Hannu Vartiainen
Microeconomics I: Mechanism Design
May 2013

1 General

- Time table
  - Week 19: Classical theory (Vartiainen)
  - Weeks 19-20: Economic (quasilinear) environments (Välimäki)
  - Weeks 20-21: Matching markets (Vartiainen)
- Requirement: take home exam
- Problem sets

This part

- General results
- Applications to matching markets
- Main material:
  - Lecture notes
  - MWG chapter 23
  - Moulin (1985), Part IV
  - Roth and Sotomayor (1995)
  - Selected articles
- Other lecture notes, from which the current ones have borrowed and which the participants of the course are urged to consult, include
  - Zvika Neeman (http://www.econ.ceu.hu/download/Syllabi/AdvMicro2.pdf)
  - Fujito Kojima (https://sites.google.com/site/fuhitokojimaeconomics)
  - Alvin Roth (http://kuznets.fas.harvard.edu/~aroth/alroth.html)

2 Introduction - methodology of mechanism design

- The goal of mechanism design theory is to understand the constraints that self interest and rationality of the agents imposes on the performance economic institutions, and to operationalize this understanding to use in applications
- Achievements:
  - Deep results on the boundaries of what can be implemented
  - Market design industry: trading mechanisms, matching markets, network design, voting mechanisms, contract design,...
- Desiderata for useful mechanisms
  - Incentives
  - Robustness
  - Stability
  - Optimal given the above constraints

2.1 Mechanism design

- This section explores general economic design problems
- Economic behavior is shaped by the institution - or game form - within which the behavior takes place
- How is the behavior of the agents has to be evaluated? Game theory!
- Game theory takes the game form as given, and asks how do the agents behave in the given game
- In economic design the question is quite the opposite: how should the one design the institution - or game form - so that the goals (whatever they are) are achieved?
- Thus the game form is not given but subject to formulation by the planner (this approach was initiated by Hurwicz 1959)
- The agents (players) take the game form as given and play it rationally (i.e., according to the chosen equilibrium)

- Agents' strategic behavior constraints what can be achieved within the institution
- When information is complete - and the planner all-mighty - the constraint imposed by players' strategic behavior becomes vacuous: the planner can always force the ideal outcome (or the best that the agents accept)
- Situation becomes more problematic when *information is incomplete*
- It may no longer be known to the planner what the optimal collective decision is $\Rightarrow$ has to design a game form that *elicits* the information from the agents
- Objective functions depend on the information of the agents
- *Mechanism design studies the means of implementing objective functions under the constraint that only the agents know the relevant information*
- Two general questions arise:
  - Which objective functions can be implemented?
  - What is the optimal objective function in the class of implementable functions?.

2.2 Implementing social choice functions

- The set of social alternatives X
- The set L of *linear* orders on X, i.e. $R \subset X \times X$ such that
  - (Complete) xRy or yRx, for all $x, y \in X$
  - (Transitive) xRy and yRz imply xRz, for all $x, y, z \in X$
  - (Antisymmetric) xRy and yRx imply $x = y$
- There is a set $\{1, ..., n\}$ of agents
- Agent's $i \in \{1, ..., n\}$ preferences are described by $\succsim_i \in L$
- Notational conventions:

$$\begin{array}{ll} \succsim & = (\succsim_1, ..., \succsim_n) \\ \succsim' & = (\succsim'_1, ..., \succsim'_n) \\ \succsim_{-i} & = (\succsim_1, ..., \succsim_{i-1}, \succsim_{i+1}, ..., \succsim_n) \\ \succsim & = (\succsim_i, \succsim_{-i}) \end{array}$$

- A **social choice function** (SCF) f associates an outcome to each profile of preferences
$$f : L^n \to X$$
- We are interested in SCFs that satisfy "nice" properties: $f(\succsim)$ is an outcome that is deemed "desirable" under the profile $\succsim$ of preferences

Definition 1 *A SCF f is **Pareto efficient** if whenever some alternative a is at the top of every individual i's ranking $\succsim_i$, then $f(\succsim_1, ..., \succsim_n) = x$:*

- Observe that this is a *weak* definition of Pareto efficiency, a stronger one would require that the selected alternative would not ever be Pareto dominated by another alternative
- Agent i knows her own preferences $\succsim_i \in L$
- The problem is that $\succsim$ is not publicly observable when the outcome $x = f(\succsim)$ is to be decided $\Rightarrow$ it needs to be **communicated**
- Under what conditions can we guarantee that the agent report her preferences truthfully?

Definition 2 *A SCF f is **strategy-proof** if, for every individual i and for every $\succsim_i \in L$,*
$$f(\succsim) \succsim_i f(\succsim_i', \succsim_{-i})$$
for all $\succsim_i' \in L$ and for all $\succsim_{-i} \in L^{n-1}$

- If SCF f is strategy-proof, then it is not in agent's interest to manipulate the collective choice

Definition 3 *A SCF f is **monotonic** if whenever $f(\succsim_1, ..., \succsim_n) = x$ and $\{y : x \succsim_i y\} \subseteq \{y : x \succsim_i' y\}$ for every i, then $f(\succsim_1', ..., \succsim_n') = x$*

- That is, if x is chosen under $\succsim$ and no other alternative passes x in any agent's ranking when moved to $\succsim'$, then x is chosen also under $\succsim'$
- Notice that because it allows the relative ranking of the other alternatives to change, monotonicity implies a type of **independence of irrelevant alternatives**

Definition 4 *A SCF f is **dictatorial** if there is an individual i such that $f(\succsim_1, ..., \succsim_n) = x$ whenever x is at the top of $\succsim_i$*

- The following result is due to Muller and Satterthwaite (1977) (for proof, see Reny, 2001)

Lemma 5 *If $\#X \geq 3$, then any Pareto efficient and monotonic SCF f is dictatorial*

- In an unrestricted domain of preferences, Pareto efficiency implies the following deeper property

Definition 6 *A SCF f is **onto** if for any $x \in X$ there is $\succsim \in L^n$ such that $f(\succsim) = x$*

Theorem 7 *(Gibbard-Satterthwaite, 1973, 1975) If $\#X \geq 3$, then a SCF is strategy proof and onto if and only if it is dictatorial*

- In words, **any** rule that is not dictatorial is sometimes subject to manipulation

Proof. It suffices to show that strategy proof + onto $\Rightarrow$ Pareto efficient + monotonic

Monotonicity: Suppose that $f(\succsim) = x$ and that $\{z : x \succsim_i z\} \subseteq \{z : x \succsim_i' z\}$ for some i. We want to show that $f(\succsim_i', \succsim_{-i}) = x$. Suppose to the contrary that $f(\succsim_i', \succsim_{-i}) = y \neq x$. By strategy-proofness, $y \in \{z : x \succsim_i z\}$ (if not, then $\succsim_i$ can manipulate). Similarly, $x \in \{z : y \succsim_i' z\}$ (if not, then $\succsim_i'$ can manipulate) or, equivalently, $y \in \{z : z \succsim_i' x\}$. Thus $y \in \{z : z \succsim_i' x\} \cap \{z : x \succsim_i' z\} = \{x\}$, a contradiction. Suppose now that $\{z : x \succsim_i z\} \subseteq \{z : x \succsim_i' z\}$ for all i. Because we can move from $\succsim = (\succsim_1, ..., \succsim_n)$ to $\succsim' = (\succsim_1', ..., \succsim_n')$ by swifting from $\succsim_i$ to $\succsim_i'$ one i at a time, and because we have shown that the SCF must remain unchanged for every such change, we must have $f(\succsim') = f(\succsim)$.

Pareto efficiency: Let x be on the top of each individual's ranking $\succsim_i$. Because f is onto, $f(\succsim') = x$ for some $\succsim'$. By monotonicity, the SCF remains equal to x when $\succsim''$ is formed from $\succsim'$ by raising x to the top of every individual's ranking and hence $f(\succsim'') = f(\succsim') = x$. Again by monotonicity, forming $\succsim$ from $\succsim''$ by changing the alternatives below x will not affect SCF, and hence $f(\succsim) = f(\succsim'') = x$. ■

- Essentially the same proof can be used for Arrow's (Im)possibility Theorem: in an unrestricted domain, the only feasible way to aggregate agents' preferences into a single collective preference profile is dictatorhip (again, see Reny 2001).

- A **random** dictator rule is also strategy-proof (also anonymous and neutral), but is likely to be inefficient

Example 8 *Let $n = 3$ and $X = \{x, y, z, w\}$. The vNM payoffs are given by*

	1	*2*	*3*
x	50	0	10
y	10	50	0
z	0	10	50
w	40	40	40

Ex ante payoff from the random dictator rule for each agent is $\frac{1}{3}\cdot 50+\frac{1}{3}\cdot 10+\frac{1}{3}\cdot 0 = 20$ *whereas choice w would give all agents payoff* 40

- Note that the conclusion is independent of the utility scales of the agents
- There are two ways to circumvent the Gibbard-Satterthwaite theorem
  - imposing restrictions on the domain of individuals' preferences
  - assuming more demanding solution concepts
- An alternative x is called a **Condorcet winner** if it beats any other alternative in **majority comparison**, i.e.

$$\#\{i : x \succ_i y\} > \#\{i : y \succ_i x\}, \quad \text{for all } y \neq x$$

Proposition 9 *Assume that n is odd and that preferences are restricted to $D \subset L$ such that for all $\succsim \in D^n$ there is a unique Condorcet winner. Then the SCF $f^W : D^n \to X$ that always chooses the Condorcet winner is strategy proof.*

Proof. Let $f^W(\succsim) = x$ and C be the set of agents that prefer x to y under $\succsim$. If $f^W(\succsim_i', \succsim_{-i}) = y \neq x$, and $y \succ_i x$, then the set of agents that prefer x to y under $(\succsim_i', \succsim_{-i})$ is a superset of C. But then y cannot be a Condorcet winner under $(\succsim_i', \succsim_{-i})$. ■

Example 10 ***Single-peaked preferences****: there is a compact set of outcomes $X \subseteq \mathbb{R}$. Agent i has single-peaked preferences over outcomes if there is a unique "ideal point" x_i such that $x < x' < x_i$ implies $x_i \succ x' \succ x$ and $x > x' > x_i$ implies $x_i \succ x' \succ x$, for $x, x' \in X$. When the agent compares between two outcomes that are both to the right or to the left of the ideal point, she strictly prefers whichever option is closest to x_i.*

Remark 11 The Median Voter Theorem: *If preferences are single peaked, then x^m that coincides with the ideal point of the* ***median voter*** *is a unique Condorcet winner.*

Corollary 12 *If preferences are single peaked, then a SCF that always chooses the median voter's ideal point is strategy proof*

Example 13 *If $\#X = 2$, then majority choice between the two candidates (the Condorcet winner) is strategy proof and nondictatorial. In fact, May's Theorem (1951) states that majority winner is the* ***only*** *anonymous (independent of the names of the agents), neutral (independent of the names of the alternatives), and positively responsive (increase in popularity does not affect negatively to the chances of becoming elected) SCF in this case.*

Example 14 ***Economic environments****: $X = A \times \mathbb{R}^n$ and that there is vNM function $v : A \to \mathbb{R}$ such that $(a,t) \gtrsim (a',t')$ iff*

$$v(a) + t \geq v(a') + t', \text{ for all } (a,t),(a',t') \in X$$

Hence, the preferences have a ***quasilinear*** *representation. The linear term, which permits transferable utility, can be interpreted as* ***money.***

As we will see later, Groves mechanisms permit dominant strategy implementation of the socially optimal allocation in A in economic environments.

2.3 General mechanisms

- The strategy-proof allocation rule relies on a seemingly restrictive assumption, that the planner simply asks the agents to reveal their preferences for the decision making purposes
- If this does not work - as suggested by the Gibbard-Satterthwaite theorem - would there be another mechanism, or a game form or an institution, that the planner could use to induce the desired outcome as a response of the agents' choices?
- What a mechanism implements depends on what we assume of the agents' information and the **equilibrium concept** they use
- Since there are infinitely many game forms that the planner could employ, it is important to understand which SCFs can be implemented by some mechanism, the implementable SCFs need to be **characterized**
- A mechanism is a strategic game without specification of preferences or information structure

Definition 15 *An* ***mechanism*** *is a pair (S,g), consisting of a message space $S = S_1 \times ... \times S_n$ and an outcome function*

$$g : S_1 \times ... \times S_n \to X$$

- Each agent i sends a message s_i in some message space S_i
- After the agents have transmitted a profile of messages $(s_1, ..., s_n) = s \in S$, the outcome function g of the mechanism determines a social allocation $x = g(s)$
- Agent transmit the messages independently and simultaneously
- There are no *a priori* restrictions on the message space $S = S_1 \times ... \times S_n$
- A mechanism defines a game form for which must choose the appropriate equilibrium concept
  - Dominant strategy equilibrium (very robust)
  - Nash equilibrium (if the preferences are commonly known)
  - Bayes-Nash equilibrium (if is preferences are only known by i)
- A **strategy** for agent i is a function

$$\sigma_i : L \to S_i$$

- Then a strategy $\sigma = (\sigma_1, ..., \sigma_n)$ together with the outcome function g leads to the mapping $g(\sigma(\cdot)) : L^n \to X$
- Graphically, we have the following commuting diagram:

$$\begin{array}{ccc} L^n & \overset{f(\cdot)}{\longrightarrow} & X \\ & \searrow^{\sigma(\cdot)} \quad \nearrow_{g(\cdot)} & \\ & S_1 \times ... \times S_n & \end{array}$$

  where $\sigma(\succsim)$ is the equilibrium message profile under $\succsim = (\succsim_1, ..., \succsim_n)$, whatever the used equilibrium notion is
- An analogy is the "market" that implements a Pareto efficient allocation through a Walrasian equilibrium (Hurwicz in the 1970s) (however consumers in a "market" are not strategic and so a market is not a game form)

2.4 Mechanism as a game form

- The central question is whether a particular a SCF (or any other objective function) $f(\cdot)$ can be induced by $g(\sigma(\cdot))$, where σ is pinned down by an appropriate equilibrium of the mechanism (S, g)

Definition 16 *A mechanism* $\Gamma = (S, g)$ ***implements*** *the objective function* f *if for all* $\succsim = (\succsim_1, ..., \succsim_n)$ *there is an* ***equilibrium*** *profile*

$$\sigma(\succsim) = (\sigma_1(\succsim_1), ..., \sigma_n(\succsim_n))$$

of the game induced by Γ *such that*

$$g(\sigma(\succsim)) = f(\succsim)$$

- Observe that the concept of implementation given in the definition above is not as strong as it might be: why not require that the condition be satisfied for *all* equilibrium action profiles?
- This stronger sense of implementation is referred to in the literature as *full* (or strong or unique) implementation
- The weaker concept used here is referred as **incentive compatibility**
- A typical research question: which SCFs (or objective functions) satisfy the incentive compatibility constraint
- At the first glance a complex problem because we have to consider all possible mechanism g on all possible domains of strategies S
- The result called revelation principle simplifies the problem remarkably

Definition 17 *A mechanism* (S, g) *is* ***direct*** *if* $S_i = L$ *and* $g = f$ *for all* i

Definition 18 *The objective function* $f(\cdot)$ *is* ***truthfully implementable*** *(or* ***incentive compatible****) if the direct mechanism*

$$\Gamma = (L^n, f)$$

has an equilibrium σ *such that* $\sigma_i(\succsim_i) = \succsim_i$ *for all* $\succsim_i \in L$, *for all* i

2.4.1 Dominant strategy implementation

- Dominant strategy implementability seemingly a generalization of strategy proofness:

Definition 19 *The strategy profile* $\sigma\left(\succsim\right)=\left(\sigma_1\left(\succsim_1\right),...,\sigma_n\left(\succsim_n\right)\right)$ *is a* ***dominant strategy equilibrium*** *of mechanism* $\Gamma=(S,g)$ *if for all* i *and all* $\succsim_i\in L$,

$$g\left(\sigma_i\left(\succsim_i\right),s_{-i}\right)\succsim_i g\left(s_i,s_{-i}\right)$$

for all $s_i\in S_i$, *for all* $s_{-i}\in S_{-i}$

- For a direct mechanism, dominant strategy implementability coincides with the notion of strategy proofness
- Will allowing arbitrary mechanisms one to implement SCFs that are not strategy proof
- First, we prove the **revelation principle** for the dominant equilibrium concept

Lemma 20 *(Revelation principle, e.g. Gibbard) Let* $\Gamma=(S,g)$ *be a mechanism that implements the SCF* $f(\cdot)$ *in dominant strategy equilibrium. Then the direct mechanism* (L^n,f) *implements* f

Proof. Let $\sigma\left(\succsim\right)$ be the profile of dominant messages under $\succsim$. By the definition of implementation,

$$g\left(\sigma\left(\succsim\right)\right)=f\left(\succsim\right),\ \text{for all}\ \succsim.$$

Since σ constitutes a dominant strategy in Γ, we have, for all $\succsim_i$,

$$g\left(\sigma_i\left(\succsim_i\right),s_{-i}\right)\succsim_i g\left(s_i,s_{-i}\right),\ \text{for all}\ s_{-i},\ \text{for all}\ s_i.$$

In particular,

$$g\left(\sigma\left(\succsim\right)\right)\succsim_i g(\sigma\left(\succsim_i',\succsim_{-i}\right)),\ \text{for all}\ \succsim_{-i},\ \text{for all}\ \succsim_i'.$$

By the definition of implementation,

$$f\left(\succsim\right)\succsim_i f\left(\succsim_i',\succsim_{-i}\right),\ \text{for all}\ \succsim_{-i},\ \text{for all}\ \succsim_i,\succsim_i'$$

Thus truthful announcement is a dominant strategy with (L^n,f). ■

- Since any SCF that is implementable in dominant strategies is, by the revelation principle, strategy-proof, Gibbard Satterhwaite implies the following corollary

Corollary 21 *Suppose that* $\#X\geq 3$. *Then the SCF* f *is onto and implementable in dominant strategies if and only if it is dictatorial.*

2.4.2 Nash implementation

- Strategy proofness is a demanding solution concept: the desired choice must be dominant strategy for the agents, irrespective of the choices of the other agents
- A less demanding - and standard - solution concept is Nash equilibrium
- However, with Nash implementation the previous notion of implementation becomes vacuous: one can always induce truthtelling in *a* Nash equilibrium!

Example 22 *Let $n \geq 3$ and let f be SCF. Choose a mechanism $\Gamma^f = ((L^n)^n, g)$ such that $g(\succsim) = f(\succsim)$ if at least $n-1$ of the announcements of $s_1, ..., s_n \in L^n$ agree with $\succsim$ at profile $\succsim$. Then $((L^n)^n, g)$ implements f in Nash equilibrium.*

- Thus the canonical mechanism Γ^f implements *any* f in Nash equilibrium
- A problem with Γ^f is that at each $\succsim$ it also entertains many other, untruthful equilibria that are not consistent with $f(\succsim)$
- A more appropriate implementation concept would require that *all* the equilibria have the desired property

Definition 23 *A mechanism $\Gamma = (S, g)$ **fully** implements the SCF f if for all $\succsim = (\succsim_1, ..., \succsim_n)$ if*

$$g(\sigma(\succsim)) = f(\succsim)$$

*for **all** the equilibrium profiles $\sigma(\succsim) = (\sigma_1(\succsim_1), ..., \sigma_n(\succsim_n))$ of the game induced by Γ, and there is at least one such equilibrium.*

- The following result is due to Maskin (1977) connects full Nash implementation to the concept of strategy-proofness

Lemma 24 *If a SCF f is fully implementable in Nash equilibrium, then it is monotonic*

Proof. Let (S, g) fully Nash implement f. Then there is an equilibrium strategy $\sigma : L \to S$ such that $g(\sigma(\succsim)) = f(\succsim)$. Suppose that $x = f(\succsim) \neq f(\succsim')$ Then there is an action profile s such that $g(s) = x$ that constitutes a Nash equilibrium under $\succsim$ but not under $\succsim'$. Thus there is an agent i and an action s_i' such that $g(s_{-i}, s_i') \succ_i' g(s)$ but such that $g(s) \succsim_i g(s_{-i}, s_i')$. Letting $g(s_{-i}, s_i') = y$ it follows that $y \succ_i' x$ but such that $x \succsim_i y$ and f is monotonic. ■

- Careful analysis of game forms that intuitively work well may lead to surprising findings

Example 25 *(Solomon's predicament) Two women came to King Solomon, each arguing that a certain baby is hers. Solomon ordered the baby to be cut in half, and each half be given to one woman. One of the women said "Yes, neither I nor the other woman will have the baby". The other one "Oh Lord, give the baby to her, just don't kill him!", upon which Solomon declared her the true mother and gave the child to her. Solomon's judgment became known throughout all of Israel and was considered an example of profound wisdom.*

Formally, the set of feasible outcomes is

$$X = \begin{cases} x_1 \text{ (give the baby to 1)} \\ x_2 \text{ (give the baby to 2)} \\ z \text{ (cut the baby in half)} \end{cases}$$

and the preferences are given by

$$\begin{array}{ll} \succsim & \text{(1 is the real mother):} \quad x_1 \succ_1 x_2 \succ_1 z, \quad x_2 \succ_2 z \succ_2 x_1 \\ \succsim' & \text{(2 is the real mother):} \quad x_2 \succ'_2 x_1 \succ'_2 z, \quad x_1 \succ'_1 z \succ'_1 x_2 \end{array}$$

Solomon's game form

		2	
		Mine	*Not mine*
1	*Mine*	z	x_2
	Not mine	x_1	z

But the unique Nash equilibrium of Solomon's game always allocates the child to the ***wrong*** *mother!*

- Would there be *any* game form that would allow Solomon always allocate the child to the right mother?
- Solomon's SCF is defined by $f(\succsim) = x_1$, $f(\succsim') = x_2$ which is *not* monotonic, i.e. not Nash implementable
- To see this, note that $x = f(\succsim) \neq f(\succsim')$ but there does not exist an alternative w and a player i such that $x \succsim_i w$ and $w \succ'_i x$
- Since monotonicity is a necessary condition also for styrategy-proofness, it follows from Gibbard-Satterthwaite that:

Corollary 26 *(Muller and Satterthwaite 1977) If $\#X \geq 3$, then a SCF $f : L^n \to X$ is fully Nash implementable and onto if and only if it is dictatorial*

- However, in restricted domain, i.e. in a subset of preferences of L^n, further SCFs become fully Nash implementable

- To characterize them, we define the notion of *no-veto power*, which alludes to an assumption that there is no agent who can prevent an alternative becoming implemented when all other agents want to implement it

Definition 27 *A SCF f has* ***no-veto power*** *if $x \in f(\succsim)$ whenever $x \succsim_i y$ for all y for at least $n-1$ agents i*

- Maskin's (1977) theorem:

Theorem 28 *(Maskin 1977) Let $n \geq 3$. Let $D \subseteq L^n$ be a domain of preferences. A SCF $f : D \to X$ is fully Nash implementable if it is monotonic and has no-veto power*

Proof. Construct a mechanism (S, g) such that $S_i = D \times X \times \mathbb{N}$ for all i with a typical element (p_i, x_i, k_i), and

$$g(s) = \begin{cases} f(\succsim), & \text{if } (p_i, k_i) = (\succsim, 0), \text{ for all } i \\ x_i, & \begin{cases} \text{if } (p_j, k_j) = (\succsim, 0) \neq (p_j, k_j), \text{ for all } j \neq i, \\ \text{and } f(\succsim) \succsim_i x_i \end{cases} \\ x_i, & \begin{cases} \text{if neither of the above cases apply,} \\ \text{and } k_i > k_j \text{ for all } j \neq i \end{cases} \end{cases}$$

If $(p_i, k_i) = (\succsim, 0)$, for all i, and the true preference profile is $\succsim'$, then necessarily $\succsim = \succsim'$ since otherwise (by monotonicity) there is an agent i and x_i such that $f(\succsim) \succsim_i x_i$ and $x \succ'_i f(\succsim)$. Thus $f(\succsim) = f(\succsim')$ becomes implemented.

If $(p_j, k_j) = (\succsim, 0) \neq (p_j, k_j)$, for all $j \neq i$, and $f(\succsim) \succsim_i x_i$, and the true preference profile is $\succsim'$, then x_i must be top ranked in $\succsim'_j$ for all $j \neq i$. By no-veto power $x_i = f(\succsim')$ becomes implemented.

In all other cases the integer game is triggered, and no Nash equilibrium exists.

■

- The integer construction of the mechanism in Maskin's theorem is somewhat artificial and often criticized
- The mechanism relies on there not being highest integer, i.e. that when the game is triggered, the agents are maximizing in an open set, which does not have a solution
- However, an integer game (or an equivalent construction) is needed for the mechanism to block undesired equilibria
- Without infinite and open ended message space, undesirable mixed strategies cannot be avoided
- In particular, with full implementation it no longer suffices to focus on direct mechanisms $\Rightarrow$ existence not enough, the main problem is to get rid of the undesirable equilibria

2.4.3 Virtual implementation

- Abreu and Sen (1991) demonstrated that the restrictive conclusions of the impossibility results á la Muller and Satterthwaite/Maskin are, in a sense, a knife edge result: one can approach practically *any* SCF arbitrarily close in a probabilistic sense with a SCF that *is* fully Nash implementable
- Let us restrict focus on SCFs that are **ordinal**, depend only on the agents preferences over the set of alternatives
- Assume that the agents' preferences have a vNM extension over lotteries ΔX (= set of lotteries on X), i.e. there is a collection U_i of utility functions $u_i : X \to \mathbb{R}$ such that for any $\succsim_i \in L$ there is a unique $u_i \in U$ such that

$$\begin{array}{rcl} \sum_x \ell(x)u_i(x) & \geq & \sum_x \ell'(x)u_i(x) \quad \text{iff} \quad \ell \succsim_i \ell', \\ \text{for all } \ell, \ell' & \in & \Delta X \end{array}$$

- Denote the degenerate lottery that implements $x \in X$ with probability 1 by 1_x

Definition 29 *A SCF $f : U^n \to \Delta X$ is **virtually Nash implementable** if there is another function $f^\varepsilon : U^n \to \Delta X$, ε-close to f under each $u \in U^n$, that is Nash implementable*

- If f is virtually implementable, then one can approach it arbitrarily close with a Nash implementable choice function

Lemma 30 *(Abreu and Sen 1991) Probabilistic SCF*

$$f^\lambda(u) = \lambda \cdot f(u) + (1-\lambda) \cdot \frac{1}{\#X} \sum_x 1_x$$

is monotonic, for any $f : U^n \to \Delta X$ and for any $\lambda \in (0, 1]$

Proof. Let $\ell = f^\lambda(u)$. Then

$$u_i(\ell) = \lambda \cdot u_i(f(u)) + (1-\lambda) \cdot \frac{1}{\#X} \sum_x u_i(x)$$

Take $u' \neq u$, i.e. there are $y, z \in X$ such that $u_i(y) \geq u_i(z)$ and $u'_i(z) > u'_i(y)$. Construct a lottery $\ell^{y,z}$ such that

$$\ell^{y,z} = \lambda \cdot f(u) + (1-\lambda) \cdot \frac{1}{\#X} \sum_{x \neq z} 1_x + \frac{1}{\#X} \cdot 1_y$$

Then

$$\begin{aligned} u_i(\ell) - u_i(\ell^{y,z}) &= (1-\lambda)\frac{1}{\#X} \cdot (u_i(y) - u_i(z)) \\ u_i'(\ell) - u_i'(\ell^{y,z}) &= (1-\lambda)\frac{1}{\#X} \cdot (u_i'(y) - u_i'(z)) \end{aligned}$$

Since $u_i(\ell) \geq u_i(\ell^{y,z})$ and $u_i'(\ell) < u_i'(\ell^{y,z})$, f is monotonic. ■

- Note that the lemma holds for arbitrarily low positive λ, and for any basic choice function f
- By the characterization result concerning Nash implementation, without monotonicity the only restriction on implementability is the no-veto power

Corollary 31 *(Abreu and Sen 1991) Let $n \geq 3$.* ***Any*** *SCF f is virtually Nash implementable if it satisfies no-veto power*

- No-veto power is a weak condition, met in almost any reasonable environment

3 Market design

- By the Gibbard-Satterthwaite and other impossibility results we know that no mechanism works well in all domains
- Luckily, many relevant economic domains are characterized by restrictions that permit useful tailoring of mechanisms
- In particular, the existence of a price mechanism, or transferability of utility, is very useful (Juuso's lectures)
- But many natural allocation problems cannot be solved by using price mechanisms
  - School choice
  - Health care
  - Labor markets
  - Military drafts
- In such circumstances, the aggregation questions do not have a natural solution, and the choice of the allocation mechanism will crucially affect the outcome

- **Market design**: how to design mechanisms that allocate resources in a desirable way (even in the absence of price mechanism)
- Aim is to provide applicable, well functioning, and robust methods to allocate resources
- "Optimal" market design can be quite context sensitive, the details of the markets matter for the design
- Useful approach: start with a generic, well functioning mechanism and tailor it to the context
- Recently market design techniques have been applied successfully to problems such as student placement in schools, labor markets where workers and firms are matched, and organizing organ donation network

3.1 Two-sided matching theory

- Authoritative reference: Roth and Sotomayor (1990)
- Two sided market: The agents belong to one of two disjoint sets
- Matching: exchange or association bilateral
  - An agent is only associated to agents in the set he/she does not belong to
  - The only aspect that affects the payoff of an agent is identity of the agents he/she is associated to
- Gale and Shapley (1962) proposed a theory of stable matching

Example 1: Medical intern placement

- Medical students in many countries work as residents (interns) at hospitals .
- In the U.S. around 25 000 medical students and 1 000 hospitals are matched
- Beginning around 1900, market was decentralized, and suffered from **unraveling** of appointment dates (Roth and Peranson 1999)
  - 1900-1945 contracts up to 2 years in advance of graduation
  - 1945-1952 chaotic recontracting, congestion, and mismatch because students' quality and interests were unknown at the matching stage

  - 1952- 1972 National Intern Matching Program initiated with high rates of orderly partipation
  - 1972- 1995 declining rate of participation (particularly among married couples)
  - 1995- 1998 Market experienced a crisis of confidence with fears of substantial decline in orderly participation

- What makes a clearinghouse, in particular NIMP in the 50s and 60s, successful?
- A matching is "stable" if there aren't a doctor and residency program, not matched to each other, who would both prefer to be
- Hypothesis: successful clearinghouses produce stable matching
- Married couples make the situation problematic

Example 2: School choice

- In many countries (including Finland) there is freedom in choosing the school to your children
- School authorities take into account preferences of children (and their parents)
- Because school seats are limited (for popular schools), school authorities should decide who is admitted
- Typical goals of school authorities are: (1) efficient placement, (2) fairness of outcomes, (3) easy for participants to understand and use, etc.
- For example, in Finland 100.000 students and 1 000 educational institutions participate the secondary school allocation mechanism every year
- Nontrivial design problem, how should the match be design (e.g. Abdulkadiroglu and Sönmez, 2003)

Example 3: Organ donation

- Kidney exchange is a preferred method to save kidney-disease patients.
- There are kidney shortages, and willing donor may be incompatible with the patient

- S/he may not be willing to donor to another patient that is compatible
- Kidney exchange tries to solve this by matching donor-patient pairs
- What is a "good way" to match donor-patient pairs? (Roth, Sönmez and Unver, 2007)

3.1.1 One-to-one matching model

- **The marriage market** by Gale and Shapley (1962) (suggested reading Roth and Sotomayor 1990, Ch. 2) is defined by a triple $(M, W, \succsim)$
- M is a finite set of "men" and W is a finite set of "women"
- Each man can be matched to at most one woman, and vice versa (so the model is called "one-to-one matching")
- Each man m has preferences $\succsim_m$ over women and being matched to himself (denoted by m) and each women w has preferences $\succsim_m$ over men and being to herself (w).
  - $w \succ_m w'$ means man m strictly prefers woman w to woman w'
  - $m \succ_w m'$ means woman w strictly prefers man m to woman m'
  - $w \succsim_m m$ means w is **acceptable** to m and $m \succsim_w w$ means m is **acceptable** to w

Definition 32 *A **matching** μ is a one-to-one mapping from $M \cup W$ to itself such that, for all $m \in M$ and for all $w \in W$,*

- $\mu(m) \in W \cup \{m\}$
- $\mu(w) \in M \cup \{w\}$
- $\mu(m) = w$ *iff* $\mu(w) = m$
- A matching μ is **individually rational** if $\mu(m)$ is acceptable for all men m and $\mu(w)$ is acceptable for all women w
- A matching μ is **blocked**
  - by man m if $m \succ_m \mu(m)$ and by woman w if $w \succ_w \mu(w)$
  - by pair (m, w) if $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$

- That is, μ is blocked if either m, w are not matched with one another at μ but prefer each other to their assignment at μ, or if there is m or w whose assignment at μ is not acceptable for him or her
- A matching μ that is blocked is unstable in a sense that m and w have a mutual motivation to disrupt the functioning of μ, to become matched with one another

Definition 33 *A matching μ is **stable** if it is not blocked*

- Unstable matchings are dominated by coalitions consisting of individuals or pairs, and so unstable matchings are not in the **core** of the market
- In this model, also the converse is true:

Remark 34 *The core of the marriage market equals stable matchings*

- Existence?

3.2 Deferred Acceptance algorithm (Gale and Shapley 1962)

- **Men proposing** DA, women proposing version by switching the gender roles

Step 0: Each man m proposes to his most preferred acceptable woman. If there is none, he does not make a proposal.

Each woman w holds tentatively the most preferred acceptable offer and rejects the rest.

Step k: Each man m rejected at stage $k-1$ proposes his most preferred acceptable woman who hasn't yet rejected him. If there is none, he does not make a proposal.

Each woman w holds tentatively her most preferred acceptable offer to date and rejects the rest.

Stop: When no further proposals are made, match each woman to the man (if any) whose proposal she is holding.

- Assume true preferences (incentives are discussed later)
- Arbitrarily break ties if some preferences are not strict

Theorem 35 *(Gale and Shapley 1962) A stable matching exists for every marriage market*

Proof. DA stops in finite time. The resulting matching μ is

(i) not blocked by an individual because at each step of the algorithm, no man proposes to an unacceptable woman and no woman holds an offer of an unacceptable man.

(ii) not blocked by any pair (m, w): if $w \succ_m \mu(m)$, then m proposed to w and was rejected at some step of DA. Since w's tentative match only improves as the algorithm proceeds, the match $\mu(w)$ at the end of DA is still at least as good for w as m. So w does not strictly benefit from blocking μ with m. ■

Example 36 *Let $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2\}$, and their preferences given by (documenting only acceptable alternatives)*

$$\begin{aligned}
&\succ_{m_1}: w_1, w_2\\
&\succ_{m_2}: w_1, m_2\\
&\succ_{m_3}: w_2, w_1\\
&\succ_{w_1}: m_3, m_2, m_1\\
&\succ_{w_2}: m_1, m_3
\end{aligned}$$

Following the steps of the DA algorithm, the resulting matching

$$\mu = \{(m_1, w_2), (m_2, m_2), (m_3, w_1)\}$$

is stable

- Stability is theoretically appealing, but does it matter in real life?
- Roth (1984) showed that the NIMP algorithm is equivalent to a (hospital-proposing) DA, so NIMP produces a stable matching
- Roth (1991) studied British medical match, where different regions use different matching mechanisms. He found that stable mechanisms are successfully used (and is still in use) but most unstable mechanisms were abandoned after a short period of time.
- Over time, more and more markets using matching mechanisms are discovered and documented, and more and more markets are adopting DA and other matching mechanisms, providing even more data points

	Stable	Still in use
NRMP	**Yes**	**Yes (new design 98-)**
Edinburgh ('69)	**Yes**	**Yes**
Cardiff	**Yes**	**Yes**
Edinburgh ('67)	**No**	**No**
Newcastle	**No**	**No**
Sheffield	**No**	**No**
Cambridge	No	Yes
London Hospital	No	Yes
Medical Specialities	**Yes**	**Yes**
Canadian Lawyers	**Yes**	**Yes**
Dental Residencies	**Yes**	**Yes**
Osteopaths (-'94)	**No**	**No**
Osteopaths ('94-)	**Yes**	**Yes**
Reform rabbis	**Yes**	**Yes**
NYC highschool	**Yes**	**Yes**

- Two-sidedness is important!

Example 37 *Consider the one-sided "roommate problem" in which 4 potential roommates can be matched with anyone else. Preferences given by*

$$\succ_1: 2, 3, 4$$
$$\succ_2: 3, 1, 4$$
$$\succ_3: 1, 2, 4$$
$$\succ_4: \textit{any preferences}$$

No stable matching exists

- Different stable matchings may benefit different market participants
- In particular, each version of DA favors one side of the market at the expense of the other side

Definition 38 *A stable matching μ is M-**optimal** (W-optimal) if every man (woman) likes μ at least as well as any other stable matching*

Theorem 39 *(Roth and Sotomayor 1990) When all men and women have strict preferences, the matching μ_M produced by the men proposing DA is the M-optimal stable matching*

- The W-optimal stable matching μ_W is the matching μ_W produced by DA when the women propose

- Thus an $M-$optimal and a W-optimal stable matching exists

Proof. Terminology: w is *achievable* for m if there is some stable matching μ such that $\mu(m) = w$. It suffices to show that no man is rejected by an achievable woman in any step of DA. For contradiction, suppose a man is rejected by an achievable woman. Consider the first step in which a man, say m, is rejected by an achievable woman, say w (let μ be a stable matching where $\mu(m) = w$). This means that some other man m' proposed to w in DA and replaced m as the partner of w at this step. Since this is the first step of DA where a man is rejected by an achievable woman, and $\mu(m) = w \neq \mu(m')$, necessarily $w \succ_{m'} \mu(m')$. Also we have $m' \succ_w m = \mu(w)$ since m' displaces m at w in DA. This means that pair (m', w) blocks μ. ■

- Moreover, μ_M is W-**pessimal**, that is, every woman weakly disprefers it to any stable matching, and vice versa
- This point is part of the policy debate related to many matching markets in (e.g. the old NIMP algorithm was hospital-proposing): why should one worry about the preferences of the institutions?
- Medical students argued that the old NIMP favored hospitals at the expense of students and called for reconsideration of the mechanism

Theorem 40 *("Rural Hospital Theorem", Roth and Sotomayor 1990) Let the preferences of the agents be strict. The set of men and women that are unmatched is the same for all stable matchings.*

Proof. Let μ_M be the M-optimal stable matching and μ be an arbitrary stable matching. Since μ_M is M-optimal, all the men that are matched in μ are matched in μ_M. Since μ_M is W-pessimal, all the women that are matched in μ_M are matched in μ. But the number of matched men and women are the same in any matching. This means that the same set of men and women are matched in μ_M and μ. ■

Theorem 41 *(Weak Pareto optimality for the men) There is no individually rational matching μ (stable or not) such that $\mu(m) \succ_m \mu_M(m)$ for all $m \in M$*

Proof. If μ were such a matching, then it matches every man to a woman. Hence $\mu(M) = M$. Moreover, every man m is rejected under DA by $\mu(m)$. Hence all women in $\mu(M)$ are matched under μ_M, i.e., $\mu_M(\mu(M)) = M$. Hence all m would have been matched under μ_M and $\mu_M(M) = \mu(M)$. DA stops as soon as every woman in $\mu_M(M)$ has an acceptable proposal. By assumption, $\mu(w)$ is acceptable for w. Since $\mu(m) \succ_m \mu_M(m)$ for all $\mu \in M$, and the fact that no w has rejected $m = \mu(w)$ contradict the assumption that DA results in $\mu_M(m)$. ■

- Weak Pareto optimality relies on a strong blocking notion
- A weaker one, assuming that no man's payoff becomes worse and some increase, is not met by the DA (exercise)

3.2.1 Incentives

- Let's consider strategic behavior in centralized matching mechanisms, in which participants submit a list of stated preferences
- By the **revelation principle**, some of the results will apply to decentralized markets also, in which agents have different sets of strategies
- Consider a marriage market $(M, W, \succsim)$ whose outcome will be determined by a centralized clearinghouse, based on a list of preferences that players will state ("reveal")
- If the profile of stated preferences is $\succsim'$, the algorithm employed by the clearinghouse or a mechanism f produces a matching $f(\succsim')$
- Given M, W, the mechanism f produces a matching for all $\succsim'$
- If the produced matching is stable with respect to the preference profile $\succsim'$, for any $\succsim'$, we say that f is a stable matching mechanism
- A matching mechanism f can be interpreted as a social choice function, and hence can be examined for strategy-proofness

Theorem 42 *(Impossibility, Roth 1982) No stable and strategy-proof matching mechanism exists*

Proof. One example for which nonstable matching mechanism induces a dominant strategy is sufficient. Consider an example with 2 agents on each side with true preferences

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

There are two stable matchings, $\mu = \{(m_1, w_1), (m_2, w_2)\}$ and $\mu' = \{(m_1, w_2), (m_2, w_1)\}$. Players m_1, m_2 prefer the former whereas w_1, w_2 the latter.

To see that μ is **not** strategy-proof, if the preferences of w_2 are replaced with $\succsim'_{w_2}$ such that

$$\succ'_{w_2}: m_1,$$

then μ' is the only viable stable matching. But then it is not in the best interest of w_2 to report her preferences truthfully.

To see that μ' is not strategy-proof, replicate the above with the roles of m_1 and w_2 switched. ■

- In particular, DA is **not** strategy-proof
- The proof of the impossibility theorem leaves open the possibility that situations in which some participant can profitably manipulate his preferences are rare

Theorem 43 *(Roth and Sotomayor 1990) Let preferences of the agents be strict. When there is more than one stable matching, then at least one agent can profitably misrepresent his or her preferences, assuming the others tell the truth.*

- The misrepresenting agent can manipulate the mechanism in such a way that s/he becomes matched to his/her most preferred achievable mate under the true preferences at every stable matching under the misrepresented preferences
- The proof amounts to noting that in a $W-$optimal stable matching μ_w any woman w would benefit from misrepresenting her preferences by removing the men below $\mu_W(w)$ from her acceptable preferences
- Thus there is no way to organize the market so as to achieve a stable matching without occasionally exposing some of the agents to the question of gaming the system
- But not all agents are in a similar position to misrepresent their preferences, for example public organizations
- Can we say something about who has the incentives to manipulate the mechanism?

3.2.2 Men's incentives in the M-optimal stable mechanism

Theorem 44 *(Dubins and Freedman 1981, Roth 1982) The mechanism that yields the $M-$optimal stable matching in the marriage market is strategy-proof for the men*

- That is, it is dominant strategy for every man to reveal his preferences truthfully
- Can be extended to (weak) group strategy-proofness: a group of men cannot *strictly* profit from jointly manipulating the mechanism (Hatfield and Kojima)
- The next lemma elaborates the property of $M-$optimal matchings and pairwise stability: any other individually rational (not necessarily stable) matching is either disliked by all men or there is there is a man that dislikes the new matching that blocks the new matching with a woman who is associated to a man who likes the new matching

Lemma 45 (Blocking Lemma) *(Gale and Sotomayor 1985) Let μ be any individually rational matching with respect to strict preferences $\succ$ and let M' be the set of men who prefer μ to μ_M. If M' is nonempty there is a pair (m, w) which blocks μ such that m is in $M \backslash M'$ and w is in $\mu(M')$*

Proof.

Case 1 $\mu(M') \neq \mu_M(M')$. Choose $w \in \mu(M') \backslash \mu_M(M')$,. Then m' such that $w = \mu(m')$ prefers w to $\mu_M(m')$ so w prefers $\mu_M(w) = m$ to m'. But m is not in M' since w is not in $\mu_M(M')$, hence m prefers w to $\mu(m)$ (since preferences are strict), so (m, w) blocks μ

Case 2 $\mu(M') = \mu_M(M') = W'$. Let w be the woman in W' who receives the last proposal from an acceptable member of M' in DA. Since all w in W' have rejected acceptable men from M', w had some man m engaged when she received this last proposal. We show that (m, w) is the desired blocking pair. First, m is not in M' for if so, after having been rejected by w, he would have proposed again to a member of W' contradicting the fact that w received the last such proposal. But m prefers w to his mate under μ_M and since he is no better off under μ, he prefers w to $\mu(m)$. On the other hand, m was the last man to be rejected by w so she must have rejected her mate under μ before she rejected m and hence she prefers m to $\mu(w)$, so (m, w) blocks μ.

■

Theorem 46 *(Demange, Gale, and Sotomayor 1987) Let preferences be strict. Let $\succsim$ be the true preferences of the agents, and let $\succsim'$ differ from $\succsim$ in that some coalition C of men and women misrepresent their preferences. Then there is no matching μ', stable for $\succsim'$, which is preferred to every stable matching μ under the true preferences $\succsim$ by all members of C.*

Proof. Let the subset $\bar{M} \cup \bar{W}$ of men and women misrepresent their preferences and are strictly better off under μ, stable w.r.t. $\succ'$,than under any stable matching w.r.t. $\succ$. μ must be individually rational with respect to $\succ$. Then

$$\begin{array}{c} \mu(m) \succ_m \mu_M(m) \text{ for every } m \in \bar{M} \\ \mu(w) \succ_w \mu_W(w) \text{ for every } w \in \bar{W} \end{array} \tag{1}$$

where μ_M and μ_W are the M and W-optimal stable matchings.

It suffices to show that $\bar{M} \cup \bar{W}$ is empty. If $\bar{M}$ is not empty we can apply the Blocking Lemma to the market $(M, W, \succ)$, since by (1) $\bar{M}$ is a subset of M', thus there is a pair (m', w') which blocks μ under $\succ$ such that $\mu_M(m') \succ_{m'} \mu(m')$ and w' is in $\mu(M')$. Then also $\mu_M(w') \succ_{w'} \mu(w')$, since otherwise w' and $\mu(w')$ would block μ_M. Clearly m' and w' are not in $\bar{M} \cup \bar{W}$ and so are not misrepresenting their preferences, so they will also block μ under $\succ$, contradicting that μ is stable under $\succ$. Hence $\bar{M}$ must be empty.

A symmetric argument applies to the emptiness of $\bar{W}$. ■

- To see that the above theorem implies strategy-proofness of the $M-$optimal ($W-$optimal) DA for men (women), let C consist of a single man m
- The result says that no matter which stable matching will result from a the misrepresentation of preferences of m, the outcome is not profitable for at least some member of of C, i.e. m

3.2.3 Nash equilibria

- If strategy-proofness cannot be combined with stability, what about Nash equilibria?
- Let μ be any stable mechanism
- Construct a strategy such that each woman w in $\mu(M)$ lists only $\mu(w)$ as her acceptable man and each man states his true preferences

Theorem 47 *(Gale and Sotomayor 1985) Let all preferences are strict. Construct μ, a stable matching for $(M, W, \succsim)$.Then the above strategy constitutes a Nash equilibrium in the game induced by the M-optimal stable matching mechanism (and μ_M is the matching that results).*

- For a proof, note that any man m can only possibly be matched with $\mu(m)$, hence any deviation from the truthful strategy would lead to no-change or a worse outcome

- Furthermore, every equilibrium misrepresentation by the women nevertheless yields a matching that is stable with respect to the true preferences
- However, Nash equilibrium is a demanding solution concept, especially in large markets where the identities or the preferences of the other participants are likely to be unknown

3.3 Many-to-one matching - the college admission model

- In many distributional problems, the aim is to allocate agents to different locations that an accommodate many agents
  - Workers to firms
  - School choice
  - Military drafting
  - Health care
- In such a situation, one party of the marriage model associates to many members of the other party, who each is associated to a single member of the first party, "many-to-one matching problem"
- How to extend the marriage markets to cover also this case?
- **College admission** model of Gale and Shapley 1962 is defined by the triple $(C, S, \succsim)$, where
  - C is a finite set of colleges and S is finite set of students
  - each student s has preferences $\succsim_s$ over colleges C and no placement, $\emptyset$, and each college c has has preferences, or priorities, $\succsim_c$ over students S and a minimal acceptance criterion $\emptyset$
- Each college $c \in C$ can be matched to at most $q_c \in \mathbb{N}$ students, and and each student s can be matched to one college (hence "many-to-one matching")

Definition 48 *A **matching** μ is a one-to-many mapping from $C \cup S$ to $S \cup C \cup \{\emptyset\}$ such that*

- $\mu(s) \in C \cup \{\emptyset\}$, *for all* $s \in S$
- $\mu(c) \subseteq S \cup \{\emptyset\}$, *for all* $c \in C$
- $\mu(s) = c$ *iff* $\mu(s) = c$, *for all* $s \in S$ *and for all* $c \in C$

- A matching μ is **individually rational** if $\mu(c) \subseteq \{s : s \succsim_c \emptyset\}$ for all c and $\mu(s) \in \{c : c \succsim_s \emptyset\}$
- A matching μ is **blocked** if it is not individually rational, or there is a pair (c, s) such that
  - $s \succ_c s'$ for some $s' \in \mu(c)$ or $s \succ_c \emptyset$ and $\#\mu(c) < q_c$
  - $c \succ_s \mu(s)$
- That is, μ is blocked if either s is not matched to c at μ but prefer each other to their assignment at μ, or if there is s or c whose assignment at μ is not acceptable for him or her
- Note that this model assumes implicitly that there are no externalities between students from the view of colleges
- This could be an issue especially in the context of firm/worker allocation problems
- Formally, preferences of the colleges (defined over 2^S) are **responsive** if, for any set of students $T \subseteq S$ and any students s and s' in $S \backslash T$,
  - $T \cup \{s\} \succ_c T \cup \{s'\}$ if and only if $\{s\} \succ_c \{s'\}$, and
  - $T \cup \{s\} \succ_c T$ if and only if s is acceptable to c

Proposition 49 *When the college preferences are responsive, a matching is in the core if and only if it is (pairwise) stable*

- **Student proposing** DA

Step 0: Each student s proposes to his most preferred acceptable college. If there is none, he does not make a proposal.

Each college c holds tentatively the most preferred acceptable offers up to its quota, and rejects the rest.

Step k: Each student s rejected at stage $k-1$ proposes his most preferred acceptable college who hasn't yet rejected him. If there is none, he does not make a proposal.

Each college c holds tentatively the most preferred acceptable offers it has received to date up to its quota, and rejects the rest.

Stop: When no further proposals are made, match each college to the students (if any) whose proposal it is holding.

- To interpret the college admission problem using the marriage market model, replace college c by q_c distinct colleges denoted by c_1,c_2, ..., $c_{\#q_c}$
  - Each of the subcolleges c_i has c's preferences over students, and quota $q_{c_i} = 1$
  - Each student's s preference list is modified by replacing c, wherever it appears on his list, by the string c_1,c_2, ..., $c_{\#q_c}$, over which the student is indifferent
- The reinterpreted model is a one-to-one matching market
- CAVEAT: assuming away externalities may hinder seriously the functioning of the mechanism, e.g. married couples/NIMP

Lemma 50 *A matching of the college admissions problem is stable if and only if the corresponding matching of the related marriage market are stable*

Theorem 51 *(Gale and Shapley 1962) A stable matching exists for every college admission problem*

- The many-to-one counterpart of the Rural Hospital Theorem:

Theorem 52 *Let colleges' preferences over students be strict. Then all colleges fill the same number of positions across stable matchings. Any student unmatched in any one stable matching is unmatched in all stable matching.*

- So any college that fails to fill all of its positions at some stable matching will not be able to fill any more positions at any other stable matching
- Thus not only will the colleges fill the same number of positions in all stable matchings but they will fill them with exactly the same students
- Immediately from the related marriage model:

Theorem 53 *(Roth) The student-optimal stable matching procedure is strategy-proof for the students*

- ...but not for the colleges
- However, college preferences are often more easily detectable and their behavior enforceable
- Moreover, the student optimal matching is also Pareto optimal for the students

3.3.1 Couples

Example 54 *Matching with couples: There are two colleges c_1, c_2, both with quota 1, and one single student s and one student couple (m, w). Preferences are*

$$\succ_s: c_1, c_2$$
$$\succ_{(m,w)}: (c_1, c_2)$$
$$\succ_{c_1}: m, s$$
$$\succ_{c_2}: s, w$$

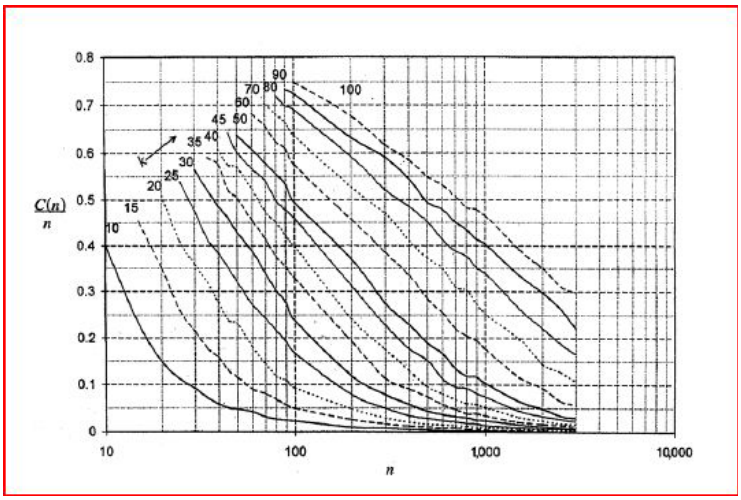
There is no stable matching. So, what should we do?

- The new (and current) NRMP algorithm, called the Roth-Peranson algorithm (1998), is based on student-proposing DA, but try to accommodate couples
- The algorithm (basic idea Roth and Vande Vate, 1989) allows couples to express preferences on *pairs* of hospital programs
- Sketch of the Roth-Peranson algorithm:
  1. Run DA without couples, and then add couples one at a time
  2. If someone is displaced, then such an agent is allowed to apply later in the algorithm.

3.3.2 Large markets

- Manipulation of the (student proposing) DA might be beneficial for the colleges since there is a feedback loop: by changing their intake at one step a college can affect the composition of the students left in the market which may have an effect on the rejection recisions of the other colleges and, hence, the applications that the deviating college may have
- But the relative magnitude of the such a manipulative act should be sensitive to the size of the market: if there are many students, one college's decision will not affect the average composition of the still available students
- Simulation on randomly generated data (Kojima 2012)
- Simple model: n colleges, n students
- Preferences drawn independently and uniformly
- Each student applies to k hospitals

- $C(n)$ = number of students matched differently at school-proposing and college-proposing DAs.



- Suggests two things:

1. Since the student proposing DA is strategy-proof for the students, and college proposing DA is strategy-proof for the colleges, and their differences vanish as the market becomes larger, the large market should be approximately manipulation free

2. Since the student proposing DA is student-optimal stable mechanism and the college proposing DA is student-pessimal, and their differences vanish as the market becomes larger, in a large market the choice of a stable mechanism should not make a big difference

- Theory model (Kojima and Pathak 2009):

- Finite sets S of students and C of colleges

- Each student can be matched to at most one college, and college c can be matched with at most q_c students (no couples)

- College c's vNM payoff from a match $\mu : C \cup S \to C \cup S \cup \{\emptyset\}$ is additively separable

$$u_c(\mu) = \sum_{s \in \mu(c)} u_c(s)$$

- Timing of the game: Students and colleges submit their preference lists and quotas simultaneously

- DA is applied under the reported preferences

Theorem 55 *(Kojima and Pathak, 2009) The expected proportion of colleges that can manipulate DA when others are truthful goes to zero as the number of colleges goes to infinity*

Theorem 56 *(Kojima and Pathak, 2009) The expected proportion of colleges that are matched to the same set of students in all stable matchings goes to one as the number of colleges goes to infinity*

- DA is strategy-proof for students, so truthtelling is an optimal strategy for students
- Strategic rejection by a college causes a chain of application and rejections
- Some of the rejected students may apply to the manipulating college, and the college may be made better off if these new applicants are desirable
- In a large market, there is a high probability that there will be many colleges with vacant positions, so the students who are strategically rejected (or those who are rejected by them and so on) are likely to apply to those vacant positions and be accepted
- So the manipulating college is unlikely to be made better off in a large market and the DA is "approximately" manipulation free

4 House allocation problem

- So far we talked about two-sided matching:
  - men and women
  - students and colleges
  - doctors and hospitals
- When we discuss allocations students to schools: do schools really have preferences?
- Often the institutions do not, per se, prefer agent over another but their priorities are given by outside
- For example, school seats or intern placements may be viewed as objects to be allocated

- However, now also different agents may have different rights concerning the houses, they e.g. may **own** one

  => additional restrictions on feasible matchings

- House allocation problem (Hylland and Zeckhauser 1979) is defined by a triple $(N, H, \succsim)$ where

  - N is a set of agents
  - H is a set of goods (houses) with $\#H = \#N$
  - Each agent $i \in A$ has strict preferences $\succsim_i$ over houses: $h \succsim_i h'$ and $h' \succsim_i h$ if and only if $h = h'$

- Possible applications:

  - public housing
  - organ allocation
  - office allocation
  - school choice problems

- Matching μ is a function from H to A, specifying who receives what house: $\mu(i)$ is the house agent i receives in matching μ

- A matching μ is Pareto-efficient if there is no other matching $\mu' \neq \mu$ such that $\mu'(i) \succsim_i \mu(i)$ for every agent $i \in A$ (hence there is at least one i such that $\mu'(i) \succ_i \mu(i)$)

- A matching mechanism ϕ specifies, for each preference profile $\succsim = (\succsim_i)_{i \in A}$ a matching μ

- A matching mechanism ϕ is **strategy proof** if revealing preferences truthfully is a dominant strategy

- A mechanism is Pareto-effcient if $\phi(\succsim) = \mu$ is Pareto-efficient for every $\succsim$

4.1 Serial dictator

- A **serial dictatorship** mechanism (or priority mechanism) specifies an order over agents, and then lets the first agent receive her favorite good, the next agent receive her favorite good among remaining objects, etc.

- When members of one party of the two-sided markets do not have preferences - or if they are indifferent over their matches - Gale and Shapley's DA is equivalent to a serial dictator
- Easy to implement: decide the order (randomly, or using some existing priority such as seniority) and let applicants choose according to the order
- No agent cannot profit from choosing anything but the best alternative in the remaining ones
- A mechanism is **group strategy-proof** if no group of agents can jointly misreport preferences in a way to make some member strictly better off, while no one in the group is made worse off
- Group strategy-proofness implies strategy-proofness

Theorem 57 *Serial dictatorship is group strategy-proof*

Proof. In any group C of agents there is an agent, say i, that is first to choose from the remaining goods. For i a deviation cannot be strictly profitable ■

Theorem 58 *Serial dictatorship is Pareto-optimal*

Proof. Suppose there is matching μ' that Pareto-dominates μ from a serial dictatorship. Consider the agent i with the highest priority who receives a strictly better object under μ' than under μ. It has to be that

1. There exists an agent j who receives $\mu(j) = \mu'(i)$ who chooses before i, else i would have picked $\mu'(i)$ under μ.
2. $\mu'(j) = \mu(j)$, because $\mu'(j) \succsim_j \mu(j)$ by assumption that μ' Pareto dominates μ, and it can't be that $\mu'(j) \succsim_j \mu(j)$ by definition of i, a contradiction

■

- Are there other mechanisms that are (group) strategy-proof and Pareto-optimal?
- A mechanism is neutral if the names of the objects (houses) will not affect the choice of the mechanism: let $\rho : H \to H$ be a permutation of houses, and let $\succsim^\rho$ be the preference profile such that $h \succsim_i h'$ iff $\rho(h) \succsim_i^\rho \rho(h')$, then for any neutral mechanism $\phi(\succsim) = \rho(\phi(\succsim^\rho))$

Theorem 59 *(Svensson 1998) A mechanism is group strategy-proof and neutral if and only if is a serial dictatorship*

4.2 Housing market

- In many market design contexts, the aim is to improve the current allocation of objects
- How to achieve good market allocation without a price mechanism?
- A housing market (Shapley and Scarf,1974) is a tuple $(N, H, \succsim, \omega)$ such that
  - N is a set of agents
  - H is a set of goods (houses) with $\#H = \#N$
  - Each agent $i \in N$ has **strict** preferences $\succsim_i$ over houses: $h \succsim_i h'$ and $h' \succsim_i h$ if and only if $h = h'$
  - $\omega : N \to H$ is the initial allocation of the houses, i.e. i's initial house is $\omega(i)$
- A matching $\mu : N \to H$ specifying the final allocation of the houses: $\mu(i)$ is the house that agent i receives in μ

Definition 60 *A matching μ is in the **core** if there is **no** coalition of agents C and a submatching $\mu^C : C \to H$ such that*

1. $\mu^C(C) = \mu(C)$
2. $\mu^C(i) \succ_i \mu(i)$ *for all* $i \in C$

Definition 61 *A matching μ is **individually rational** if $\mu(i) \succsim_i \omega(i)$ for all $i \in N$*

- Any core matching is individually rational and Pareto-optimal
- Existence?
- The proof is by employing Gale's **Top trading cycles algorithm** (attributed to David Gale by Shapley and Scarf)

Step 0: Each agent points to the owner of his favorite house (perhaps to herself)

Since there are finite number of agents, there is at least one cycle

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

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 (perhaps to herself)

Every agent in a cycle is assigned the house of the agent he points to and removed from the market with his assignment

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

Stop: When there is no unassigned agent left

Theorem 62 *(Roth and Postlewaite1977) The outcome of Gale's TTC algorithm is the unique matching in the core of a housing market.*

Proof. Sufficiency: Let μ be the resulting matching of TTC from an initial allocation ω. Suppose there is a coalition of agents B that block μ with a matching μ'. Denote by C_k the agents that are removed in TTC at step k. Since agents in C_0 obtain their most preferred house, they cannot be members of B. Similarly, since agents in C_1 obtain their most preferred house of those agents that are not members of C_0, agents in C_1 cannot be members of B. Continuing this way implies that B cannot contain any agent that is removed by TTC.

Necessity: Agents who leave in Step 0 have to receive their top choices for otherwise they will form a blocking coalition. Subject to that, agents who leave in Step 2 have to receive their top choices among the remaining choices for otherwise they will form a blocking coalition. Proceeding in a similar way, each agent should receive her outcome under Gale's TTC algorithm. ■

- Moreover, matching resulting from the TTC is the unique Walrasian allocation: If the algorithm terminates in T steps, here are the possible equilibrium prices
  - the price of each house that leaves the algorithm in Step 0 is T
  - the price of each house that leaves the algorithm in Step 1 is $T-1$

    ...
  - the price of each house that leaves the algorithm in Step T is 1
- Does the mechanism that always implements TTC have good incentive properties?

Theorem 63 *(Roth 1982) The TTC algorithm is strategy-proof.*

Proof. Suppose an agent leaves TTC with her assignment in Step k. She cannot stop the formation of cycles that form before Step k by misrepresenting her preferences. (these cycles only depend on preferences of agents who are in those cycles). At or after k she receives the best of the remaining houses. So she cannot receive a better assignment through a preference manipulation ■

- So TTC has many desirable properties
- Are there others?

Theorem 64 *(Ma 1994) TTC is the only matching mechanism that is Pareto-optimal, individually rational, and strategy-proof*

Proof. (Sketch) Let μ be the TTC matching, i.e. the core under $\succsim$. Construct the following transformation $\succsim_i'$ of $\succsim_i$ for each agent i

$$\succsim_i\text{: } h, ..., \mu(i), h', ..., \omega(i), h'', ..., h'''$$
$$\succsim_i'\text{: } h, ..., \mu(i), \omega(i), h', ..., h'', ..., h'''$$

where ω is the initial allocation. Since μ results from TTC under $\succsim$, and hence is in the core, it is also Pareto-efficient under $\succsim$. Thus $f(\succsim') = \mu$ if f is Pareto-optimal and individually rational. By strategy-proofness, also $f(\succsim) = \mu$ (recall that strategy-proofness implies monotonicity). ■

4.2.1 House allocation with existing tenants

- Many markets, e.g. campus housing, the problem is a mix between housing markets and housing allocation
  - Some agents are existing tenants, who can stay in their current room but can participate in the matching
  - Others are newcomers, who do not have their room currently
- How should one distribute the new vacant houses, maximizing all agents' preferences at the same time honoring the existing tenants' rights to their houses?
- A housing market with existing tenants is again a tuple $(N^{old}, N^{new}, H, \succsim, \omega)$
  - the sets of agents N^{old} existing "old" tenants N^{new} newcomers
  - the set of houses H with $\#H \geq \#N^{old}$
  - agent $i \in N^{old} \cup N^{new}$ has **strict** preferences $\succsim_i$ over houses
  - $\omega : N^{old} \to H$ is the initial allocation of the houses among the existing tenants
- A matching $\mu : N^{old} \cup N^{new} \to H \cup \{\emptyset\}$ specifies the final allocation of the houses: if $\mu(i) \in H$, then i assigned with the house $\mu(i)$, and if $\mu(i) = \emptyset$, then i is not assigned a house

- Two special cases:
  - all agents are existing tenants ($\#N^{new} = 0$) and there is no vacant house ($\#H = \#N^{old}$) = housing market => Gale's top trade cycle
  - all agents are newcomers ($\#N^{old} = 0$) = house allocation problem => serial dictator
- The mixed case requires a novel solution
- **The you request my house - I get your turn** (YRMH-IGYT) mechanism: form an ordering $\pi : \mathbb{N} \to N$ of the agents

Step 0: Let the agent $\pi(1)$ receive his top choice, agent $\pi(2)$ his top choice among the remaining houses and so on, until someone requests the house of an existing tenant

(a) If $\pi(k)$ requests an existing tenant's house who has already received a house, then proceed the assignment to the agent $\pi(k+1)$

(b) If $\pi(k)$ requests an existing tenant's house who has not already received a house, then assign the tenant's house with $\pi(k)$, give the tenant a right to choose a house, and then proceed the assignment to the agent $\pi(k+1)$

- If there is no cycle, then match the agents to their assigned houses
- If there is a cycle, then match the agents in the cycle with their assigned houses and move to the next step

Step k: Repeat the previous step's procedure with the ordering π on the unmatched agents

Stop: When there are no agents or houses to be matched

- The YRMH-IGYT mechanism generalizes previous important mechanisms:
  1. Serial dictatorship when there are no existing tenants: Without existing tenants, the "you request my house..." contingency does not happen, so the mechanism coincides with serial dictatorship
  2. Gale's TTC if all agents are existing tenants and there is no vacant house: In that 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

- YRMH-IGYT can be interpreted as a variant of Gale's TTC in which all vacant houses (and houses whose initial owners are already assigned houses) point to the highest priority agents rather than the owners of the houses

Theorem 65 *(Abdulkadiroglu and Sönmez 1999) Any YGMH-IGYT mechanism is individually rational, strategy-proof, and Pareto-optimal*

- individual rationality:
  - whenever some agent points to a house of an existing tenant, the latter is promoted to the top of the priority
  - whenever an agent is in the top of the priority ordering, she can guarantee her house by forming a cycle of herself and her house
- A mechanism is **consistent** if the match is unchanged if the mechanism is implemented on a subproblem after one removes some agents and their assignments

Theorem 66 *(Sönmez and Unver 2005) A mechanism is Pareto-optimal, individually rational, strategy-proof, weakly neutral, and consistent if and only if it is a YRMH-IGYT mechanism*

5 Applications

- By now we understand well properties of some good mechanisms in one-sided and two-sided matching markets
- In the two-sided markets, student proposing Deferred Acceptance by Gale and Shapley is
  - stable
  - student optimal in the class of stable mechanisms
  - weakly Pareto-efficient for the students
  - strategy-proof for the students
  - approximately strategy-proof for the schools in large markets
- Moreover, DA is essentially the only mechanism having these properties
- In the one-sided markets, Top trade cycle by Gale is
  - in the core

  - strategy-proof
  - Pareto-efficient

- Moreover, TTC is essentially the only mechanism having these properties
- These mechanisms serve as a benchmark for the design of mechanisms in applications
- The design strategy is to modify them as needed to account for context dependent complications
- In this lecture, we shall explore how this have been done in
  - organ donation
  - school choice

5.1 Kidney exchange

- Transplant is an important treatment of serious kidney diseases
- There are 90 00 patients on the waiting list for cadaver kidneys in the U.S.
- In 2010 almost 11 000 transplants of cadaver kidneys performed of which 6 300 from living donors in the U.S., 5 000 patients died while on the waiting list and more than 2 000 others were removed from the list as (too sick)
- Sometimes donors are incompatible with their intended recipient => possibility of exchange
- Buying and selling kidneys is illegal => donation is the only source of transplant
- An example of *repugnant* trade
- For a successful transplant, the donor kidney needs to be **compatible** with the patient
  - blood type A->A,AB; B->B,AB; O->A,B,AB,O; AB->AB
  - proteins in the tissue
- Problem: a living donor only wants to give away a kidney if it helps a certain patient

  => donors compatible to an unknown patient not always willing

- How to increase the number and quality of transplant?
- Two simple mechanisms:

1. **A paired exchange**:
    - match two patient-donor pairs where the donor of pair 1 is incompatible with the patient of pair 1 but is compatible with the patient of pair 2, and vice versa
    - In such a case, donor 1 can give her kidney to the patient 2 and the donor 2 can give his kidney to the patient 1 in return
2. **A list exchange**:
    - the donor of the incompatible pair donates his/her kidney to someone on the waiting list, and
    - the patient of the incompatible pair is placed at the top of the waiting list

- In 2004, the Renal Transplant Oversight Committee of New England approved the establishment of a clearinghouse for kidney exchange.
- Roth, Sönmez and Unver as well as doctors design the clearinghouse.
- Desiderata
    - Efficiency (Pareto efficiency, maximizing the number of transplantation)
    - Incentives (strategy-proofness)
    - fairness
- A **kidney exchange model** (Roth, Sönmez and Unver 2004) is defined by
    - A set of donor-patient (kidney-transplant) pairs $\{(k_1, t_1), ..., (k_n, t_n)\}$
    - A strict preference $\succsim_i$over $\{k_1, ..., k_n\} \cup \{w\}$ for each t_i, where w is priority in the waitlist (in exchange of donating kidney k_i)
- A **matching** is a function $\mu : \{t_1, ..., t_n\} \to \{k_1, ..., k_n\} \cup \{w\}$ that specify which patient obtains which kidney (or waitlist)
- We assume w can be matched with any number of patients
- A **mechanism** is a procedure to select a matching for each problem

- With these assumptions, the kidney exchange problem can be interpreted as a the problem of house allocation with existing tenants:

$$\begin{aligned} \text{donor} &= \text{occupied house} \\ \text{waitlist} &= \text{vacant house} \\ \text{patient} &= \text{tenant} \end{aligned}$$

- Hence, a promising solution is **YRMH-IGYT** mechanism (a.k.a. TTC mechanism)
- At each step $t = 0, 1, ...$
  - Let the agent with the top priority receive her first choice kidney, the second agent his top choice among the remaining kidney and so on, until someone requests the kidney of a paired donor.
  - If the paired patient whose paired donor is requested has already received a kidney, then proceed the assignment to the next agent

    Otherwise, insert the paired patient at the top of the priority order and proceed with the procedure
  - If at any step a cycle forms, assign these kidneys by letting them exchange, and then proceed with the algorithm.
- Stop, when all kidneys or patients are matched
- Recall from the previous lecture that the YRMH-IGYT is Pareto efficient, strategy-proof, and individually rational
- However, YRMH-IGYT may not be feasible since
  1. only pairwise exchanges may be possible (at least initially) since all surgeries should be conducted simultaneously (contracting is illegal)
  2. patients may have dichotomous preferences (0-1 preferences), that is, all compatible kidneys are equally good and all incompatible kidneys are equally bad, at least as first approximation
- Now we can think the donor-patient pairs as the players, and ask whether two such players can be matched pairwisely in a meaningful way
- A matching is Pareto efficient if there is no other matching that makes every patient weakly better off and at least one patient strictly better off

- A mechanism is strategy-proof if no pair benefits by misreporting who is mutually compatible with them
- Consider the following **priority mechanism** (serial dictatorship):

Step 0: Order pairs in some priority ordering (could be random or favor waiting time, etc.)

If there is any matching in which the top priority pair is matched, then match that pair. Otherwise, skip that pair.

Step 1: Match the second-top priority pair if there is such a matching that also match the first pair (if they were matched in the previous step), then match the pair. Otherwise, skip that pair

Step k: Match the kth top priority pair if there is such a matching that also match all the pairs that were matched in previous steps, then match the pair. Otherwise, skip that pair.

Stop: When there are no pairs to be matched.

Theorem 67 *(Roth, Sönmez, and Unver 2004) The priority mechanism is Pareto efficient and strategy-proof*

- With dichotomous preferences, the priority mechanism is also stable
- However, with richer preferences this need not hold - recall the formal similarity of the model to the roommate problem
- An exchange involving more than two pairs may be difficult, but may not be infeasible.
- How much efficiency gain can we obtain through larger exchanges?

Example 68 *A pair is denoted as type x-y if the patient and donor are A,B,or O blood-types x and y, respectively. Consider a population composed of O-B, O-A, A-B, A-B, B-A (blood-type incompatible), and A-A, A-A, A-A, B-O (positive crossmatch). Assume there is no tissue rejection between patients and other patients' donors.*

- *If only two-way exchanges are possible:*
  (A-B,B-A), (A-A,A-A), (O-B,B-O)
- *If three-way exchanges are also feasible:*
  (A-B,B-A); (A-A,A-A,A-A); (B-O,O-A,A-B)

- *The three-way exchanges allow*

  1. *an odd number of A-A pairs to be transplanted (instead of only an even number with two-way exchanges), and*
  2. *O-type donors can facilitate three transplants rather than two*

- Would it help to have four-way exchanges?

  - in the above example, no
  - in general, maybe but rarely

- More than five-way exchanges practically useless (Roth, Sönmez, and Unver 2007)

- Conclusion: good mechanism achievable as soon as transplantation technology allows four-way exchanges instead of two

5.2 School choice

- In many countries, children were automatically sent to a school in their neighborhoods

- Recently, many cities employ school choice programs: school authorities take into account preferences of children and their parents

- Typical goals of the authorities are: (1) efficient placement, (2) fairness of outcomes, (3) easy for participants to understand and use, etc.

- Abdulkadiroglu and Sönmez (2003) showed that placement mechanisms used in many cities such as Boston are flawed, and proposed a new mechanism

- Finite sets S of students and C of schools

- Each student s can be matched to at most one school, and each school c can admit at most q_c students

- Each student s has strict preferences $\succ_s$ over schools and being unmatched (denoted by $\emptyset$).

- For each school, there is a (for now, strict) priority order $\succ_c$over students

- The outcome is a **matching**, which specifies which student attends which school

- A matching is **stable** if it not blocked
  - by a student who is matched to an unacceptable school, or by a school that takes in unacceptable students
  - by a student-school pair that would rather be matched with one another than their current matches
- In school choice, stability can be understood as a fairness criterion
- No blocking pair means **no justified envy**: there is no situation in which student s is matched to a worse school than c, and c admits another student who has lower priority at than s does
- The old **Boston mechanism** is defined as follows

Step 0: Each student submits a preference ranking of the schools

Step 1: Each school considers 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: For each school still with available seats, each school considers the students who have listed it as their kth 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 kth choice

Stop: When there is no school with available seats, or when every student has been assigned to a school

- Boston mechanism encourages manipulation: even if a student has a very high priority at a school, unless she lists it as her top choice she loses her priority to students who have top ranked that school.
- The Boston mechanism is unstable, i.e., does not eliminate justified envy: priorities are lost unless the school is ranked as the top choice
- Boston mechanism may produce an inefficient matching given students may behave strategically => many students end up unassigned
- Abdulkadiroglu, Roth, and Sönmez (2004): Of the 15.135 students analyzed, 19% (2910) listed two overdemanded schools as their top two choices, and about 27% (782) of these ended up unassigned

- Such behavior is clearly a bad choice, and people suffer from not being sophisticated enough to game the system (in this sense, strategy-proofness can also be interpreted as a fairness criterion)
- Natural replacement: **student proposing DA**
  - stable
  - student optimal in the class of stable mechanisms
  - (weakly) Pareto-efficient for the students
  - strategy-proof for the students
  - approximately strategy-proof for the schools in large markets
- Moreover, as schools are merely goods to be consumed rather than players, it makes sense to take only into account the welfare of the students
- Further, priorities at schools are often decided by law and hence they cannot behave strategically
- However, **student proposing DA** is not (strongly) Pareto-efficient, i.e. there could be another matching that would not decrease payoff of any student but would increase some

Example 69 *Let $S = \{i, j, k\}, C = \{a, b\}$, and student preferences be*

$$\succ_i: b, a$$
$$\succ_j: a$$
$$\succ_k: a, b$$

and both schools have one position and priorities are

$$\succ_a: i, j, k$$
$$\succ_b: k, i$$

Student proposing DA results in $\mu = \{(i, a)(j, \emptyset), (k, b)\}$ which is weakly less preferred by every student and strictly preferred by some student to $\mu' = \{(i, b), (j, \emptyset), (k; a)\}$, hence DA is inefficient!

- Hence, the student-proposing DA may not produce a Pareto-efficient matching - indifference matter!
- Since the student-proposing DA is Pareto dominant among stable matchings, no stable matching is Pareto efficient

- In school choice, stability may be desirable but may not be indispensable: it e.g. depends on the school districts
- Can we improve the efficiency at the expense of stability (=fairness)?
- The TTC algorithm (Abdulkadiroglu and Sönmez 2003):

1. Assign a counter for each school that keeps track of how many seats are still available at the school. Initially set the counters equal to the capacities of the schools
2. Each student "points to" her favorite school. Each school points to the student who has the top priority.
3. There is at least one cycle (why?). 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 are unchanged
4. Repeat above steps for the remaining school seats and students

- Thus in the two-sided markets TTC allows students to trade priorities, starting with the students with highest priorities

Theorem 70 *(Abdulkadiroglu and Sönmez 2003) The TTC mechanism is Pareto-efficient and strategy-proof*

Example 71 *In the previous example, TTC results in matching* $\mu' = \{(i, b), (j, \emptyset), (k, a)\}$, *which is Pareto efficient. However, this matching is* ***not*** *stable:* (j, a) *is a blocking pair.*

- Instability of TTC a result of two-sidedness
- E.g., Boston and New York City have explicitly worked with economists (Boston: Abdulkadiroglu, Roth, Sönmez) and designed their school choice mechanisms
- Since priorities are set by law for Boston schools, Abdulkadiroglu et al. recommended not only DA but also TTC, for efficiency reasons
- However the school system finally chose DA: policy makers were not appealed by the idea of trading
- Student proposing DA was implemented in Boston in 2006 and is still in use

- Similar experiences with the New York school match
- After the new design:
  - 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
  - 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