Matching Theory

Mihai Manea

MIT

Based on slides by Fuhito Kojima.

Market Design

- Traditional economics focuses mostly on decentralized markets.
- Recently, economists are helping to design economic institutions for centralized markets.
  - placing students in schools
  - matching workers to firms in labor markets
  - matching patients to compatible organ donors
  - allocating space, positions, tasks
  - auctioning electromagnetic spectrum, landing slots at aiports
- The economics of market design analyzes and develops institutions. Practical solutions require attention to the details and objectives of concrete markets.

Hospitals and Residents

- Graduating medical students are hired as residents at hospitals.
- In the US more than 20,000 doctors and 4,000 hospitals are matched through a clearinghouse, the National Resident Matching Program.
- Doctors and hospitals submit preference rankings and the clearinghouse uses an algorithm to assign positions.
- Some centralized markets succeed, while others fail. What makes a good matching mechanism?

School Choice

- School districts use centralized student placement mechanisms.
- School districts take into account the preferences of students and decide the priorities each school assigns to students.
- What is a desirable student placement mechanism? Walking distance, siblings, affirmative action, test scores. . .

Kidney Exchange

- Some patients who need a kidney find a willing donor. The patient may be incompatible with the donor, in which case a direct transplant is not feasible.

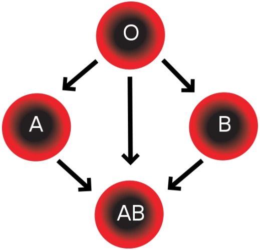

Figure : Blood type compatibility

- A kidney exchange matches two (or more) incompatible donor-patient pairs and swaps donors.
- How to design efficient kidney exchange mechanisms? Incentive and fairness requirements?

One-to-One Matching

The Marriage Problem

A **one-to-one matching** or **marriage problem** (Gale and Shapley 1962) is a triple (M, W, R), where

- $M = \{m_1, ..., m_p\}$ is a set of **men**
- $W = \{w_1, ..., w_q\}$ is a set of **women**
- $R = (R_{m_1}, \ldots, R_{m_p}, R_{w_1}, \ldots, R_{w_q})$ is a **preference profile**.

For $m \in M$, R_m is a **preference relation** over $W \cup \{m\}$.
For $w \in W$, R_w is a **preference relation** over $M \cup \{w\}$.

P_m, P_w denote the **strict preferences** derived from R_m, R_w.

In applications men and women correspond to students and schools, doctors and hospitals, etc. Extend theory to the case where a woman can be matched to multiple men, **many-to-one matching**.

Preferences

Consider a man m

- $wP_m w'$: man m prefers woman w to woman w'
- $wP_m m$: man m prefers woman w to being single
- $mP_m w$: woman w is **unacceptable** for man m

Similar interpretation for women.

Assumption: All preferences are *strict*.

Matchings

The outcome of a marriage problem is a **matching**. A matching is a function $\mu : M \cup W \to M \cup W$ such that

- $\mu(m) \in W \cup \{m\}, \forall m \in M$
- $\mu(w) \in M \cup \{w\}, \forall w \in W$
- $\mu(m) = w \iff \mu(w) = m, \forall m \in M, w \in W$.

Assumption: There are no **externalities**. Agent $i \in M \cup W$ prefers a matching μ to a matching ν iff $\mu(i) P_i \nu(i)$.

Stability

A matching μ is **blocked by an agent** $i \in M \cup W$ if $iP_i\mu(i)$. A matching is **individually rational** if it is not blocked by any agent.

A matching μ is **blocked by a man-woman pair** $(m, w) \in M \times W$ if both m and w prefer each other to their partners under μ, i.e.,

$$wP_m\mu(m) \ \ \& \ \ mP_w\mu(w).$$

A matching is **stable** if it is not blocked by any agent or pair of agents.

Stability and the Core

Proposition 1

The set of stable matchings coincides with the core of the associated cooperative game.

Proof.

A matching μ is in the core if there exists no matching ν and coalition $S \subset M \cup W$ such that $\nu(i) P_i \mu(i)$ and $\nu(S) \subset S$... □

The Deferred Acceptance (DA) Algorithm

Theorem 1 (Gale and Shapley 1962)

Every marriage problem has a stable matching.

The following **men-proposing deferred acceptance algorithm** yields a stable matching.

Step 1. Each man proposes to his first choice (if acceptable). Each woman *tentatively* accepts her most preferred acceptable proposal (if any) and rejects all others.

Step $k \geq 2$**.** Any man rejected at step $k - 1$ proposes to his next highest choice (if any). Each woman *tentatively* accepts her most preferred acceptable proposal to date and rejects the rest.

The algorithm terminates when there are no new proposals, in *finite* time.

Each woman is matched with the man whose proposal she holds (if any) at the last step. Any woman who has never tentatively accepted someone or any man who has been rejected by all acceptable women remains single.

Example

$$\underbrace{\begin{aligned} P_{m_1} &: w_2 \succ w_1 \succ w_3 \succ m_1 \\ P_{m_2} &: w_1 \succ w_2 \succ w_3 \succ m_2 \\ P_{m_3} &: w_1 \succ w_2 \succ w_3 \succ m_3 \end{aligned}}_{\text{Men's Preferences}}$$

$$\underbrace{\begin{aligned} P_{w_1} &: m_1 \succ m_3 \succ m_2 \succ w_1 \\ P_{w_2} &: m_2 \succ m_1 \succ m_3 \succ w_2 \\ P_{w_3} &: m_2 \succ m_1 \succ m_3 \succ w_3 \end{aligned}}_{\text{Women's Preferences}}$$

The resulting matching is

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

- Women get weakly better off and men get weakly worse off as the algorithm proceeds.
- The algorithm eventually stops, producing a matching μ.
- μ is stable
    - μ cannot be blocked by any individual agent, since men never propose to unacceptable women and women immediately reject unacceptable men.
    - Suppose the pair (m, w) blocks μ. Then $wP_m\mu(m)$ implies that m proposed to w in the algorithm and, as they are not matched with each other, w rejected m in favor of someone better. But w gets weakly better throughout the algorithm, hence $\mu(w)P_wm$, which contradicts the assumption that (m, w) blocks μ.

Stable Mechanisms in Real Markets

- Stability is theoretically appealing, but is it relevant in applications?
- Roth (1984) showed that the NRMP algorithm is equivalent to a (hospital-proposing) DA algorithm, so NRMP produces a stable matching.
- Roth (1991) studied the British medical match, where various regions use different matching mechanisms. Stable mechanisms outlast unstable ones.

Evidence from the Medical Match

Market	Stable	Still in use
NRMP	yes	yes (new design 98-)
Edinburgh ('69)	yes	yes
Cardiff	yes	yes
Birmingham	no	no
Edinburgh ('67)	no	no
Newcastle	no	no
Sheffield	no	no
Cambridge	no	yes
London Hospital	no	yes

Men-Optimal Stable Matching

Theorem 2 (Gale and Shapley 1962)

*There exists a **men-optimal stable matching** that every man weakly prefers to any other stable matching. Furthermore, the men-proposing deferred acceptance algorithm delivers the men-optimal stable matching.*

Proof.

We say that w is **achievable** for m if there is some stable matching μ with $\mu(m) = w$. For a contradiction, suppose a man is rejected by an achievable woman at some stage of the deferred acceptance algorithm.

Consider the first step of the algorithm in which a man m is rejected by an achievable woman w. Let μ be a stable matching where $\mu(m) = w$. Then w tentatively accepted some other man m' at this step, so (i) $m' P_w m$. Since this is the first time a man is rejected by an achievable woman, (ii) $w P_{m'} \mu(m')$.

By (i) and (ii), (m', w) blocks μ, contradicting the stability of μ. □

The Opposing Interests of Men and Women

Analagous to the men-optimal stable matching, there is a women-optimal stable matching (obtained by a version of the deferred acceptance algorithm where women propose).

- μ^M: men-optimal stable matching
- μ^W: women-optimal stable matching

Theorem 3 (Knuth 1976)

μ^W is the worst stable matching for each man. Similarly, μ^M is the worst stable matching for each woman.

Example with 2 men, 2 women, with "reversed" preferences

Proof of Opposing Interests

Suppose there is a man m and stable matching μ such that $\mu^W(m) P_m \mu(m)$.

Then m is not single under μ^W. Let $w = \mu^W(m)$. Clearly, $w \neq \mu(m)$, so $m \neq \mu(w)$.

By the definition of μ^W, $m = \mu^W(w) P_w \mu(w)$. But then (m, w) blocks μ, yielding the desired contradiction.

Relevance of Opposing Interests

The result shows that different stable matchings may benefit different market participants. In particular, each version of the deferred acceptance algorithm favors one side of the market at the expense of the other.

This point was part of a policy debate in NRMP in the 90s. The previous NRMP algorithm had hospitals proposing. Medical students argued that the system favors hospitals over doctors and called for the doctor-proposing version of the mechanism.

Agents Matched at Stable Matchings

$\tilde{\mu}(W) := \mu(W) \cap M$: set of men who are matched under μ
$\tilde{\mu}(M) := \mu(M) \cap W$: set of women who are matched under μ

Theorem 4 (McVitie and Wilson 1970)

The set of matched agents is identical at every stable matching.

Proof.

Let μ be an arbitrary stable matching.

- $|\tilde{\mu}^M(W)| \geq |\tilde{\mu}(W)| \geq |\tilde{\mu}^W(W)|$, since any man matched under μ (μ^W) is also matched under μ^M (μ)
- similarly, $|\tilde{\mu}^W(M)| \geq |\tilde{\mu}(M)| \geq |\tilde{\mu}^M(M)|$
- obviously, $|\tilde{\mu}^M(W)| = |\tilde{\mu}^M(M)|$ & $|\tilde{\mu}^W(W)| = |\tilde{\mu}^W(M)|$, hence all inequalities hold with equality
- in particular, $|\tilde{\mu}^M(W)| = |\tilde{\mu}(W)|$; since any man matched under μ is also matched under μ^M, we get $\tilde{\mu}^M(W) = \tilde{\mu}(W)$
- analogous argument for women □

Relevance of the Result

One motivation is the allocation of residents to hospitals in rural areas. Rural hospitals are not attractive to residents and have difficulties filling their positions. It has been argued that the matching mechanism should be adjusted so that more doctors go to rural areas. The theorem shows that this is not feasible if stable matchings are implemented.

Also, if some students were matched at some stable matchings and not others, they would find it unfair if one of the matchings that do not include them is selected.

Join and Meet

Definition 1

For any matchings μ and μ', the function $\mu \vee^M \mu' : M \cup W \to M \cup W$ (**join** of μ and μ') assigns each man the more preferred of his two assignments under μ and μ' and each woman the less preferred.

$$\mu \vee^M \mu'(m) = \begin{cases} \mu(m) & \text{if } \mu(m) R_m \mu'(m) \\ \mu'(m) & \text{if } \mu'(m) P_m \mu(m) \end{cases}$$

$$\mu \vee^M \mu'(w) = \begin{cases} \mu(w) & \text{if } \mu'(w) R_w \mu(w) \\ \mu'(w) & \text{if } \mu(w) P_w \mu'(w) \end{cases}$$

$\mu \wedge^M \mu' : M \cup W \to M \cup W$ (**meet** of μ and μ') is defined analogously, by reversing preferences.

Example

$$
\begin{array}{llll}
P_{m_1}: & w_1\ w_2\ w_3\ m_1 & P_{w_1}: & m_2\ m_3\ m_1\ w_1 \\
P_{m_2}: & w_2\ w_3\ w_1\ m_2 & P_{w_2}: & m_3\ m_1\ m_2\ w_2 \\
P_{m_3}: & w_2\ w_1\ w_3\ m_3 & P_{w_3}: & m_1\ m_2\ m_3\ w_3
\end{array}
$$

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

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

The join and meet of μ and μ' are

$$
\mu \vee^M \mu' = \begin{pmatrix} m_1 & m_2 & m_3 & w_1 & w_2 & w_3 \\ w_1 & w_2 & w_2 & m_1 & m_2 & m_3 \end{pmatrix}
$$

$$
\mu \wedge^M \mu' = \begin{pmatrix} m_1 & m_2 & m_3 & w_1 & w_2 & w_3 \\ w_3 & w_1 & w_3 & m_2 & m_3 & m_1 \end{pmatrix}
$$

Neither is a matching!

The Lattice Structure of the Set of Stable Matchings

Theorem 5 (Conway)

*If μ and μ' are **stable** matchings, then $\mu \vee^M \mu'$ and $\mu \wedge^M \mu'$ are*

1. *matchings*
2. *stable.*

We prove the result for the join. The proof for the meet is similar.

Proof of Part 1.

$\mu \vee^M \mu'$ is a **matching**

- The sets of single agents under μ and μ' are identical (Theorem 4), hence also identical under $\mu \vee^M \mu'$.
- If a man-woman pair are matched to each other under both μ and μ', this also holds under $\mu \vee^M \mu'$.
- Consider a man m with different mates under μ and μ'. W.l.o.g., assume $w := \mu(m) P_m \mu'(m)$. Then $\mu \vee^M \mu'(m) = w$.
- We need to show that $\mu \vee^M \mu'(w) = m$. Else, $m = \mu(w) P_w \mu'(w)$ and hence (m, w) blocks μ', contradicting its stability. □

Proof of Part 2.

$\mu \vee^M \mu'$ is **stable**

- For a contradiction, suppose that (m, w) blocks $\mu \vee^M \mu'$. W.l.o.g., assume $\mu \vee^M \mu'(w) = \mu(w)$.
- Then

$$mP_w[\mu \vee^M \mu'(w)] = \mu(w) \text{ and } wP_m[\mu \vee^M \mu'(m)]R_m\mu(m),$$

so (m, w) blocks μ, contradicting its stability. □

Discuss median stable matchings.

Strategic Behavior

- We know many desirable properties of stable matchings, given information about the preferences of market participants.
- But in reality, preferences are private information, so the clearinghouse needs to rely on the participants' reports.
- Do participants have incentives to state their preferences truthfully?

Direct Mechanisms

Fix M and W, so that each preference profile R defines a marriage problem.

$\mathcal{R}_i$: set of all preference relations for agent i
$\mathcal{R} = \prod_{i \in M \cup W} \mathcal{R}_i$: set of all preference profiles
$\mathcal{R}_{-i}$: set of all preferences for all agents except i
$\mathcal{M}$: set of all matchings

A **mechanism** is a systematic procedure which determines a matching for every marriage problem. Formally, a mechanism is a function $\varphi : \mathcal{R} \to \mathcal{M}$.

Stable Mechanisms

A mechanism φ is **stable** if $\varphi(R)$ is stable for each $R \in \mathcal{R}$.

φ^M (φ^W) : the mechanism that selects the men-(women-)optimal stable matching for each problem

Preference Revelation Games

Each mechanism φ induces a **preference revelation game** for every preference profile R where

- the set of **players** is $M \cup W$
- the **strategy** space for player i is $\mathcal{R}_i$
- the **outcome** is determined by the mechanism—if agents *report* R', the outcome is $\varphi(R')$
- i's **preferences** over outcomes are given by his *true* preference R_i.

A mechanism φ is **strategy-proof** if, for every (true) preference profile R, truthful preference revelation is a **(weakly) dominant strategy** for every player in the induced preference revelation game.

Formally, a mechanism φ is strategy-proof if

$$\varphi(R_{-i}, R_i)(i)\, R_i\, \varphi(R_{-i}, R_i')(i),$$

$$\forall i \in M \cup W,\ \forall R_i, R_i' \in \mathcal{R}_i, \forall R_{-i} \in \mathcal{R}_{-i}.$$

φ^M (φ^W) is Not Strategy-Proof

- Let $M = \{m_1, m_2\}$, $W = \{w_1, w_2\}$ and

$$\begin{aligned} P_{m_1} &: w_1, w_2, m_1 \\ P_{m_2} &: w_2, w_1, m_2 \\ P_{w_1} &: m_2, m_1, w_1 \\ P_{w_2} &: m_1, m_2, w_2. \end{aligned}$$

- When each agent reports his true preferences, φ^M produces $\varphi^M(R) = \{(m_1, w_1), (m_2, w_2)\}$.
- If w_1 instead reports

$$P'_{w_1} : m_2, w_1, m_1$$

then φ^M produces $\varphi^M(R') = \{(m_1, w_2), (m_2, w_1)\}$, which w_1 prefers to $\varphi^M(R)$.

- Hence w_1 has incentives to misreport her preferences and the deferred acceptance mechanism is not strategy-proof.

Incompatibility of Stability and Strategy-Proofness

Theorem 6 (Roth 1982)

*There exists no mechanism that is both **stable** and **strategy-proof**.*

Proof.

Consider the following 2 men, 2 women problem

$$\begin{array}{llll} R_{m_1}: & w_1 \; w_2 \; m_1 & R_{w_1}: & m_2 \; m_1 \; w_1 \\ R_{m_2}: & w_2 \; w_1 \; m_2 & R_{w_2}: & m_1 \; m_2 \; w_2 \end{array}$$

In this problem there are only two stable matchings,

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

Let φ be any stable mechanism. Then $\varphi(R) = \mu^M$ or $\varphi(R) = \mu^W$.

next slide...

Proof (Continuation).

Suppose that $\varphi(R) = \mu^M$. If w_1 misrepresents her preferences to be

$$R'_{w_1} : m_2, w_1, m_1$$

then μ^W is the unique stable matching for the manipulated economy (R_{-w_1}, R'_{w_1}). Since φ is stable, it must be that $\varphi(R_{-w_1}, R'_{w_1}) = \mu^W$. But then φ is not strategy-proof, as $\mu^W \, P_{w_1} \, \mu^M$.

If, on the other hand, $\varphi(R) = \mu^W$ then m_1 can report false preferences

$$R'_{m_1} : w_1, m_1, w_2$$

and ensure that his favorite stable matching μ^M is selected by φ, since it is the only stable matching for the manipulated economy (R_{-m_1}, R'_{m_1}). □

A Stronger Negative Result

Every stable matching is Pareto efficient (proof?) and individually rational.

Theorem 7 (Alcalde and Barbera 1994)

There exists no mechanism that is ***Pareto efficient, individually rational*** *and* ***strategy-proof****.*

Proof.

For R from the previous proof, any efficient and individually rational mechanism φ satisfies $\varphi(R) = \mu^M$ or $\varphi(R) = \mu^W$. Suppose $\varphi(R) = \mu^M$. Moreover, $\varphi(R_{-w_1}, R'_{w_1}) \in \{\mu^W, \mu\}$, where $\mu = \{(m_2, w_2)\}$.

If $\varphi(R_{-w_1}, R'_{w_1}) = \mu^W$, we obtain a contradiction as before.

Suppose $\varphi(R_{-w_1}, R'_{w_1}) = \mu$. Consider $R'_{w_2} : m_1, w_2, m_2$. The only efficient and individually rational matching at (R_M, R'_W) is μ^W, so $\varphi(R_M, R'_W) = \mu^W$. But then $m_1 = \varphi_{w_2}(R_M, R'_W)\ R_{w_2}\ \varphi_{w_2}(R_{-w_1}, R'_{w_1}) = m_2$, and w_2 has incentives to report R'_{w_2} at (R_{-w_1}, R'_{w_1}). □

Incentives of Men Under φ^M

Theorem 8 (Dubins and Freedman 1981, Roth 1982)

Truth-telling is a weakly dominant strategy for all men under the men-optimal stable mechanism.

Similarly, truth-telling is a weakly dominant strategy for all women under the women-optimal stable mechanism.

Many-to-One Matching

College Admissions Problems

A **college admissions problem** (Gale and Shapley 1962) is a 4-tuple (C, S, q, R) where

- $C = \{c_1, ..., c_m\}$ is a set of **colleges**
- $S = \{s_1, ..., s_n\}$ is a set of **students**
- $q = (q_{c_1}, \ldots, q_{c_m})$ is a vector of college **capacities**
- $R = (R_{c_1}, \ldots, R_{c_m}, R_{s_1}, \ldots, R_{s_n})$ is a list of **preferences**.

R_s: preference relation over colleges and being unassigned, i.e., $C \cup \{s\}$
R_c: preference relation over sets of students, i.e., 2^S
$P_c(P_s)$: strict preferences derived from $R_c(R_s)$

College Preferences

Suppose colleges have rankings over individual students. How should they compare between sets of students?

If T is a set consisting of c's 2^{nd} & 4^{th} choices and T' consists of its 3^{rd} & 4^{th} choices, then $T \; P_c T'$. If T'' contains c's 1^{st} & 5^{th}, then $T \; ?_c T''$. Multiple P_c's are consistent with the same ranking of singletons, but this is not essential for the definition of stability.

R_c is **responsive** (Roth 1985) if

- whether a student is acceptable for c
- the relative desirability of two students

do not depend on other students in the class.

Responsiveness

Formally, R_c is responsive if

1. for any $T \subset S$ with $|T| < q_c$ and $s \in S \setminus T$,

$$(T \cup \{s\})\, P_c\, T \quad \Longleftrightarrow \quad \{s\}\, P_c\, \emptyset$$

2. for any $T \subset S$ with $|T| < q_c$ and $s, s' \in S \setminus T$,

$$(T \cup \{s\})\, P_c\, (T \cup \{s'\}) \quad \Longleftrightarrow \quad \{s\}\, P_c\, \{s'\}.$$

Matchings

The outcome of a college admissions problem is a **matching**.
Formally, a matching is a *correspondence* $\mu : C \cup S \rightrightarrows C \cup S$ such that

1. $\mu(c) \subseteq S$ with $|\mu(c)| \leq q_c$ for all $c \in C$ (we allow $\mu(c) = \emptyset$),
2. $\mu(s) \subseteq C \cup \{s\}$ with $|\mu(s)| = 1$ for all $s \in S$, and
3. $s \in \mu(c) \iff \mu(s) = \{c\}$ for all $c \in C$ and $s \in S$.

Stability

A matching μ is **blocked by a college** $c \in C$ if there exists $s \in \mu(c)$ such that $\emptyset \, P_c \, \{s\}$.

A matching μ is **blocked by a student** $s \in S$ if $s \, P_s \, \mu(s)$.

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

1. $c \, P_s \, \mu(s)$ and
2. either
    1. there exists $s' \in \mu(c)$ such that $\{s\} \, P_c \, \{s'\}$ or
    2. $|\mu(c)| < q_c$ and $\{s\} \, P_c \, \emptyset$.

A matching is **stable** if it is not blocked by any agent or pair.

Stable Matchings and the Core

(c, s) blocks a matching μ if $c \, P_s \, \mu(s)$ and either

- there exists $s' \in \mu(c)$ such that $\{s\} \, P_c \, \{s'\}$, which means that the coalition $\{c, s\} \cup \mu(c) \setminus \{s'\}$ can *weakly block* μ in the associated cooperative game, or
- $|\mu(c)| < q_c$ and $\{s\} \, P_c \, \emptyset$, hence the coalition $\{c, s\} \cup \mu(c)$ can *weakly block* μ in the cooperative game.

A coalition **weakly blocks** an outcome of a cooperative game if it has a feasible action that makes every member weakly better off, with at least one strict preference → **core defined by weak domination** (contained in the standard core). This is the right concept of stability in many-to-one settings, as colleges may block a matching by admitting new students while holding on to some old ones.

Proposition 2 (Roth 1985)

The weak domination core coincides with the set of stable matchings.

The Deferred Acceptance Algorithm

The following **student applying deferred acceptance algorithm** yields a stable matching.

Step 1. Each student "applies" to her first choice college. Each college *tentatively* accepts the most preferred acceptable applicants up to its quota and rejects all others.

Step $k \geq 2$**.** Any student rejected at step $k - 1$ applies to his next highest choice (if any). Each college considers both the new applicants and the students held at step $k - 1$ and *tentatively* accepts the most preferred acceptable applicants from the combined pool up to its quota; the other students are rejected.

The algorithm terminates when there are no new applications, in *finite* time.

The Correspondence between Many-to-One and One-to-One Matchings

Many (but not all) results for the marriage problem extend to the college admissions problem. The following trick is useful in proofs.

For any college admissions problem (C, S, q, R), construct the **related marriage problem** as follows.

- "Divide" each college c into q_c distinct "seats" $c^1, \ldots, c^{q_c}$. Each seat has unit capacity and ranks students according to c's preferences over singletons. (This is feasible when R_c is responsive, and hence consistent with a unique ranking of students. . . but not for more general preferences.) C^* denotes the resulting set of college seats.
- For any student s, extend her preferences to C^* by replacing each college c in her original preferences R_s with the block $c^1, \ldots, c^{q_c}$, in this order.

Example

The college admissions problem defined by
$C = \{c_1, c_2\}, q_{c_1} = 2, q_{c_2} = 1, S = \{s_1, s_2\}$ and

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

R_{s_1}	R_{s_2}
c_1	c_2
c_2	c_1

$$\mu = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}$$

is transformed into the marriage problem with $M = \{c_1^1, c_1^2, c_2\}, W = S$ and

$R_{c_1^1}$	$R_{c_1^2}$	R_{c_2}
s_2	s_2	s_1
s_1	s_1	s_2

R_{s_1}	R_{s_2}
c_1^1	c_2
c_1^2	c_1^1
c_2	c_1^2

$$\mu^* = \begin{pmatrix} c_1^1 & c_1^2 & c_2 \\ s_2 & c_1^2 & s_1 \end{pmatrix}.$$

Stability Lemma

In the related marriage problem

- each seat at a college c is an individual unit that has preferences consistent with P_c
- students rank seats at different colleges as they rank the respective colleges, whereas seats at the same college are ranked according to their index.

Given a matching for a college admissions problem, it is straightforward to define the **corresponding matching** for its related marriage problem: for any college c, assign the students matched to c in the original problem to seats at c, such that students ranked higher by P_c get lower indexed seats.

Lemma 1 (Roth 1985)

A matching in a college admissions problem is stable if and only if the corresponding matching for the related marriage problem is stable.

Further Results

The Stability Lemma can be used to extend many results from marriage problems to college admissions.

- The student (college) proposing deferred acceptance algorithm produces the student-(college-)optimal stable matching.
- Opposing interests, lattice structure.
- The rural hospital theorem also extends. The following stronger version holds.

Theorem 9 (Roth 1986)

Any college that does not fill all its positions at some stable matching is assigned precisely the same set of students at every stable matching.

Preferences over Stable Matchings

Any two classes to which a college can be stably matched are ranked in the following strong sense.

Theorem 10 (Roth and Sotomayor 1989)

If μ and ν are stable matchings such that $\mu(c)\ P_c\ \nu(c)$, then

$$\{s\}\ P_c\ \{s'\}, \forall s \in \mu(c), s' \in \nu(c) \setminus \mu(c).$$

The set of stable matchings depends only on colleges' ranking of individual students. The same is true about preferences over stable matchings.

Corollary 1

Suppose the preferences P_c and P'_c are consistent with the same ranking of individuals and P_{-c} is a preference profile for $C \cup S \setminus \{c\}$. Let Σ denote the common set of stable matchings for (P_c, P_{-c}) and (P'_c, P_{-c}). Then

$$\mu(c)\ P_c\ \nu(c) \implies \mu(c)\ P'_c\ \nu(c), \forall \mu, \nu \in \Sigma.$$

Some Properties Do Not Extend to Many-to-One Settings

Not all properties carry over to many-to-one matchings, especially those concerning incentives.

- No stable mechanism is strategy-proof for colleges (Roth 1985). In particular, even the college-proposing deferred acceptance rule is not strategy-proof for colleges. Intuition: a college is like a coalition of players in terms of strategies.
- On the contrary, student-proposing deferred acceptance is still strategy-proof for students. Why?
- Colleges may benefit simply by misreporting capacities. Sonmez (1997) shows that no stable mechanism is immune to misreporting capacities.

Incentives for Colleges under Stable Mechanisms

Consider the college admissions problem with
$C = \{c_1, c_2\}, q_{c_1} = 2, q_{c_2} = 1, S = \{s_1, s_2\}$ and the following preferences

R_{c_1}	R'_{c_1}	R_{c_2}
$\{s_1, s_2\}$	$\{s_2\}$	$\{s_1\}$
$\{s_2\}$	$\emptyset$	$\{s_2\}$
$\{s_1\}$		

R_{s_1}	R_{s_2}
c_1	c_2
c_2	c_1

.

Each of the problems (R_{c_1}, R_{-c_1}) and (R'_{c_1}, R_{-c_1}) has a unique stable matching,

$$\begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix} \text{ and, respectively, } \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}.$$

Hence college c_1 benefits from reporting R'_{c_1} instead of R_{c_1} under any stable mechanism (including the college-optimal stable one).

Manipulation of φ^C via Capacities

Consider the college admissions problem with
$C = \{c_1, c_2\}, q_{c_1} = 2, q_{c_2} = 1, S = \{s_1, s_2\}$ and the following preferences

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

Let $q'_{c_1} = 1$ be a potential capacity manipulation by college c_1. We have

$$\varphi^C(R, q) = \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix} \text{ and } \varphi^C(R, q'_{c_1}, q_{c_2}) = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}.$$

Hence c_1 benefits under φ^C by underreporting its number of seats.

14.16 Strategy and Information
Spring 2016