Social Choice

Algorithmic Game Theory

Winter 2024/25

Voting

Impossibility Results

Structured Preferences

Kidney Exchange

Stable Matching

Definitions

- Set A of candidates (or outcomes, alternatives)
- Set $\mathcal{N}$ of n voters (or players)
- Set L of possible preferences (total orders of A)
- Each voter i has a preference (or preference order) $\succ_i \in L$ on the candidates A

- A social welfare function is a function $F : L^n \to L$.
- A social choice function is a function $f : L^n \to A$.

A social choice function outputs only a single winner, a social welfare function outputs a complete ranking of all candidates.

Majority Rule with Two Candidates

Direct Election: candidates $A = \{a, b\}$.

Voter	Preference
1	a $\succ_1$ b
2	b $\succ_2$ a
3	a $\succ_3$ b

Majority Rule / Condorcet Winner: rank $a \succ b$ if a majority of voters prefer a to b. Here: $a \succ b$

Majority Rule for two candidates implements many desirable properties:

- Represents the majority of preferences
- Each candidate is in the position he/she appears most often
- Strategic voting is not profitable:
  A voter with majority preference changes his vote: Can only get worse.
  A voter without majority preference cannot change the outcome by changing his vote.

Condorcet Paradox

Three candidates: $A = \{a, b, c\}$.

Voter	Preference
1	a $\succ_1$ b $\succ_1$ c
2	b $\succ_2$ c $\succ_2$ a
3	c $\succ_3$ a $\succ_3$ b

Majority Rule yields a cycle:
2 voters prefer a over b, 2 prefer b over c, and 2 prefer c over a...

The instance shows that the collective preference can be contradictive (cyclic, non-transitive) although every single voter has a well-defined preference.

The example is called Condorcet Paradox and is attributed to the Marquis de Condorcet around 1785.

Plurality Rule

As an example for a social welfare function, let us examine the Plurality Rule. The winner is the candidate with largest number of first-place rankings. We break ties w.r.t. alphabet.

Voter	Preference
1	a $\succ_1$ b $\succ_1$ c
2	c $\succ_2$ a $\succ_2$ b
3	b $\succ_3$ c $\succ_3$ a

Plurality: $f(\succ_1, \succ_2, \succ_3) = a$

Voter	Reported Preference
1	a $\succ_1$ b $\succ_1$ c
2	c $\succ_2$ a $\succ_2$ b
3	c $\succ_3$ b $\succ_3$ a

Plurality: $f(\succ_1, \succ_2, \succ_3) = c$

Strategic voting is profitable for the third voter!

How can we avoid strategic voting? A trivial way is to choose one voter as a dictator who dictates the outcome in his vote. But is there a different way?

2000 USA Presidential Election

Outcome of the 2000 USA presidential election was eventually decided in the state of Florida. In Florida, the plurality rule is used.

Final Results:

Candidate	Party	Votes
Bush	Republicans	$2,912,790$
Gore	Democrats	$2,912,253$
Nader	Green	$97,488$

General assumption: Majority of Nader voters prefer Gore over Bush. The existence of Nader flipped the election's outcome.

List of Properties of Social Welfare Functions F

Two desirable properties of social welfare functions F:

- Unanimity: If $a \succ_i b$ for all $i \in \mathcal{N}$, then $a \succ b$, where $\succ = F(\succ_1, \ldots, \succ_n)$.
- Independence of Irrelevant Alternatives (IIA): The social preference between any two candidates a and b depends only on the voters' preferences between a and b.

  Formally:
  For every $a, b \in A$ and every $\succ_1, \ldots, \succ_n, \succ'_1, \ldots, \succ'_n \in L$, we denote by $\succ = F(\succ_1, \ldots, \succ_n)$ and $\succ' = F(\succ'_1, \ldots, \succ'_n)$.
  If $a \succ_i b \Leftrightarrow a \succ'_i b$ for all i, then $a \succ b \Leftrightarrow a \succ' b$.

Note: The Plurality Rule violates IIA!

Dictatorship is an example of a social welfare function with both properties: In a dictatorship, for some player $i \in \mathcal{N}$,

$$\succ_i = F(\succ_1, \ldots, \succ_n)$$

for all $\succ_1, \ldots, \succ_n$. Player i is called the dictator.

Arrow's Theorem

Theorem (Arrow, 1950)

Every social welfare function over a set of $|A| \geq 3$ candidates that satisfies unanimity and IIA is a dictatorship.

Suppose F is a social welfare function that satisfies unanimity and IIA.

Lemma (Pairwise Neutrality)

Suppose $\succ_1, \ldots, \succ_n$ and $\succ'_1, \ldots, \succ'_n$ are two preference profiles, and $\succ = F(\succ_1, \ldots, \succ_n)$ and $\succ' = F(\succ'_1, \ldots, \succ'_n)$. If for every voter i we have $a \succ_i b \Leftrightarrow c \succ'_i d$, then $a \succ b \Leftrightarrow c \succ' d$.

We omit the proof of the lemma and proceed to show the theorem.

Who is the Dictator?

Proof (Theorem):
Pairwise neutrality shows that a social welfare function that satisfies unanimity and IIA has a general underlying approach of determining a global preference. This approach is similar for all preference orders and all pairwise comparisons of elements. This insight can be used to show that, in fact, the approach boils down to having one dictator determine the entire ordering.

Let $a \neq b$ and $c \neq d$.

- If there are no voters with $a \succ_i b$, then $b \succ a$.
- If there are n voters with $a \succ_i b$, then $a \succ b$.
- First change at voter i^*:

1	...	$i^* - 1$	i^*	...	n	Result
$a \succ_i b$			$b \succ_i a$			$b \succ a$
$a \succ_i b$				$b \succ_i a$		$a \succ b$

Claim: i^* is a dictator!

i^* is the Dictator

- i^* is a dictator if $c \succ_{i^*} d \Rightarrow c \succ d$ for all $c \neq d \in A$.
- Consider an arbitrary set of preferences with $c \succ_{i^*} d$ and $e \in A$ with $e \neq c$ and $e \neq d$.
- Move third element e s.t. it appears as below in $\succ_i$:

1	e	$\dots$					
$\dots$	e	$\dots$					
i^*	$\dots$	c	$\dots$	e	$\dots$	d	$\dots$
$\dots$	$\dots$						e
n	$\dots$						e

- Because of IIA the movement of e does not change the order of c and d in $\succ$.
- (c, e) appears exactly as (a, b) previously. By pairwise neutrality we know $c \succ e$. The same argument shows $e \succ d$.
- Thus, $c \succ d$. Note that c and d are arbitrary candidates. Hence, i^*'s preference determines all pairwise rankings and therefore the entire output. □ (Theorem)

Properties of Social Choice Functions f

- f can be strategically manipulated by voter i if for some $\succ_1, \ldots, \succ_n$ and some $\succ'_i$ we have that $a \succ_i b$ where $b = f(\succ_1, \ldots, \succ_n)$ and $a = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n)$. f is called incentive compatible (IC) or strategyproof if it cannot be manipulated.
- f is monotone if $f(\succ_1, \ldots, \succ_n) = a \neq b = f(\succ_1, \ldots, \succ'_i, \ldots, \succ_n)$ implies that $a \succ_i b$ and $b \succ'_i a$.
- f is surjective or onto A if for every candidate $a \in A$ there is a set of preferences such that a is the winner.
- f has always-a-winner (AAW) if it yields a winner for every set of preferences.
- f fulfills the Condorcet-winner criterion (CWC) if f outputs the unique Condorcet winner if there is one, i.e., a candidate a such that for every $b \neq a$ a majority of voters ranks $a \succ_i b$.
- Independence of Irrelevant Alternatives (IIA): If $f(\succ_1, \ldots, \succ_n) = a$, then by changing preferences without changing the order of a and b, we cannot make b the winner.

Some Social Choice Functions

- Condorcet: If there is a unique Condorcet winner, then a is the winner.
- Borda Count awards points to each candidate: For each voter, top gets $|A|-1$ points, second $|A|-2$ points, $\dots$, last gets 0 points. A candidate with the highest score wins (if there are several, break ties arbitrarily).
- Dictatorship: The first-placed candidate of some fixed voter $i \in \mathcal{N}$ wins. Player i is called the dictator.

	AAW	CWC	IIA	IC	Monot.	Surj.
Condorcet	-	✓	✓	-	-	✓
Plurality	✓	-	-	-	-	✓
Borda Count	✓	-	-	-	-	✓
Dictatorship	✓	-	✓	✓	✓	✓

Gibbard-Satterthwaite Theorem

Proposition

A social choice function is IC if and only if it is monotone.

Proof: Exercise. □

Theorem (Gibbard 1973; Satterthwaite 1975)

Suppose f is a social choice function onto A with $|A| \geq 3$. f is IC if and only if f is a dictatorship.

The proof works by extending a social choice function f to a social welfare function F that satisfies unanimity and IIA. The result then follows by contradiction from Arrow's Theorem.

Single-Peaked Preferences

While the Gibbard-Satterthwaite Theorem is devastating, it requires the generality of preferences. If preferences are restricted and structured, a richer class of IC social choice rules exist.

Let us consider a set of outcomes that can be located along a line. We assume that $A = [0, 1]$.

Definition

A preference order $\succ_i$ over A is single-peaked if there is a point $p_i \in A$ such that for all $x \in A \backslash \{p_i\}$ and $\lambda \in [0, 1)$ we have

$$(\lambda x + (1 - \lambda)p_i) \succ_i x \ .$$

As application consider the problem of deciding a tempature value for a shared office. Every employee has an optimal value and would like the temperature to be as close as possible to that.

Order Mechanisms

For single-peaked preferences the Gibbard-Satterthwaite theorem does not apply.

k-th Order Mechanism for Single-Peaked Preferences:

- Suppose k is a number in $\{1, \ldots, n\}$
- Collect only the peaks $p_1, \ldots, p_n$ of the voters.
- Sort the peaks increasingly from 0 to 1 und and output the k-th largest peak

Proposition

For every fixed $k \in \{1, \ldots, n\}$, the k-th order mechanism is incentive compatible. If $n \geq 2$, it is not a dictatorship.

Order Mechanisms

Proof:
Let p be the outcome if all voters report their order truthfully. If $p_i > p$, voter i cannot change the outcome with $p_i' > p_i$. If he lies a peak $p_i' \leq p$, it results in a worse outcome $p' \leq p$. The argument for $p_i < p$ is similar. Non-dictatorship is obvious. □

The most prominent rule is the median mechanism with $k = \lfloor (n+1)/2 \rfloor$. Note that taking the average of the peaks $\sum_{i=1}^{n} p_i/n$ is not IC.

By the same argument as above, every k-th order mechanism remains IC if – in addition to the reported peaks – we consider any number of apriori fixed outcomes $y_j \in [0, 1]$ and include them into the sorting. The mechanisms chooses the k-th largest location of $\{p_1, \ldots, p_n, y_1, \ldots, y_m\}$.

Order Mechanisms

For every fixed k, the k-th order mechanism is anonymous, i.e., it satisfies $f(\succ_1, \ldots, \succ_n) = f(\succ'_1, \ldots, \succ'_n)$ if $(\succ_1, \ldots, \succ_n)$ is a permutation of $(\succ'_1, \ldots, \succ'_n)$.

Theorem (Moulin 1980; Ching 1997)

Suppose p_i are the reported peaks. A social choice rule f for is incentive compatible, surjective, and anonymous for single-peaked preferences if and only if f is a k-th order mechanism over a set $\{p_1, \ldots, p_n, y_1, \ldots, y_m\}$, where $y_j \in [0, 1]$ are fixed outcomes.

The result is a complete characterization for anonymous IC mechanisms. Anonymity is required, because every dictatorship is not a k-th order mechanism but surjective and IC (and non-anonymous).

House Allocation

Matching with Preferences over Objects

- n agents and n houses
- Assumption: Agent i owns house i
  (not really necessary, simplifies analysis)
- Agent i has a strict and total preference order $\succ_i$ over houses
  (getting no house is least preferred)
- Assign one house to every agent

The set A of outcomes contains all partial matchings of houses to agents. The preference of an agent only applies to the house that the agent obtains. Thus, all matchings in which agent i gets the same house are equivalent for her.

What are IC mechanisms? Do they have further good properties?

Top-Trading-Cycles

$G = (V, E)$ is a directed graph:

- V is the set of remaining agents with their houses
- E is the set of directed edges:

$$(i, \ell) \in E \Leftrightarrow \ell \in V \text{ owns best remaining house for } i \in V$$

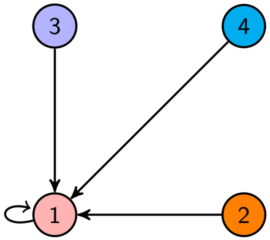
Top-Trading-Cycles (TTC) Mechanism:

1. Query preferences of agents
2. **while** $V \neq \emptyset$
3. Compose the edge set E as described above
4. Compute all directed cycles $C_1, \ldots, C_h$ in G
   (Self-loop is a cycle, all cycles are distinct)
5. **for** every edge (i, ℓ) in a cycle $C_1, \ldots, C_h$ **do**
6. Assign house ℓ to agent i.
7. Remove all agents in $C_1, \ldots, C_h$ from V

Example

Agent	Preference			
1	1	2	3	4
2	1	3	2	4
3	1	4	2	3
4	1	2	4	3

Graph G in round 1:

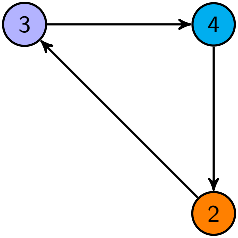


Assignment:
Agent 1 – House 1

Example

Agent	Preference		
2	3	2	4
3	4	2	3
4	2	4	3

Graph G in round 2:



Assignment:
Agent 2 – House 3
Agent 3 – House 4
Agent 4 – House 2

Top-Trading-Cycles – Analysis

Observations:

- Every agent in G has out-degree 1. There is at least one cycle in G, and all cycles in G are distinct.
- Let V_k be the set of agents that are removed in round k of TTC. Every agent in V_k gets the house that she likes best – given that houses of agents in $V_1 \cup \ldots \cup V_{k-1}$ are gone.
- Agent $i \in V_k$ gets her best house in the k-th round. The owner of this house is also in V_k and, thus, also gets her best house.

Theorem (Roth 1982)

The TTC Mechanism is incentive compatible.

Proof:
Consider agent $i \in V_j$ with true preference $\succ_i$. If i tells the truth, she gets her best remaining house in round j.

Top-Trading-Cycles – Analysis

Consider any house of an agent in $V_1 \cup \ldots \cup V_{j-1}$ that i likes better.
Agent i can get none of these houses:

- In round $k = 1, \ldots, j-1$ no agent $\ell \in V_k$ wants the house of i, since otherwise i would be in a cycle with ℓ.
- No agent $\ell \in V_k$ wants the house of i in any round $< k$, since otherwise ℓ would still want it in round k.

The houses of agents in $V_1 \cup \ldots \cup V_{j-1}$ remain out of reach for i, no matter which preference order i reports to the mechanism. Thus, since i must wait until round j, TTC delivers the best (remaining) house when i reports truthfully. □

TTC is an IC mechanism, but there are many more.

Why would TTC be preferable to other IC mechanisms?

Top-Trading-Cycles – Core Assignment

Let M be an assignment of houses – agent i gets house $M(i)$.
Let M_S be an assignment that results from M if a coalition $S \subseteq \mathcal{N}$ takes their initial houses and redistributes them among themselves.

Definition

A set of agents $S \subseteq \mathcal{N}$ is a blocking coalition for M if there is an assignment M_S such that

- for every agent $j \in S$ is M_S at least as good: $M_S(j) \succeq_j M(j)$
- for at least one agent $i \in S$ is M_S strictly better: $M_S(i) \succ_i M(i)$

We say that an assignment M without blocking coalition is in the core.

Assignments in the core satisfy an optimality condition: No subset of agents want to remove their houses from the mechanism and redistribute them among themselves. In particular, every agent i gets a house that is at least as good as the one i started with initially.

Top-Trading-Cycles – Core Assignment

Theorem (Roth, Postlewaite 1977)

The TTC Mechanism computes the unique core assignment.

Proof:

Induction: Only the TTC assignment can be in the core, but no other one.

- Start: Every $i \in V_1$ gets the best house overall. V_1 is a blocking coalition for every assignment that does not distribute the houses in V_1 as TTC.
- Assume: The houses of agents in $V_1, \ldots, V_{j-1}$ must be assigned as in TTC.
- Step: Given the assumption, every $i \in V_j$ gets the best (remaining) house. V_j is a blocking coalition for every assignment that distributes the houses in $V_1, \ldots, V_{j-1}$ as TTC, but the ones in V_j not as TTC.

Hence: Either TTC assignment is in the core or core is empty.

Top-Trading-Cycles – Core Assignment

Let M be the TTC assignment and M_S the result of a redistribution of houses of agents in S. Agent $i \in S$ initially owns house i. If S with M_S is blocking for M, then i must get some house – otherwise she is worse off than in M.

Hence, the redistribution among S consists of a set of cycles. We distinguish 2 cases:

- There is a cycle that contains agents from V_j and V_ℓ with $j < \ell$:
  At least one agent $i \in V_j$ receives a house from an agent in V_ℓ and hence is worse off than in M.
- Each cycle contains only agents of one set V_j:
  In M every agent $i \in V_j$ gets her best house among the ones of agents in V_j. No agent can have a more preferred house than in M.

As a result, the TTC assignment is in the core. It is the unique assignment with this property. □

Random Serial Dictatorship

Random Serial Dictatorship (RSD) Mechanism:

1. Consider agents in uniform random order
2. Query preferences of agents
3. Let V be the set of all houses
4. **for** $i = 1, 2, 3, \ldots, n$ **do**
5. Assign to agent i her best house h from V
6. Remove h from V

This mechanism is an ordered version of a dictatorship.
The following result can be shown similarly as for TTC.

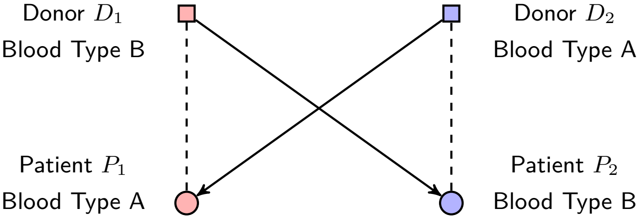
Theorem

For every chosen ordering of agents the RSD mechanism is incentive compatible.

Is the RSD assignment in the core? For all orderings of agents? For none or some?

Kidney Donation

Many countries have (plans for) kidney exchange programs. Patients often have a relative or a friend who is willing to donate a kidney but does not fit to the patient (e.g., in terms of blood type). The goal is organ exchange: Two (patient,donor)-pairs exchange the donation organ if they both fit to the corresponding patient of the other pair.



Top-Trading-Cycles for Kidney Exchange

This is essentially a house allocation:

- "Houses" are organs, "agents" are patients
- Preferences over organs = Prob. of successful transplantation
- TTC Mechanism is IC and in the core

Remarks:

- No legal obligation to donation. All operations of a cycle are conducted simultaneously to avoid incentive problems: Donor D_i drops out as soon as patient P_i receives a kidney.
- Complicated: long cycles = many simultaneous operations. If, however, a single kidney donation is present, then the cycle becomes a path. Then we first remove the kidney of donor D_i before patient P_i receives the transplant.
- Usually rather binary preferences: Kidney suitable for patient or not.

Organ Donation via Matching

Kidney Exchange via Matching in a Graph $G = (V, E)$:

- For patient P_i suppose E_i is set of compatible donors
- Every pair (P_i, D_i) is a vertex $v_i \in V$
- We examine simple exchanges, i.e., cycles of length 2: Edge $\{v_i, v_j\} \in E$ if $D_i \in E_j$ and $D_j \in E_i$
- Patient can lie about reporting set E_i
- Obviously: Patient only has incentive to lie $F_i \subseteq E_i$

Matching-Mechanism for Kidney Exchange:

- Query sets F_i of patients
- Construct graph G as above, where $E = \{\{v_i, v_j\} \mid D_i \in F_j, D_j \in F_i\}$
- Compute a matching M in G with maximum cardinality
- Kidney exhange according to edges in M

Maximum Matching with Priority Lists

The maximum matching is not unique. Different maximum matchings distribute the kidneys to different patients. We must compute the matching in a "monotone" fashion. This is achieved by prioritizing patients. This is a common approach, e.g., using wait lists for organ donation.

Maximum Matching with Priority Lists

1. M_0 is the set of all maximum matchings of G
2. **for** $i = 1, 2, 3, \dots, n$ **do**
3. Let Z_i be the set of all matchings in M_{i-1} in which vertex v_i is matched
4. **if** $Z_i \neq \emptyset$ **then** $M_i \leftarrow Z_i$; **else** $M_i \leftarrow M_{i-1}$.
5. **return** arbitrary matching from M_n

Theorem

The matching mechanism with priority lists for kidney exchange is incentive compatible for every fixed priority list.

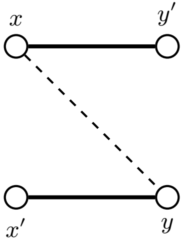
Recall: Stable Matching

- Set $\mathcal{X}$ of m men, set $\mathcal{Y}$ of n women
- Each $x \in \mathcal{X}$ has preference order $\succ_x$ over all $y \in \mathcal{Y}$.
- Each $y \in \mathcal{Y}$ has preference order $\succ_y$ over all $x \in \mathcal{X}$.
- For every person being unmatched is least preferred.
- For matching M let $M(x) \in \mathcal{Y}$ be the partner of man $x \in \mathcal{X}$ in M, and $M(y) \in \mathcal{X}$ the partner of woman $y \in \mathcal{Y}$ in M.
- $M(x) = *$ if x unmatched in M, and $M(y) = *$ similarly.

Stable Matching

When is a matching stable? What is a hazard to stability?

- In M a pair $\{x, y\}$ is a blocking pair if and only if x and y prefer each other to $y' = M(x)$ and $x' = M(y)$, respectively.
- M is a stable matching if and only if it admits no blocking pair.



Theorem (Gale, Shapley 1962)

The Deferred-Acceptance algorithm computes a stable matching in at most $O(nm)$ iterations.

Deferred-Acceptance Algorithm

Algorithm 1: Deferred Acceptance (DA) Algorithm with Man-Proposal

Initialize $\succ'_x = \succ_x$ for all $x \in \mathcal{X}$
while *there is an unmatched man* $x \in \mathcal{X}$ *with* $\succ'_x \neq \emptyset$ **do**
    Every man $x \in \mathcal{X}$ proposes to topmost woman in $\succ'_x$
    Every woman $y \in \mathcal{Y}$ keeps most preferred man from proposals $A_y(S)$
    y rejects all other men from $A_y(S)$
    If his current proposal is rejected, man x removes top-entry from $\succ'_x$

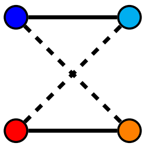
Suppose in the DA algorithm the men make their proposals sequentially in some order instead of simultaneously.

- Are there several stable matchings?
- Can the algorithm compute several (all?) stable matchings?

Uniqueness

(2, 1) A

(1, 2) B



1 (A, B)

2 (B, A)

Stable matchings are not always unique. In this case there is a man-optimal matching (solid) and a woman-optimal (dashed) one. In an optimal matching, every agent on one side gets simultaneously matched to the best partner.

$h(x) \in \mathcal{Y}$ is the best woman s.t. $\exists$ stable matching M' with $M'(x) = h(x)$.
$h(y) \in \mathcal{X}$ is the best man s.t. $\exists$ stable matching M' with $M'(y) = h(y)$.

Definition

A stable matching M is

- man-optimal if $M(x) = h(x)$ for every $x \in \mathcal{X}$.
- woman-optimal if $M(y) = h(y)$ for every $y \in \mathcal{Y}$.

Optimality

In an optimal matching every agent on one side simultaneously gets matched to the best partner that the agent could obtain in any possible stable matching.

It is unclear if it is even possible to fulfill this criterion.

The following theorem shows this property and even more: No matter in which order the proposals are made in the DA algorithm, it always computes one of the two unique optimal matchings.

Theorem

The DA algorithm with man-proposal always computes the unique man-optimal stable matching. With woman-proposal it always computes the unique woman-optimal stable matching.

Optimality

Proof: By contradiction to stability:

- Let M be the matching of the algorithm with man-proposal.
- For every man x consider the best stable partner $h(x)$.
- Suppose M is not man-optimal, then there is at least one man x with $h(x) \succ_x M(x)$.
- Consider the first iteration, in which a man x is rejected by $h(x)$. Let M' be a stable matching which contains $(x, h(x))$.
- $h(x)$ rejects man x only because she has a better proposal of man $i \succ_{h(x)} x$.
- Since this is the very first iteration with such a rejection, man i is still proposing to women ranked above $h(i)$.
- M' stable, so i's partner in M' is worse: $h(x) \succ_i h(i) \succeq_i M'(i)$
- $(i, h(x))$ is a blocking pair in M'. M' not stable – contradiction $\square$

Proposals and Incentives

Theorem

The DA algorithm with man-proposal is incentive compatible for the men.

Proof:

We use the following notation. The liar wlog. is man number 1.

- True preference: $\pi = (\succ_1, \ldots, \succ_n)$, algorithm computes M
- Man 1 lies: $\pi' = (\succ', \succ_2, \ldots, \succ_n)$, algorithm computes M'

If lying is profitable, then $M'(1) \succ_1 M(1)$. We will show that in this case M' is not stable for π' – contradiction.

The following claim shows that within the set of men that profit from the lie the assigned partners get permuted.

Proposals and Incentives

Claim

Let $R = \{x \mid M'(x) \succ_x M(x)\}$ be the men that profit in M'. For every man $x \in R$ and his new partner $y = M'(x)$, the old partner $x' = M(y)$ of y in M is also in R, i.e., $x' \in R$.

Proof:

- $x' = 1$: Clear, $x' \in R$ by assumption that 1 wants to lie
- $x' \neq 1$: Since $x \in R$ we have $y \succ_x M(x)$. Then $x' \succ_y x$, since otherwise M has blocking pair $\{x, y\}$.
- $\Rightarrow M'(x') \succ_{x'} y$, since otherwise M' has blocking pair $\{x', y\}$.
- $\Rightarrow M'(x') \succ_{x'} M(x')$ and hence $x' \in R$. $\square$ (Claim)

The set T of partners of the men that profit is the same in both matchings M und M': $T = \{y \mid M(y) \in R\} = \{y \mid M'(x) \in R\}$.

Proposals and Incentives

Since all men of R improve in M', all women of T deteriorate, since otherwise M has a blocking pair.

Now consider computation of M with true profile π:

- Let x be the last man of R that makes a proposal.
- This proposal goes to his final partner $y = M(x) \in T$.
- No further proposals of men in R, so all but x already matched as in M. Hence: y has rejected $M'(y)$ in a previous round.
- y can only hold a proposal of a man $x' \notin R$ with $x' \succ_y M'(y)$.
- $x' \notin R$ implies $M(x') \succeq_{x'} M'(x')$, and rejection of y implies $y \succ_{x'} M(x')$.
- Thus, $y \succ_{x'} M'(x')$ and $x' \succ_y M'(y)$.
- Since $x' \notin R$, it holds $x' \neq 1$. Thus, M' has blocking pair for π'. A contradiction. $\square$ (Theorem)

Not IC for Passive Side

DA algorithm with woman-proposal not IC for men:

	B	~~A~~	**A**
	A	**C**	B
	C	B	C
Women:	1	2	3

~~B~~	~~A~~	~~A~~
A	**C**	**B**
C	B	C
1	2	3

Men:	A	B	C
	1	3	1
	3	**1**	3
	~~2~~	2	**2**

A	B	C
1	**3**	1
~~2~~	~~1~~	3
~~3~~	2	**2**

Literature

- Arrow. A Difficulty in the concept of Social Welfare. 1950.
- Gibbard. Manipulation of Voting Schemes. A general result. Econometrica 41(4):587–601, 1973.
- Satterthwaite. Strategy-proofness and Arrow's Conditions. Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. Journal of Economic Theory 10(2):187–217. 1975.
- Moulin. On strategy-proofness and single peakedness. Public Choice 35(4):437–455, 1980.
- Ching. Strategy-proofness and "median voters". International Journal of Game Theory 26(4):473–490, 1997.
- Roth. Incentive compatibility in a market with indivisible goods. Economic Letters 9(2):127–132, 1982.
- Shapley, Scarf. On Cores and Indivisibility. Journal of Mathematical Economics 1(1):23–27, 1974.

Literature

- Roth, Postlewaite. Weak versus strong domination in a market with indivisible goods. Journal of Mathematical Economics 4(2):131–137, 1977.
- Roth, Sönmez, Ünver. Pairwise kidney exchange. Journal of Economic Theory 125(2):151–188, 2005.
- Gale, Shapley. College Admissions and the Stability of Marriage. American Mathematical Monthly 69:9–14, 1962.
- Dubins, Freedman. Machiavelli and the Gale-Shapley Algorithm. American Mathematical Monthly 88(7):485–494, 1981.
- Roth. The Economics of Matching: Stability and Incentives. Mathematics of Operations Research 7(4):617–628, 1982.

- Brandt, Conitzer, Endriss, Lang, Procaccia. Handbook of Computational Social Choice (Chapter 14)
- Nisan, Vazirani, Roughgarden, Tardos. Algorithmic Game Theory, 2007. (Chapters 9 and 10)
- Roughgarden. Twenty Lectures on Algorithmic Game Theory, 2016. (Chapters 9.4 and 10)