questions, *International Journal of Game Theory* **36** (2008), 537–569. 12

Alvin E. Roth and Elliott Peranson, The redesign of the matching market for American physicians: Some engineering aspects of economic design, *American Economic Review* **89:4** (1999), 748–780.

Alvin E. Roth and Andrew Postlewaite, Weak versus strong domination in a market with indivisible goods, *Journal of Mathematical Economics* **4** (1977), 131–137. 59, 60

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver, Kidney exchange, *The Quarterly Journal of Economics* **119:2** (2004), 457–488. 176, 177, 179, 181, 182, 184

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver, Pairwise kidney exchange, *Journal of Economic Theory* **125** (2005), 151–188.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver, Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences, *American Economic Review* **97:3** (2007), 828–851.

Alvin E. Roth and Marilda Sotomayor, The college admissions problem revisited, *Econometrica* **57:3** (1989), 559–570. 36, 39, 40

Mark A. Satterthwaite and Hugo Sonnenschein, Strategy-proof allocation mechanisms at differentiable points, *Review of Economic Studies* **48** (1981), 587–597. 31

Lloyd Shapley and Herbert Scarf, On cores and indivisibility, *Journal of Mathematical Economics* **1** (1974), 23–37. 51, 53

Lloyd Shapley and Martin Shubik, The assignment game I: the core, *International Journal of Game Theory* **1** (1972), 111–130.

Tayfun Sönmez, Manipulation via capacities in two-sided matching markets, *Journal of Economic Theory* **77** (1997), 197–204. 46

Tayfun Sönmez, Can pre-arranged matches be avoided in two-sided matching markets?, *Journal of Economic Theory* **86:1** (1999), 148–156.

Tayfun Sönmez and M. Utku Ünver, House allocation with existing tenants: an equivalence, *Games and Economic Behavior* **52** (2005), 153–185. 100

Tayfun Sönmez and M. Utku Ünver, House allocation with existing tenants: a characterization, *Games and Economic Behavior* **69** (2010), 425–445. 99

Lars-Gunnar Svensson, Queue allocation of indivisible goods, *Social Choice and Welfare* **11** (1994), 223–230.

Lars-Gunnar Svensson, Strategy-proof allocation of indivisible goods, *Social Choice and Welfare* **19** (1999), 557–567. 76, 77

Qianfeng Tang and Jingsheng Yu, A new perspective on Kesten's school choice withconsent idea, *Journal of Economic Theory* **154** (2014), 543–561. 154, 155, 156, 161

M. Utku Ünver, Dynamic kidney exchange, *Review of Economic Studies* 77 (2010), 372–414.

Özgür Yilmaz, The probabilistic serial mechanism with private endowments, *Games and Economic Behavior* **69:2** (201), 475–491.

Lin Zhou, On a conjecture by Gale about one-sided matching problems, *Journal of Economic Theory* **52** (1990), 123–135. 122