Matching and Market Design

Theory and Practice

Xiang Sun

February 25, 2018

Contents

Acknowledgement

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

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`.
- Summer Institute 2016 Methods Lectures, NBER. Available at `http://www.nber.org/econometrics_minicourse_2016/`

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), Jun Cheng (成军), Yongkang Hu (胡永康), Liucong Huang (黄柳淙), Yikuan Huang (黄怡宽), Zhijiang Huang (黄志疆), Cong Jiang (蒋璁), Yiwen Jin (金怡文), Jia'en Li (李佳恩), Qin Li (李擎), Xiao Lin (林潇), Haomin Liu (刘昊旻), Yuting Liu (刘雨婷), Jie Lu (陆劼), Xiaojie Mao (毛小介), Keji Mei (梅珂佶), Tianchen Song (宋天辰), Xunkang Tian (田珣康), Rui Wang (汪瑞), Xiao Wang (王筱), Ya Wen (文雅), Kaiwen Wu (吴锴文), 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 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.
- Because school seats are limited (for popular schools), school districts should decide who is admitted.
- How should school districts decide placements of students in schools?

1.6 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?

1.7 Targets:

- Efficiency: Pareto efficiency, individual optimality, ordinal efficiency, *ex ante* efficiency, *ex post* efficiency, *etc.*
- Fairness: stability, anonymity, envy-freeness, equal treatment of equals, *etc.*
- Incentives: strategy-proofness, nonbossiness, *etc.*
- Easy for participants to understand and use.

1.8 Reading:

- Information for the Public: Stable matching: Theory, evidence, and practical design.
- Scientific Background: Stable allocations and the practice of market design.
- Roth (2015).
- Sakai (2013).

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

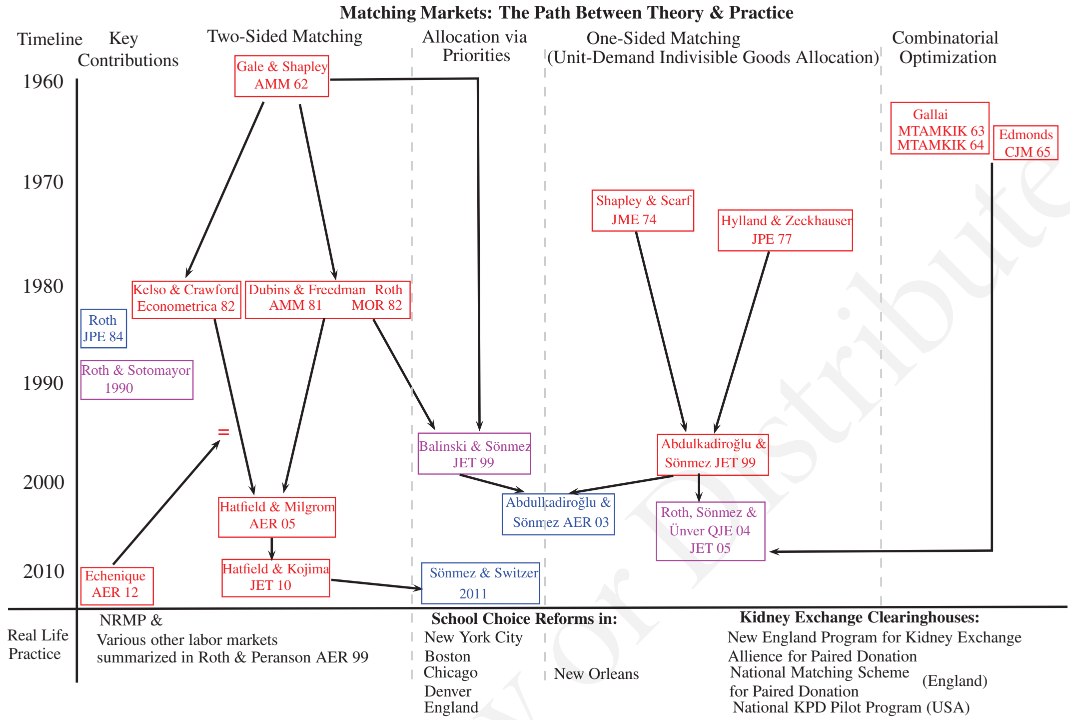


Figure 1.1: Overview (Taken from Sönmez's lecture notes).

Two-sided matching

1.9 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.

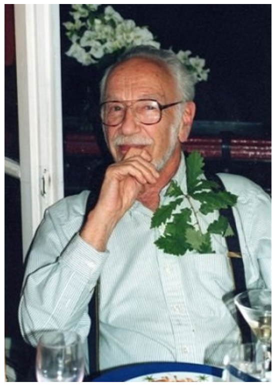

(b) David Gale.

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.

1.10 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

1.11 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.

1.12 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.

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.

1.13 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

1.14 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.

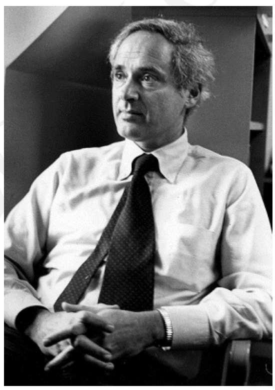


Figure 1.6: Herbert Scarf.

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.

In 1994, Jinpeng Ma provided an axiomatic characterization (known by MA's characterization) of top trading cycles algorithm.

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

The TTC algorithm is the only mechanism that satisfies individual rationality, Pareto efficiency and strategy-proofness for the classic Shapley-Scarf model. This makes the TTC a natural choice for other related situations.

1.15 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

1.16 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

1.17 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.

1.18 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 (婚姻问题) 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 the 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 In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, a matching (配对) is a outcome, and is defined by 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 $\mu(w) = m$ (*i.e.*, a matching is mutual: you are matched with me if and only if I am matched with you).

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 fixed 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.

The relation $\succsim_M$ gives a partial order on the set of stable matchings; see 2.37.

☞ 2.7 A matching μ is Pareto efficient[1] (帕累托有效) 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[2] (个人理性) 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 (稳定) if it is not blocked by any individual or any pair.

Roughly speaking, a matching is stable if there are no individuals or pairs of individuals who can profitably deviate from it.

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.

[1] In general, Pareto efficiency or Pareto optimality is a state of allocation of resources from which it is impossible to reallocate so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse off.

[2] In general, individual rationality constraints are said to be satisfied if a mechanism leaves all participants at least as well off as they would have been if they hadn't participated. They are also called participation constraints or rational participation constraints.

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) Case 1: If $\nu(i_0) = i_0$, then μ is blocked by the individual i_0. Contradiction.

(3) Case 2: 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.

$\square$

$$
\begin{tikzcd}
\nu(w) = m \arrow[r, "\nu"] & w = \nu(m) \arrow[d, "\mu"] \\
& \mu(w)
\end{tikzcd}
$$

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

2.14 Question: Does a stable matching always exists? How to get a stable matching?

2.3 Deferred acceptance algorithm

2.15 Men-proposing deferred acceptance algorithm.

Step 1: (a) Each man m proposes to his first choice (if he has any acceptable choices).

(b) 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: (a) 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).

(b) 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.

2.16 Question: Why do we call this algorithm the "deferred acceptance" algorithm? Hint: Compare it with the Boston mechanism 8.21.

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:

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

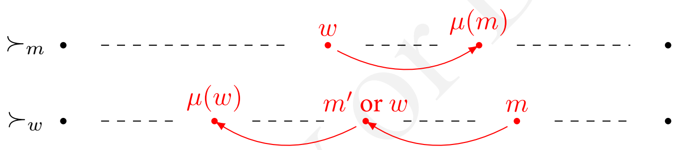
2.18 Observation: As the algorithm proceeds, the tentative partners of a man is weakening, and the tentative partners of a woman is improving.

2.19 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 (m' in the figure) 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 $m \succ_w \mu(w)$.

□

2.20 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.

2.21 Theorem on optimality (Theorem 2 in Gale and Shapley (1962)): The matching determined by men-proposing deferred acceptance algorithm is at least as 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.

(1) For contradiction, suppose that a man is rejected by an achievable woman.

(2) Consider the first step (say Step k) in which a man (call him m) is rejected by an achievable woman (call her w).

(3) Then w keeps some other man m' at this step, so $m' \succ_w m$.

(4) Let μ be a stable matching where $\mu(m) = w$.

(5) Since this is the first step of DA where a man is rejected by an achievable woman, $w \succ_{m'} \mu(m')$. Otherwise,

- Case 1: $\mu(m') \succ_{m'} w$, then m' is rejected by an achievable woman $\mu(m')$ before Step k.
- Case 2: $\mu(m') = w = \mu(m)$, which leads to $m = m'$. Contradiction.

(6) Thus, (m', w) blocks μ, contradicting the stability of μ.

□

2.22 Remark: Theorem 2.21 says that different stable matchings may benefit different participants. In particular, each version of deferred acceptance algorithm favors one side at the expense of the other side.

2.23 Remark: Intuitively, men may have different (individually) optimal matchings, since they have different preferences. However, restricting to the set of stable matchings, the stable matching resulting from men-proposing deferred acceptance algorithm is optimal for every man.

2.24 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.

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

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

It turns out that the stable matchings are not unique.

2.26 If some individuals may be indifferent between possible mates, *i.e.*, some individuals' preferences is not strict, Theorem 2.21 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 I

✌ 2.27 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 inequality 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 inequality 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.

$\square$

2.28 Remark: Decomposition theorem (Theorem 2.27) 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:

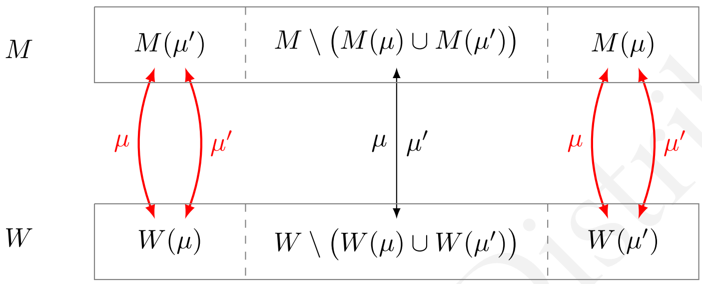


Figure 2.1: Decomposition theorem

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

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 $\mu \succ_W \mu'$.

$\square$

2.30 Corollary: When all the 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.

2.31 Rural hospital theorem[3] (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 that m is matched under μ' but not under μ. Then $m \in M(\mu')$.

(2) By decomposition theorem (Theorem 2.27), μ maps $M(\mu')$ to $W(\mu)$.

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

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

□

2.32 Direct proof:

Proof. (1) Let μ^M be the man-optimal stable matching and μ be an arbitrary stable matching.

(2) Since μ^M is man-optimal, all the men that are matched in μ are matched in μ^M.

(3) Since μ^M is woman-pessimal, all the women that are matched in μ^M are matched in μ (why?).

(4) But for any given matching, the number of matched men and women are the same to each other (why?).

(5) So the same set of men and women are matched in μ^M and μ (exercise: complete the argument).

□

For an alternative proof, see Ciupan, Hatfield and Kominers (2016).

2.33 Remark: One motivation is the allocation of residents in rural hospitals. Hospitals in rural areas cannot fill positions for residents, and some people argue that the matching mechanisms should be changed so that more doctors end up in rural hospitals. But the theorem says that it is impossible as long as stable matchings are implemented.

If some men were matched in some stable matching and not in others, the latter may be unfair to them. The theorem says that there is no need to worry.

2.34 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.

2.35 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.

Exercise: Provide several examples (as simple as possible) to illustrate the points above.

2.36 Lattice theorem (Conway): When all the 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 statement 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 2.27), 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.

(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 $w \succ_m \lambda(m) = \mu(m) \succsim_m \mu'(m)$.

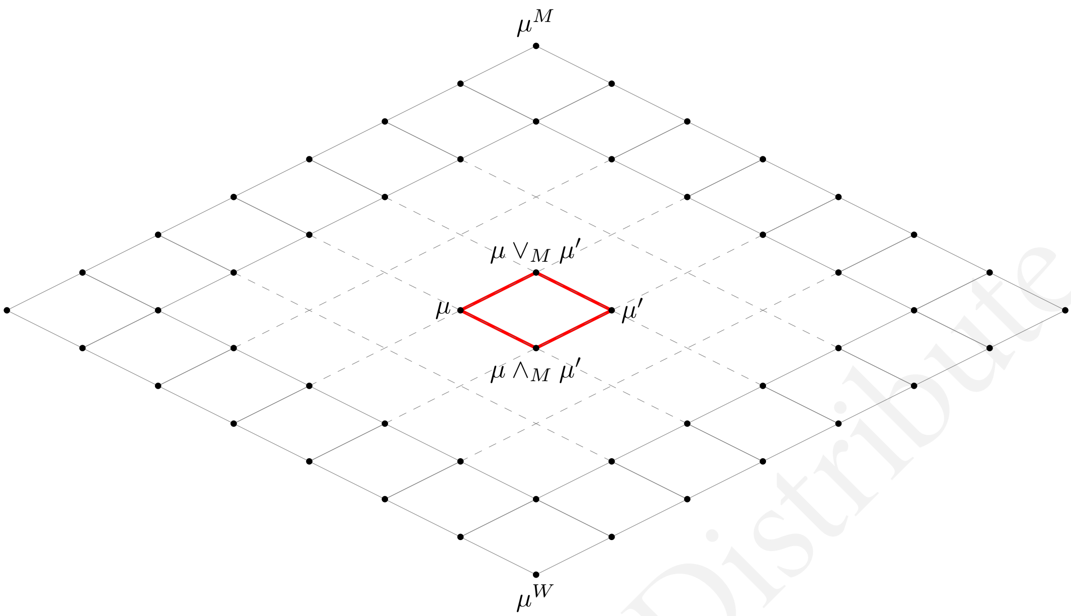
- If $w \in W(\mu)$, then $m \succ_w \lambda(w) = \mu'(w)$, and hence μ' is blocked by (m, w).
- If $w \in W \setminus W(\mu)$, then $m \succ_w \lambda(w) = \mu(w)$, and hence μ is blocked by (m, w).

(7) Therefore, λ is a stable matching.

□

2.37 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 2.36) implies that the set of stable matchings is a lattice under $\succsim_M$ (defined in 2.6), dual to $\succsim_W$.



2.38 To compute all the stable matchings, see McVitie and Wilson (1971), Irving and Leather (1986) and Section 3.2 of Roth and Sotomayor (1989).

2.39 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 that 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; otherwise her waiting list is full, and some man is rejected at the last step.

(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 $\mu^M(M) = \mu(M)$.

$\square$

2.40 Remark: There is no other matching, stable or not, that all men prefer to μ^M.

We have already studied the sense in which it is as good a stable matching as the men can achieve, but now we want to ask whether there might not be some other unstable matching that all the men would prefer. If so, then we might conclude that, even at the man-optimal stable matching, the men collectively "pay a price" for stability. However, this turns out not to be the case.

2.41 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 two women, and their preferences are as follows:

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

Table 2.4

Then

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

Nevertheless

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

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

2.5 Properties of stable matchings II

2.42 Definition: In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, we say that a matching μ' weakly dominates another matching μ if there exists a coalition $\emptyset \neq A \subseteq M \cup W$, such that $\mu'(i) \succsim_i \mu(i)$ and $\mu'(i) \in A$ for any $i \in A$ and $\mu'(i_0) \succ_{i_0} \mu(i_0)$ for some $i_0 \in A$.

A matching μ is in the core if there exists no matching μ' which weakly dominates μ.

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

Proof. "$\Rightarrow$": Assume that μ is in the core.

(1) If μ is blocked by an individual i, then it is weakly dominated by any matching μ' with $\mu'(i) = i$ via the singleton coalition $\{i\}$.

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

"$\Leftarrow$": Assume that μ is a stable matching.

(1) If μ is not in the core, then μ is weakly dominated by some matching μ' via a coalition A. Hence, there exists $i_0 \in A$ such that $\mu'(i_0) \succ_{i_0} \mu(i_0)$.

(2) For notational simplicity, denote $i_0 = m$.

(3) Since μ is individually rational, $\mu'(m) \succ_m \mu(m) \succsim_m m$, and hence $\mu'(m) \in W$. Denote $\mu'(m)$ by w.

(4) Since $w \in A$, we have $\mu'(w) \succsim_w \mu(w)$.

(5) Clearly, $\mu'(w) \neq \mu(w)$; otherwise, $\mu'(m) = \mu(m)$. Thus, $\mu'(w) \succ_w \mu(w)$.

(6) The matching μ is blocked by (m, w). It is a contradiction.

□

2.44 Remark: There is another version of core.

In a marriage problem $\Gamma = \langle M, W, \succsim \rangle$, we say that a matching μ' dominates another matching μ 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.

Exercise: Show that the set of stable matchings, the core, and the core defined via strict domination are the same.

2.45 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.

2.46 Blocking lemma (Hwang (unknown), 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.

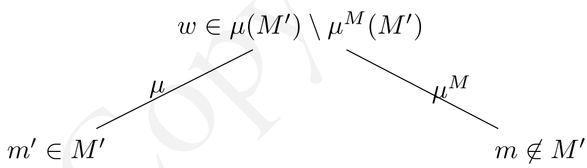


Figure 2.2

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 $\mu(w)$.

$\square$

2.47 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 $\mu^M(w) \succsim_w \mu(w)$.

2.48 *Proof of Theorem 2.45.* (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 2.47.

(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.

$\square$

2.6 Extension: Extending the men's preferences

2.49 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. $\mu^W[\succsim']$).

2.50 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 $\succsim' \rhd_M \succsim$.

2.51 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 prefer 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.

□

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

Consider the Example 2.49. Let

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

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

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

2.53 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.

□

2.54 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 2.53), $\mu^M[\succsim] \vee_M \mu^M[\succsim']$ under $\succsim$ is stable for $\langle M, W, \succsim \rangle$.

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

(3) Also by lattice lemma (Lemma 2.53), $\mu^W[\succsim] \vee_W \mu^W[\succsim']$ under $\succsim$ is stable for $\langle M, W, \succsim' \rangle$.

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

$\square$

2.55 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 $\mu^W[\succsim]$.

2.7 Extension: Adding another woman

2.56 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 $\mu^M[\Gamma]$.

2.57 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 2.54 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 $\mu^M[\Gamma] \succsim_W \mu^M[\Gamma']$.

□

2.58 Remark: Theorem 2.57 states that when new women enter, no man is hurt under the man-optimal matchings.

2.59 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. $\mu(S) = \{w_1, w_2, \ldots, w_k\}$.

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

Figure 2.4

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

(iv) By induction, we have

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

□

2.60 Remark: There exist some men who are in fact helped in quite a clear way (unless the new women remain unmatched): They are better off at every stable matching in the new market than they were at any stable matching of the old market. Furthermore (unless these men were all previously unmatched), there are some women who are similarly harmed by the entry of new women into the market.

2.8 Incentive compatibility I

☞ 2.61 A (direct) mechanism (机制) φ is a systematic procedure that determines a matching for each marriage problem $\langle M, W, \succsim\rangle$.

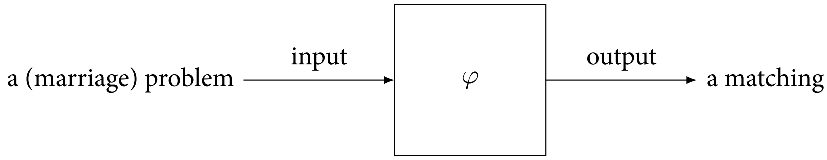


Figure 2.5: A mechanism

We have already studied two typical mechanisms which select the man-optimal and woman-optimal stable matchings, denoted by DA^M and DA^W, respectively. We call them the man-optimal stable mechanism and the woman-optimal stable mechanism, respectively.

For the sake of convenience, we shall use "the men-proposing deferred acceptance algorithm" interchangeably with "the man-optimal stable mechanism".

2.62 Question: What is the difference between a matching and a mechanism?

☞ 2.63 A mechanism φ is stable if it always selects a stable matching.[4]

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

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

2.64 Let $\mathcal{P}_i$ denote the set of all preferences for $i \in M \cup W$, $\mathcal{P} = \mathcal{P}_{m_1} \times \cdots \times \mathcal{P}_{m_p} \times \mathcal{P}_{w_1} \times \cdots \times \mathcal{P}_{w_q}$ 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.

2.65 We have learned properties of stable matching, given information about preferences of participants. But in reality, preferences are private information, so the clearinghouse should ask participants. Do people have incentives to tell the truth?

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 $\varphi\colon \mathcal{P} \to \mathcal{M}$.

☞ 2.66 A mechanism φ is strategy-proof[5] (抗策略操作) if for each marriage problem $\langle M, W, \succsim\rangle$, for each $i \in M \cup W$, and for each $\succsim_i' \in \mathcal{P}_i$, we have

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

[4] Table 1 in Roth (2002) shows that unstable matching algorithms tend to die out while stable algorithms survive the test of time.

[5] In general, a mechanism is strategy-proof if it is a weakly-dominant strategy for every individual to reveal his/her private information.

2.67 Example: Deferred acceptance algorithm is not strategy-proof.

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

The outcome of men-proposing deferred acceptance algorithm is

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

However, w_1 can be better off if she misreports her preference $\succ'_{w_1} : m_2$. The new outcome is

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

2.68 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, $\{m_1, m_2, \ldots, m_p\}$. 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.

✌ 2.69 Impossibility theorem (Theorem 3 in Roth (1982b)): There exists no mechanism that is both stable and strategy-proof. In other words, for any stable mechanism φ, there exist a marriage problem $\langle M, W, \succsim\rangle$, a person $i \in M \cup W$, and a preference $\succsim_i'$ such that

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

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 2.8

(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 $(\succsim_{-m_1}, \succsim_{m_1}')$.

□

2.70 Remark: No perfect mechanism exists.

2.71 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.

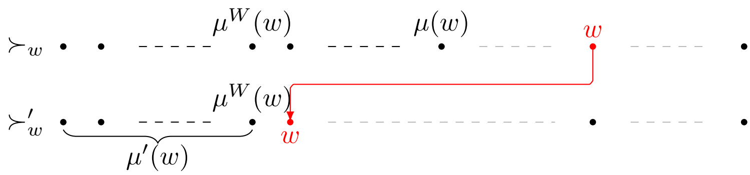
2.72 Theorem: When any stable mechanism is applied to a marriage problem in which preferences are strict and there is more than one stable matching, then at least one individual can profitably misreport his or her preference, assuming that the others tell the truth.

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

(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.

- It is obvious that μ^W is individually rational under the new preference profile, since $\mu^W(w) \succsim'_w w$ and $\mu^W(i) \succsim_i i$ for each $i \neq w$.
- It is trivial that μ^W is not blocked by a pair which does not contain w under the new preference profile; otherwise μ^W is blocked by this pair under the original preference profile.
- If μ^W is blocked by a pair (m, w) under the new preference profile, then $m \succ'_w \mu^W(w)$ and $w \succ_m \mu^W(m)$. Thus, $m \succ_w \mu^W(w)$ and $w \succ_m \mu^W(m)$, which means that μ^W is blocked by the pair (m, w) under the original preference profile.

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

(7) It follows from rural hospital theorem (Theorem 2.31) that w is not single under μ' (μ^W and μ' are two stable matchings under the new preference profile).

(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. That is, $\mu'(w) \succsim_w \mu^W(w)$.

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

- It is obvious that μ' is individually rational under the original preference profile, since $\mu'(w) \succsim_w \mu^W(w) \succ_w w$ and $\mu'(i) \succsim_i i$ for each $i \neq w$.
- It is trivial that μ' is not blocked by a pair which does not contain w under the original preference profile; otherwise μ' is blocked by this pair under the new preference profile.
- If μ' is blocked by a pair (m, w) under the original preference profile, then $m \succ_w \mu'(w)$ and $w \succ_m \mu'(m)$. Thus, $m \succ'_w \mu'(w)$ and $w \succ_m \mu'(m)$, which means that μ' is blocked by the pair (m, w) under the new preference profile.

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

(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.

$\square$

2.73 Question: What is the difference between Theorems 2.69 and 2.72?

2.74 Proposition: If φ is a stable mechanism, and μ is a stable matching in $\langle M, W, \succ\rangle$, then for each $i \in M \cup W$, there exists $\succ_i'$ such that $\varphi[\succ_i', \succ_{-i}](i) = \mu(i)$.

Proof. (1) Let $\succ_i'$: $\mu(i), i$.

(2) Note that μ is also stable at $(\succ_i', \succ_{-i})$.

(3) If i is matched at $\succ$, then i is also matched at $(\succ_i', \succ_{-i})$.

(4) Since only $\mu(i)$ is acceptable to i at $\succ_i'$ and φ is stable, $\varphi[\succ_i', \succ_{-i}](i) = \mu(i)$.

$\square$

The proposition implies that a man/woman can misreport to obtain any stable assignment under a stable mechanism.

2.75 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 2.9

(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 $\varphi[\succsim^1] = \mu_2^1$.

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
w_2	w_1		m_2

Table 2.10

(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:

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

- 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 2.11

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 2.12

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:

m_1	m_2	w_1	w_2
w_1	w_2	m_2	m_1
		m_1	m_2

Table 2.13

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 $\succsim^5_{m_2}$.

□

2.9 Incentive compatibility II

2.76 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.

Intuition: Men are not punished when applying to preferred women. This is in a contrast with the Boston mechanism.

Proof. It is a corollary of theorem of limits on successful manipulation (Theorem 2.86).

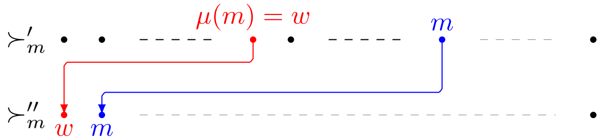
We provide an alternative proof as follows:

(1) In the marriage problem $\langle M, W, \succ\rangle$, suppose that man m misreports $\succ'_m$. Let $\text{DA}^M[\succ'_m, \succ_{-m}] = \mu$. It is sufficient to show that by truthfully reporting $\succ_m$, m will be weakly better off.

(2) Case 1: If $\mu(m) = m$ or $m \succ_m \mu(m)$, nothing needs to be proved.

(3) Case 2: Suppose that $\mu(m) = w$.

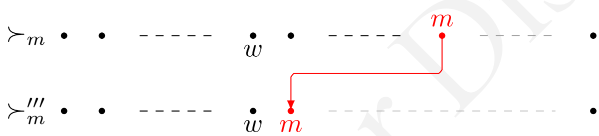
(4) Suppose m reports $\succ''_m$: w, m, *i.e.*, only w is acceptable to him.



(i) At $(\succ''_m, \succ_{-m})$, μ is still stable due to less desires.

(ii) Since m is matched to w under μ, rural hospital theorem (Theorem 2.31) implies that m being unmatched will be unstable at $(\succ''_m, \succ_{-m})$.

(5) Consider $\succ'''_m$: $\ldots, w, m$, which is obtained by truncating the true preference from w.



(i) m being unmatched will also be unstable at $(\succ'''_m, \succ_{-m})$: If a matching making m single is stable under $(\succ'''_m, \succ_{-m})$, then it is also stable under $(\succ''_m, \succ_{-m})$.

(ii) Therefore, under $\text{DA}^M[\succ'''_m, \succ_{-m}]$, m is matched to some woman weakly better than w.

(iii) As the DA procedure is the same under $(\succ'''_m, \succ_{-m})$ and $(\succ_m, \succ_{-m})$, m will be weakly better off by truthfully reporting $\succ_m$.

□

✌ 2.77 Remark: Deferred acceptance algorithm is the unique stable and one-sided strategy-proof mechanism; see Theorem 8.34.

2.78 Remark: The men-proposing deferred acceptance algorithm is group strategy-proof for men.

2.79 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 $\mu^M[\succsim''](m) = \mu^M[\succsim'](m)$.

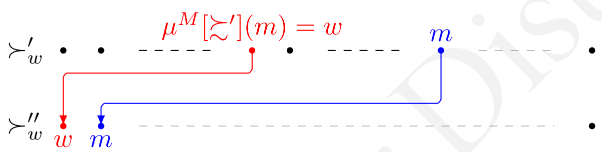
Proof. (1) Clearly the matching $\mu^M[\succsim']$ is stable under the preference profile $\succsim''$.

(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 $\mu^M[\succsim'](m) = \mu^M[\succsim''](m)$.

$\square$

2.80 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.



2.81 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.

□

2.82 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.

2.83 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 2.86). □

2.84 Remark: Theorem 2.83 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.

2.85 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.

2.86 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 2.46) 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.

□

2.87 Remark: Theorem 2.86 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 $\succsim$.

2.10 Non-bossiness

☞ 2.88 Definition: A mechanism φ is said to be non-bossy (不专横)[6] if, for each marriage problem $\langle M, W, \succ\rangle$, for each $i \in M \cup W$, and for each $\succ'_i \in \mathcal{P}_i$,

$$\varphi[\succ'_i, \succ_{-i}](i) = \varphi[\succ](i) \text{ implies } \varphi[\succ'_i, \succ_{-i}] = \varphi[\succ].$$

2.89 Example: Deferred acceptance algorithm is not non-bossy.

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

[6] The concept of non-bossiness is due to Satterthwaite and Sonnenschein (1981). A mechanism is "non-bossy" if whenever a change in an individual's preference does not bring about a change in his assignment, then it does not bring about a change in anybody's assignment. See Thomson (2014).

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 2.14

The men-proposing DA outcome is

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

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

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

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

2.90 Theorem (Theorem 1 in Kojima (2010)): There exists no stable mechanism that is non-bossy for marriage problems.

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 2.15

(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 non-bossy.

(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

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

so φ is not non-bossy.

□

2.91 A rough idea is to note that the men-proposing DA is not non-bossy, 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 non-bossiness (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.

2.92 Exercise: Find a non-bossy mechanism for marriage problems.

Chapter 3

College admissions

Contents

3.1 The formal model

3.1 In a college admissions 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: In a college admissions problem, a matching is the outcome, 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.

Example: Suppose that there are three students $\{1,2,3\}$ and a college c has three quotas. Then the college c should have a ranking over the groups of students: $\{1,2,3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\emptyset$.

3.5 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).

3.6 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).

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

3.7 Example: Suppose that there are two students $\{1, 2\}$ and a college c has two quotas. The following preference $\succ_c$ is not responsive:

c
$\{1, 2\}$
$\{1\}$
$\emptyset$

Table 3.1

3.2 Stability

3.8 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.

3.9 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 $s \succ_c s'$ (justifiable envy), or
  – $|\mu(c)| < q_c$ and $s \succ_c \emptyset$ (wasteful).

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

3.11 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.

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 3.2

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

The only individually rational matchings without unemployment are

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

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

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

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

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

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 $(c_2, \{s_1, s_3\})$.

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

3.13 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.

3.14 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

3.15 The importance of Proposition 3.14 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.

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

3.17 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.

3.18 Example: 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_2	s_1
c_2	c_1	s_1	s_2

Table 3.3

The related marriage problem is as follows: Three seats $C^* = \{c_1^1, c_1^2, c_2\}$ and three students $\{s_1, s_2, s_3\}$. The preferences are given by

s_1	s_2	s_3	c_1^1	c_1^2	c_2
c_1^1	c_2	c_1^1	s_2	s_2	s_1
c_1^2	c_1^1	c_1^2	s_1	s_1	s_2
c_2	c_1^2				

Table 3.4

3.19 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.

In the college admissions problem above, consider a matching

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

Then we have a corresponding matching for the related marriage problem

$$\begin{bmatrix} c_1^1 & c_1^2 & c_2 \\ s_1 & s_3 & s_2 \end{bmatrix}.$$

3.20 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.

Proof. Exercise. □

3.4 Deferred acceptance algorithm and properties of stable matchings

3.21 College-proposing deferred acceptance algorithm.

Step 1: (a) 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).

(b) 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: (a) 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).

(b) 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."

3.22 Student-proposing deferred acceptance algorithm.

Step 1: (a) Each student proposes to her top-choice individually rational college (if she has one).

(b) 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: (a) 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).

(b) 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."

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

Proof. It is a consequence of theorem on stability in marriage problem (Theorem 2.45) and Lemma 3.20. □

3.24 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 $k_\ell = \min\{q_\ell, a_\ell\}$.

3.25 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 c_k.

□

3.26 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.

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

Proof. It follows from Theorem 2.39 and Lemma 3.20. □

3.28 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 3.5

(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 colleges strictly prefer μ' to μ^C.

□

3.29 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 2.39 and Lemma 3.20 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.

This result is also consistent with the fact that DA is not strongly Pareto optimal; see Example 2.41.

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

Proof. The proof is immediate via Theorem 2.31 and Lemma 3.20. □

3.31 Lemma (Lemma 3 in Roth and Sotomayor (1989)): Suppose that 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 3.30, 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.

□

3.32 Remark: The proof of Lemma 3.31 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 $j > i$.

3.33 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 3.31.

3.34 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 3.30 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 3.31 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 $j > i$.

□

3.35 Exercise: Find a non-trivial example to illustrate the above result does not necessarily hold for colleges which fill quotas at some stable matching.

Hint: Consider the example in the proof of Theorem 2.69.

3.36 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 3.31, $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j)$ 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:

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

□

3.37 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 3.34 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 3.31 $\bar{\mu}(c^j) \succ_c \bar{\mu}'(c^j) = s'$; otherwise $\mu'(c) \succ_c \mu(c)$, which contradicts the fact $\mu(c) \succ_c \mu'(c)$.

(6) The decomposition lemma (Lemma 2.51) 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.

□

3.38 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 $j \in \mu'(c) \setminus \mu(c)$.

3.39 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, $\mu(c) \succ_c \mu'(c)$.)

3.40 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. □

3.41 Example: Let the preferences over individuals be given by

and let the quotas be $q_{c_1} = 3$, $q_{c_j} = 1$ for $j = 2, \ldots, 5$. 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 3.6

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

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

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

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

3.42 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 3.31 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 2.29, which implies $\mu' \succ_S \mu$.

□

3.43 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.

3.44 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 $\mu \wedge_C \mu'$.

3.45 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 3.31.

(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 $\mu \vee_C \mu'$.

□

3.46 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.

3.47 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 2.51) $\bar{\mu}$ 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 3.31 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 $\mu'(s) \succsim_s \mu(s)$.

□

3.48 Theorem: Suppose that $\succ' \rhd_C \succ$ 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 results are obtained if $\succ' \rhd_S \succ$.

Proof. (1) Suppose that $\succ' \rhd_C \succ$.

(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 2.54.

□

3.49 Theorem: Suppose that 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 that 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 2.57.

□

3.50 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,

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

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

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

Proof. Part 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 $\mu'(c) = \mu(c) \cup s \setminus s'$.

Part 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 that 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 that 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).

□

3.52 Remark: There is another version of core.

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,

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

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

Exercise: Find a college admission problem such that the core and the core defined via strict domination are not the same.

3.6 Incentive compatibility

3.53 Throughout this section we fix $S = \{s_1, s_2, \dots, s_p\}$, and $C = \{c_1, c_2, \dots, c_r\}$, so each pair of preference profile and quota profile defines a college admissions problem.

3.54 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 agents except i.

Let $\mathcal{Q}_c$ denote the set of all quotas for college c, $\mathcal{Q} = \mathcal{Q}_{c_1} \times \mathcal{Q}_{c_2} \times \cdots \times \mathcal{Q}_{c_r}$ denote the set of all quota profiles, and $\mathcal{Q}_{-c}$ denote the set of all quota profiles for all schools except c.

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

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

☞ 3.56 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.

3.57 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

☞ 3.58 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 $\succsim_{-i} \in \mathcal{P}_{-i}$,

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

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

Proof. It follows immediately from Theorem 2.69. □

3.60 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 2.75. □

3.61 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 2.76. □

3.62 Remark: Deferred acceptance algorithm is the unique stable and one-sided strategy-proof mechanism; see Theorem 8.34.

3.63 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 3.7

(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 that 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).

□

3.64 Remark: A college is like a coalition of players in terms of strategies.

3.65 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.

3.66 Roth (1984) showed that the algorithm independently discovered by the National Residency Matching Program (NRMP) in the United States was equivalent to the college-optimal stable mechanism. Roth (1991) observed that several matching mechanisms that have been used in Britain for hospital-intern matching were unstable and as a result were abandoned, while stable mechanisms survived. This key observation helped to pin down stability as a key property of matching mechanisms in the college admissions framework. Roth and Peranson (1999) introduced a new design for the NRMP matching mechanism based on the student-optimal stable mechanism. Interestingly, the replacement of the older stable mechanism with the newer mechanism was partially attributed to the positive and negative results in Theorems 3.61 and 3.63, respectively.

3.6.2 Capacity manipulation

☞ 3.67 In a college admission problem $\langle S, C, q, \succ\rangle$, a college c manipulates a mechanism φ via capacities 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 capacity manipulation if it can never be manipulated via capacities.

3.68 Example: The college-optimal stable mechanism is not immune to capacity manipulation:

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:

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 3.8

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

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

(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.

□

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

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

(2) Let ϕ be a stable mechanism, $C = \{c_1, c_2\}$ and $S = \{s_1, s_2, s_3\}$,

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 3.9

$q_{c_1} = q_{c_2} = 2$ and $q'_{c_1} = q'_{c_2} = 1$.

(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.

□

3.70 Exercise: Is there a stable mechanism that is immune to capacity manipulation for college admissions problems with two colleges and two students?

3.71 Remark: In one-to-one matching, DA cannot be manipulated by an agent if and only if there is a unique stable partner. The statement is false in many-to-one matching.

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

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

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

Proof. Omitted. □

3.74 Remark: Example 3.68 shows that the college-optimal stable mechanism is capacity manipulable even under strongly monotonic preferences.

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

3.76 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. See Chapter 9 for the definition of acyclicity.

Proof. Omitted. □

3.7 Comparison of marriage problems and college admissions

3.77 Comparison of marriage problems and college admissions problems:

	Marriage problems	College admissions (with responsive preferences)
Existence of stable matchings	$\surd$	$\surd$
One-sided individual optimality	$\surd$	$\surd$
One-sided weakly Pareto optimality	$\surd$	$\surd$ (s) and $\times$ (c)
Rural hospital theorem	$\surd$	$\surd$
Two-sided strategy-proofness	$\times$	$\times$
One-sided strategy-proofness	$\surd$	$\surd$ (s) and $\times$ (c)

3.8 National resident/intern matching program

3.78 Students who graduate from medical schools in US are typically employed as residents (interns) at hospitals, where they comprise a significant part of the labor force.

In the early twentieth century, the market for new doctors was largely decentralized. During the 1940s, competition for medical students forced hospitals to offer residencies/internships increasingly early, sometimes several years before a student would graduate. This so-called unraveling had many negative consequences. Matches were made before students could produce evidence of how qualified they might become, and even before they knew what kind of medicine they would like to practice.

The market also suffered from congestion: when an offer was rejected, it was often too late to make other offers.

3.79 In response to the failure of the US market for new doctors, a centralized clearinghouse was introduced in the early 1950s. This institution is now called the National Resident Matching Program (NRMP) or National Intern Matching Program (NIMP).

3.80 NRMP algorithm.

Initial editing of ranking lists: Each hospital ranks the students who have applied to it and each student ranks the hospital to which he has applied.

These ranking lists are mailed to the central clearinghouse, where they are edited by removing from each hospital's ranking list any student who has marked that hospital as unacceptable, and by removing from each student's ranking list any hospital which has indicated he is unacceptable.

The edited lists are thus ranking lists of acceptable alternatives.

Matching phase: $1:1$ **step:** Check to see if there are any students and hospitals which are top-ranked in one another's ranking. (If a hospital has a quota of q then the q highest students in its ranking are top-ranked.) If no such matches are found, the matching phase proceeds to the $2:1$ step; otherwise the algorithm proceeds to the tentative assignment and update phase.

$k:1$ **step:** Seek to find student-hospital pairs such that the student is top-ranked on the hospital's ranking and the hospital is k-th ranked by the student. If no such matches are found, the matching phase proceeds to the $(k+1):1$ step; otherwise the algorithm proceeds to the tentative assignment and update phase.

Tentative assignment and update phase:

- When the algorithm enters the tentative assignment and update phase from the $k:1$ step of the matching phase, the $k:1$ matches are tentatively made; *i.e.*, each student who is a top-ranked choice of his k-th choice hospital.
- The rankings of students and hospitals are then updated in the following way:
  - Any hospital which a student s_j ranks lower than his tentative assignment is deleted from his ranking. (So the updated ranking of a student s_j tentatively assigned to his k-th choice now lists only his first k choices.)
  - Any student s_j is deleted from the ranking of any hospital which was deleted from s_j's ranking. (So the updated ranking of each hospital now include only those applicants who have not yet been tentatively assigned to a hospital they prefer.)
- When the rankings have been updated in this way, the algorithm returns to the start of the matching phase. Any new tentative matches found in the matching phase replace prior tentative matches involving the same student.

End: The algorithm terminates when no new tentative matches are found, at which point tentative matches become final.

3.81 Example: Consider the problem consisting of two hospitals $\{h_1, h_2\}$, each with a quota of one, and three students $\{s_1, s_2, s_3\}$. The preferences are given by

h_1	h_2	s_1	s_2	s_3
s_1	s_1	h_1	h_1	h_1
s_2	s_2	h_2		h_2
s_3	s_3			

Table 3.10

The edited lists are:

h_1	h_2	s_1	s_2	s_3
s_1	s_1	h_1	h_1	h_1
s_2	$\cancel{s_2}$	h_2		h_2
s_3	s_3			

Table 3.11

In $1:1$ step, one tentative match (h_1, s_1) is found. Then the algorithm proceeds to tentative assignment and update phase. The updated lists are

h_1	h_2	s_1	s_2	s_3
s_1	$\cancel{s_1}$	h_1	h_1	h_1
s_2	$\cancel{s_2}$	$\cancel{h_2}$		h_2
s_3	s_3			

Table 3.12

The algorithm returns to the matching phase. In $1:1$ step, no new tentative match. In $2:1$ step, one tentative match (h_2, s_3) is found. Then the algorithm proceeds to tentative assignment and update phase, but there is no new update for rankings.

The outcome is

$$\begin{bmatrix} h_1 & h_2 & \\ s_1 & s_3 & s_2 \end{bmatrix}.$$

3.82 Roth (1984) showed that the NRMP algorithm is equivalent to a (hospital-proposing) DA algorithm, so NRMP produces a stable matching. Roth (1984) argued that the success of NRMP was due to the fact that it produced stable matchings.

3.83 Several issues led to the redesign NRMP algorithm:

- The NRMP algorithm favors hospitals at the expense of students.
- Both students and hospitals may have incentives to manipulate the NRMP algorithm.
- NRMP has special features, called "match variations". An example is couples.

3.84 Theorem (Theorem 10 in Roth (1984)): In a market in which some agents are couples, the set of stable outcomes may be empty.

Proof. Consider the problem consisting of two hospitals $\{h_1, h_2\}$, each with a quota of one, one single student s and one couple (m, w). The preferences are given by

h_1	h_2	s	(m,w)
m	s	h_1	(h_1, h_2)
s	w	h_2	

Table 3.13

In this market, no stable matching exists. □

3.85 Remark: The rural hospital theorem also fails in the market above.

3.86 In 1995, Roth was hired by the board of directors of NRMP to direct the design of a new algorithm. The new algorithm (which is called Roth-Peranson algorithm), designed by Roth and Peranson (1999), is a student-proposing algorithm modified to accommodate couples: potential instabilities caused by the presence of couples are resolved sequentially, following the instability-chaining algorithm of Roth and Vate (1990).

For details of the new NRMP algorithm, see Roth and Peranson (1999).

3.9 New York City high school match

3.87 Main reference: Abdulkadiroğlu *et al.* (2005a) and Abdulkadiroğlu *et al.* (2009).

3.88 Background: Over 90,000 students enter high schools each year.

The old NYC system was decentralized:

- Each student can submit a list of at most 5 schools.
- Each school obtains the list of students who listed it, and independently make offers.
- There were waiting lists (run by mail), and 3 rounds of move waiting lists.

3.89 Problems with the old system:

- The system left 30,000 children unassigned to any of their choices and they are administratively assigned.
- Strategic behavior by schools: school principals were concealing capacities.

3.90 In New York City, schools behave strategically.

Deputy Chancellor of Schools (NYT 19 November 2004):

> Before you might have had a situation where a school was going to take 100 new children for 9th grade, they might have declared only 40 seats and then placed the other 60 children outside the process.

Unlike Boston, the market seems to be really two-sided, *i.e.*, we should treat both students and schools are strategic players.

3.91 Since NYC is a two-sided matching market, the student-proposing DA is the big winner:

- DA implements a stable matching (probably more important for NYC than for Boston.)
- DA is strategy-proof for students: it is a dominant strategy for every student to report true preferences.
- There is no stable mechanism that is strategy-proof for schools.
- When the market is large, it is almost strategy-proof for schools to report true preferences; Kojima and Pathak (2009): Recall there are 90000 students and over 500 public high schools in New York City.

3.92 Abdulkadiroğlu *et al.* (2009) and NYC Department of Education changed the mechanism to the student-proposing DA, except for some details:

- Students can rank only 12 schools.
- Seats in a few schools, called specialized high schools (such as Stuyvesant and Bronx High School of Science), is assigned in an earlier round, separately from the rest.
- Some top students are granted to get into a school when they rank the school as their first choices.
- All unmatched students in the main round will be assigned in the supplementary round, where the random serial dictatorship is used.

These features come from historical constraints and could not be changed.

This make it technically incorrect to use standard results in two-sided matching, but they seem to be small enough a problem (it may be interesting to study if this is true and why or why not.)

3.93 Effect of changes in the mechanism:

- Over 70,000 students were matched to one of their choice schools: an increase of more than 20,000 students compared to the previous year match.
- An additional 7,600 students matched to a school of their choice in the third round.
- 3,000 students did not receive any school they chose, a decrease from 30,000 who did not receive a choice school in the previous year.

Part II

One-sided matching

Chapter 4

Housing market

Contents

4.1 The former model

4.1 Housing market model was 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 houses in order to get a better one.

☞ 4.2 Definition: Formally, a housing market is a quadruple $\langle A, H, \succ, e\rangle$ 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 Like a pure exchange economy, in a housing market, agents can trade the houses among themselves according to certain rules and attempt to make themselves better off.

Example: Let $A = \{a_1, a_2, a_3, a_4\}$ and let h_i be the occupied house of agent i. Let the preference profile $\succ$ be given as:

a_1	a_2	a_3	a_4
h_4	h_3	h_2	h_3
h_3	h_4	h_4	h_2
h_2	h_2	h_1	h_1
h_1	h_1	h_3	h_4

Table 4.1

These four agents can trade the houses and get the following (Pareto) improved reallocation

$$\mu_1 = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ h_4 & h_3 & h_1 & h_2 \end{bmatrix}.$$

They also have the following (Pareto) improved reallocation

$$\mu_2 = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ h_4 & h_3 & h_2 & h_1 \end{bmatrix}.$$

What are desirable outcome of such a reallocation process? What allocative mechanisms are appropriate for achieving desirable outcomes?

☞ 4.4 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.

☞ 4.5 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

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

☞ 4.6 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.

In Example 4.3, the matchings μ_1 and μ_2 are individually rational.

☞ 4.7 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.

In Example 4.3, the matchings μ_1 and μ_2 are Pareto efficient.

4.8 In Example 4.3, if houses are assigned according to μ_1, then agents 2 and 3 will not attend this reallocation process. Instead, they will trade with each other; that is, agent 2 gets house 3 and agent 3 gets house 2. Clearly, this trade benefits agent 3 and does not hurt agent 2, compared with μ_1.

In other words, matching μ_1 is blocked by the coalition $\{2, 3\}$ and the trade between them. Such a matching is not good enough, and a core matching, defined in the following paragraphs, is required to exclude such blocks.

4.9 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 $b \in B$. That is, $\{\mu(a) \mid a \in B\} = \{h_b \mid b \in B\}$.

☞ 4.10 Definition: A matching μ is in the core[1] 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 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}.

[1] It was also called strong core in the literature. In game theory, the core is the set of feasible allocations that cannot be improved upon by a subset (a coalition) of the economy's consumers. A coalition is said to improve upon or block a feasible allocation if the members of that coalition are better off under another feasible allocation that is identical to the first except that every member of the coalition has a different consumption bundle that is part of an aggregate consumption bundle that can be constructed from publicly available technology and the initial endowments of each consumer in the coalition.

4.11 Remark: It is clear that a core matching is Pareto efficient (take $B = A$) and individually rational (take $B = \{a\}$ for some $a \in A$).

☞ 4.12 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 $(\mu, \boldsymbol{p}) \in \mathcal{M} \times \mathbb{R}^n_{++}$ is a competitive equilibrium (or a Walrasian 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}.

4.13 Remark: The market clear condition trivially holds since each matching is required to be a bijection. Furthermore, in a competitive equilibrium $(\mu, \boldsymbol{p})$, for each agent a, the price of her final house $p_{\mu(a)}$ equals the price of her initial house p_{h_a}. (Exercise)

4.14 Proposition: If each agent's preference is strict, then any competitive equilibrium allocation is in the core.

Proof. (1) Let $(\mu, \boldsymbol{p})$ be a competitive equilibrium. Suppose that μ is not in the core.

(2) Then there is a coalition $B \subseteq A$ and a B-matching ν such that $\nu(a) \succsim_a \mu(a)$ for all $a \in B$ and $\nu(a_0) \succ_{a_0} \mu(a_0)$ for some $a_0 \in B$.

(3) Since μ is a competitive equilibrium matching, $p_{\nu(a)} \geq p_{h_a} = p_{\mu(a)}$ for all $a \in B$ and $p_{\nu(a_0)} > p_{h_{a_0}} = p_{\mu(a_0)}$ (Here we need to assume each agent's preference to be strict).

(4) Since ν is a B-matching, $\sum_{a \in B} p_{\nu(a)} = \sum_{a \in B} p_{h_a}$.

(5) Thus,

$$\sum_{a \in B} p_{\mu(a)} < \sum_{a \in B} p_{\nu(a)} = \sum_{a \in B} p_{h_a} = \sum_{a \in B} p_{\mu(a)},$$

which leads to a contradiction.

□

It is well known that any competitive equilibrium allocation is in the core for exchange economies with divisibilities.

4.15 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

4.16 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 attributed this algorithm to David Gale.

4.17 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 TTC.

4.18 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 t is the last step, then
$$H = H^0 \cup H^1 \cup H^2 \cup \cdots \cup H^t.$$
- $G' = \langle B, \succ\rangle$: the 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 $\langle A, H, \succ, e\rangle$.
- $\mathrm{Br}_a(H')$ where $a \in A$ and $H' \subseteq H$: agent a's favorite house among H'. Then for each $a \in A^k$, we have
$$\mathrm{Br}_a\left(H \setminus \cup_{\ell=1}^{k-1} H^\ell\right) = \mathrm{TTC}(a).$$
- $a \xrightarrow{G'} b$ 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.

4.19 *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 her 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.

□

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

(1) Price vector $\boldsymbol{p}$ is defined as follows:

- for any a and 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.

□

4.21 Example of the 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 4.2

Step 1:

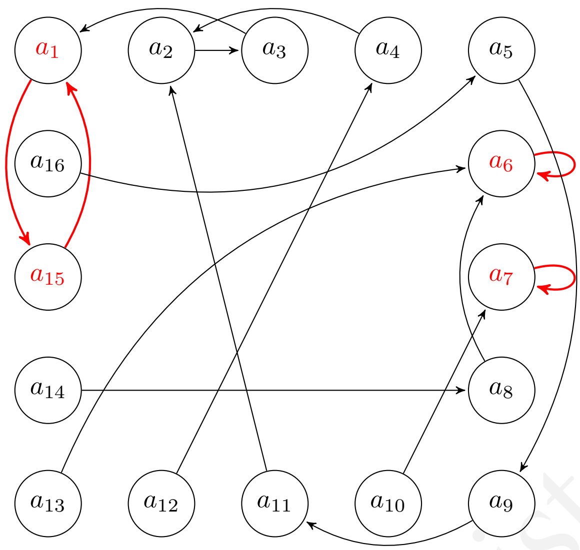


Figure 4.1: Step 1

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

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 4.3

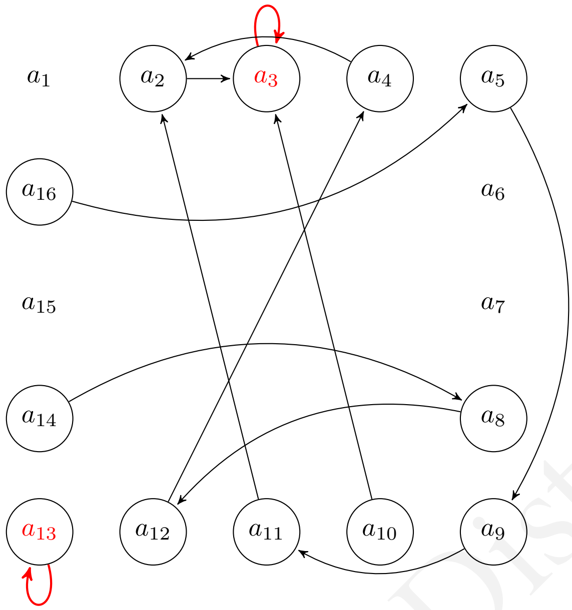


Figure 4.2: Step 2

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

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 4.4

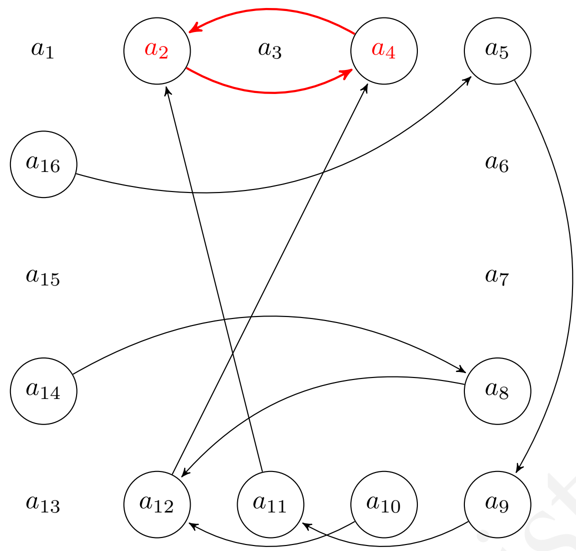


Figure 4.3: Step 3

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

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 4.5

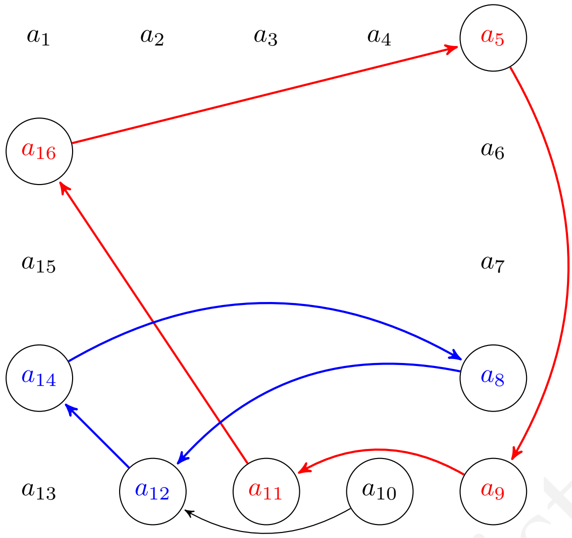


Figure 4.4: Step 4

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

Step 5: The reduced preferences are as follows:

a_{10}
h_{10}

Table 4.6

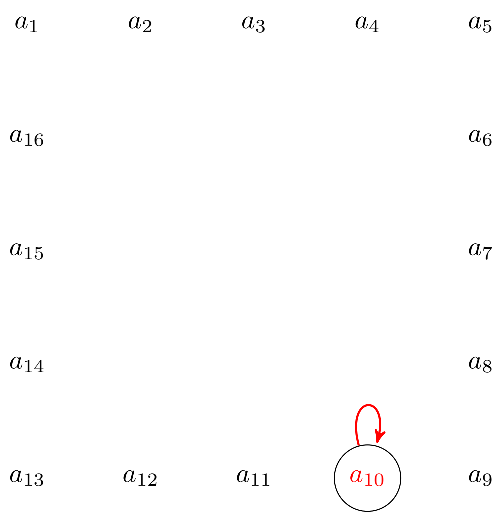


Figure 4.5: Step 5

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

The outcome is

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

4.22 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.

□

4.23 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 4.16 implies that no matching weakly dominates a competitive equilibrium matching (or core matching). Then apply Lemma 4.22. □

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

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

4.25 Remark: 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

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

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 that 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 headed 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 tail and agent $a_{in_i}^k$ as the head 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 tails (and each of them points to a different one), which in turn converts these chains into one or several cycles.

4.3 Incentive compatibility

☞ 4.26 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

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

4.27 Theorem (Theorem in Roth (1982a)): The core mechanism TTC is strategy-proof.

Intuition: Once being pointed by others, an agent never loses the chain pointing to her, so she can get the house any later time if she wants.

For the proof, we need the following three lemmas.

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

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

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.

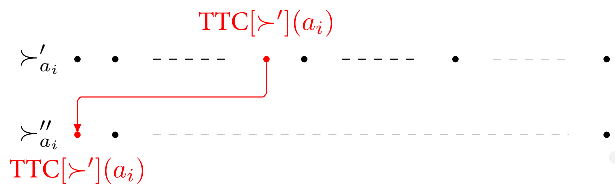
□

4.29 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. 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. This lemma follows by induction. □

4.30 Simple misreport manipulation 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 $\text{TTC}[\succ'](a_i)$ is a_i's favorite house under $\succ''_i$. Then we have

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

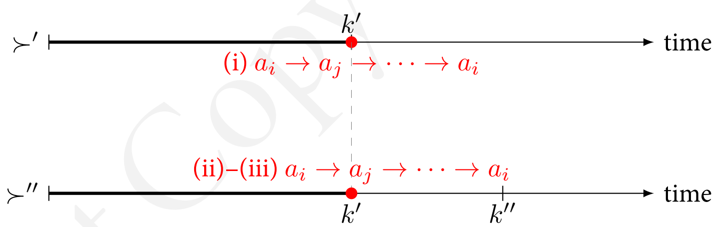


Proof. (1) Let k' be the step at which agent a_i with house $h_j \triangleq \text{TTC}[\succ'](a_i)$ is removed from the market $\langle A, H, \succ', e\rangle$. That is, $a_i, a_j \in B^{k'}[\succ']$.

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

(3) Let k'' be the step at which agent a_i with house $\text{TTC}[\succ''](a_i)$ 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'.

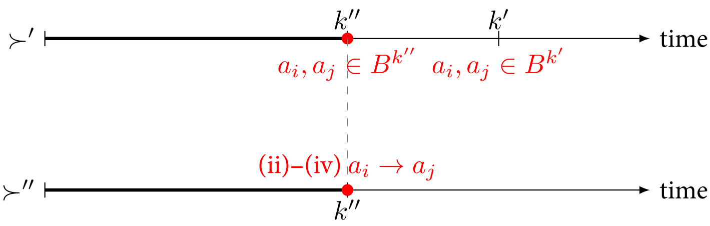
(ii) Then Lemma 4.29 implies that $B^{k'}[\succ'] = B^{k'}[\succ'']$. Hence, $a_i, a_j \in B^{k'}[\succ'] = B^{k'}[\succ'']$.

(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 $\text{TTC}[\succ''](a_i) = h_j = \text{TTC}[\succ'](a_i)$ 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 4.29 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 $\langle A, H, \succ'', e\rangle$.

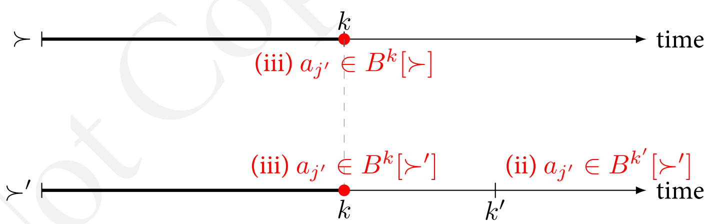
(v) Hence h_j is exactly the house which is removed with agent a_i at Step k'' in the market $\langle A, H, \succ'', e\rangle$, that is, $\mathrm{TTC}[\succ''](a_i) = h_j = \mathrm{TTC}[\succ'](a_i)$.

$\square$

4.31 *Proof of Theorem 4.27.* 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 a_i is removed from the market. Let $h_j = \mathrm{TTC}[\succ](a_i)$ and $h_{j'} = \mathrm{TTC}[\succ'](a_i)$. We will see that $h_{j'} \succ_{a_i} h_j$ is impossible.

Lemma 4.30 implies that it is sufficient to consider a preference $\succ'_{a_i}$ that ranks $h_{j'}$ first.

Case 1: $k' \geq k$.



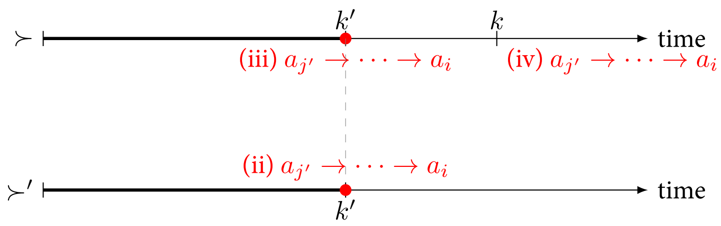
(1) Lemma 4.29 implies that $B^\ell[\succ] = B^\ell[\succ']$ for $1 \leq \ell \leq k$.

(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 4.29 implies that $B^\ell[\succ] = B^\ell[\succ']$ for $1 \le \ell \le k'$.

(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 $a_{n_m} = a_i$ is not removed st Step k in the market $\langle A, H, \succ, e \rangle$, Lemma 4.28 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.

$\square$

4.32 Definition: A mechanism φ is group strategy-proof, if for each housing market $\langle A, H, \succ, e \rangle$, 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.

4.33 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 $\mathrm{TTC}[\succ'](a_i) \succ_{a_i} \mathrm{TTC}[\succ](a_i)$, then there exist agents $a_j \in A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$ and agent $a_\ell \in A^k[\succ] \cup A^{k+1}[\succ] \cup \cdots \cup A^t[\succ]$ such that

$$h_\ell \succ'_{a_j} \mathrm{TTC}[\succ](a_j).$$

Proof. (1) Assume the contrary. Then

$$\text{TTC}[\succ](a_j) \succsim'_{a_j} h_\ell,$$

for all $a_j \in A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$ and $a_\ell \in A^k[\succ] \cup A^{k+1}[\succ] \cup \cdots \cup A^t[\succ]$.

(2) It is clear that the equalities above can not hold; otherwise $\text{TTC}[\succ](a_j) = h_\ell$ due to the strictness of preferences.

(3) Since each $\text{TTC}[\succ](a_j) = h_m$ for some $a_m \in A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$, it follows from the top trading cycle algorithm that

$$A^1[\succ'] \cup A^2[\succ'] \cup \cdots \cup A^{k'-1}[\succ'] = A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$$

for some k'.

(4) Since $\text{TTC}[\succ'](a_i) \succ_{a_i} \text{TTC}[\succ](a_i)$, $\text{TTC}[\succ'](a_i)$ must have been taken in an earlier trading cycle under $\succ$.

(5) Thus, $\text{TTC}[\succ'](a_i) = h_j$ for some $a_j \in A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$.

(6) For preference profile $\succ'$, a_i and a_j are in the same cycle , thus a_i is in $A^1[\succ'] \cup A^2[\succ'] \cup \cdots \cup A^{k'-1}[\succ']$.

(7) But $A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ] = A^1[\succ'] \cup A^2[\succ'] \cup \cdots \cup A^{k'-1}[\succ']$ and a_i is not in $A^1[\succ] \cup A^2[\succ] \cup \cdots \cup A^{k-1}[\succ]$. A contradiction.

□

4.34 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.

4.35 Theorem (Theorem in Bird (1984)): TTC is group strategy-proof.

Proof. (1) Assume that 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 $\text{TTC}[\succ'](a_i) \succ_{a_i} \text{TTC}[\succ](a_i)$, 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 $q \geq k$ 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.

□

4.36 Remark: We have shown a stronger result: 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 $a \in B$,

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

4.4 Axiomatic characterization of top trading cycles algorithm

4.37 Theorem (Theorem 1 in Ma (1994)): The core mechanism TTC is the only mechanism that is individually rational, Pareto efficient, and strategy-proof.

4.38 *Proof of Theorem 4.37.* (1) Suppose that there is another mechanism φ satisfying the three conditions.

(2) Fix a housing market $\langle A, H, \succ, e\rangle$.

(3) Let A^1 be the set of agents matched in Step 1 of TTC for $\langle A, H, \succ, e\rangle$. We first show that for any agent $a \in A^1$, $\varphi[\succ](a) = \mathrm{TTC}[\succ](a)$.

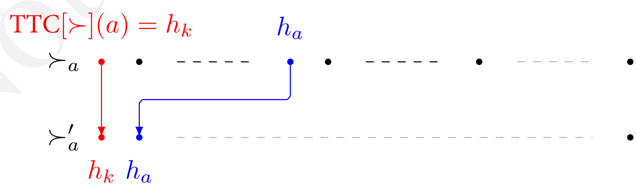
(4) Suppose not, then $\varphi[\succ](a)$ is worse. That is, $\mathrm{TTC}[\succ](a) \succ_a \varphi[\succ](a)$.

(5) Since TTC is individually rational, $\mathrm{TTC}[\succ](a) \succsim_a h_a$.

(6) If $\mathrm{TTC}[\succ](a) = h_a$, we have a contradiction with individual rationality of φ; that is, $h_a \succ_a \varphi[\succ](a)$.

(7) Thus, a trades with others under TTC at $\succ$. Assume that the trading cycle is $a \to k \to \cdots \to 1 \to a$.

(8) Consider a new preference $\succ'_a$: h_k, h_a.



(9) Then $\mathrm{TTC}[\succ] = \mathrm{TTC}[\succ'_a, \succ_{-a}]$ and $\mathrm{TTC}[\succ'_a, \succ_{-a}](a) = \mathrm{TTC}[\succ](a) = h_k$.

(10) Since φ is individual rational, a must be assigned h_k or h_a under $\varphi[\succ'_a, \succ_{-a}]$.

(11) 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:

$$\varphi[\succ'_a, \succ_{-a}](a) = h_k = \mathrm{TTC}[\succ](a) \succ_a \varphi[\succ](a).$$

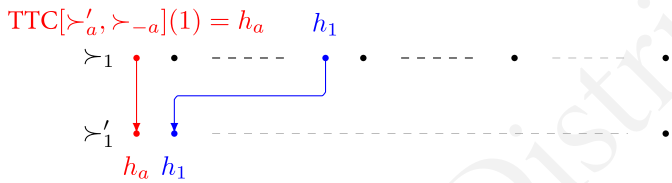
(12) Thus, $\varphi[\succ'_a, \succ_{-a}](a) = h_a$, which is not $h_k = \text{TTC}[\succ'_a, \succ_{-a}](a)$.

(13) Summary:

$$\begin{aligned} &\text{TTC}[\succ'_a, \succ_{-a}] = \text{TTC}[\succ], \\ &\varphi[\succ'_a, \succ_{-a}](a) = h_a. \end{aligned}$$

(14) Since $\varphi[\succ'_a, \succ_{-a}](a) = h_a$, we have $\varphi[\succ'_a, \succ_{-a}](1) \neq h_a = \text{TTC}[\succ'_a, \succ_{-a}](1)$. Thus, $\text{TTC}[\succ'_a, \succ_{-a}](1) = h_a \succ_1 \varphi[\succ'_a, \succ_{-a}](1)$.

(15) Consider a new preference $\succ'_1 : h_a, h_1$.



(16) Similarly, at $[\succ'_a, \succ'_1, \succ_{-a-1}]$, agent 1 is assigned h_a under TTC ($a \to k \to \cdots \to 1 \to a$ is still a cycle), but is assigned h_1 under φ ($\varphi[\succ'_a, \succ'_1, \succ_{-a-1}](1) = h_a = \text{TTC}[\succ'_a, \succ_{-a}](1) \succ_1 \varphi[\succ'_a, \succ_{-a}](1)$).

(17) Summary:

$$\begin{aligned} &\text{TTC}[\succ'_a, \succ'_1, \succ_{-a-1}] = \text{TTC}[\succ'_a, \succ_{-a}] = \text{TTC}[\succ], \\ &\varphi[\succ'_a, \succ'_1, \succ_{-a-1}](a) = h_1. \end{aligned}$$

(18) By induction, at $\succ' = [\succ'_a, \succ'_1, \ldots, \succ'_k]$, $\text{TTC}[\succ'] = \text{TTC}[\succ]$, but $\varphi[\succ'](i) = h_i$ for each $i \in \{a, 1, \ldots, k\}$, violating the Pareto efficiency of φ.

(19) By induction on the steps of cycles, we complete the proof.

□

4.39 Theorem 4.37 is "robust" via the following three examples.

4.40 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:

Then both

$$\text{TTC}[\succ] = \begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_1 & h_3 \end{bmatrix} \text{ and } \mu = \begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_3 & h_1 \end{bmatrix}$$

are individually rational, and Pareto efficient under $\succ$.

a_1	a_2	a_3
h_2	h_1	h_1
h_3	h_3	h_3
h_1	h_2	

Table 4.7

Define a mechanism for this market

$$\varphi[\succ'] = \begin{cases} \mu, & \text{if } \succ' = \succ; \\ \mathrm{TTC}[\succ'], & \text{otherwise.} \end{cases}$$

Now φ is not strategy-proof.

4.41 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.

4.42 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 4.8

$$\varphi[\succ] = \begin{bmatrix} a_1 & a_2 \\ h_2 & h_1 \end{bmatrix} \neq \begin{bmatrix} a_1 & a_2 \\ h_1 & h_2 \end{bmatrix} = \mathrm{TTC}[\succ],$$

and is not individually rational.

Chapter 5

House allocation

Contents

5.1 The former model

5.1 The house allocation problem was 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 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 SD^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, SD^f, is Pareto efficient.

Proof. (1) Suppose that there is a matching ν that Pareto dominates $\mathrm{SD}^f[\succ]$.

(2) Consider the agent $a = f(i)$ with the highest priority who obtains a strictly better house in ν than in $\mathrm{SD}^f[\succ]$.

(3) Then $\nu(a) = \mathrm{SD}^f[\succ](b)$ for some agent $b = f(j)$ with $j < i$.

(4) By assumption, a is the agent with highest priority such that $\nu(a) \succ_a \mathrm{SD}^f[\succ](a)$, so $\nu(b) \succ_b \mathrm{SD}^f[\succ](b)$ is impossible.

(5) Since ν Pareto dominates $\mathrm{SD}^f[\succ]$, $\nu(b) \succsim_b \mathrm{SD}^f[\succ](b)$.

(6) Therefore, $\nu(b) = \mathrm{SD}^f[\succ](b)$, which leads to a contradiction.

□

5.8 Core from assigned endowments μ, denoted by TTC^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,

$$\mathrm{TTC}^\mu[\succ] = \mathrm{TTC}[\succ, \mu].$$

5.9 Theorem (Lemma 1 in Abdulkadiroğlu and Sönmez (1998)): For any house allocation problem $\langle A, H, \succ \rangle$, for any ordering f, and for 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.

Given a house allocation problem $\langle A, H, \succ \rangle$, let $\mathrm{SD}^{\mathcal{F}} = \{\nu \in \mathcal{M} \mid \mathrm{SD}^f[\succ] = \nu \text{ for some } f \in \mathcal{F}\}$, and $\mathrm{TTC}^{\mathcal{M}} = \{\nu \in \mathcal{M} \mid \mathrm{TTC}^\mu[\succ] = \nu \text{ for some } \mu \in \mathcal{M}\}$. Then it suffices to show

$$\mathrm{TTC}^{\mathcal{M}} = \mathrm{SD}^{\mathcal{F}} = \mathscr{E}.$$

5.10 *Proof of Theorem 5.9, Step 1:* "$\mathrm{TTC}^{\mathcal{M}} \subseteq \mathrm{SD}^{\mathcal{F}}$".

(1) Let $\nu \in \mathrm{TTC}^{\mathcal{M}}$. Then there exists $\mu \in \mathcal{M}$ with $\nu = \mathrm{TTC}^\mu[\succ]$.

(2) Let Step t be the last step of top trading cycles algorithm and let $\{A^1, A^2, \ldots, A^t\}$ be the cycle structure.

(3) For each $k = 1, 2, \ldots, t$ and each $a \in A^k$, we have

$$\mathrm{Br}_a(H \setminus \cup_{\ell=0}^{k-1} H^\ell) = \mathrm{TTC}^\mu[\succ](a) = \nu(a).$$

(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 $\mathrm{SD}^f[\succ](f(i)) = \nu(f(i))$.

(6) By top trading cycles algorithm and the construction of f, we have

$$\mathrm{SD}^f[\succ](f(1)) = \mathrm{Br}_{f(1)}(H) = \mathrm{TTC}^\mu[\succ](f(1)) = \nu(f(1)).$$

(7) Suppose that $\mathrm{SD}^f[\succ](f(j)) = \nu(f(j))$ for all $j = 1, 2, \ldots, i-1$ where $2 \leq i \leq n$.

(8) Let $f(i) \in A^k$. We have the following:

- By top trading cycles algorithm, we have

$$\mathrm{Br}_{f(i)}(H \setminus \cup_{\ell=0}^{k-1} H^\ell) = \mathrm{TTC}^\mu[\succ](f(i)) = \nu(f(i)).$$

- 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,

$$\nu(f(i)) = \mathrm{Br}_{f(i)}(H \setminus \cup_{\ell=0}^{k-1} H^\ell) \succsim_{f(i)} \mathrm{Br}_{f(i)} \left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right) \succsim_{f(i)} \nu(f(i)),$$

and hence

$$\nu(f(i)) = \mathrm{Br}_{f(i)} \left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right).$$

(10) It follows that

$$\nu(f(i)) = \mathrm{Br}_{f(i)} \left(H \setminus \cup_{j=1}^{i-1} \nu(f(j))\right) = \mathrm{Br}_{f(i)} \left(H \setminus \cup_{j=1}^{i-1} \mathrm{SD}^f[\succ](f(j))\right) = \mathrm{SD}^f[\succ](f(i)).$$

□

5.11 *Proof of Theorem 5.9, Step 2: "$\varphi^{\mathcal{F}} \subseteq \mathscr{E}$".* See Proposition 5.7. □

5.12 *Proof of Theorem 5.9, Step 3: "$\mathscr{E} \subseteq \mathrm{TTC}^{\mathcal{M}}$".*

(1) Let $\mu \in \mathscr{E}$. Consider the mechanism TTC^μ.

(2) Since $\mathrm{TTC}^\mu[\succ] = \mathrm{TTC}[\succ, \mu]$, TTC^μ is individually rational. That is, for all $a \in A$, $\mathrm{TTC}^\mu[\succ](a) \succsim_a \mu(a)$.

(3) Since μ is Pareto efficient and the preference profile is strict, we have $\mathrm{TTC}^\mu[\succ] = \mu$, which in turn implies $\mu \in \mathrm{TTC}^{\mathcal{M}}$, completing the proof of "$\mathscr{E} \subseteq \mathrm{TTC}^{\mathcal{M}}$."

□

5.13 Theorem (Theorem 1 in Abdulkadiroğlu and Sönmez (1998)): For any house allocation problem $\langle A, H, \succ\rangle$, 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 $\mathcal{M}^\nu = \{\mu \in \mathcal{M} \mid \mathrm{TTC}^\mu[\succ] = \nu\}$ and $\mathcal{F}^\nu = \{f \in \mathcal{F} \mid \mathrm{SD}^f[\succ] = \nu\}$.

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 $\tilde{A}[\mu] = \{A^1[\mu], A^2[\mu], \ldots, A^{t_\mu}[\mu]\}$ for the housing market $\langle A, H, \succ, \mu\rangle$.

(2) For all $k = 2, 3, \ldots, t_\mu$, partition $A^k[\mu]$ into its chains as in Remark 4.25.

(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 head.

  (ii) Order the chains based on the index of the endowments of the tails of the chains (starting the chain whose tail 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 $\mathrm{TTC}^\mu[\succ] = \nu$.

(2) By top trading cycles algorithm, for each $k = 1, 2, \ldots, t_\mu$, for each $a \in A^t[\mu]$, we have

$$\mathrm{Br}_a\left(H \setminus \cup_{\ell=0}^{k-1} H^\ell\right) = \mathrm{TTC}^\mu[\succ](a) = \nu(a).$$

(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 $\mathrm{SD}^{f(\mu)}$, 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 that $f = f(\mu) = f(\mu')$ orders the agents as $a_1, a_2, \dots, a_n$.

(2) Let

$$\tilde{A}[\mu] = \Big\{ \underbrace{\{a_1, \dots, a_{m_1}\}}_{A^1[\mu]}, \underbrace{\{a_{m_1+1}, \dots, a_{m_2}\}}_{A^2[\mu]}, \dots, \underbrace{\{a_{m_{k-1}+1}, \dots, a_{m_k}\}}_{A^k[\mu]}, \dots, \underbrace{\{a_{m_t-1}, \dots, a_n\}}_{A^t[\mu]} \Big\},$$

$$\tilde{A}[\mu'] = \Big\{ \underbrace{\{a_1, \dots, a_{m'_1}\}}_{A^1[\mu']}, \underbrace{\{a_{m'_1+1}, \dots, a_{m'_2}\}}_{A^2[\mu']}, \dots, \underbrace{\{a_{m'_{k-1}+1}, \dots, a_{m'_k}\}}_{A^k[\mu']}, \dots, \underbrace{\{a_{m'_{t'}-1}, \dots, a_n\}}_{A^{t'}[\mu']} \Big\}.$$

We want to show that $t = t'$ and $A^k[\mu] = A^k[\mu']$ for all $k = 1, 2, \dots, t$. We proceed by induction.

(3) Suppose that $A^1[\mu] \neq A^1[\mu']$. 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

$$\mathrm{Br}_{a_{m'_1+1}}(H) = \mathrm{TTC}^{\mu}[\succ](a_{m'_1+1}) = \nu(a_{m'_1+1}).$$

(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 head of her chain, and hence $a_{m'_1+1} \in U^2[\mu']$.

(7) Therefore

$$\mathrm{Br}_{a_{m'_1+1}}(H) \neq \mathrm{Br}_{a_{m'_1+1}}(H \setminus H^1[\mu']) = \mathrm{TTC}^{\mu'}[\succ](a_{m'_1+1}) = \nu(a_{m'_1+1}),$$

which leads to a contradiction.

(8) Therefore $A^1[\mu] = A^1[\mu']$.

(9) Suppose that $A^{\ell}[\mu] = A^{\ell}[\mu']$ for all $\ell = 1, 2, \dots, k-1$ 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

$$\mathrm{Br}_{a_{m'_k+1}}(H \setminus \cup_{\ell=0}^{k-1} H^{\ell}[\mu]) = \mathrm{TTC}^{\mu}[\succ](a_{m'_k+1}) = \nu(a_{m'_k+1}).$$

(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 head 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

$$\nu(a_{m'_k+1}) = \mathrm{TTC}^{\mu'}[\succ](a_{m'_k+1}) \in H^{k+1}[\mu'],$$

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 head 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 tails 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, tails 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,

$$\sum_{\nu \in \mathrm{SD}^{\mathcal{F}}} |\mathcal{F}^\nu| \geq \sum_{\nu \in \mathrm{TTC}^{\mathcal{M}}} |\mathcal{M}^\nu|.$$

(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 Neutrality

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, $g \succ_a^\sigma h$ 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

This means that the "real" outcome of a neutral mechanism is independent of the names of the indivisible goods.

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 $\succ_2$, then $\varphi[\succ](1)$ is the best element in $\{b, c\}$ according to $\succ_1$, $\varphi[\succ](2) = a$ and $\varphi[\succ](3)$ is the remaining element. If all other cases, $\varphi[\succ](1)$ is the best element in H according to $\succ_1$, $\varphi[\succ](2)$ is the best element in $H \setminus \{\varphi[\succ](1)\}$ according to $\succ_2$ 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.

Consider the following preference profile $\succ$:

1	2	3
a	a	a
b	b	b
c	c	c

Table 5.1

Then the matching produced by φ is

$$\mu = \begin{bmatrix} 1 & 2 & 3 \\ b & a & c \end{bmatrix}.$$

Now consider the permutation σ: $\sigma(a) = b$, $\sigma(b) = c$ and $\sigma(c) = a$. Then $\sigma\big(\varphi[\succ](2)\big) = b$. On the other hand, the new preference profile $\succ^\sigma$ is as follows:

1	2	3
b	b	b
c	c	c
a	a	a

Table 5.2

Then the matching produced by φ is

$$\mu' = \begin{bmatrix} 1 & 2 & 3 \\ b & c & a \end{bmatrix}.$$

Thus, $\varphi[\succ^\sigma](2) = c \neq b = \sigma\big(\varphi[\succ](2)\big)$.

5.24 Example: One-sided DA is not neutral.

i	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 5.3

The matching produced by DA is

$$\mu = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

If we exchange the labels of a and b, the problem becomes

i	j	k	a	b
a	b	b	i	k
b		a	j	i
			k	

Table 5.4

The matching produced by DA is

$$\mu = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

Thus, one-sided DA is not neutral.

☞ 5.25 Definition: A mechanism φ is non-bossy if for any $\succ$, $a \in A$ and $\succ'_a$,

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

5.26 Lemma (Lemma 1 in Svensson (1999)): Let φ be a strategy-proof and non-bossy 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 non-bossiness 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.

□

5.27 Theorem (Theorem 1 in Svensson (1999)): A mechanism φ is strategy-proof, non-bossy 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 5.26, we have $\varphi[\succ''] = \varphi[\succ']$.

□

5.28 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 8.18 and Theorem 5.27. □

5.5 Consistency

5.29 For any problem $\Gamma = \langle A, H, \succ\rangle$, any non-empty 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.

☞ 5.30 Definition: A mechanism φ is consistent[1] if for each problem $\Gamma = \langle A, H, \succ\rangle$ and for each non-empty subset A' of A, one has

$$\varphi[\Gamma](a) = \varphi[r^{\varphi[\Gamma]}_{A'}(\Gamma)](a) \text{ for each } a \in A'.$$

[1] A mechanism is consistent if the assignment is unchanged if the mechanism is implemented on a sub-problem after one removes some agents and their assignment.

A mechanism φ is pairwise consistent if for any problem $\Gamma = \langle A, H, \succ \rangle$, any non-empty subset A' of A with even cardinality, and any allocation μ, one has

$$\varphi[\Gamma](a) = \varphi[r_{A'}^{\varphi[\Gamma]}(\Gamma)](a) \text{ for each } a \in A'.$$

5.31 Example: DA is not consistent.

i	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 5.5

☞ 5.32 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.

✌ 5.33 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

5.34 Question: How about the fairness of simple serial dictatorship and core from assigned endowments?

5.35 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.

5.36 Random priority (or random serial dictatorship):

Phase 1: Draw each orderings of the agents with equal probability.

Phase 2: Run simple serial dictatorship according to the selected ordering.

Mathematically, random priority is defined as

$$\mathrm{RP}[\succ] = \frac{1}{n!}\sum_{f\in\mathcal{F}} p^{\mathrm{SD}^f[\succ]} \text{ for each } \succ,$$

where $p^{\mathrm{SD}^f[\succ]}$ is the lottery that assigns probability 1 to matching $\mathrm{SD}^f[\succ]$.

5.37 Core from random endowments:

Phase 1: Draw each initial assignment with equal probability.

Phase 2: Run TTC according to the selected initial assignment.

Mathematically, core from random endowments is defined as

$$\varphi^{\mathrm{cre}}[\succ] = \frac{1}{n!}\sum_{\mu\in\mathcal{M}} p^{\mathrm{TTC}^\mu[\succ]} \text{ for each } \succ,$$

where $p^{\mathrm{TTC}^\mu[\succ]}$ is the lottery that assigns probability 1 to matching $\mathrm{TTC}^\mu[\succ]$.

5.38 Theorem (Theorem 2 in Abdulkadiroğlu and Sönmez (1998)): Random priority and core from random endowments coincide.

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.

The model is a generalization of both the housing market and the house allocation problem.

☞ 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 decided. The ordering may be randomly chosen from a given distribution of orderings or may favor some subgroup of agents (for example, seniors over juniors).

(2) Houses in G are assigned to these agents based on the simple serial dictatorship induced by f under the reported preference profile.

6.10 Problems of random serial dictatorship with squatting rights:

- Since this algorithm does not guarantee each existing tenant a house that is at least as good as her own, it may be not individual rational.
- Some of agents may choose to stay "Out" (*i.e.*, use their squatting rights), and this may result in the loss of potentially large gains from trade. Thus, the resulting matching may not be Pareto efficient.

6.11 Exercise: How about the strategy-proofness of the random serial dictatorship with squatting rights?

6.2.2 Random serial dictatorship with waiting list

6.12 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.

6.13 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 available houses for agents a_1 and a_2 both are $\emptyset$. The set of acceptable available 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

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ h_3 & h_1 & h_4 \end{bmatrix}.$$

6.14 Random serial dictatorship with waiting list is inefficient.

Consider the example in the previous item. The outcome is Pareto dominated by

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ h_2 & h_3 & h_1 \end{bmatrix}.$$

6.15 Exercise: Is random serial dictatorship with waiting list individually rational and strategy-proof?

6.16 Question: How about the algorithm when agents are not removed?

6.2.3 MIT-NH4 mechanism

6.17 The following mechanism is used at the residence NH4 of MIT.

6.18 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.

6.19 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 $H_V = \{h_5\}$. 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

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_5 & h_2 & h_3 & h_4 & h_1 \end{bmatrix}.$$

6.20 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

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_3 & h_2 & h_5 & h_4 & h_1 \end{bmatrix} \text{ and } \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_4 & h_2 & h_5 & h_3 & h_1 \end{bmatrix}.$$

6.21 Exercise: Is the MIT-NH4 mechanism individually rational and strategy-proof?

6.22 Question: Is there any other way to resolve the squatting conflict? In particular, how about the way that a_4 is assigned h_5? Hint: Compare with TTC^f in 6.23.

6.3 Top trading cycles algorithm

6.23 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 TTC^f to denote the top trading cycles mechanism induced by the ordering f.

6.24 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:

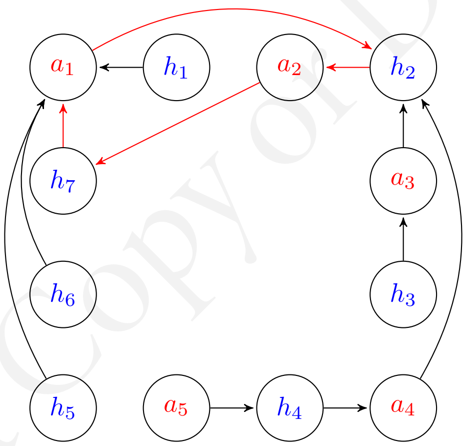


Figure 6.1: Step 1

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

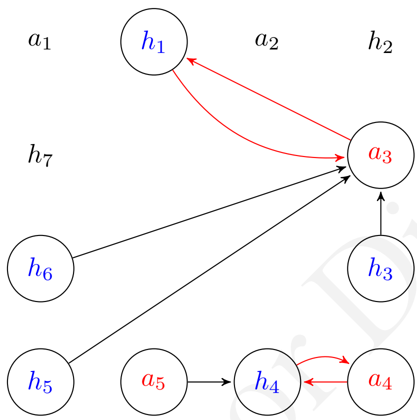


Figure 6.2: Step 2

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

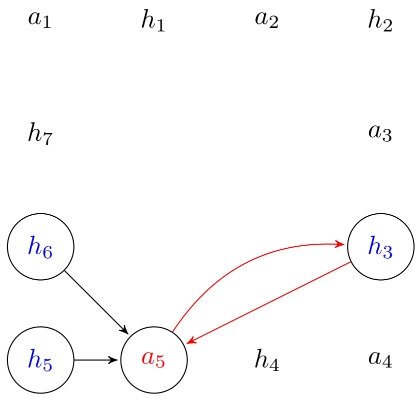


Figure 6.3: Step 3

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:

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ h_2 & h_7 & h_1 & h_4 & h_3 \end{bmatrix}$$

6.25 Theorem (Proposition 1 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism TTC^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 TTC^f is Pareto efficient.

□

6.26 Theorem (Proposition 2 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism TTC^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.

□

6.27 Theorem (Theorem 1 in Abdulkadiroğlu and Sönmez (1999)): For any ordering f, the induced top trading cycles mechanism TTC^f is strategy-proof.

Proof. The proof is analogous to the proof of Theorem 4.27. □

6.28 There is another version of TTC.

Top trading cycles algorithm, induced by a given initial endowment μ.

Phase 1: Construct an initial allocation μ by

- assigning each existing tenant her own house,
- randomly assigning the vacant houses to newcomers with uniform distribution.

Phase 2: Run TTC for the induced housing market to determine the final outcome.

We use TTC^μ to denote the top trading cycles mechanism induced by the initial endowment μ.

6.29 Unless otherwise mentioned, TTC always refers to TTC with an ordering rather than TTC with an initial endowment.

6.30 It is clear that TTC^μ is Pareto efficient, individual rational, and strategy-proof.

6.31 Exercise: What is the difference between TTC^f and TTC^μ.

Hint: There is a hidden bias in TTC^μ. In TTC^μ, an initial allocation is constructed by assigning each existing tenant her current house and randomly assigning vacant houses to newcomers. This might be interpreted as granting property rights of vacant houses to newcomers. Therefore existing tenants who also have claims on vacant houses give up these claims under TTC^μ.

Consider the following house allocation with existing tenants: $A_E = \{a_1, a_2\}$, $A_N = \{a_3\}$, $H_O = \{h_1, h_2\}$, and $H_V = \{h_3\}$. Here the existing tenant a_i occupies the house h_i for $i = 1, 2$. 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_3	h_3	h_3
h_2	h_2	h_2
h_1	h_1	h_1

Table 6.6

Then $\text{TTC}^f[\succ] = \begin{bmatrix} a_1 & a_2 & a_3 \\ h_3 & h_2 & h_1 \end{bmatrix}$.

On the other hand, the unique possible initial endowment μ is $\begin{bmatrix} a_1 & a_2 & a_3 \\ h_1 & h_2 & h_3 \end{bmatrix}$, and the resulting matching $\mathrm{TTC}^{\mu}[\succ]$ is $\begin{bmatrix} a_1 & a_2 & a_3 \\ h_1 & h_2 & h_3 \end{bmatrix}$. For agent a_1, the outcome under $\mathrm{TTC}^f[\succ]$ is better than the outcome under $\mathrm{TTC}^{\mu}[\succ]$.

6.32 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. If $\varphi[\succ](f(i)) \succ_{f(i)} \mathrm{TTC}^f[\succ](f(i))$ for some $\succ$ and i, then there exists $\succ'$ and $j < i$ such that $\mathrm{TTC}^f[\succ'](f(j)) \succ'_{f(j)} \varphi[\succ'](f(j))$.

6.33 Interpretation:

- As far as agent $f(1)$ is concerned, TTC^f assigns him a house that is at least as good as the assignment of any Pareto efficient, individual rational, and strategy-proof mechanism at all preference profiles.
- Next consider all Pareto efficient, individual rational, and strategy-proof mechanisms that perform equally well for agent $f(1)$. TTC^f assigns agent $f(2)$ a house that is at least as good as the assignment of any such mechanism at all preference profiles.
- In general, consider all Pareto efficient, individual rational, and strategy-proof mechanisms that perform equally well for agents $f(1)$, $f(2)$, ..., $f(k)$ where $k < |A|$. TTC^f assigns agent $f(k+1)$ a house that is at least as good as the assignment of any such mechanism at all preference profiles.

6.34 Remark: There are many applications where agents are naturally ordered based on their seniority. Let f denote this ordering. Then Theorem 6.32 shows that there is no Pareto efficient, individually rational and strategy-proof mechanism which always better respects the seniority of the agents than TTC^f.

6.4 You request my house—I get your turn algorithm

6.35 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.

6.36 The YRMH-IGYT algorithm generalizes simple serial dictatorship and TTC:

- The YRMH-IGYT algorithm coincides with simple serial dictatorship when there are no existing tenants: Without existing tenants, the "you request my house ..." contingency simply does not happen, so the mechanism coincides with simple serial dictatorship.
- The YRMH-IGYT algorithm coincides with TTC when all agents are existing tenants and there is no vacant house: In this case, an agent's request always points to a house owned by someone, and the assignment of a house happens if and only if there is a cycle made of existing tenants.

6.37 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 6.7

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.

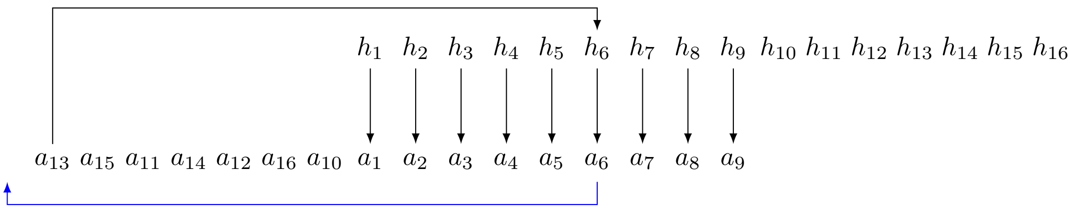


Figure 6.4: Step 1

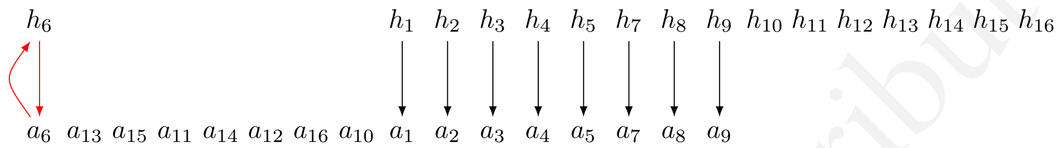


Figure 6.5: Step 2

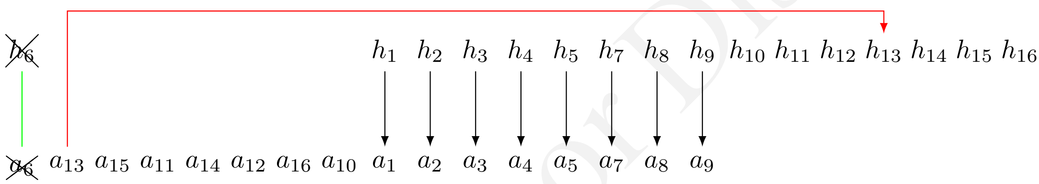


Figure 6.6: Step 3

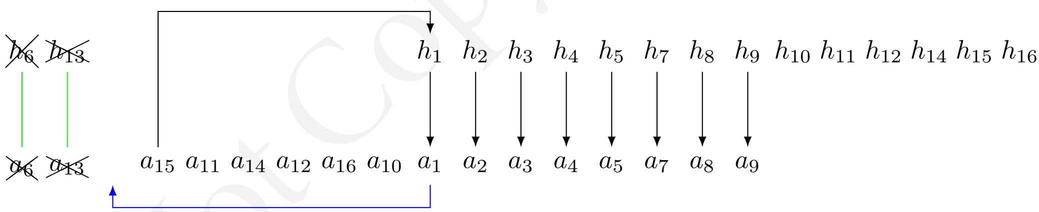


Figure 6.7: Step 4

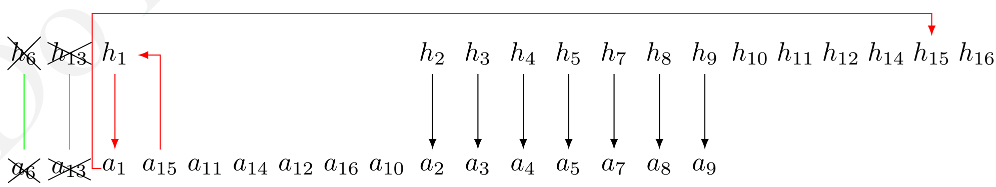


Figure 6.8: Step 5

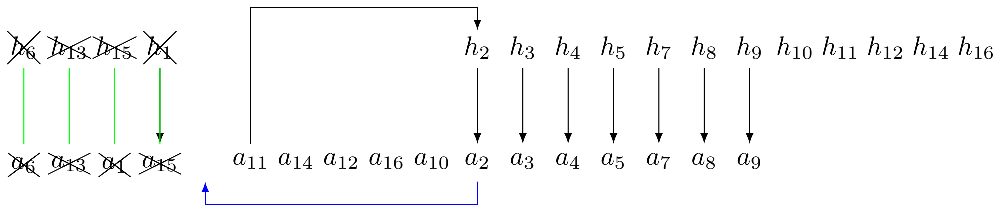


Figure 6.9: Step 6

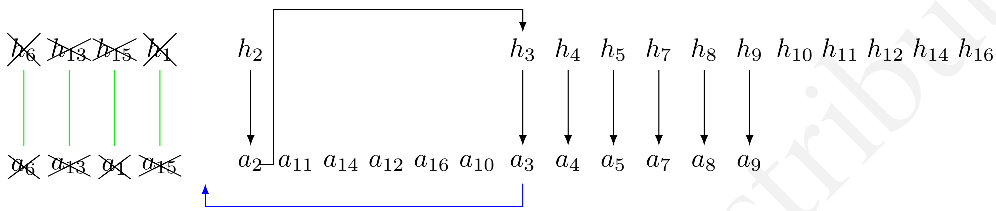


Figure 6.10: Step 7

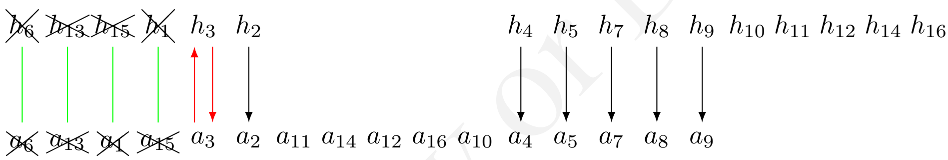


Figure 6.11: Step 8

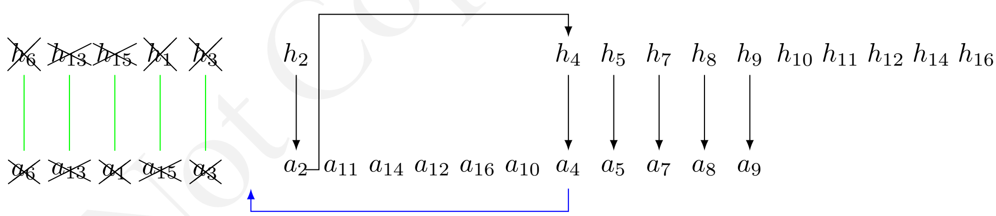


Figure 6.12: Step 9

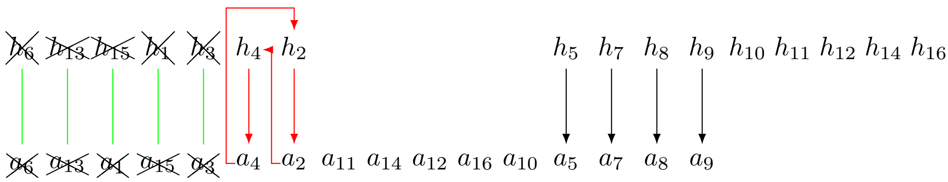


Figure 6.13: Step 10

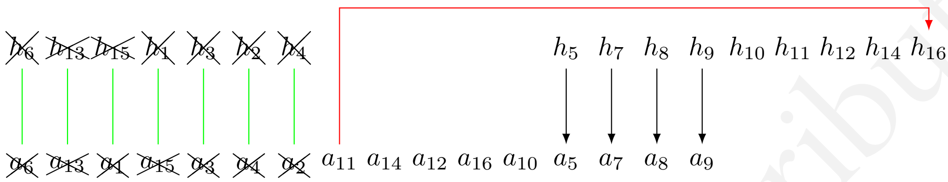


Figure 6.14: Step 11

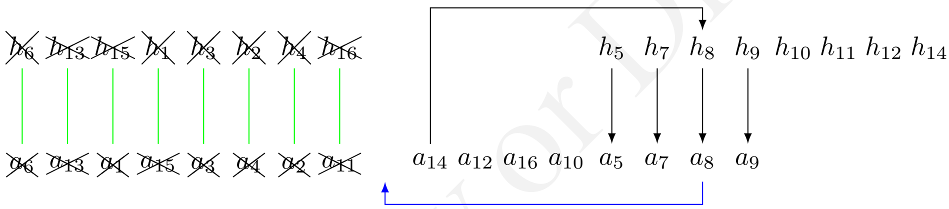


Figure 6.15: Step 12

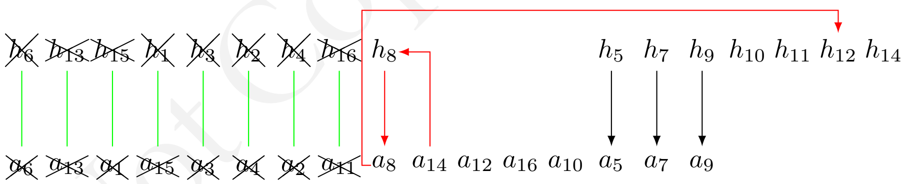


Figure 6.16: Step 13

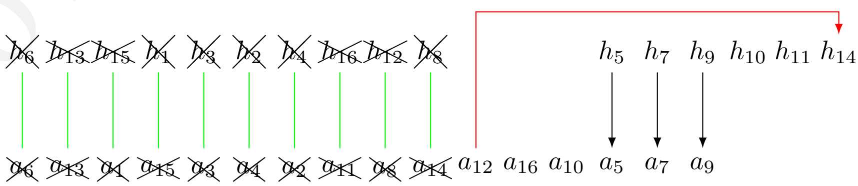


Figure 6.17: Step 14

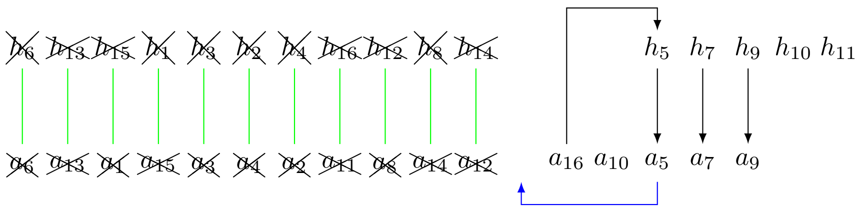


Figure 6.18: Step 15

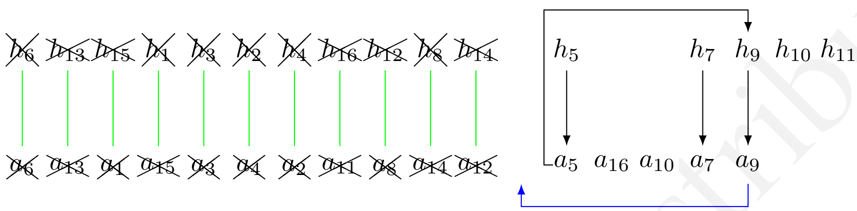


Figure 6.19: Step 16

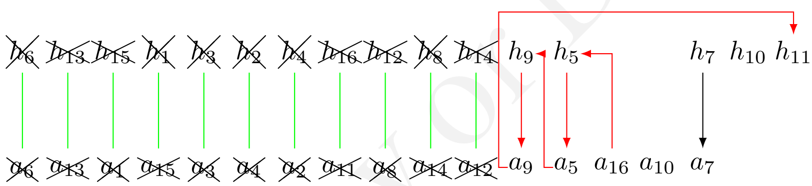


Figure 6.20: Step 17

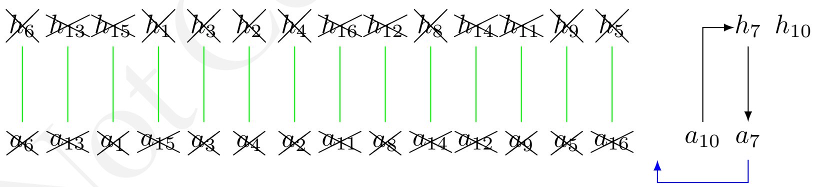


Figure 6.21: Step 18

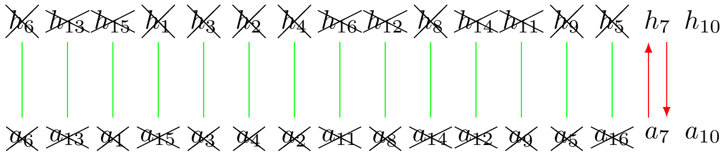


Figure 6.22: Step 19

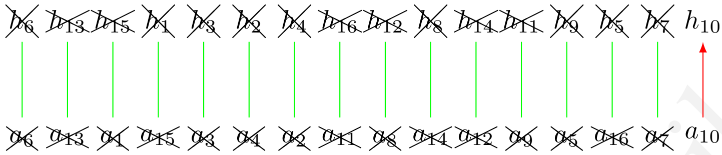


Figure 6.23: Step 20

The outcome of the algorithm is

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

6.38 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

6.39 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 $\sigma^{-1}(g) \succ_a \sigma^{-1}(h)$.

☞ 6.40 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

$$\varphi[\succ^\sigma](a) = \sigma\big(\varphi[\succ](a)\big) \text{ for any } a \in A.$$

6.41 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.

☞ 6.42 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

$$\varphi[\Gamma](a) = \varphi\left[r^{\varphi[\Gamma]}_{A',H'}(\Gamma)\right](a) \text{ for each } a \in A'.$$

✌ 6.43 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

6.44 Here we assume that $|A_E| = n$ and $|A_N| = |H_V| = m$.

6.45 Let $\mathcal{M}^* = \{\mu \in \mathcal{M} \mid \mu(a) = h_a$ for all $a \in A_E\}$ be the set of matchings which assign each existing tenant her current house.

Core from random endowments, φ^{cre}, is defined as

$$\varphi^{\text{cre}} = \frac{1}{m!} \sum_{\mu \in \mathcal{M}^*} \text{TTC}^{\mu} .$$

6.46 Let $\mathcal{F}^* = \{f$ is an ordering of agents $\mid f^{-1}(a) < f^{-1}(a')$ for all $a \in A_N$ and $a' \in A_E\}$.

Define a new mechanism as follows

$$\psi = \frac{1}{n!m!} \sum_{f \in \mathcal{F}^*} \text{TTC}^{f} .$$

6.47 Theorem (Theorem 1 in Sönmez and Ünver (2005)): φ^{cre} and ψ are equivalent.

Proof. Omitted. □

6.48 The TTC induced by initial endowments is equivalent to an extreme case of TTC induced by orderings where newcomers are randomly ordered first and existing tenants are randomly ordered next.

6.49 Question: Let $\mathcal{F}$ be the set of all orderings. Are $\frac{1}{(m+n)!} \sum_{f \in \mathcal{F}} \text{TTC}^f$ and $\frac{1}{n!m!} \sum_{f \in \mathcal{F}^*} \text{TTC}^f$ equivalent?

Chapter 7

Random assignment mechanism

Contents

7.1 Random assignment problem

☞ 7.1 A random assignment problem, denoted by $\Gamma = \langle N, O, \succ\rangle$, consists of

- $N = \{1, 2, \ldots, n\}$ is a finite set of agents,
- $O = \{o_1, o_2, \ldots, o_n\}$ 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 $a = b$.

☞ 7.2 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 deterministic assignments.

An example for deterministic assignment:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here agent 1 gets object 1, agent 2 gets object 3, agent 3 gets object 2, and agent 4 gets object 4.

☞ 7.3 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.

Let $\mathcal{R}$ denote the set of random assignments.

An example for random assignment:

$$P = \begin{pmatrix} \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \end{pmatrix}.$$

Here agent 1 gets object 1 with probability $\frac{5}{12}$, object 2 with probability $\frac{1}{12}$, object 3 with probability $\frac{5}{12}$, and object 4 with probability $\frac{1}{12}$.

7.4 For each agent, a lottery of objects[1] is a probability distribution over the set of objects.

[1] In expected utility theory, a lottery is a discrete distribution of probability on a set of states of nature. The elements of a lottery correspond to the probabilities that each of the states of nature will occur.

Since there are n objects, a lottery can be written as a n-dimensional vector such that the j-th component is the probability that agent receives the j-th object.

For each random assignment P, the i-th row P_i is clearly an agent i's lottery of objects.

☞ 7.5 A random assignment mechanism is a procedure the assigns a random assignment P for each random assignment problem $\langle N, O, \succ\rangle$.

7.6 A von Neumann-Morgenstern utility function[2] u_i is a real-valued mapping from O to $\mathbb{R}$.

We extend the domain of u_i to the set of lotteries as follows. Agent i's expected utility for the lottery P_i is

$$u_i(P_i) = \sum_{o\in O} P_{io} \cdot u_i(o) = P_i \cdot u_i,$$

where $u_i = \big(u_i(o_1), u_i(o_2), \cdots, u_i(o_n)\big)$.

We say that u_i is consistent/compatible with $\succ_i$ when $u_i(a) > u_i(b)$ if and only if $a \succ_i b$.[3]

Example: There are three objects $\{a, b, c\}$ and agent 1's preference is $a \succ b \succ c$. Then $(1, \frac{1}{3}, 0)$ and $(1, \frac{2}{3}, 0)$ are two consistent utility functions.

7.2 Random priority mechanism

7.7 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 a preference profile $\succ$, the corresponding simple serial dictatorship assignment is denoted by $\mathrm{SD}^f[\succ]$, defined as usual.

7.8 Random priority (or random serial dictatorship):

Step 1: Draw each orderings of the agents with equal probability.

Step 2: Run simple serial dictatorship according to the selected ordering.

Mathematically, random priority is defined as

$$\mathrm{RP} = \frac{1}{n!} \sum_{f\in\mathcal{F}} \mathrm{SD}^f .$$

[2]In decision theory, the von Neumann-Morgenstern utility theorem shows that, under certain axioms of rational behavior (completeness, transitivity, continuity, and independence), a decision-maker faced with risky (probabilistic) outcomes of different choices will behave as if he is maximizing the expected value of some function defined over the potential outcomes at some specified point in the future. This function is known as the von Neumann-Morgenstern utility function. The theorem is the basis for expected utility theory.

[3]In economics, an ordinal utility function is a function representing the preferences of an agent on an ordinal scale. The ordinal utility theory claims that it is only meaningful to ask which option is better than the other, but it is meaningless to ask how much better it is or how good it is. All of the theory of agent decision-making under conditions of certainty can be, and typically is, expressed in terms of ordinal utility.

7.9 Core from random endowments:

Step 1: Draw each initial assignment with equal probability.

Step 2: Run TTC according to the selected initial assignment.

Mathematically, core from random endowments is defined as

$$\varphi^{\text{cre}} = \frac{1}{n!} \sum_{\mu \in \mathcal{M}} \text{TTC}^{\mu} .$$

7.10 Theorem (Theorem 2 in Abdulkadiroğlu and Sönmez (1998)): Random priority and core from random endowments are equivalent.

7.11 Example: There are four agents $\{1, 2, 3, 4\}$ and 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 7.1

The matching produced by RP is

$$P = \begin{pmatrix} \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12} \end{pmatrix}.$$

7.12 The summary of RP:

- Easy to implement and widely used in practice.
- *Ex post* efficient (but not *ex ante* efficient or ordinally efficient).
- Fair (equal treatment of equals).
- Strategy-proof.

7.13 Other related mechanisms: Pathak and Sethuraman (2011).

7.3 Simultaneous eating algorithm and probabilistic serial mechanism

7.14 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:

$$W = \left\{ \omega_i \colon [0,1] \to \mathbb{R}_+ \;\middle|\; \omega_i \text{ is measurable and } \int_0^1 \omega_i(t)\,\mathrm{d}t = 1 \right\}.$$

7.15 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 pre-specified speeds.

For each $o \in O' \subseteq O$, let $N(o, O') = \{i \in N \mid o \succ_i b \text{ for all } b \in O', b \neq o\}$—the set of agents who are eating 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 $t^0 = 0$, $O^0 = O$, $P^0 = \mathbf{0}$ (the $n \times n$ matrix of zeros).

Step k: Suppose that t^0, O^0, P^0, …, t^{k-1}, O^{k-1}, P^{k-1} are already defined.

- For each $o \in O^{k-1}$, 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 instant $t = t^{k-1}$, and $t^k(o)$ specifies the time instant when the object o will be eaten away given that no new agent enters.

- 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 time instant is denoted as

t^k. Note that, at the time 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 time 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 time 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 $\int_0^1 \omega_i(s)\, \mathrm{d}s \leq 1$ guarantees that $P^k_{io} \leq 1$.

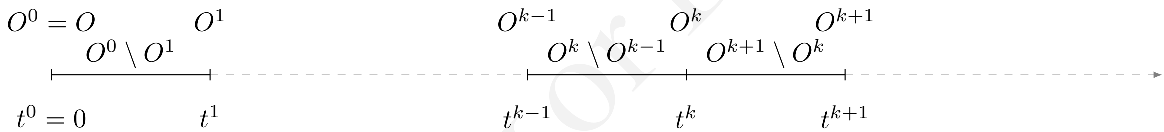


Figure 7.1

7.16 By the construction, $O^k \subsetneq O^{k-1}$ for each k, $O^n = \emptyset$, 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 $\succ$: $P_\omega[\succ] = P^n$.

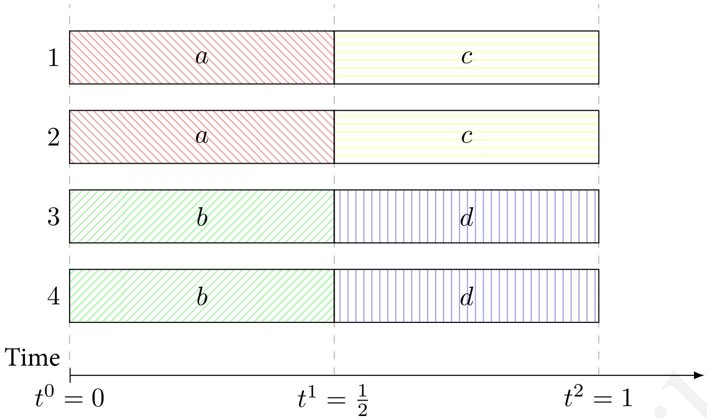
7.17 The probabilistic serial mechanism PS: Simultaneous eating algorithm with uniform eating speeds $\omega_i(t) = 1$ for all $i \in N$, all $t \in [0, 1]$.

7.18 Example: There are four agents $\{1, 2, 3, 4\}$ and 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 7.2

The process of PS is illustrated below



Thus, PS produces the matching

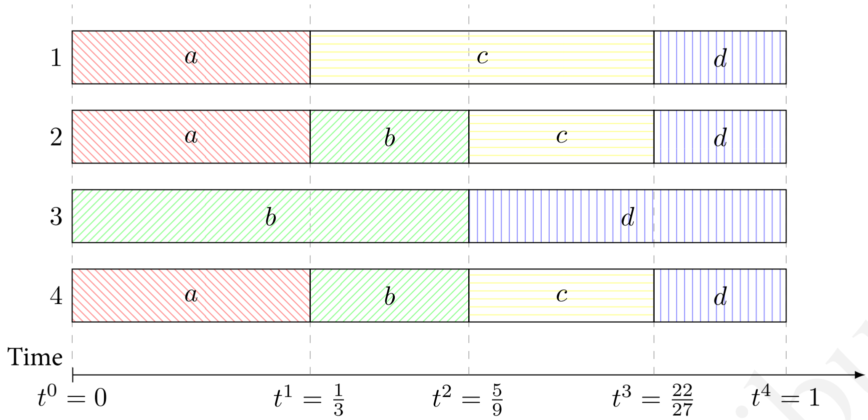
$$Q = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

7.19 Example: There are four agents $\{1, 2, 3, 4\}$ and four objects $\{a, b, c, d\}$. The preferences are as follows:

1	2	3	4
a	a	b	a
c	b	a	b
b	c	d	c
d	d	c	d

Table 7.3

The process of PS is illustrated below



Thus, PS produces the matching

$$Q = \begin{pmatrix} \frac{1}{3} & 0 & \frac{13}{27} & \frac{5}{27} \\ \frac{1}{3} & \frac{2}{9} & \frac{7}{27} & \frac{5}{27} \\ 0 & \frac{5}{9} & 0 & \frac{4}{9} \\ \frac{1}{3} & \frac{2}{9} & \frac{7}{27} & \frac{5}{27} \end{pmatrix}.$$

7.20 Summary of PS:

- Easy to implement.
- Ordinally efficient (not *ex ante* efficient).
- Fair (envy-free, equal treatment of equals).
- Not strategy-proof.

7.4 Efficiency

7.4.1 Basics

☞ 7.21 Given a preference profile $\succ$, a deterministic assignment X Pareto dominates another deterministic assignment Y at $\succ$ if

$$X_i \succsim_i Y_i \text{ for all } i \in N \text{ and } X_{i_0} \succ_{i_0} Y_{i_0} \text{ for some } i_0 \in N,$$

where X_i denotes the object agent i receives under X.

A deterministic assignment X is Pareto efficient at $\succ$ if there is no deterministic assignment that Pareto dominates it at $\succ$.

☞ 7.22 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. That is, there is no random assignment Q such that

$$Q_i \cdot u_i \geq P_i \cdot u_i \text{ for all } i \in N \text{ and } Q_{i_0} \cdot u_{i_0} > P_{i_0} \cdot u_{i_0} \text{ for some } i_0 \in N.$$

- A random assignment P is *ex post* efficient at $\succ$, if it is a convex combination of Pareto efficient deterministic assignments (at $\succ$). That is, it takes the form

$$P = \sum_{\gamma \in \Gamma} \alpha_\gamma \cdot X_\gamma,$$

where $\{\alpha_\gamma\}_{\gamma \in \Gamma}$ is a convex system of weights and each X_γ is a Pareto efficient deterministic assignment at $\succ$.

By Theorem 5.9, P is *ex post* efficient at $\succ$ if and only if it takes the form

$$P = \sum_{f \in \mathcal{F}} \alpha_f \cdot \mathrm{SD}^f[\succ] \text{ for some convex system of weights } \alpha_f,$$

where SD^f is the simple serial dictatorship induced by the ordering f.

7.23 Example: There are four agents $\{1, 2, 3, 4\}$ and 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 7.4

The random assignment $Q = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$ is *ex post* efficient since $Q = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} +$ $\frac{1}{2}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, where $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$ are Pareto efficient at $\succ$.

7.24 Question: In the example above,

(i) is it possible that Q has a form $\sum_{\gamma\in\Gamma}\alpha_\gamma\cdot X_\gamma$ such that $\{\alpha_\gamma\}_{\gamma\in\Gamma}$ is a convex system of weights and some X_γ is not a Pareto efficient deterministic assignment at $\succ$?

(ii) is the random assignment $P=\begin{pmatrix}\frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12}\\ \frac{5}{12} & \frac{1}{12} & \frac{5}{12} & \frac{1}{12}\\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12}\\ \frac{1}{12} & \frac{5}{12} & \frac{1}{12} & \frac{5}{12}\end{pmatrix}$ *ex post* efficient?

7.4.2 Ordinal efficiency

☞ 7.25 Given agent i's preference $\succ_i$, a lottery P_i first-order stochastically dominates another lottery Q_i with respect to $\succ_i$, denoted by $P_i \succsim_i^{sd} Q_i$, if

$$\sum_{k:\ o_k\succsim_i o_j} P_{ik} \geq \sum_{k:\ o_k\succsim_i o_j} Q_{ik} \text{ for all } j.$$

That is, P_i first-order stochastically dominates Q_i if and only if

- the probability of receiving the favorite object is at least as much in P_i as in Q_i, and in general,
- for any j, the probability of receiving one of top j favorite objects is at least as much in P_i as in Q_i.

7.26 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 $u_i=\big(u_i(o_1),u_i(o_2),\ldots,u_i(o_n)\big)$.

Moreover, $P_i\neq Q_i$ implies that the corresponding inequality is strict.

Proof. "$\Rightarrow$": Suppose that P_i first-order stochastically dominates Q_i with respect to $\succ_i$.

(1) Without loss of generality, we assume that $o_1\succ_i o_2\succ_i\cdots\succ_i o_n$.

(2) Then we have

$$\sum_{k=1}^{j} P_{ik}\geq\sum_{k=1}^{j} Q_{ik} \text{ for all } j=1,2,\ldots,n.$$

(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

$$u_i\cdot P_i=\sum_{k=1}^{n} u_i(o_k)P_{ik}$$

$$
\begin{aligned}
&= u_i(o_n)\sum_{k=1}^{n} P_{ik} + [u_i(o_{n-1}) - u_i(o_n)]\sum_{k=1}^{n-1} P_{ik} + [u_i(o_{n-2}) - u_i(o_{n-1})]\sum_{k=1}^{n-2} P_{ik} + \cdots \\
&\quad + [u_i(o_j) - u_i(o_{j+1})]\sum_{k=1}^{j} P_{ik} + \cdots + [u_i(o_1) - u_i(o_2)]\sum_{k=1}^{1} P_{ik} \\
&\geq u_i(o_n)\sum_{k=1}^{n} Q_{ik} + [u_i(o_{n-1}) - u_i(o_n)]\sum_{k=1}^{n-1} Q_{ik} + \cdots + [u_i(o_1) - u_i(o_2)]\sum_{k=1}^{1} Q_{ik} \\
&= u_i \cdot Q_i
\end{aligned}
$$

"$\Leftarrow$": Suppose that $u_i \cdot P_i \geq u_i \cdot Q_i$ for any von Neumann-Morgenstern utility function u_i consistent with $\succ_i$.

(1) Without loss of generality, we assume $o_1 \succ_i o_2 \succ_i \cdots \succ_i o_n$. Then it suffices to show that

$$\sum_{k=1}^{j} P_{ik} \geq \sum_{k=1}^{j} Q_{ik} \text{ for all } j = 1, 2, \ldots, n.$$

(2) Assume that $1 \leq \ell \leq n$ is the first number such that $\sum_{k=1}^{\ell} P_{ik} < \sum_{k=1}^{\ell} Q_{ik}$.

(3) Take $\varepsilon > 0$ and construct a von Neumann-Morgenstern utility function u_i such that

$$0 < u_i(o_j) - u_i(o_{j+1}) \begin{cases} < \varepsilon, & \text{if } j \neq \ell \\ > \frac{n-1}{\sum_{k=1}^{\ell}(Q_{ik} - P_{ik})}\varepsilon & \text{if } j = \ell \end{cases}.$$

(4) Therefore, we have

$$u_i \cdot P_i - u_i \cdot Q_i < \varepsilon \cdot \sum_{j \neq \ell} \sum_{k=1}^{j} [P_{ik} - Q_{ik}] - (n-1)\varepsilon < 0,$$

which contradicts the hypothesis.

□

☞ 7.27 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.

The random assignment P is ordinally efficient at $\succ$ if it is not ordinally dominated at $\succ$ by any other random assignment.

In environments where only ordinal preferences can be used, ordinal efficiency is a natural efficiency concept.

7.28 Given a preference profile $\succ$ and a random assignment P, we define a binary relation $\tau(P, \succ)$ on O as follows:

$$a\tau(P, \succ)b \Leftrightarrow \text{there exists } i \in N \text{ such that } a \succ_i b \text{ and } P_{ib} > 0.$$

7.29 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.

Proof. "$\Rightarrow$": Suppose that P is ordinally efficient.

(1) Assume that the relation $\tau(P, \succ)$, denoted τ for simplicity, has a cycle:

$$o_K \tau o_{K-1} \tau \cdots \tau o_2 \tau o_1 = o_K.$$

(2) Without loss of generality, we assume that the objects o_k, $k = 1, 2, \ldots, K-1$ are all different.

(3) By definition of τ, we can construct a sequence $i_1, i_2, \ldots, i_{K-1}$ in N such that:

$$P_{i_k o_k} > 0 \text{ and } o_{k+1} \succ_{i_k} o_k \text{ for all } k = 1, 2, \ldots, K-1.$$

(4) Choose $\delta > 0$ such that

$$\delta \leq P_{i_k o_k} \text{ for all } k = 1, 2, \ldots, K-1.$$

(5) Define a matrix $\Delta = (\delta_{io})$ as follows:

$$\begin{cases} \delta_{i_k o_k} = -\delta, & \text{for } k = 1, 2, \ldots, K-1, \\ \delta_{i_k o_{k+1}} = \delta, & \text{for } k = 1, 2, \ldots, K-1, \\ 0, & \text{otherwise.} \end{cases}$$

(6) Define a matrix $Q = P + \Delta$.

(7) By construction, Q is a bistochastic matrix and hence a random assignment.

(8) 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}.

"$\Leftarrow$": Suppose that the relation $\tau(P, \succ)$ is acyclic.

(1) Assume that 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.

□

☞ 7.30 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 $n \geq 4$.

7.31 *Proof of Proposition 7.30, Statement (i).* Part 1: Suppose that P is *ex ante* efficient at u. We want to show that P 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 7.26, we have $u_i \cdot Q_i \geq u_i \cdot P_i$ 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.

Part 2: Suppose that $n = 2$ and P is ordinally efficient at $\succ$. We want to show that P is *ex ante* efficient at u.

(1) Suppose that $P = \begin{pmatrix} x & 1-x \\ 1-x & x \end{pmatrix}$ is not *ex ante* efficient at u.

(2) Then there exists $Q = \begin{pmatrix} y & 1-y \\ 1-y & y \end{pmatrix}$ such that $u_i \cdot P_i \leq u_i \cdot Q_i$ for all i and $u_{i_0} \cdot P_{i_0} < u_{i_0} \cdot Q_{i_0}$ for some i_0.

(3) By simple calculation, we have

$$(u_{11} - u_{12})(x - y) \leq 0, \ (u_{22} - u_{21})(x - y) \leq 0.$$

(4) Case 1: agent 1 prefers o_1 to o_2. Then $u_{11} > u_{12}$, $x < y$, and $u_{22} > u_{21}$. In this case, P is stochastically dominated by $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. It is a contradiction.

(5) Case 2: agent 1 prefers o_2 to o_1. Then $u_{11} < u_{12}$, $x > y$, and $u_{22} < u_{21}$. In this case, P is stochastically dominated by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. It is a contradiction.

Part 3: Suppose that $n \geq 3$. We want to show that in some particular case P is ordinally efficient at $\succ$, but is not *ex ante* efficient at u.

(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) Claim: Every random assignment here is ordinally efficient.

(4) Suppose that a random assignment R is not ordinally efficient, and is stochastically dominated by R'.

(5) Then $R \neq R'$, and $\sum_{k:\, o_k \succsim_i o_j} R'_{ik} \geq \sum_{k:\, o_k \succsim_i o_j} R_{ik}$ for all i and j.

(6) Then,

$$\sum_i \sum_{k:\, o_k \succsim_i o_j} R'_{ik} \geq \sum_i \sum_{k:\, o_k \succsim_i o_j} R_{ik} \text{ for all } j.$$

(7) Since three agents have the same ordinal preference, we have

$$\sum_i \sum_{k:\, o_k \succsim_i o_j} R'_{ik} = \sum_{k:\, o_k \succsim_i o_j} \sum_i R'_{ik} = \sum_{k:\, o_k \succsim_i o_j} 1 = \sum_{k:\, o_k \succsim_i o_j} \sum_i R_{ik} = \sum_i \sum_{k:\, o_k \succsim_i o_j} R_{ik} \text{ for all } j$$

which leads to a contradiction.

□

7.32 *Proof of Proposition 7.30, Statement (ii).*

Part 1: If P is ordinally efficient at $\succ$, then it is *ex post* efficient at $\succ$.

(1) Suppose that 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 the stochastic dominance is preserved by convex combinations).

Part 2: When $n \leq 3$, if P is *ex post* efficient at $\succ$, then it is ordinally efficient at $\succ$. (Question. Hint: Consider different preference profiles and check the corresponding *ex post* efficient assignments)

Part 3: When $n \geq 4$, P may not be ordinally efficient at $\succ$, if it is *ex post* 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 7.5

(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 $P = \sum_{f \in \mathcal{F}} \frac{1}{n!} \mathrm{SD}^f[\succ]$, so P is *ex post* efficient.

$\square$

7.33 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.

7.4.3 Efficiency of RP and PS

7.34 Recall Theorem 5.9 (Lemma 1 in Abdulkadiroğlu and Sönmez (1998)): Simple serial dictatorship is Pareto efficient. Thus, RP is *ex post* efficient.

7.35 RP is not ordinally efficient or *ex ante* efficient: See the example in the proof of Proposition 7.30, Statement (ii), Part 3.

7.36 Theorem (Theorem 1 in Bogomolnaia and Moulin (2001)): Fix a preference profile $\succ$.

(i) For every profile of eating speed functions $\omega = (\omega_i)_{i \in N}$, the random assignment $P_\omega[\succ]$ is ordinally efficient.

(ii) Conversely, for every ordinally efficient random assignment P at $\succ$, there exists a profile $\omega = (\omega_i)_{i \in N}$ such that $P = P_\omega[\succ]$.

7.37 Intuition: At each instant, everyone is eating her favorite available object. If agent i likes a better than b but eats b, then a was already eaten away.

7.38 *Proof of Theorem 7.36, Statement (i).*

(1) Assume that for some ω, $P_\omega[\succ]$ is not ordinally efficient.

(2) By Proposition 7.29, 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 time instant in simultaneous eating algorithm when the agent i_k starts to acquire good o_k, i.e., the least t for which $P^t_{i_k o_k} \neq 0$.

(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 $o_0 = o_{K+1}$.

$\square$

7.39 *Proof of Theorem 7.36, Statement (ii).*

(1) Let P be an ordinally efficient assignment.

(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) For each $k \geq 1$, let

$$\bar{O}^k = \bar{O}^{k-1} \setminus B^k \text{ and } B^{k+1} = \{o \in \bar{O}^k \mid \nexists b \in \bar{O}^k \text{ such that } b\tau o\}.$$

It is clear that this sequence will stop in finite steps. Suppose that this sequence stops at Step K, for which $\bar{O}^K = \emptyset$ and $B^K = \bar{O}^{K-1}$.

Note that $\{B^1, B^2, \ldots, B^K\}$ forms a partition of O.

(4) For all $k = 1, 2, \ldots, K$, when $\frac{k-1}{K} \leq t \leq \frac{k}{K}$,

$$\omega_i(t) \triangleq \begin{cases} K \cdot P_{io}, & \text{if } o \in B^k \text{ and } i \in N(o, \bar{O}^{k-1}), \\ 0, & \text{otherwise.} \end{cases}$$

We will check that P is the result of the simultaneous eating algorithm with eating speeds ω and that $\bar{O}^0, \bar{O}^1, \ldots, \bar{O}^K$ coincide with $O^0, O^!, \ldots, O^K$ from this algorithm.

(5) Claim: For each $k = 1, 2, \ldots, K$, and for any o and o' in B^k, if $P_{io} > 0$, then $P_{io'} = 0$. Assume that $P_{io'} > 0$. Without loss of generality, assume that $o \succ_i o'$. Then $o\tau o'$, which contradicts the fact that $o' \in B^k$.

This claim guarantees that each agent eats at most one object between $\frac{k-1}{K}$ and $\frac{k}{K}$.

(6) Claim: For each $k = 1, 2, \ldots, K$, if $o \in B^k$ and $P_{io} > 0$, then $i \in N(o, \bar{O}^{k-1})$. Assume that $i \notin N(o, \bar{O}^{k-1})$. Then there exists $o' \in \bar{O}^{k-1}$ such that $o' \succ_i o$. Thus, $o'\tau o$ and $o \notin B^k$, which leads to a contradiction.

This claim guarantees that agent i with $P_{io} > 0$ will eat object $o \in B^k$ between $\frac{k-1}{K}$ and $\frac{k}{K}$.

(7) Therefore, from $\frac{k-1}{K}$ to $\frac{k}{K}$, only objects in the set B^k will be eaten in the simultaneous eating algorithm.

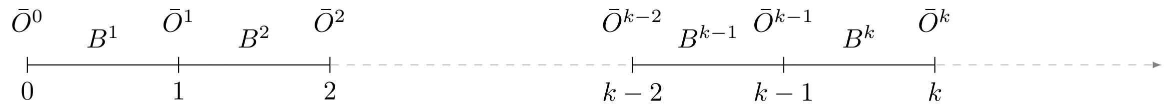


Figure 7.2

(8) 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 $K \cdot P_{io}$, 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 $\sum_i K \cdot P_{io} \cdot \frac{1}{K} = 1$.

(9) Hence, $t^1 = \frac{1}{K}$, $O^1 = \bar{O}^1$, and P^1 is as follows:

$$P^1_{io} = \begin{cases} P_{io}, & \text{if } o \in B^1, \\ 0, & \text{if } o \notin B^1. \end{cases}$$

(10) We proceed by induction. Suppose that

$$t^{k-1} = \frac{k-1}{K},\ O^{k-1} = \bar{O}^{k-1},\ \text{and } P^{k-1}_{io} = \begin{cases} P_{io}, & \text{if } o \in B^1 \cup B^2 \cup \cdots \cup B^{k-1}, \\ 0, & \text{otherwise.} \end{cases}$$

(11) 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 $P^{k-1}_{io} = 0$.

(12) Therefore, we have

$$\sum_{i \in N(o,\bar{O}^{k-1})} \int_{\frac{k-1}{K}}^{t} \omega_i(s)\,\mathrm{d}s + \underbrace{\sum_{i \in N} P^{k-1}_{io}}_{=0}$$
$$= \begin{cases} \displaystyle\sum_{i \in N(o,\bar{O}^{k-1})} \int_{\frac{k-1}{K}}^{t} K \cdot P_{io}\,\mathrm{d}s = [Kt-(k-1)] \cdot \underbrace{\sum_{i \in N(o,\bar{O}^{k-1})} P_{io}}_{=1} = Kt-(k-1), & \text{if } o \in B^k, \\ 0, & \text{if } o \notin B^k. \end{cases}$$

(13) So,

$$t^k(o) = \begin{cases} \frac{k}{K}, & \text{if } o \in B^k, \\ +\infty, & \text{otherwise.} \end{cases}$$

(14) Thus, $t^k = \frac{k}{K}$, $O^k = \bar{O}^k$, and

$$P^k_{io} = \begin{cases} P_{io}, & \text{if } o \in B^1 \cup B^2 \cup \cdots \cup B^k, \\ 0, & \text{otherwise.} \end{cases}$$

□

7.5 Fairness

7.5.1 Anonymity

7.40 A mechanism φ is anonymous if the mapping $\succ \mapsto \varphi[\succ]$ is symmetric from the n preferences to the n assignments.

7.41 Remark: In view of Theorem 7.36, 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.

7.42 Proposition (Lemma 4 in Bogomolnaia and Moulin (2001)): Fix at profile of eating speeds ω. Let φ be the mechanism derived from ω at all profiles. φ is anonymous if and only if it coincides with PS.

7.43 *Proof.* We only prove "only if" direction.

(1) Let φ be mechanism derived from ω. Suppose that φ is anonymous.

(2) We fix a preference profile $\succ$, and let $P = \varphi[\succ]$.

(3) 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 $0 = x_0 < x_1 \leq x_2 \leq \cdots \leq x_k \leq \cdots \leq x_n = 1$.

(4) 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.

(5) 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 $x_k < x_n$.

(6) 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 φ 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 φ 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 φ.

(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 φ, 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 φ, and hence they will end up at y_k with the same partial assignment as at x_k under PS.

(7) In the following, we will check that such $y_1, y_2, \ldots, y_n$ exist.

(8) Define

$$\bar{t}_i(k) = \max\left\{t \,\middle|\, \int_0^t \omega_i(s)\,\mathrm{d}s \geq x_k\right\}, \quad \underline{t}_i(k) = \min\left\{t \,\middle|\, \int_0^t \omega_i(s)\,\mathrm{d}s \geq x_k\right\}$$

$$\bar{t}(k) = \min_i \bar{t}_i(k), \qquad \underline{t}(k) = \max_i \underline{t}_i(k),$$

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.

(9) Proceed by induction on k. Suppose that under φ 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 $y_k \in [\underline{t}(k), \bar{t}(k)]$.

(10) In the following, we will show that $\underline{t}(k) > \bar{t}(k)$ is impossible by contradiction.

(11) Since $|N(o_k)| \geq 2$ whenever $x_k < x_n$, we focus on the case such that $|N(o_k)| \geq 2$.

(12) 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 7.3

(13) We have

$$\sum_{i \in N(o_k)} \int_{y_{k-1}}^{\bar{t}(k)} \omega_i(s)\,\mathrm{d}s \qquad \text{at } \bar{t}(k) < \underline{t}(k) = \underline{t}_2(k), \text{ agent 2 is still eating } o_k$$

$$
\begin{aligned}
&< |N(o_k)| \cdot (x_k - x_{k-1}) && \text{amount of } o_k \text{ left after } y_{k-1} \\
&< \sum_{i \in N(o_k)} \int_{y_{k-1}}^{\underline{t}(k)} \omega_i(s)\, \mathrm{d}s && \text{at } \underline{t}(k) > \underline{t}_1(k) \text{, agent 1 starts to eat another object}
\end{aligned}
$$

(14) For any object o_j with $j > k$, we have

$$\sum_{i \in N(o_j)} \int_{y_{k-1}}^{\bar{t}(k)} \omega_i(s)\, \mathrm{d}s \leq |N(o_j)| \cdot (x_k - x_{k-1}) \leq \text{amount of } o_j \text{ left after } y_{k-1}.$$

Moreover, the equality is possible only if $x_j = x_k$.

(15) 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 $s^1, s^2 \in \big(\bar{t}(k), \underline{t}(k)\big)$.

(16) 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.

(17) By induction hypothesis, all agents get exactly the same partial assignment at x_{k-1} under PS and at y_{k-1} under φ.

(18) 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 φ.

□

7.5.2 Envy-freeness

7.44 A random assignment P is envy-free at a profile $\succ$ if $P_i \succsim_i^{sd} P_j$ for all $i, j \in N$.

A random assignment P is weakly envy-free at a profile $\succ$ if for all $i, j \in N$,

$$P_j \succsim_i^{sd} P_i \Rightarrow P_i = P_j.$$

7.45 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 $n \geq 3$.

7.46 Intuition: At each instant, everyone is eating her favorite available object. So everyone has chance to eat a better (from her viewpoint) object than anyone else, so at the end, no one envies assignment someone else.

7.47 *Proof of 7.45, Statement (i).* (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

$$P^{k_1}_{io_1} = t^{k_1} \geq P^{k_1}_{jo_1} \text{ for all } j \in N.$$

(iii) Since o_1 is fully allocated at instant k_1, we have

$$P_{io_1} = P^{k_1}_{io_1} \geq P^{k_1}_{jo_1} = P_{jo_1} \text{ for all } j \in N.$$

(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

$$P_{io_1} + P_{io_2} = P^{k_2}_{io_1} + P^{k_2}_{io_2} = t^{k_2} \geq P^{k_2}_{jo_1} + P^{k_2}_{jo_2} = P_{jo_1} + P_{jo_2} \text{ for all } j \in N.$$

(6) Repeating this argument we find that P_i first-order stochastically dominates P_j at $\succ_i$, as desired.

$\square$

7.48 *Proof of 7.45, Statement (ii).* (1) Let $\succ$ be a preference profile at which $P_2 \succ_1^{sd} P_1$, 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.

$$\begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_1 & \cancel{o_1} \\ & \nwarrow \\ o_1? & o_1 \end{pmatrix} \end{array}$$

(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 $P_2 \succ_1^{sd} P_1$, 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 $\varphi^{\bar{f}}$,

$$\begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} o_1 & \cancel{o_1} \\ \cancel{o_1} & o_1 \end{pmatrix} \end{array} \tag{7.1}$$

- or none of 1, 2 gets o_1 in any of φ^f or $\varphi^{\bar{f}}$.

$$\begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} \varphi^f \\ \varphi^{\bar{f}} \end{array} & \begin{pmatrix} \cancel{o_1} & \cancel{o_1} \\ \cancel{o_1} & \cancel{o_1} \end{pmatrix} \end{array} \tag{7.2}$$

(6) Consider o_2:

(i) If 2 gets o_2 in $\varphi^{\bar{f}}$, then by Equation (7.2), 1 can not get o_1 in φ^f, and hence 1 gets o_2 in φ^f.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & \quad & 2 \end{array} \\ \begin{array}{c}\varphi^f\\ \\ \varphi^{\bar{f}}\end{array} & \begin{pmatrix} o_2 & & \\ & \nwarrow & \\ & & o_2 \end{pmatrix}\end{array}$$

(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 (7.1), 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.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & \quad & 2 \end{array} \\ \begin{array}{c}\varphi^f\\ \\ \varphi^{\bar{f}}\end{array} & \begin{pmatrix} o_1 & \leftarrow & o_2 \\ & \searrow & \\ o_2 & \leftarrow & o_1 \end{pmatrix}\end{array}$$

(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.

$$\begin{array}{cc} & \begin{array}{ccc} 1 & \quad & 2 \end{array} \\ \begin{array}{c}\varphi^f\\ \\ \varphi^{\bar{f}}\end{array} & \begin{pmatrix} o_k & & \\ & \nwarrow & \\ & & o_k \end{pmatrix}\end{array}$$

(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.

$$\begin{array}{c} \\ \varphi^f \\ \\ \varphi^{\bar f} \end{array} \begin{array}{c} \begin{array}{ccc} 1 & & 2 \end{array} \\ \left(\begin{array}{ccc} o_\ell & \leftarrow & o_k \\ & \searrow & \\ o_k & \leftarrow & o_\ell \end{array}\right)\end{array}$$

(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 $P_{2o_k} = P_{1o_k}$.

□

7.49 *Proof of 7.45, Statement (iii).* 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.

□

7.50 *Proof of 7.45, Statement (iv).* (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 7.6

(2) It is clear that

$$\mathrm{RP} = \mathrm{RP}[\succ] = \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{5}{6} & \frac{1}{6} \end{pmatrix}$$

(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

$$u_1 \cdot \mathrm{RP}_3 = \frac{5}{6}u_1(b) + \frac{1}{6}u_1(c) = 7.5 > 6.5 = \frac{1}{2}u_1(a) + \frac{1}{6}u_1(b) + \frac{1}{3}u_1(c) = u_1 \cdot \mathrm{RP}_1 .$$

That is, in the RP assignment, agent 1 envy the allocation of agent 3.

(4) By Proposition 7.26, $\text{RP}_1 \succ_1^{sd} \text{RP}_3$ does not hold. Hence, RP is not envy-free.

□

7.5.3 Equal treatment of equals

7.51 Definition: A mechanism $\varphi\colon \succ \mapsto \varphi[\succ]$ has the property "equal treatment of equals" if

$$\succ_i = \succ_j \Rightarrow \varphi_i[\succ] = \varphi_j[\succ].$$

7.52 Proposition: PS and RP have the property "equal treatment of equals."

7.6 Incentive compatibility

7.53 Definition: A mechanism φ is strategy-proof if for each random assignment problem $\langle N, O \succ \rangle$, for each $i \in N$, and for each $\succ_i'$, we have

$$\varphi_i[\succ] \succ_i^{sd} \varphi_i[\succ_{-i}, \succ_i'].$$

7.54 Definition: A mechanism φ is weakly strategy-proof if for each random assignment problem $\langle N, O \succ \rangle$, for each $i \in N$, and for each $\succ_i'$, we have

$$\varphi_i[\succ_{-i}, \succ_i'] \succ_i^{sd} \varphi_i[\succ] \Rightarrow \varphi_i[\succ_{-i}, \succ_i'] = \varphi_i[\succ].$$

7.55 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 $n \geq 3$.

7.56 *Proof of Proposition 7.55, Statement (i).* For any ordering f, the priority mechanism $\succ \mapsto \varphi^f[\succ]$ is obviously strategy-proof. This property is preserved by convex combinations. □

7.57 *Proof of Proposition 7.55, Statement (ii).* (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 $P_1^* \succ_1^{sd} P_1$ 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 $P_1^* \succ_1^{sd} P_1$ 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.

$\square$

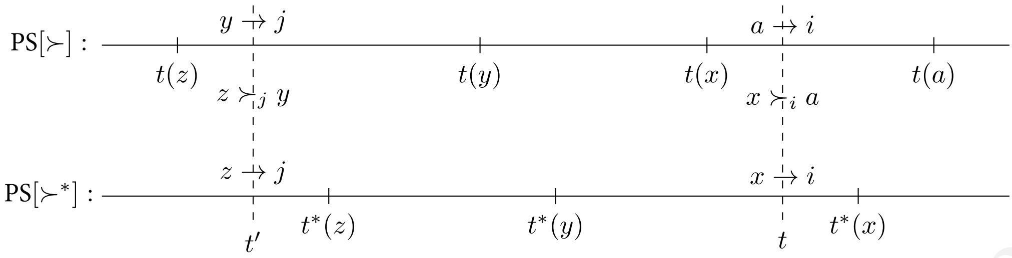


Figure 7.4

7.58 *Proof of Claim.*

(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 $t(z) \leq t' < t^*(z)$.

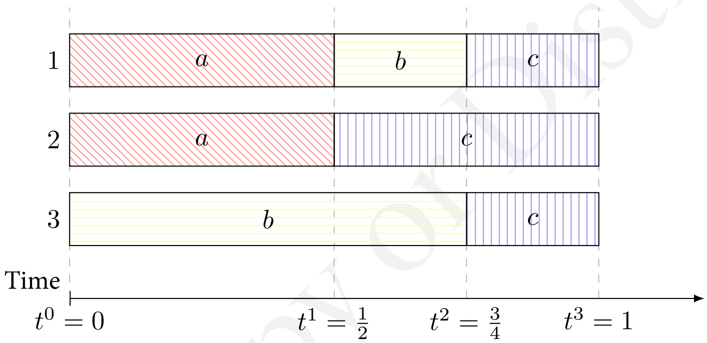
(11) Since $t(z) \leq t' < t(y) \leq t(x) < t(a)$, we have $z = a$, and hence $z \in B$, which contradicts the definition of y.

$\square$

7.59 *Proof of Proposition 7.55, Statement (iv).* There are three goods $\{a, b, c\}$, three agents $\{1, 2, 3\}$. The preference profile is as follows:

1	2	3
a	a	b
b	c	a
c	b	c

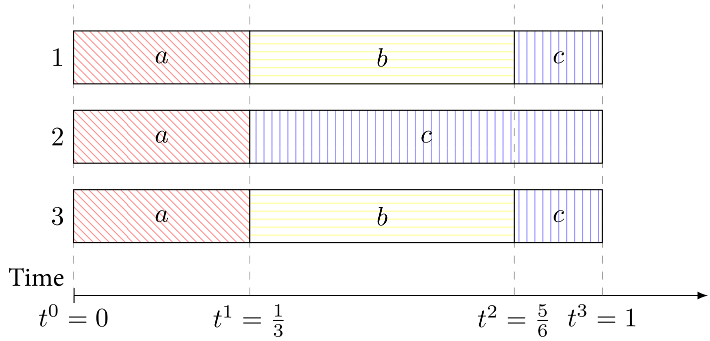
The process of PS is illustrated below



Thus, PS produces the matching

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

However, if agent 3 misreports her preference as $a \succ'_3 b \succ'_3 c$, then the process of PS is illustrated below



Thus, PS produces the matching

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}.$$

Agent 3 may be better off: Consider the following consistent von Neumann-Morgenstern utility $u_3 = (9, 10, 0)$. Then $P_3 \cdot u_3 = \frac{15}{2} < 8 = Q_3 \cdot u_3$.

By Proposition 7.26, $P_3 \succsim_3^{sd} Q_3$ does not hold. Hence, PS is not strategy-proof. □

7.60 Question: Is there a random assignment problem such that the assignment of some agent i_0 under the true preference profile is first-order stochastically dominated by the assignment under the misreported preference profile?

7.7 RP vs PS

7.61 Comparison of RP and PS:

	RP	PS
Ex ante efficiency	$\times$	Question
Ordinal efficiency	$\times$	$\surd$
Ex post efficiency	$\surd$	$\surd$
Envy-freeness	$\times$	$\surd$
Weak envy-freeness	$\surd$	$\surd$
Equal treatment of equals	$\surd$	$\surd$
Strategy-proofness	$\surd$	$\times$
Weak strategy-proofness	$\surd$	$\surd$

Table 7.7

7.62 Proposition (Proposition 2 in Bogomolnaia and Moulin (2001)):

- Fix $n = 3$. RP is characterized by the combination of three axioms: ordinal efficiency, equal treatment of equals, and strategy-proofness.
- PS is characterized by the combination of three axioms: ordinal efficiency, envy-freeness, and weak strategy-proofness.

7.63 Theorem (Theorem 1 in Kesten (2009)):

7.8 Impossibility results

7.64 Theorem (Zhou (1990)): Incompatibility of *ex ante* efficiency, equal treatment of equals, and strategy-proofness.

7.65 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.

7.66 Theorem (Proposition 1 in Erdil (2014)): If a strategy-proof mechanism is non-wasteful, then it is not (FSD) dominated by any other strategy-proof mechanism.

7.67 Theorem (Theorem 1 in Martini (2016)): Let there be $n \geq 4$ agents and $m \geq 3$ objects. Then there is no mechanism that is strategy-proof, non-wasteful and satisfies equal treatment of equals.

7.9 Large markets

7.68 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.

7.69 Che and Kojima (2010) show that PS and RP are asymptotically equivalent, as the size of the market increases.

7.70 Manipulations have two effects: (1) given the same set of available objects, reporting false preferences may prevent the agent from eating his most preferred available object; (2) reporting false preferences can affect expiration dates of each good. (1) always hurts the manipulating agent, while (2) can benefit the agent.

Intuitively, the effect (2) becomes small as the market becomes large.

7.10 Implementing random assignments

☞ 7.71 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.

An example for lottery assignment:

$$p = \frac{5}{12}\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix} \oplus \frac{1}{12}\begin{pmatrix}0&1&0&0\\0&0&0&1\\1&0&0&0\\0&0&1&0\end{pmatrix} \oplus \frac{5}{12}\begin{pmatrix}0&0&1&0\\1&0&0&0\\0&0&0&1\\0&1&0&0\end{pmatrix} \oplus \frac{1}{12}\begin{pmatrix}0&0&0&1\\0&1&0&0\\0&0&1&0\\1&0&0&0\end{pmatrix}.$$

7.72 Relationship between random assignments and lottery assignments.

We associate to each lottery assignment p a random assignment P is the following way:

$$P = \sum_{X\in\mathcal{D}} p(X)\cdot X.$$

On the other hand, 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.

7.73 Lemma: Let P be a bistochastic matrix that is not a permutation matrix. Then it can be written as a convex combination of two bistochastic matrices,

$$P = \lambda P^1 + (1-\lambda)P^2,$$

where P^1 and P^2 has the following properties:

(i) If P_{io} is an integer, then P^1_{io} and P^2_{io} are integers.

(ii) P^1 and P^2 has at least one more integer entry than P.

Proof. See Hylland and Zeckhauser (1979). □

7.74 The equivalence implies that any random assignment is induced by a lottery assignment. Thus, any random assignment can be implemented.

Part III

School choice

Chapter 8

Introduction to school choice

Contents

8.1 The former model

☞ 8.1 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 $s = s'$.

8.2 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 can be regarded as 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 $\succ_s$.

It is an important issue that the priority of each school is weak. However, we will only consider the strict priorities in this chapter unless otherwise mentioned.

☞ 8.3 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 $s \in S$, $|\mu^{-1}(s)| \leq q_s$.

Let $\mathcal{M}$ denote the set of all the matchings.

☞ 8.4 A mechanism is a systematic procedure that determines a matching for each school choice problem.

8.5 In school choice problem, we allow only students to report preferences, and schools' priorities are exogenously given and publicly known.[1]

Thus, when I, S, q, and $\succ$ are given, a mechanism $\varphi^{\succ}$, or simply φ, becomes a function

$$\varphi\colon \mathcal{P}^{|I|} \to \mathcal{M},$$

where $\mathcal{P}$ is the set of all the possible preferences for students.

8.6 Typical goals of school authorities are:

- efficient placement,
- fairness of outcomes,
- strategy-proof,
- easy for participants to understand and use, *etc*.

☞ 8.7 In a school choice problem $\langle I, S, q, P, \succ\rangle$, 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 it always selects a Pareto efficient matching for each school choice problem.

[1] In many school districts, schools are not allowed to submit their own preferences; Instead, school priorities are set by law.

Exercise: Compare this definition of Pareto efficient matchings with that in Definitions 2.7. Why is there a difference?

8.8 A mechanism (Pareto) dominates another mechanism if for every school choice problem, the outcome of the first weakly dominates that of the latter, with strict dominance for at least one school choice problem.

8.9 In a school choice problem $\langle I, S, q, P, \succ \rangle$, a matching μ is individually rational if no student prefers being unmatched to her assignment.

A mechanism φ is individually rational if it always selects a individually rational matching for each school choice problem.

8.10 In a school choice problem $\langle I, S, q, P, \succ \rangle$, 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 it always selects a non-wasteful matching for each school choice problem.

8.11 We say that student i desires school s at μ if $s P_i \mu(i)$.

In a school choice problem $\langle I, S, q, P, \succ \rangle$, 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 it always selects a matching that eliminates justified envy for each school choice problem.

8.12 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.

8.13 Remark: In school choice, stability can be understood as a fairness criterion.

8.14 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.

8.15 A mechanism φ is strategy-proof if no student can benefit from misreporting for each school choice problem, *i.e.*, truth-telling is a weakly dominant strategy for all students under the mechanism φ.

Formally, for each $\langle I, S, q, P, \succ \rangle$, for each $i \in I$, and for each P'_i, we have

$$\varphi[P_i, P_{-i}](i) R_i \varphi[P'_i, P_{-i}](i).$$

8.16 A mechanism φ is group strategy-proof if for any $\langle I, S, q, P, \succ\rangle$, there do not exist $J \subseteq I$ and $P'_J = (P'_i)_{i\in J}$ such that

$$\varphi[P'_J, P_{-J}](i) R_i \varphi[P](i) \text{ for all } i \in J \text{ and } \varphi[P'_J, P_{-J}](j) P_j \varphi[P](j) \text{ for some } j \in J.$$

8.17 A mechanism φ is non-bossy if for each $\langle I, S, q, P, \succ\rangle$, for each $i \in I$, and for each P'_i,

$$\varphi[P](i) = \varphi[P'_i, P_{-i}](i) \text{ implies } \varphi[P] = \varphi[P'_i, P_{-i}].$$

Non-bossiness ensures that students can not be bossy, that is, change the matching for others, by reporting different preferences, without changing their own.

8.18 Theorem (Lemma 1 in Pápai (2000)): A mechanism φ is group strategy-proof if and only if it is strategy-proof and non-bossy.

Proof. It is obvious that group strategy-proofness implies strategy-proofness and non-bossiness. So it suffices to show the other direction.

(1) Suppose that the mechanism φ is strategy-proof and non-bossy.

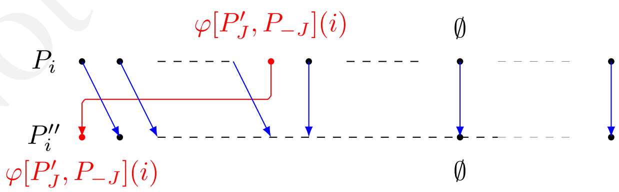
(2) Let $\langle I, S, q, P, \succ\rangle$, $J \subseteq I$, and P'_J be such that for all $i \in J$,

$$\varphi[P'_J, P_{-J}](i) R_i \varphi[P](i).$$

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

$$o_i(P_{-i}) \triangleq \left\{s \in S \colon \text{there exists } P'''_i \text{ such that } \varphi[P'''_i, P_{-i}](i) = s\right\}.$$

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 $\varphi[P_1'', P_{-1}](1) = \varphi[P](1)$.

(6) By non-bossiness, 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 $s \in S$, $s R_i'' \varphi[P_J', P_{-J}](i)$ implies $s R_i' \varphi[P_J', P_{-J}](i)$.

(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 non-bossiness, $\varphi[P_i'', P_{J\setminus\{i\}}', P_{-J}] = \varphi[P_J', P_{-J}]$.

(13) By the similar argument 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.

□

8.19 Remark: Theorem 8.18 is a general result for one-sided matchings.

8.2 Boston school choice mechanism (immediate acceptance mechanism)

8.20 The most commonly used school choice mechanism is that used by the Boston Public School until 2005.

The Boston mechanism attempts to assign as many students as possible to their first choice school, and only after all such assignments have been made does it consider assignments of students to their second choices, and so on.

8.21 The Boston mechanism.[2]

Round 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.)

Round 2: Each student submits a preference ranking of the schools.

Round 3: The final round 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.

8.22 In Boston, students have priorities at schools set by the school system:

(i) Students who already attend the school,

(ii) Students who live in a walk zone and have their siblings already attending the school,

(iii) Students whose siblings are already attending the school,

(iv) Students who live in a walk zone,

(v) All other students.

Priorities are weak, *i.e.*, there are many students in each priority class: This is going to be important but for now let us ignore the issue.

8.23 Example: Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, where $I = \{i, j, k\}$, $S = \{a, b\}$, $q_a = q_b = 1$, and the preferences and priorities are as follows:

[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	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 8.1

The procedure of the Boston mechanism is

Step	1	End
a	$j, \cancel{k}$	j
b	i	i
$\emptyset$	k	k

Table 8.2

Student i is on the list of school b, and students j and k are on the list of schools a where j has higher priority. So i is assigned to b, j is assigned to a, and k remains unmatched.

The resulting matching is

$$\mu = \begin{bmatrix} i & j & k \\ b & a & \emptyset \end{bmatrix}.$$

8.24 The Boston mechanism is not necessarily stable.

Consider Example 8.23. The matching μ is blocked by the pair (k, b).

8.25 The Boston mechanism is not strategy-proofness.

Consider Example 8.23, if k misreports her preference as $P'_k\colon b, a, \emptyset$ instead, the Boston mechanism produces the following matching

$$\mu' = \begin{bmatrix} i & j & k \\ \emptyset & a & b \end{bmatrix},$$

and student k benefits from submitting a false preference.

8.26 As seen in Example 8.23, a student (for example, k) who ranks a school (b) as her second choice loses her priority to students (i) who rank it as their first choice. Thus, it is risky for the student to use her first choice at a highly sought-after school if she has relatively low priority there. If she does not receive her first choice, she might drop far down list.

Besides, the Boston mechanism gives students incentive to misreport their preferences by improving the ranking of schools in their choice lists for which they have high priority. Chen and Sönmez (2006) found the experimental evidence on preference manipulation under Boston mechanism.

8.27 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.

Advice from the West Zone Parents Group[3] meeting (27 October 2003)

> One school choice strategy is to find a school you like that is undersubscribed and put it as a top choice, OR, find a school that you like that is popular and put it as a first choice and find a school that is less popular for a "safe" second choice.

8.28 Abdulkadiroğlu *et al.* (2006) found that of the 15135 students, 19% (2910) listed two over-demanded schools as their top two choices, and about 27% (782) of these ended up unassigned.

8.29 Since priorities are set by law for Boston schools, Abdulkadiroğlu *et al.* (2006) recommended not only DA but also TTC: remember TTC is more efficient than DA.

However, the school system finally chose DA: the story says the idea of "trading priorities" in TTC did not appeal to policy makers.

For Boston Public School system, the Boston mechanism was replaced by DA in 2006.

8.30 Question: How about the efficiency of the Boston mechanism?

8.3 Deferred acceptance algorithm and student-optimal stable mechanism

8.31 In a school choice problem $\langle I, S, q, P, \succ\rangle$ with given strict priorities $\succ$, let DA (or $\text{DA}^{\succ}$ in some environments) denote the student-optimal stable mechanism, which is produced by Gale and Shapley's student-proposing deferred acceptance algorithm.

[3] This group is a well-informed group of approximately 180 members who meet regularly prior to admissions time to discuss Boston school choice for elementary school, recommends two types of strategies to its members.

8.32 Theorem: For each school choice problem, DA produces a stable matching, which is also at least as good for every student as any other stable matching.

Proof. Recall Theorem 3.23 and Corollary 3.26. □

The welfare is maximized by student-proposing DA, subject to stability.

8.33 Theorem: DA is strategy-proof.

Proof. Recall Theorem 3.61. □

8.34 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.

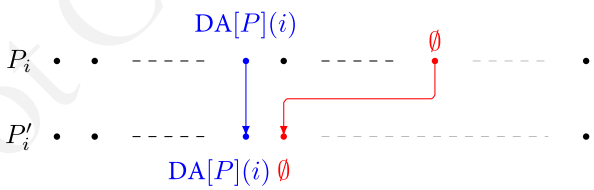
(1) Suppose that $\varphi(\neq \text{DA})$ is another stable and strategy-proof mechanism in school choice problems.

(2) Thus, there exists a school choice problem $\langle I, S, q, P, \succ\rangle$ such that $\varphi[P] \neq \text{DA}[P]$.

(3) There will then be some student $i \in I$ who is not assigned to her optimal school $\text{DA}[P](i)$.

(4) It is clear that $\text{DA}[P](i) P_i \varphi[P](i)$, and hence $\text{DA}[P](i) \neq \emptyset$.

(5) Consider a new preference P_i' of i: P_i' keeps the same ranking among schools and sets the schools behind $\text{DA}[P](i)$ unacceptable.



(6) Clearly, $\text{DA}[P]$ is stable under $[P_i', P_{-i}]$.

(7) By Theorem 3.30, we know that the set of students remaining unassigned is the same at all stable matchings for the given preference profile $[P_i', P_{-i}]$.

(8) Since $\varphi[P_i', P_{-i}]$ is another stable matching under $[P_i', P_{-i}]$, $\varphi[P_i', P_{-i}](i) \neq \emptyset$.

(9) Thus, we have $\varphi[P_i', P_{-i}](i) P_i' \emptyset$, and hence $\varphi[P_i', P_{-i}](i) R_i' \text{DA}[P](i)$.

(10) Since P_i' and P_i share the same ranking among schools from the top-ranked school to $\text{DA}[P](i)$, we have $\varphi[P_i', P_{-i}](i) R_i \text{DA}[P](i)$.

(11) Since $\mathrm{DA}[P](i) P_i \varphi[P](i)$, 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'. It contradicts the fact that φ is strategy-proof.

$\square$

8.35 The major drawback of DA is its lack of Pareto efficiency.

Consider the school choice problem $\langle I, S, q, P, \succ \rangle$, where $I = \{i, j, k\}$, $S = \{a, b\}$, $q_a = q_b = 1$, and the preferences and priorities are as follows:

i	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 8.3

The procedure of DA is

Step	1	2	3	End
a	$j, \cancel{k}$	j	$\cancel{j}, i$	i
b	i	$\cancel{i}, k$	k	k
$\emptyset$	k	i	j	j

Table 8.4

and the resulting matching is

$$\mu = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

It is clear that μ is Pareto dominated by the matching

$$\mu' = \begin{bmatrix} i & j & k \\ b & \emptyset & a \end{bmatrix}.$$

The efficiency of DA will be detailedly discussed in Chapter 9.

8.36 Remark: DA is strategy-proof and stable, but not Pareto efficient. Are there mechanisms that improve the efficiency of students without sacrificing the other two properties?

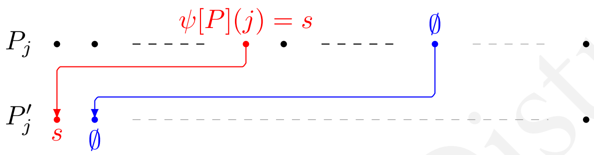
- Stability will be lost for sure, since DA produces the student-optimal stable matching.

- Strategy-proofness will also be lost, due to the following impossibility result.

8.37 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 stable mechanism, then there is no strategy-proof mechanism that dominates φ.

Proof. (1) Suppose that there exists a strategy-proof mechanism ψ that dominates φ. Then there exists a school choice problem $\langle I, S, q, P, \succ\rangle$ such that $\psi[P](i) R_i \varphi[P](i)$ for all $i \in I$ and $\psi[P](j) P_j \varphi[P](j)$ for some $j \in I$.

(2) Let $s = \psi[P](j)$. Consider a new preference P'_j of j: P'_j: $s, \emptyset$.



(3) Since φ is strategy-proof, $\varphi[P'_j, P_{-j}](j) = \emptyset$; otherwise $\varphi[P'_j, P_{-j}](j) = s$ will lead j to misreport P'_j when her true preference is P_j:

$$\varphi[P'_j, P_{-j}](j) = s = \psi[P](j) P_j \varphi[P](j).$$

(4) Since $\varphi[P'_j, P_{-j}]$ is Pareto dominated by $\psi[P'_j, P_{-j}]$, the same set of students is matched; see Lemma 8.38.

(5) Thus, $\psi[P'_j, P_{-j}](j) = \emptyset$.

(6) 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.

□

8.38 Lemma (Lemma 1 in Erdil and Ergin (2008) and Claim in Abdulkadiroğlu *et al.* (2009)): In a school choice problem $\langle I, S, q, P, \succ\rangle$, suppose that μ is a stable matching[4] 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:

[4] Indeed, we only require that μ is individually rational (for students) and non-wasteful.

(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 non-wastefulness 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 individual rationality 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 individual rationality of μ.

□

This result implies that any Pareto improvement upon a stable matching must be through trading cycles.

8.39 Corollary: Given strict school priorities, no Pareto efficient and strategy-proof mechanism dominates DA.

8.40 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.

8.41 The efficiency improvement of DA will be detailedly discussed in Chapter 10.

8.42 DA was implemented in Boston in 2006 and is in use. Its variation is used in New York City.

8.4 Top trading cycles mechanism

8.43 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_1 points to s_2, ..., s_k points to i_k, i_k points to s_1.) 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.

8.44 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.

8.45 Theorem (Proposition 3 in Abdulkadiroğlu and Sönmez (2003)): TTC is Pareto efficient.

Proof. Recall Theorem 4.16. □

8.46 Theorem (Proposition 4 in Abdulkadiroğlu and Sönmez (2003)): TTC is strategy-proof.

Proof. Recall Theorem 4.27. □

8.47 TTC does not completely eliminate justified envy.

Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, 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 8.5

The matching produced by TTC is

$$\mu = \begin{bmatrix} i & j & k \\ b & \emptyset & a \end{bmatrix}.$$

It is clear that k violates j's priority at school a, since $j \succ_a k$ and $\mu(k) = aP_j\emptyset = \mu(j)$.

8.48 Remark: Although TTC is Pareto efficient and DA is not, the two are not Pareto ranked in general.

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

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 8.6

The outcomes of DA and TTC are

$$\mu^{\mathrm{DA}} = \begin{bmatrix} i & j & k \\ s_2 & s_3 & s_1 \end{bmatrix} \text{ and } \mu^{\mathrm{TTC}} = \begin{bmatrix} i & j & k \\ s_2 & s_1 & s_3 \end{bmatrix},$$

where neither matching Pareto dominates the other one.

8.49 For school choice problems, TTC and DA are 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.

8.50 Question: How to improve the fairness of TTC?

Hint (Hakimov and Kesten (2014)): Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, where $I = \{i, j, k\}$, $S = \{a, b\}$, $q_a = 1$, $q_b = 2$, and the priorities for the schools and the preferences of the students are given as follows

i	j	k	a	b
b	a	a	i	k
a	b	b	j	j
			k	i

Table 8.7

When we apply TTC to this problem, student i who has the highest a-priority, exchanges one slot at school a in return for one slot at school b from student k who has the highest b-priority. This allocation is Pareto efficient. However, the priority of student j for school a is violated by student k, *i.e.*, j has justified envy over k.

TTC gives student k ownership over both slots of school b before student j enters the market. But then student i has no choice but to trade with student k, which in turn leads to the violation of the priority of student j for school a. However, had student i traded his right for one slot at school a with student j for his right for one slot at school b, there would not be any priority violations. Indeed, such a trade would have led to the Pareto efficient and stable allocation underlined in the above profile.

8.5 Case study: Chinese college admissions

8.51 Main reference: Chen and Kesten (2017).

8.52 To alleviate the problem of high-scoring students not being accepted by any universities, the parallel mechanism was proposed by Zhenyi Wu (吴振一). A Chinese parallel mechanism was first implemented in Hunan tier 0 college admissions in 2001. From 2001 to 2012, variants of the mechanism have been adopted by 28 provinces to replace Boston mechanisms; Wu and Zhong (2014).

8.53 Chinese parallel mechanism with a parameter $e \in \{1, 2, \ldots, \infty\}$,[5] denoted by φ^e:

Round 1: **Step** 1: Each student applies to his first choice. Each school s considers its applicants. Those students with the highest s-priority are tentatively assigned to school s up to its quota. The rest are rejected.

Step k: Each rejected student, who is yet to apply to his e-th choice school, applies to his next choice. If a student has been rejected from all his first e choices, then he remains unassigned in this round and does not make any applications until the next round.
Each school s considers its applicants. Those students with highest s-priority are tentatively assigned to school s up to its quota. The rest of the applicants are rejected.

End: The round terminates whenever each student is either assigned to a school or is unassigned in this round, *i.e.*, he has been rejected by all his first e choice schools. At this point, all tentative assignments become final and the quota of each school is reduced by the number of students permanently assigned to it.

Round t: **Step** 1: Each unassigned student from the previous round applies to his $\big((t-1)e+1\big)$-st choice school. Each school s considers its applicants. Those students with the highest s-priority are tentatively assigned to school s up to its quota. The rest of the applicants are rejected.

Step k: Each rejected student, who is yet to apply to his $\big((t-1)e+e\big)$-th choice school, applies to his next choice. If a student has been rejected from all his first $\big((t-1)e+e\big)$ choices, then he remains unassigned in this round and does not make any applications until the next round.
Each school s considers its applicants. Those students with the highest s-priority are tentatively assigned to school s up to its quota. The rest of the applicants are rejected.

End: The round terminates whenever each student is either assigned to a school or is unassigned in this round, *i.e.*, he has been rejected by all his first $(t-1)e+e$ choice

[5] For example, $e=2$ for Heilongjiang, $e=3$ for Jiangsu, $e=4$ for Anhui, $e=5$ for Hebei, $e=6$ for Hainan, $e=10$ for Tibet.

schools. At this point, all tentative assignments become final and the quota of each school is reduced by the number of students permanently assigned to it.

End: The algorithm terminates when each student has been assigned to a school. At this point, all the tentative assignments become final.

8.54 Remark: There are two limiting cases:

- The Chinese parallel mechanism with a parameter $e = 1$ is equivalent to the Boston mechanism.
- The Chinese parallel mechanism with a parameter $e = \infty$ is equivalent to DA.

8.55 Proposition (Proposition 1 in Chen and Kesten (2017)): Within the family Chinese parallel mechanisms, that is, $e \in \{1, 2, \ldots, \infty\}$,

(i) there is exactly one member that is Pareto efficient; this is the Boston mechanism;

(ii) there is exactly one member that is strategy-proof; this is the DA mechanism;

(iii) there is exactly one member that is stable; this is the DA mechanism.

8.56 Theorem (Theorem 1 in Chen and Kesten (2017)): For any e, φ^e is more manipulable than $\varphi^{e'}$, where $e' > e$.

A mechanism ψ is said to be manipulable at a problem $\langle P, \succ\rangle$ if there exists some student j such that ψ is manipulable by student j at $\langle P, \succ\rangle$. We consider mechanism φ to be more manipulable than mechanism ψ if (i) at any problem ψ is manipulable, then φ is also manipulable; and (ii) there is at least one problem at which φ is manipulable but ψ is not.

8.57 Proposition (Proposition 2 in Chen and Kesten (2017)): Let $e' > e$.

(i) If $e' = ke$ for some $k \in \mathbb{N} \cup \{\infty\}$, then $\varphi^{e'}$ is more stable than φ^e.

(ii) If $e' \neq ke$ for any $k \in \mathbb{N} \cup \{\infty\}$, then $\varphi^{e'}$ is not more stable than φ^e.

A mechanism φ to be more stable than mechanism ψ if (i) at any problem ψ is stable, φ is also stable; and (ii) there is at least one problem at which φ is stable but ψ is not.

Chapter 9

Acyclicity

Contents

9.1 Cycles and efficiency of deferred acceptance algorithm

9.1 Consider the school choice problem $\langle I, S, q, P, \succ\rangle$ in Example 8.35, 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 9.1

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 cyclic, since j may block a potential matching between i and k without affecting his own position, that is

$$i \succ_a j \succ_a k \succ_b i.$$

Because of such a cycle, in DA,

(1) k applies to her favorite a but j displaces k,

(2) k is forced to apply to her second choice b, displacing i from his favorite b,

(3) i is forced to apply to his second choice a, displacing j.

In the end, j is displaced by school a anyway, with the result being just causing more rejections and making i and k worse off.

☞ 9.2 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 (possibly empty) 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.

9.3 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.

9.4 For any problem $\Gamma = \langle I, S, q, P, \succ \rangle$, any $I' \subseteq I$, and any matching μ, the reduced problem of Γ with respect to I' and q' under μ is

$$r^{\mu}_{I'}(\Gamma) = \langle I', S, q', P_{I'}, \succ|_{I'} \rangle,$$

where $q'_s = q_s - |\mu^{-1}(s) \setminus I'|$.

It 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.

☞ 9.5 Definition: A mechanism φ is consistent is consistent if for any problem $\Gamma = \langle I, S, q, P, \succ \rangle$, for any nonempty subset $I' \subseteq I$, and for any $i \in I'$,

$$\varphi[\Gamma](i) = \varphi\left[r^{\varphi[\Gamma]}_{I'}(\Gamma)\right](i).$$

9.6 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.

9.7 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.

9.8 This theorem is bad news for school systems, because most priority structures are cyclic.

9.9 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 $U_s(i) \triangleq \{j \in I \mid j \succ_s i\}$.

9.10 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.

□

9.11 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.

□

9.12 *Proof of Theorem 9.7, Part 1: "acyclicity implies Pareto efficiency".* It follows immediately from two lemmas above. □

9.13 *Proof of Theorem 9.7, Part 2: "Pareto efficiency implies consistency".*

(1) Assume DA is not consistent.

(2) Then, there is $\langle I, S, q, P, \succ\rangle$ and $\emptyset \neq I' \subsetneq I$ 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 3.26, 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.

$\square$

9.14 *Proof of Theorem 9.7, Part 3: "consistency implies group strategy-proofness".*

(1) By Corollary 3.26, DA is strategy-proof.

(2) By Theorem 8.18, 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

$$\mu = \text{DA}[I, S, q, P, \succ], \text{and } \nu = \text{DA}[I, S, q, P'_i, P_{-i}, \succ].$$

(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,

$$\mu|_{I\setminus\{i\}} = \text{DA}[I, S, q, P, \succ]|_{I\setminus\{i\}} = \text{DA}[r^{\mu}_{I\setminus\{i\}}(I, S, q, P, \succ)],$$
$$\nu|_{I\setminus\{i\}} = \text{DA}[I, S, q, P'_i, P_{-i}, \succ]|_{I\setminus\{i\}} = \text{DA}[r^{\nu}_{I\setminus\{i\}}(I, S, q, P'_i, P_{-i}, \succ)].$$

(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 $\mu = \nu$.

$\square$

9.15 *Proof of Theorem 9.7, 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,
- 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.

i	j	k
b	a	a
a		b

Table 9.2

(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.

□

9.16 *Proof of Theorem 9.7, 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 9.3

- 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.

□

9.17 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. □

9.2 Robust stability

9.18 In school choice problems, DA is both stable (fair) and strategy-proof. This makes it a good mechanism.

What about a combined manipulation? That is, first misreport preferences and then file for a re-matching?

This issue is intended to model appeals processes: In NYC, about 5000 students out of 90000 file for appeals under DA; 300 among them are from those who were matched to their first choices.

9.19 Consider the school choice problem $\langle I, S, q, P, \succ\rangle$, 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 9.4

The matching produced by DA is

$$\mu = \begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

Suppose that j misreports that $\emptyset$ is her first choice. Then the matching of DA is

$$\mu' = \begin{bmatrix} i & j & k \\ b & \emptyset & a \end{bmatrix}.$$

Because $j \succ_a k$, j could ask to be admitted to a; if granted, j is made better off.

9.20 A mechanism φ is immune to combined manipulations if for any school choice problem $\langle I, S, q, P, \succ \rangle$, there exist no $i \in I$, $s \in S$, and P_i' such that

- $sP_i\varphi[P](i)$, and
- $i \succ_s i'$ for some $i' \in \varphi[P_i', P_{-i}](s)$ or $|\varphi[P_i', P_{-i}](s)| < q_s$. (a student first misrepresents her preferences and then blocks the matching that is produced by the centralized mechanism)

☞ 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) φ is immune to combined manipulations.

9.21 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 the unique stable and strategy-proof mechanism for school choice problems; see Theorem 8.34.

(2) It suffices to show that DA is not immune to combined manipulations.

(3) Consider a problem with $I = \{i, j, k\}$, $S = \{a, b\}$, $q_a = q_b = 1$, and

i	j	k	a	b
b	a	a	i	k
a		b	j	i
			k	

Table 9.5

(4) Under the true preferences (P_j, P_{-j}), the DA produces

$$\begin{bmatrix} i & j & k \\ a & \emptyset & b \end{bmatrix}.$$

(5) Now consider a false preference $P_j' : \emptyset$. Then, under (P_j', P_{-j}), DA produces

$$\begin{bmatrix} i & j & k \\ b & \emptyset & a \end{bmatrix}.$$

(6) Since $aP_j\emptyset = \mathrm{DA}[P_j, P_{-j}](j)$ and $j \succ_a k \in \mathrm{DA}[P'_j, P_{-j}](a)$, DA is not robustly stable. More specifically, student j has incentives to first report P'_j and then block $\mathrm{DA}[P'_j, P_{-j}]$.

$\square$

9.22 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.

9.23 Once the priority structure $(\succ, q)$ is acyclic, DA is the unique robustly stable mechanism.

This theorem seems to be bad news for school systems: most priority structures violate acyclicity.

9.24 *Proof of Theorem 9.22, 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 9.6

It is easy to see that $\mathrm{DA}[P](j) = \emptyset$.

(3) Now consider a false preference of student j, $P'_j : \emptyset$.

(4) We have $\mathrm{DA}[P'_j, P_{-j}](k) = a$. Since

$$aP_j\emptyset = \mathrm{DA}[P](j) \text{ and } j \succ_a k \in \mathrm{DA}[P'_j, P_{-j}](a),$$

DA is not robustly stable.

$\square$

9.25 *Proof of Theorem 9.22, Part 2: "acyclicity implies robust stability".* Prove by contradiction.

(1) Assume that 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

- $cP_s\,\mathrm{DA}[P](s)$;
- $s \succ_c s'$ 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

$$P''_s\colon c, \emptyset, \quad P'' = (P''_s, P_{-s}).$$

(ii) If $\mathrm{DA}[P''](s) = c$. Since we have

$$\mathrm{DA}[P''](s) = cP_s\,\mathrm{DA}[P](s),$$

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

$$cP''_s\emptyset = \mathrm{DA}[P''](s). \tag{9.1}$$

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 (9.1) 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

$$P''_s\colon \emptyset, \quad P'' = (P''_s, P_{-s}).$$

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.

□

9.26 Remark: Given that DA is the unique stable and strategy-proof mechanism (see Theorem 8.34), this theorem implies that, given the market, there exists a robustly stable mechanism if and only if the priority structure is acyclic.

9.27 Afacan (2012) complemented the above results by considering group robustly stability that involves combined manipulations by groups of students.

As in the case with Pareto efficiency and group strategy-proofness, there could be (at least) two definitions of group robust stability, requiring that there is no group manipulation causing

- strict improvement for everyone in the manipulating coalition, or
- weak improvement for everyone, with at least one strict.

For the first concept (weaker requirement), it turns out that acyclicity is also a necessary and sufficient condition for group robust stability.

For the second concept (stronger requirement), the mechanism may be manipulable even with acyclic priority structures.

Chapter 10

Efficiency improvement on student-optimal stable mechanism

Contents

10.1 When the priority structure contains cycles, DA is not Pareto efficient as shown in Theorem 9.7. In Remark 8.36 and Corollary 8.39, 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.

10.1 Efficiency-adjusted deferred acceptance algorithm

10.2 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

The matching produced by DA is

$$\begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix},$$

and the procedure is

i	j	k	s_1	s_2
s_2	s_1	s_1	i	k
s_1		s_2	j	i
			k	

Table 10.1

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 10.2

10.3 In Example 10.2, 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.

10.4 Coming back to Example 10.2, suppose that 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 10.3

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 10.4

☞ 10.5 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 $l \in \{t, t+1, \dots, t'-1\}$, then we call student i an interrupter for school s and the pair (i, s) an interrupting pair of Step t'.

10.6 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 10.2. 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.

10.7 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 $\beta(i_1) P_{i_1} \alpha(i_1)$.

(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 α, 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 $(i_1, i_2, \dots, i_k)$ $(k \geq 2)$, 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:

$$\alpha(i_{\ell+1}) = \beta(i_\ell) P_{i_\ell} \alpha(i_\ell) = \beta(i_{\ell-1}).$$

(9) Let $i_\ell \in \{i_1, i_2, \ldots, i_k\}$ 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 that she is placed at the end of DA.

(10) Then the student $i_{\ell-1}$ in the above cycle who prefers school $\alpha(i_\ell)$ to the school $\alpha(i_{\ell-1})$ she is placed at under α was rejected from $\alpha(i_\ell)$ at an earlier step.

(11) Then, when student i_ℓ applies to school $\alpha(i_\ell)$, all the seats are already full and because student $\alpha(i_\ell)$ is placed at this school at the end of DA, some student i' is rejected.

(12) Thus, student i' is an interrupter for school $\alpha(i_\ell)$.

□

Consider an interrupting pair (i, s): it is possible that student i's rejection from school s (at Step t' according to the 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. (Exercise)

10.8 Efficiency-adjusted deferred acceptance mechanism (EADAM):

Round 0: Run DA for the school problem.

Round k: (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.

10.9 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:

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 10.5

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 10.6

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 10.7

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	$\cancel{i_4}, i_5$	i_5	$i_1, \cancel{i_5}, \cancel{i_6}$	i_1	i_1
s_2	i_1	$\cancel{i_1}, i_4$	i_4	i_4	i_4
s_3	$i_2, \cancel{i_3}$	i_2	i_2	i_2	i_2
s_4	i_6	$i_3, \cancel{i_6}$	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 10.8

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	$\cancel{i_4}, i_6$	i_6	$i_1, \cancel{i_6}$	i_1	i_1
s_2	i_1	i_1	i_1	$\cancel{i_1}, i_4$	i_4	i_4	i_4
s_3	$i_2, \cancel{i_3}$	i_2	i_2	i_2	i_2	i_2	i_2
s_4	i_6	$i_3, \cancel{i_6}$	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 10.9

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	$i_2, \cancel{i_3}$	i_2	i_2	i_2
s_4	i_6	$i_3, \cancel{i_6}$	i_3	i_3
s_5	i_5	i_5	i_5, i_6	i_5, i_6
$\emptyset$	i_3	i_6		

Table 10.10

End: There are no interrupting pairs; hence we stop.

10.10 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 $|I| \cdot |S|$ times, giving us a computationally simple polynomial-time algorithm.

10.11 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	k	

Table 10.11

The procedure of DA is

Step	1	2	3	4	5	End
s_1	$i, \cancel{j}$	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 10.12

The outcome of DA for this problem is not Pareto efficient. There are two interrupting pairs within the algorithm: (i, s_1) and (j, s_2).

Now consider the revised problem where we remove school s_1 from student i's preferences and school s_2 from those of student j. 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 10.13

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 was identified as an interrupter at Step 3 before student j, who was identified at Step 4. Thus, let us then consider the revised problem where we only remove school s_1 from student i'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 10.14

Once again, there is no change in the outcome. Hence, this approach does not work either.

10.12 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.

10.13 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 different school s_2^{r-1} (in Round $(r-1)$).

(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 different school s_3^{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 $(r-1)$.

□

10.14 Corollary (Corollary 1 in Kesten (2010)): If all students consent, then EADAM selects the Pareto efficient matching which eliminates justified envy whenever it exists.

10.15 Proposition (Proposition 3 in Kesten (2010)): The placement of a student does not change whether she consents or not.

This result makes sure that the students do not have incentive to not consent.

10.2 Simplified efficiency-adjusted deferred acceptance algorithm

10.16 When will a student consent to give up their own hope to help others? The simple answer is: when a student find herself cannot be Pareto improved anymore.

So, which students are Pareto unimprovable?

☞ 10.17 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.

10.18 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:

The DA procedure is

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 10.15

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	$i_2, \cancel{i_4}$	i_2	i_2
s_4					i_4	i_4
$\emptyset$	i_3	i_1	i_2	i_4		

Table 10.16

and the resulting matching is

$$\begin{bmatrix} i_1 & i_2 & i_3 & i_4 \\ s_1 & s_4 & s_2 & s_3 \end{bmatrix}.$$

Thus, school s_4 is underdemanded at the DA matching, since it never rejects any student throughout the DA procedure.

☞ 10.19 Definition: A school is tier-0 underdemanded at matching μ if it is underdemanded at μ.

For any positive integer k, 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.

In the previous example, school s_3 is tier-1 underdemanded at the DA matching.

☞ 10.20 Definition: School s is essentially underdemanded at matching μ if it is tier-k underdemanded at μ for some integer $k \geq 0$.

In the previous example, it is clear that s_1 and s_2 are not essentially underdemanded.

10.21 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.

10.22 Definition: Student i is not Pareto improvable (or, simply, unimprovable) at DA$[P, \succ]$ if for every matching μ that Pareto dominates DA$[P, \succ]$, $\mu(i) = \text{DA}[P, \succ](i)$.

10.23 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.

10.24 Lemma: At the DA matching, if all the students are matched, then there exists an underdemanded school.

Proof. Let μ be the DA matching. Let the last step of DA be Step k. Consider a student i who applies $\mu(i)$ at Step k under DA. Clearly, $\mu(i)$ is an underdemanded school. □

10.25 Simplified EADAM:

Round 0: Run DA for the school choice problem.

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. If all the schools are not underdemanded at the round-$(k-1)$ DA matching, then remove the students who are unmatched.

(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.

10.26 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 10.17

Suppose that j consents.

Round 0: The process of DA 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 10.18

Round 1: No underdemanded school exists. Remove j. Rerun DA:

Step	1	End
s_1	k	k
s_2	i	i
$\emptyset$		

Table 10.19

Round 2: s_1 and s_2 are underdemanded. Remove them with the matched students.

10.27 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 10.20

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 10.21

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 10.22

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.

10.28 Revisit the above example and suppose that student i_5 does not consent.

In Round 1, after removing i_5 and i_6, we have to modify the preferences for remaining students:

s_1	s_2	s_3	s_4	i_1	i_2	i_3	i_4
i_2	i_3	i_1	i_4	s_2	s_3	s_3	$\cancel{s_1}$
i_1	$\cancel{i_6}$	$\cancel{i_6}$	i_3	s_1	s_1	s_4	s_2
$\cancel{i_5}$	i_4	i_2	$\cancel{i_6}$	s_3	$\cancel{s_5}$	s_2	s_4
$\cancel{i_6}$	i_1	i_3	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
i_4	$\vdots$	$\vdots$					
i_3							

Table 10.23

Rerun DA:

Step	1	2	3	End
s_1	$\cancel{i_4}$		i_1	i_1
s_2	i_1	$\cancel{i_1}, i_4$	i_4	i_4
s_3	$i_2, \cancel{i_3}$	i_2	i_2	i_2
s_4		i_3	i_3	i_3
$\emptyset$	i_3, i_4	i_1		

Table 10.24

Round 2: At the end of Round 1, s_4 is the only underdemanded school, and i_3 is matched with it. Remove s_4 together with i_3, and rerun DA with the rest of the schools and students. The procedure of round-2 DA is illustrated in the following table:

Step	1	End
s_1	i_2	i_2
s_2	i_4	i_4
s_3	i_1	i_1
$\emptyset$		

Table 10.25

The final matching is the round-2 DA matching.

10.29 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.

10.30 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-$(k-1)$.

10.31 Lemma (Proposition 1 in Tang and Yu (2014)): The simplified EADAM is well-defined and stops within $|S \cup \{\emptyset\}| + 1 = m + 2$ rounds.

10.32 Theorem (Theorem 1 in Tang and Yu (2014)): The simplified EADAM is Pareto efficient when all students consent and is constrained efficient otherwise.

10.33 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.

10.34 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, and hence they are Pareto unimprovable.

10.35 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.

10.3 Stable improvement cycle algorithm

10.36 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 prefers s to her assignment under μ).

10.37 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) $i_\ell = d_{\mu(i_{\ell+1})}$.

10.38 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.

✌ 10.39 Theorem (Theorem 1 in Erdil and Ergin (2008)): In a school choice problem $\langle I, S, q, P, \succ\rangle$, let μ be a stable matching. If μ is Pareto dominated by another stable matching ν, then it admits a stable improvement cycle.

10.40 *Proof of Theorem 10.39.*

(1) Suppose that μ and ν are stable matchings and that 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) Lemma 8.38 implies that $\mu(I') = \nu(I') = S'$.

(4) Thus, for each $s \in S'$, there exists a student i such that $s = \nu(i) P_i \mu(i)$, *i.e.*, i desires s at μ and is assigned to s under ν.

(5) For each $s \in S'$, let i_s denote the highest $\succ_s$-priority student among those in I' that desire s at μ.

(6) Let school $\mu(i_s)$ point to s.

(7) By Lemma 8.38, $\mu(i_s) \in S'$.

(8) Since i_s desires s at μ, $\mu(i_s) \neq s$.

(9) Thus, we can repeat this for each school $s \in S'$ and find a school $t \in S' \setminus \{s\}$ that points to s.

(10) 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 $\ell = 0, 1, \ldots, n-1$.

(11) 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.

$$\mu(i_s) \to s = \nu(i_s) \Rightarrow \mu(i_\ell) = \mu(i_{s_{\ell+1}}) = s_\ell \to s_{\ell+1} = \nu(i_{s_{\ell+1}}) = \nu(i_\ell) = \mu(i_{\ell+1}).$$

(12) 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.

(13) Suppose $i_\ell \neq j$. Thus, $j \notin I'$ and $j \succ_{\mu(i_{\ell+1})} i_\ell$.

(14) Then $\mu(j) = \nu(j)$ by Lemma 8.38.

(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 higher $\succ_{\mu(i_{\ell+1})}$-priority than i_ℓ, who is matched to $\nu(i_\ell) = \mu(i_{\ell+1})$ under ν.

□

10.41 In a school choice problem $\langle I, S, q, P, \succ\rangle$ (with strict priorities), we cannot find a stable improvement cycle for the DA matching.

However, once we remove some students who are matched with essentially underdemanded schools, there could be a stable improvement cycle.

10.42 Iterative stable improvement cycles algorithm (Wang (2015)):

Step 0: Run DA algorithm and obtain a temporary matching μ^0.

Step k: (1) Identify the schools that are underdemanded at matching μ^{k-1}, settle the matching at these schools, and remove these schools and the students matched with them.

(2) For each non-consenting student i removed with the underdemanded schools, each remaining school s that i desires, and each remaining student j such that $i \succ_s j$, remove s from j's preference.

(3) For the remaining schools and students, identify all stable improvement cycles and carry out these cycles to obtain the matching μ^k. If there is no stable improvement cycle, let $\mu^k = \mu^{k-1}$, and move forward to the next round.

End: The algorithm terminates when all schools are removed.

10.43 Theorem (Theorem 1 in Wang (2015)): For every school choice problem, the matching produced by ISIC is the same as the outcome of EADAM when all students consent.

Chapter 11

School choice with weak priorities

Contents

11.1 Weak priorities

11.1 In the context of school choice, it might be reasonable to assume that the students have strict preferences, but school priority orderings 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 five indifference classes for each school in the following order:

(i) the students who already attend the school,

(ii) the students who have siblings at that school (sibling) and are in the reference area of the school (walk zone),

(iii) sibling,

(iv) walk zone,

(v) all other students.

11.2 There are at least two ways to break all the indifference classes.

- Single tie breaking: Use one lottery to decide the ordering on all students and, whenever two students are in the same indifference class, break the tie use the ordering.
- Multiple tie breaking: Draw one lottery for each school, and whenever two students are in the same indifference class for a school, break the tie using the ordering for that particular school.

Then, one can apply DA to obtain a matching with respect to the strict priority profile derived from the original one.

11.3 Policymakers from the NYC Department of Education believed that DA with single tie breaking rule is less equitable than multiple tie breaking rule:

> If we want to give each child a shot at each program, the only way to accomplish this is to run a new random. .... I cannot see how the children at the end of the line are not disenfranchised totally if only one run takes place. I believe that one line will not be acceptable to parents. When I answered questions about this at training sessions, (it did come up!) people reacted that the only fair approach was to do multiple runs.

11.4 Simulation (Table 1 in Abdulkadiroğlu *et al.* (2009)) suggests that single tie breaking rule is better in efficiency, although it is not too clear-cut.

11.5 Abdulkadiroğlu *et al.* (2015) showed 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.

11.2 DA with tie breaking rules

11.6 Proposition: DA with any tie breaking rule is stable.

Proof. 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. □

11.7 Proposition: DA with any tie breaking rule is strategy-proof.

Proof. Straightforward. □

11.8 DA with tie breaking rules does not necessarily bring 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 11.1

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

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix} \text{ and } \mu' = \begin{bmatrix} i & j & k \\ s_2 & \emptyset & s_1 \end{bmatrix}.$$

Clearly, μ is Pareto dominated by μ'.

11.9 DA with 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 11.2

Assume that ties are broken in the order $i \succ j \succ k$ for each school. DA with this tie breaking rule finds

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & s_2 & s_3 \end{bmatrix}.$$

However, 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.

11.10 If the priorities of schools are strict, then DA produces a constraint efficient matching. However, the two examples above illustrate that DA with tie breaking rules may not bring us a constrained efficient matching, provided that the priorities of schools are not strict.

11.11 Theorem (Theorem 1 in Abdulkadiroğlu *et al.* (2009)): For any tie breaking rule, there is no mechanism that is strategy-proof and dominates DA with the given tie breaking rule.

Proof. Recall Theorem 8.37. □

In other words, whatever efficiency improvement upon DA with tie breaking rules may become non-strategy-proof.

On the other hand, we could improve the efficiency upon DA with tie breaking rules without hurting the stability.

11.3 Stable improvement cycles algorithm

11.12 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 11.3

We choose the tie breaking rule $i \succ_{s_1} j \succ_{s_1} k$.

DA with this tie breaking rule produces

$$\mu = \begin{bmatrix} i & j & k \\ s_1 & \emptyset & s_2 \end{bmatrix}.$$

Clearly, μ is Pareto dominated by $\mu' = \begin{bmatrix} i & j & k \\ s_2 & \emptyset & s_1 \end{bmatrix}$.

Notice that i desires s_2 and j and k desire s_1. Besides, j and k share the same priority at school s_1. Thus, i and k can make an exchange so that finally i gets s_2 and k gets s_1. Meanwhile, such an exchange does not violate j's priority.

11.13 In a school choice problem $\langle I, S, q, P, \succsim\rangle$ with a given matching μ, for each school $s \in S$, let D_s be the set of highest $\succsim_s$-priority students among those who desire s (*i.e.*, who prefers s to her assignment under μ).

In the Example above, $D_{s_1} = \{j, k\}$ and $D_{s_2} = \{i\}$.

☞ 11.14 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) $i_\ell \in D_{\mu(i_{\ell+1})}$.

11.15 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.

✌ 11.16 Theorem (Theorem 1 in Erdil and Ergin (2008)): In a school choice problem $\langle I, S, q, P, \succsim\rangle$, let μ be a stable matching. If μ is Pareto dominated by another stable matching ν (*i.e.* μ is not constraint efficient), then it admits a stable improvement cycle.

Proof. (1) Suppose that μ and ν are stable matchings and that 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) Lemma 8.38 implies that $\mu(I') = \nu(I') = S'$.

(4) Thus, for each $s \in S'$, there exists a student i such that $s = \nu(i) P_i \mu(i)$, *i.e.*, i desires s at μ and is assigned to s under ν.

(5) For any $s \in S'$, let D'_s denote the set of highest $\succsim_s$-priority students among those in I' that desire s at μ.

(6) Fix an arbitrary student $i_s \in D'_s$ and let school $\mu(i_s)$ point to s.

(7) By Lemma 8.38, $\mu(i_s) \in S'$.

(8) Since i_s desires s at μ, $\mu(i_s) \neq s$.

(9) Thus, we can repeat this for each school $s \in S'$ and find a school $t \in S' \setminus \{s\}$ that points to s.

(10) 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 $\ell = 0, 1, \ldots, n-1$.

(11) 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.

$$\mu(i_s) \to s = \nu(i_s) \Rightarrow \mu(i_\ell) = \mu(i_{s_{\ell+1}}) = s_\ell \to s_{\ell+1} = \nu(i_{s_{\ell+1}}) = \nu(i_\ell) = \mu(i_{\ell+1}).$$

(12) Let D_s denote the set of highest $\succsim_s$-priority students among those who desire s at μ. In the following, we will show that $i_\ell \in D_{\mu(i_{\ell+1})}$.

(13) Suppose that $i_\ell \notin D_{\mu(i_{\ell+1})}$. Thus, $D_{\mu(i_{\ell+1})}$ has no intersection with I'.

(14) For any $j \in D_{\mu(i_{\ell+1})}$, we have $j \notin I'$ and $j \succ_{\mu(i_{\ell+1})} i_\ell$.

(15) Since $j \notin I'$, $\mu(j) = \nu(j)$ by Lemma 8.38.

(16) Since j desires $\mu(i_{\ell+1})$ at μ, j also desires $\mu(i_{\ell+1})$ at ν.

(17) This contradicts the stability of ν, since j has high $\succsim_{\mu(i_{\ell+1})}$-priority than i_ℓ, who is matched to $\nu(i_\ell) = \mu(i_{\ell+1})$ under ν.

□

11.17 Stable improvement cycles algorithm:

Step 0: Run DA algorithm and obtain a temporary matching μ^0.

Step k: (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.

11.18 Starting with an arbitrary stable matching, SIC produces a constrained efficient stable matching.

11.19 SIC is not strategy-proof.

Question. Hint: Consider 11.11.

11.20 The SIC algorithm is similar to but different from TTC:

- The cycles here are stable improvement cycles; students are pointing to all schools that are better than their current match. While in TTC, each agent points to her most favorite school.
- For convenience, the algorithm is described through the pointings among schools instead of that among students. Each school may point to none or multiple other schools. Hence, each school may be involved in multiple cycles, and cycle-selection is an issue (the simple way is to randomly pick one).

11.21 EADAM and simplified EADAM can also be applied to resolve the efficiency loss resulting from weak priorities. See Kesten (2010) and Tang and Yu (2014).

11.22 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 11.4

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.

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 quotas: the number of majority students matched to school s cannot exceed the majority quota q_s^M.
- Minority reserves: 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 school choice problem with minorities is tuple $\Gamma = \langle I, S, q, P, \succ \rangle$, 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 $s \in S$, $\succ_s$ is a strict priority order over the set of students.
- For each student $i \in I$, P_i is a strict preference over S and being unmatched (being unmatched is denoted by $\emptyset$). If $sP_i\emptyset$, then school s is said to be acceptable to student i. For each $i \in I$, let R_i be the symmetric extension of P_i.

☞ 12.4 A matching μ is a mapping from I to $S \cup \{\emptyset\}$ such that $|\mu^{-1}(s)| \leq q_s$ for all $s \in S$.

☞ 12.5 A mechanism is a systematic procedure that determines a matching for each school choice problem with minority students.

☞ 12.6 A matching μ Pareto dominates matching ν if $\mu(i)R_i\nu(i)$ for all $i \in I$ and $\mu(i)P_i\nu(i)$ 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 $\mu(i)R_i\nu(i)$ for all $i \in I^m$ and $\mu(i)P_i\nu(i)$ 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.

12.2 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. For each $s \in S$, let $\boldsymbol{q}_s = (q_s, q_s^M)$.

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) $\mu(i)R_i\emptyset$ for each $i \in I$, and

(2) if $sP_i\mu(i)$, then either

(i) $i \in I^m$, $|\mu^{-1}(s)| = q_s$ and $i' \succ_s i$ 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 $i' \succ_s i$ for all $i' \in \mu^{-1}(s)$, or

(iii) $i \in I^M$, $|\mu^{-1}(s) \cap I^M| = q_s^M$, and $i' \succ_s i$ 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 is stable under majority quotes if it always selects a stable matching under majority quotes for each school choice problem with minorities.

12.10 Deferred acceptance algorithm with majority quotas.

Step 1: 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 DA 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.

Start: 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.

Step 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.

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.

End: If no student remains, terminate.

✌ 12.13 Theorem (Theorem 1 in Kojima (2012)): Under DA with majority quotas, the affirmative action may hurt all the minority students.

Proof. (1) Consider a problem without affirmative action: $I = \{i_1, i_2, i_3\}$ with $I^M = \{i_1, i_2\}$ and $I^m = \{i_3\}$, $S = \{s_1, s_2\}$, $\boldsymbol{q}_{s_1} = (2, 2)$, $\boldsymbol{q}_{s_2} = (1, 1)$, and preferences and priorities are as follows:

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

(2) DA results in

$$\mu = \begin{bmatrix} s_1 & s_2 \\ i_1, i_2 & i_3 \end{bmatrix}.$$

(3) Now consider a new problem $\tilde{\Gamma} = \langle I, S, \tilde{\boldsymbol{q}}, P, \succ \rangle$ where s_2 applies the affirmative action $\tilde{\boldsymbol{q}}_{s_2} = (2, 1)$.

(4) In $\tilde{\Gamma}$, DA results in

$$\tilde{\mu} = \begin{bmatrix} s_1 & s_2 \\ i_1, i_3 & i_2 \end{bmatrix}.$$

(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.

□

12.14 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 $\tilde{\Gamma}$.

The reason that a quota for majority students can have adverse effects on minority students is simple. Consider a situation in which a school s is mostly desired by majorities. Then having a majority quota for s decreases the number of majority students who can be assigned to s 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.

12.15 The following example illustrate the case where the affirmative action benefits everyone, including the majority students, under DA with majority quotas.

Consider the following problem without affirmative action: $I = \{i_1, i_2, i_3, i_4\}$ 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

Then DA with majority quotas results in

$$\mu = \begin{bmatrix} s_1 & s_2 & \emptyset \\ i_1, i_4 & i_3 & i_2 \end{bmatrix}$$

Consider a new problem $\tilde{\Gamma} = \langle I, S, \tilde{\boldsymbol{q}}, P, \succ \rangle$ where s_1 applies the affirmative action $\tilde{\boldsymbol{q}}_{s_1} = (2, 1)$. Then in this problem, DA with majority quotas results in

$$\tilde{\mu} = \begin{bmatrix} s_1 & s_2 & \emptyset \\ i_1, i_3 & i_4 & i_2 \end{bmatrix}$$

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.

✌ 12.16 Theorem (Theorem 3 in Kojima (2012)): Under TTC with majority quotas, the affirmative action may hurt the minority students.

Proof. (1) Consider the problem without affirmative action: $I = \{i_1, i_2, i_3, i_4\}$ with $I^M = \{i_1, i_2\}$ and $I^m = \{i_3, i_4\}$, $S = \{s_1, s_2, s_3\}$, $\boldsymbol{q}_{s_1} = (2, 2)$, $\boldsymbol{q}_{s_2} = (1, 1)$, $\boldsymbol{q}_{s_3} = (1, 1)$, and preferences and priorities are as follows:

i_1	i_2	i_3	i_4	s_1	s_2	s_3
s_1	s_1	s_3	s_2	i_1	i_2	i_4
	s_3			i_2	i_3	
				i_3		
				i_4		

Table 12.3

(2) TTC produces the matching

$$\mu = \begin{bmatrix} s_1 & s_2 & s_3 \\ i_1, i_2 & i_4 & i_3 \end{bmatrix}$$

(3) Now suppose that s_1 applies the affirmative action $\tilde{\boldsymbol{q}}_{s_1} = (2, 1)$.

(4) In the new problem, TTC produces the matching

$$\mu' = \begin{bmatrix} s_1 & s_2 & s_3 & \emptyset \\ i_1 & i_4 & i_2 & i_3 \end{bmatrix}$$

(5) Every student is weakly worse off under μ' than under μ: Student i_1 and i_4 are indifferent, whereas i_2 and i_3 are strictly worse off. Note that i_3 is a minority student.

□

This result shows that TTC with majority quotas does not guarantee that an affirmative action has an intended effect to help the minority. Thus, the difficulty of affirmative action policies is not confined to DA with majority quotas.

Another remark is that every student is made weakly worse off by the affirmative action in the example used in the proof. Thus, it is possible that the policy unambiguously hurts welfare.

12.3 Affirmative action policies with minority reserves

12.17 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.

Under minority reserves: majority students may take the seats reserved for minority students if no minority students desire those seats.

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 $s \in S$.

☞ 12.18 Definition: Given $(r_s^m)_{s\in S}$, a matching μ is stable under minority reserves if

(1) $\mu(i) R_i \emptyset$ for each $i \in I$, and

(2) if $s P_i \mu(i)$, then either

(i) $i \in I^m$, $|\mu^{-1}(s)| = q_s$ and $i' \succ_s i$ 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 $i' \succ_s i$ for all $i' \in \mu^{-1}(s)$, or

(iii) $i \in I^M$, $|\mu^{-1}(s) \cap I^m| \leq r_s^m$, and $i' \succ_s i$ 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.

12.19 Definition: A mechanism is stable under minority reserves if it always selects a stable matching under minority reserves for each school choice problem.

12.20 Deferred acceptance algorithm with minority reserves:

Step 1: 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.

12.21 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.

12.22 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 $r_s^m + q_s^M = q_s$ for each $s \in S$. Let μ be a stable matching 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.

This result implies that for any stable matching under majority quotas, there exists a stable matching under the corresponding minority reserves that Pareto dominates it.

12.23 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 DA 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 $\mu^r(i) R_i \mu(i)$.

Theorem 12.13 shows that using majority quotas may hurt all the minority students in some settings. This result shows that this is impossible with minority reserves.

12.24 Example (Example 1 in Hafalir *et al.* (2013)): On very peculiar cases, such as the example below, imposing minority reserves can make some minorities worse off while leaving the rest indifferent.

Consider the problem: $I^M=\{i_1\}, I^m=\{i_2,i_3\}, S=\{s_1,s_2,s_3\}, q_{s_1}=q_{s_2}=q_{s_3}=1$, and students' preferences and schools' priorities are given by the table Minority reserves are given

i_1	i_2	i_3	s_1	s_2	s_3
s_1	s_3	s_1	i_1	i_1	i_1
s_3	s_1	s_2	i_2	i_2	i_2
s_2	s_2	s_3	i_3	i_3	i_3

by $r^m=(0,0,0)$. In this problem, the unique stable matching is

$$\mu=\begin{bmatrix} s_1 & s_2 & s_3\\ i_1 & i_3 & i_2\end{bmatrix}.$$

However, when minority reserves are $r^m=(1,0,0)$. In the new problem, the unique stable matching is

$$\mu'=\begin{bmatrix} s_1 & s_2 & s_3\\ i_2 & i_3 & i_1\end{bmatrix}.$$

With minority reserves, i_1 gets rejected from s_1 because of the presence of minority reserves at the first step of the algorithm. Then i_1 applies to s_3 and s_3 rejects i_2 in return. Next, i_2 applies to s_1 and s_1 rejects i_3. Finally, i_3 applies to s_2, which accepts her. Therefore, the introduction of minority reserves creates a rejection chain that makes some minority students worse off. Hence an increase in the minority reserves of s_1 makes i_2 worse off and i_3 indifferent.

12.25 Top trading cycles algorithm with minority reserves:

Step 1: If a school has minority reserves, then it points to its most preferred minority student; otherwise it points to the most preferred student.

Each student points to the most preferred school if there is an acceptable school and otherwise points to herself.

There exists at least one cycle. Each student in any of the cycles is matched to the school she is pointing to (if she is pointing to herself, then she gets her outside option).

All students in the cycles and schools that have filled their capacities are removed.

Step k: If a school has not filled its minority reserves, then it points to the most preferred minority student if there is any minority student left. Otherwise, it points to the most preferred student.

Each student points to the most preferred school if there is an acceptable school and otherwise points to herself.

There exists at least one cycle. Each student in any of the cycles is matched to the school she is pointing to (if she is pointing to herself, then she gets her outside option).

All students in the cycles and schools that have filled their capacities are removed.

End: If there is no cycle, then stop.

12.26 Proposition (Proposition 5 in Hafalir *et al.* (2013)): TTC with minority reserves is Pareto efficient and strongly group strategy-proof.

12.27 Theorem (Theorem 4 in Hafalir *et al.* (2013)): Suppose that μ^r and μ are the matchings produced by TTC with or without minority reserves r^m for a given preference profile. Then there exists $i \in I^m$ such that $\mu^r(i) R_i \mu(i)$.

This result implies that we cannot make all minority students worse off by having minority reserves.

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).

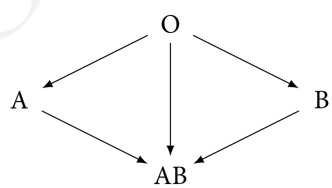
Live donation has been increasing recently.

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 donation) 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.

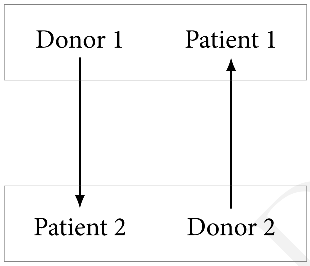


Figure 13.1: A paired exchange.

Take a look at the web page of Alliance for Paired Donation at http://paireddonation.org/.

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.

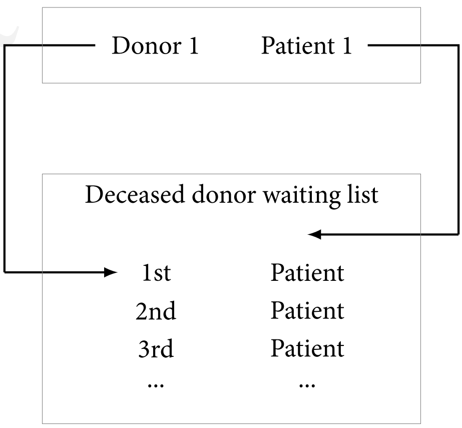


Figure 13.2: A list exchange.

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 $\{(k_1, t_1), \ldots, (k_n, t_n)\}$,
- a set of compatible kidneys $K_i \subseteq K = \{k_1, \ldots, k_n\}$ 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.

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.

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.

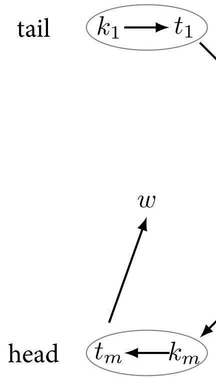


Figure 13.3: A w-chain.

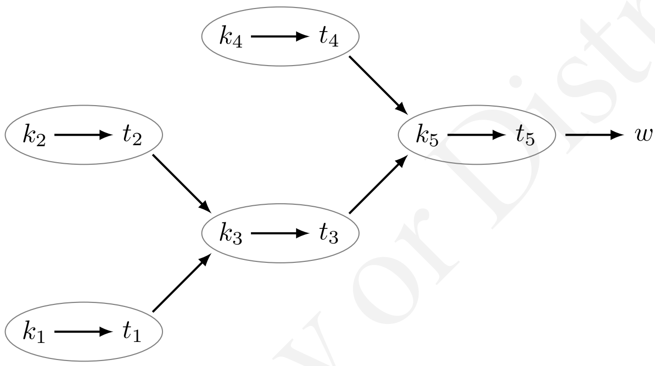


Figure 13.4: Five w-chains.

- 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:

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.

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

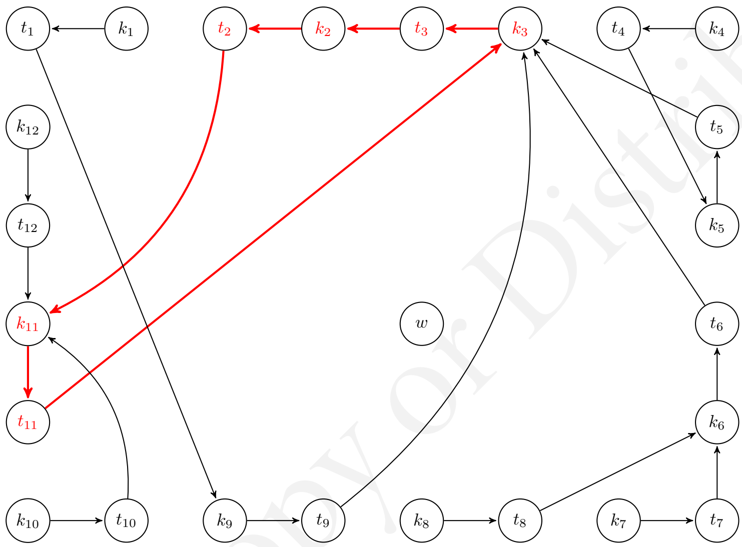


Figure 13.5: Round 1

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.

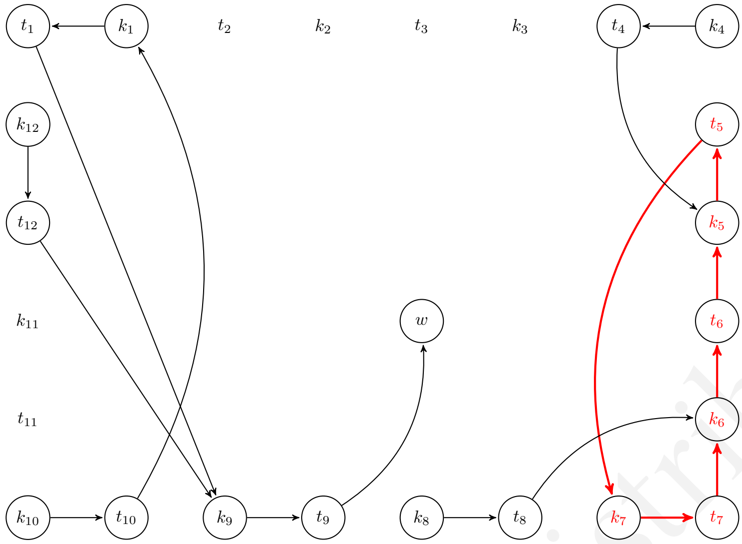


Figure 13.6: Round 2

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.

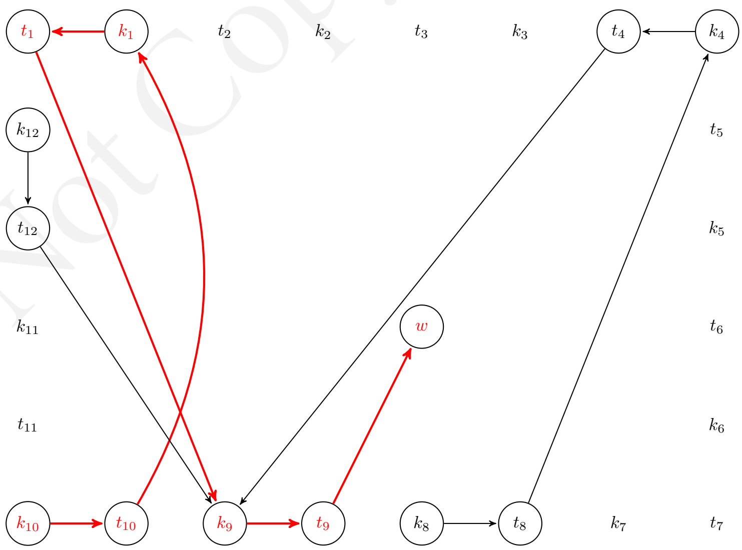


Figure 13.7: Round 3

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.

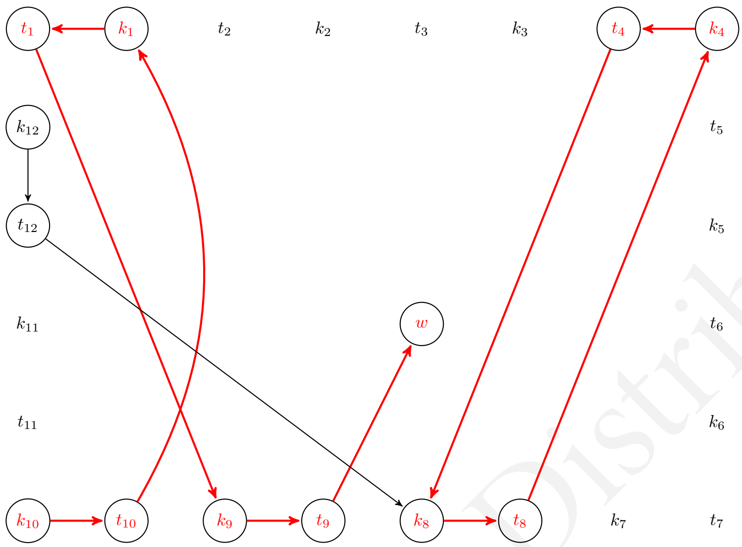


Figure 13.8: Round 4

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.

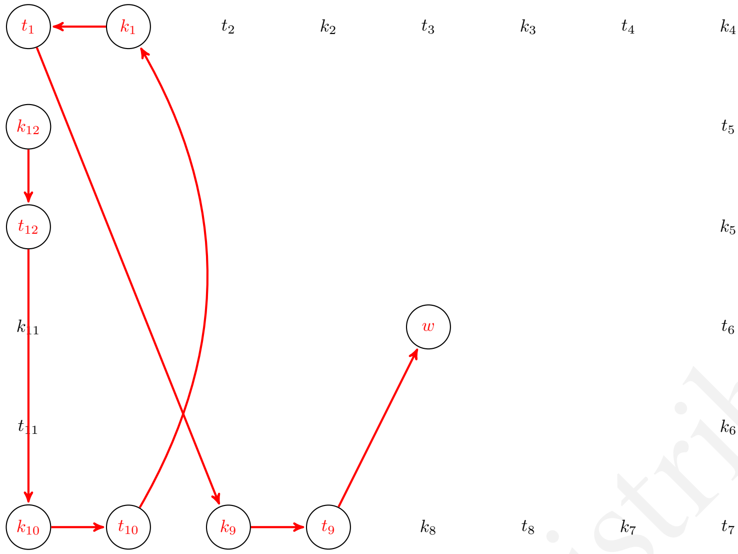


Figure 13.9: Round 5

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 wP_ik_i, construct P'_i from P_i by swapping the ranking of k_i and w.

(3) Note that $k_iP'_iw$ 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 4.27, 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 6.43: A mechanism is Pareto efficient, individually rational, strategy-proof, weakly neutral, and consistent if and only if it is a YRMH-IGYT mechanism.

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. 220

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth, The New York City high school match, *American Economic Review Papers and Proceedings* **95:2** (2005), 364–367. 72

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth, The Boston public school match, *American Economic Review Papers and Proceedings* **95:2** (2005), 368–371.

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. 72, 73, 179, 181, 220, 222

Atila Abdulkadiroğlu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez, Changing the Boston school choice mechanism, NBER Working Paper 11965. 176

Atila Abdulkadiroğlu and Tayfun Sönmez, Random serial dictatorship and the core from random endowments in house allocation problems, *Econometrica* **66:3** (1998), 689–701. 101, 103, 114, 138, 150

Atila Abdulkadiroğlu and Tayfun Sönmez, House allocation with existing tenants, *Journal of Economic Theory* **88** (1999), 233–260. 115, 124, 125, 126, 132

Atila Abdulkadiroğlu and Tayfun Sönmez, School choice: a mechanism design approach, *American Economic Review* **93:3** (2003), 729–747. 182, 229

Atila Abdulkadiroğlu and Tayfun Sönmez, Ordinal efficiency and dominated sets of assignments, *Jour-*

nal 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. 198

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. 39, 65, 177

Itai Ashlagi, Mark Braverman and Avinatan Hassidim, Stability in large matching markets with complementarities, *Operation Research* **62:4** (2014), 713–732.

Eduardo Azevedo and John William Hatfield, Existence of equilibrium in large matching markets with complementarities, working paper, 2015.

Eduardo Azevedo and Jacob Leshno, A supply and demand framework for two-sided matching markets, *Journal of Political Economy* **124:5** (2016), 1235–1268.

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. 171

Charles G. Bird, Group incentive compatibility in a market with indivisible goods, *Economics Letters* **14** (1984), 309–313. 93, 94

Anna Bogomolnaia and Hervé Moulin, A new solution to the random assignment problem, *Journal of Economic Theory* **100** (2001), 295–328. 146, 147, 150, 153, 155, 160, 165

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, Jinwoo Kim, and Fuhito Kojima, Stable matching in large economies, working paper, 2015.

Yeon-Koo Che and Youngwoo Koh, Decentralized college admissions, *Journal of Political Economy* **124:5** (2016), 1295–1338.

Yeon-Koo Che and Fuhito Kojima, Asymptotic equivalence of probabilistic serial and random priority mechanisms, *Econometrica* **78:5** (2010), 1625–1672. 165

Yeon-Koo Che and Olivier Tercieux, Efficiency and stability in large matching markets, working paper, July 2015.

Yeon-Koo Che and Olivier Tercieux, An analysis of top trading cycles in two-sided matching markets, working paper, July 2015.

Yeon-Koo Che and Olivier Tercieux, Payoff equivalence of efficient mechanisms in large matching markets, working paper, June 2015.

Yan Chen, Ming Jiang, Onur Kesten, Stephane Robin and Min Zhu, Matching in the large: An experimental study, working paper, 2016.

Yan Chen and Onur Kesten, Chinese college admissions and school choice reforms: A theoretical analysis, *Journal of Political Economy*, **125:1** (2017), 99–139. 184, 185

Yan Chen and Onur Kesten, Chinese college admissions and school choice reforms: An experimental study, working paper.

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 Tayfun Sönmez, School choice: An experimental study, *Journal of Economic Theory* **127** (2006), 202–231. 175

Andrei Ciupan, John William Hatfield and Scott Duke Kominers, An elementary proof of the Lone Wolf theorem, working paper, 2015. 20

Gabrielle Demange, David Gale and Marilda Sotomayor, A further note on the stable matching problem, *Discrete Applied Mathematics* **16** (1987), 217–222. 24, 44

Lester E. Dubins and David A. Freedman, Machiavelli and the Gale-Shapley algorithm, *American Mathematical Monthly* **88:7** (1981), 485–494. 41, 44

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, Contracts vs. salaries in matching, *American Economic Review* **102:1** (2012), 594–601.

Federico Echenique and Juan Sebastiàn Pereyra, Strategic complementarities and unraveling in matching markets, *Theoretical Economics*, forthcoming.

Federico Echenique and M. Bumin Yenmez, How to control controlled school choice, *American Economic Review*, **105:8** (2015), 2679–2694.

Lars Ehlers and Aytek Erdil, Efficient assignment respecting priorities, *Journal of Economic Theory* **145** (2010), 1269–1282.

Lars Ehlers, Isa E. Hafalir, M. Bumin Yenmez and Muhammed A. Yildirim, School choice with controlled choice constraints: Hard bounds versus soft bounds, *Journal of Economic Theory* **153** (2014), 648–683.

Ozgun Ekici, Reclaim-proof allocation of indivisible objects, *Games and Economic Behavior* **81** (2013), 1–10.

Aytek Erdil, Strategy-proof stochastic assignment, *Journal of Economic Theory* **151** (2014), 146–162. 165, 179

Aytek Erdil and Haluk Ergin, What's the matter with tie-breaking? Improving efficiency in school choice, *American Economic Review* **98** (2008), 669–689. 179, 216, 223

Haluk Ergin, Consistency in house allocation problems, *Journal of Mathematical Economics* **34** (2000), 77–97. 113

Haluk Ergin, Efficient resource allocation on the basis of priorities, *Econometrica* **70:6** (2002), 2489–2497. 188, 189, 195

Haluk Ergin and Tayfun Sönmez, Games of school choice under the Boston mechanism, *Journal of Public Economics* **90** (2006), 215–237.

Haluk Ergin, Tayfun Sönmez and M. Utku Ünver, Lung exchange, working paper, 2014.

Daniel Fragiadakis and Peter Troyan, Improving matching under hard distributional constraints, *Theoretical Economics*, forthcoming.

David Gale and Lloyd Shapley, College admissions and the stability of marriage, *American Mathematical Monthly* **69** (1962:1), 9–15. 16, 55

David Gale and Marilda Sotomayor, Some remarks on the stable matching problem, *Discrete Applied Mathematics* **11** (1985), 223–232. 19, 25, 29, 30, 32

Dan Gusfield and Robert W. Irving, *The Stable Marriage Problem: Structure and Algorithms*, Cambridge, MA: MIT Press, 1989. 207

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. 233, 234, 235

Rustamdjan Hakimov and Onur Kesten, The equitable top trading cycles mechanism for school choice, working paper, 2014. 183

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.

John William Hatfield and Fuhito Kojima, Matching with contracts: comment, *American Economic Review* **98:3** (2008), 1189–1194.

John William Hatfield and Fuhito Kojima, Group incentive compatibility for matching with contracts, *Games and Economic Behavior* **67** (2009), 745–749.

John William Hatfield and Fuhito Kojima, Substitutes and stability for matching with contracts, *Journal of Economic Theory* **145** (2010), 1704–1723.

John William Hatfield, Scott Duke Kominers, Alexandru Nichifor, Michael Ostrovsky, and Alexander Westkamp, Stability and competitive equilibrium in trading networks, *Journal of Political Economy* **121:5** (2013), 966–1005.

John William Hatfield and Paul R. Milgrom, Matching with contracts, *Econometrica* **95:4** (2005), 913–935.

J. S. Hwang, Modelling on college admissions in terms of stable marriages, working paper, Academia Sinica. 25

Aanund Hylland and Richard Zeckhauser, The efficient allocation of individuals to positions, *Journal of Political Economy* **87:2** (1979), 293–314. 99, 166

Robert W. Irving and Paul Leather, The complexity of counting stable marriages, *SIAM Journal of Computing* **15** (1986), 655–667. 22

Paula Jaramillo and Vikram Manjunath, The difference indifference makes in strategy-proof allocation of objects, *Journal of Economic Theory* **147:5** (2012), 1913–1946.

SangMok Lee, Incentive compatibility of large centralized matching markets, *Review of Economic Studies*, forthcoming.

SangMok Lee and Leeat Yariv, On the effciency of stable matchings in large markets, working paper, 2014.

Qingmin Liu, George J. Mailath, Andrew Postlewaite and Larry Samuelson, Stable matching with incomplete information, *Econometrica* **82:2** (2014), 541–587.

Yuichiro Kamada and Fuhito Kojima, Stability and strategy-proofness for matching with constraints: A problem in the Japanese medical match and its solution, *American Economic Review Papers and Proceedings 102:3* (2012), 366–370.

Yuichiro Kamada and Fuhito Kojima, Efficient matching under distributional constraints: theory and applications, *American Economic Review* **105:1** (2015), 67–99.

Yuichiro Kamada and Fuhito Kojima, General theory of matching under distributional constraints, working paper, 2015.

Yuichiro Kamada and Fuhito Kojima, Stability concepts in matching under distributional constraints, working paper, 2015.

Yuichiro Kamada and Fuhito Kojima, Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition, working paper, 2016.

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

Onur Kesten, On two competing mechanisms for priority-based allocation problems, *Journal of Economic Theory* **127** (2006), 155–171.

Onur Kesten, Why do popular mechanisms lack efficiency in random environments? *Journal of Economic Theory* **144** (2009), 2209–2226. 165

Onur Kesten, School choice with consent, *The Quarterly Journal of Economics* **125** (2010), 1297–1348. 179, 202, 204, 208, 209, 224

Onur Kesten, On two kinds of manipulation for school choice problems, *Economic Theory* **51** (2012), 677–693. 69

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, The coordinate-wise core for multiple-type housing markets is second-best incentive compatible, *Journal of Mathematical Economics* **44** (2008), 919–924.

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. 46

Fuhito Kojima, Robust stability in matching markets, *Theoretical Economics* **6** (2011), 257–267. 196, 197

Fuhito Kojima, School choice: Impossibilities for affirmative action, *Games and Economic Behavior* **75** (2012), 685–693. 230, 231

Fuhito Kojima, *Recent Developments in Matching Theory and its Practical Applications*, Advances in Economics and Econometrics: 11th world congress of Econometric Society.

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. 165

Fuhito Kojima and Parag Pathak, Incentives and stability in large two-sided matching markets, *American Economic Review* **99** (2009), 608–627. 73

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.

Fuhito Kojima, Akihisa Tamura, and Makoto Yokoo, Designing matching mechanisms under constraints: An approach from discrete convex analysis, working paper, 2015

Hideo Konishi and M. Utku Ünver, Games of capacity manipulation in hospital-intern markets, *Social Choice and Welfare* **27** (2006), 3–24. 68

Donald Knuth, *Mariages stables et leurs relations avec d'autres problèmes combinatoires: Introduction à l'analyse mathématique des algorithmes*, Les Presses de l'Université de Montréal, 1976. 18, 19

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. 254

Jinpeng Ma, Strategy-proofness and the strict core in a market with indivisibilities, *International Journal of Game Theory* **23** (1994), 75–83. 95

Mihai Manea, Asymptotic ordinal inefficiency of random serial dictatorship, *Theoretical Economics* **4:2** (2009), 165–197.

Giorgio Martini, Strategy-proof and fair assignment is wastful, *Games and Economic Behavior* **98** (2016), 172–179. 165

Andrew McLennan, Ordinal efficiency and the polyhedral separating hyperplane theorem, *Journal of Economic Theory* **105** (2002), 435–449. 150

D. G. McVitie and L.B.Wilson, Stable marriage assignment for unequal sets, *BIT Numerial Mathematics* **10** (1970), 295–309. 19

D. G. McVitie and L.B.Wilson, The stable marriage problem, *Communications of the ACM* **14** (1971), 486–492. 22

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. 191

Michael Ostrovsky, Stability in supply chain networks, *American Economic Review* **98:3** (2008), 897–923.

Szilvia Pápai, Strategyproof assignment by hierarchical exchange, *Econometrica* **68** (2000), 1403–1433. 172

Parag Pathak and Jay Sethuraman, Lotteries in student assignment: An equivalence result, *Theoretical Economics* **6** (2011), 1–17. 138

Parag 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.

Parag 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. 166

Marek Pycia and M. Utku Ünver, Incentive compatible allocation and exchange of discrete resources, working paper, 2015.

Alvin E. Roth, Incentive compatibility in a market with indivisible goods, *Economics Letters* **9** (1982), 127–132. 90, 91

Alvin E. Roth, The economics of matching: stability and incentives, *Mathematics of Operations Research* **7:4** (1982), 671–628. 22, 36, 41, 42, 43, 65

Alvin E. 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. 66, 71, 72

Alvin E. Roth, The college admissions problem is not equivalent to the marriage problem, *Journal of Economic Theory* **36** (1985), 277–288. 65

Alvin E. Roth, Common and conflicting interests in two-sided matching markets, *European Economic Review* **27** (1985), 75–96. 63

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. 58

Alvin E. Roth, A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom, *American Economic Review* **81** (1991), 415–440. 66

Alvin E. Roth, The economist as engineer: Game theory, experimentation, and computation as tools for design economics, *Econometrica* **70** (2002), 1341–1378. 35

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. 16

Alvin E. Roth, *Who Gets What—and Why: The New Economics of Matchmaking and Market Design*, Eamon Dolan/Houghton Mifflin Harcourt, June 2, 2015.
中文翻译：埃尔文·罗斯著，傅帅雄译，共享经济，机械工业出版社，2015。2

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. 66, 72

Alvin E. Roth and Andrew Postlewaite, Weak versus strong domination in a market with indivisible goods, *Journal of Mathematical Economics* **4** (1977), 131–137. 88

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver, Kidney exchange, *The Quarterly Journal of Economics* **119:2** (2004), 457–488. 243, 244, 246, 250, 251, 253

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. 22, 54, 57, 58, 59

Alvin E. Roth and John Vande Vate, Random paths to stability in two-sided matching, *Econometrica* **58** (1990), 1475–1480. 72

Toyotaka Sakai, *Market Design*, Chikumashobo Ltd., 2013.
中文翻译：坂井丰贵著，蔡晓智译，合适，后浪出版公司，江西人民出版社，2016。2

Mark A. Satterthwaite and Hugo Sonnenschein, Strategy-proof allocation mechanisms at differentiable points, *Review of Economic Studies* **48** (1981), 587–597. 45

Lloyd Shapley and Herbert Scarf, On cores and indivisibility, *Journal of Mathematical Economics* **1** (1974), 23–37. 77, 81

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. 67

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. 134

Tayfun Sönmez and M. Utku Ünver, House allocation with existing tenants: a characterization, *Games and Economic Behavior* **69** (2010), 425–445. 133

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. 108, 110, 111

Qianfeng Tang and Jingsheng Yu, A new perspective on Kesten's school choice withconsent idea, *Journal of Economic Theory* **154** (2014), 543–561. 209, 211, 212, 215, 224

Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan, Gale-Shapley stable marriage problem revisited: Strategic issues and applications, *Management Science* **47:9** (2001), 1252–1267.

William Thomson, Non-bossiness, working paper, 2014. 45

M. Utku Ünver, Dynamic kidney exchange, *Review of Economic Studies* **77** (2010), 372–414.

Ziwei Wang, Improving on the optimal stable matching: a trading cycles algorithm, working paper, 2015. 217

Binzhen Wu and Xiaohan Zhong, Matching mechanisms and matching quality: Evidence from a top university in China, *Games and Economic Behavior* **84** (2014), 196–215. 184

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. 165