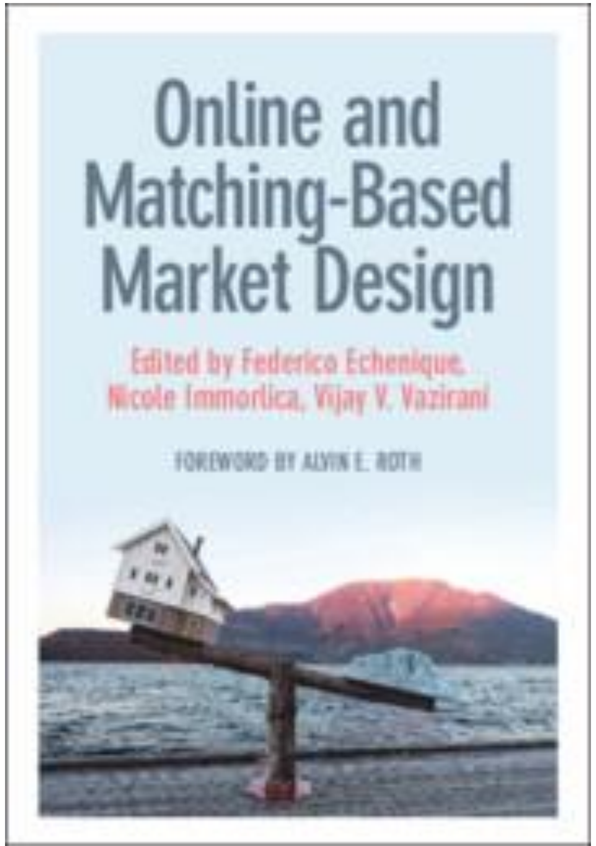
Simple proofs of important results in market design

Al Roth

Online and Matching-Based Market Design, Edited by Federico Echenique, Nicole Immorlica and Vijay V. Vazirani



Simple proofs of important results in market design

- My foreword to the book concentrated on two of the founding papers of matching and market design: **Gale and Shapley (1962)** and **Shapley and Scarf (1974)**.
- Each introduced an important algorithm: deferred acceptance (DA) and top trading cycles (TTC), respectively.
- In addition, each included fundamental theorems that could be proved very simply, sometimes essentially verbally.
- Some important subsequent theorems also allow very simple proofs (although the simple proofs were seldom the first to be discovered).

Gale and Shapley ‘62— on verbal proofs

- “The argument is carried out not in mathematical symbols but in ordinary English; there are no obscure or technical terms. Knowledge of calculus is not presupposed. In fact, one hardly needs to know how to count. Yet any mathematician will immediately recognize the argument as mathematical, while people without mathematical training will probably find difficulty in following the argument, though not be- cause of unfamiliarity with the subject matter.”

Gale and Shapley: One to one matching: The marriage model

PLAYERS: Men = $\{m_1, \ldots, m_n\}$ Women = $\{w_1, \ldots, w_p\}$

PREFERENCES (complete and transitive):

$P(m_i) = w_3, w_2, \ldots m_i \ldots$ $[w_3 >_{mi} w_2]$

$P(w_j) = m_2, m_4, \ldots w_j \ldots$

If agent k (on either side of the market) prefers to remain single rather than be matched to agent j, i.e. if $k >_k j$, then j is said to be *unacceptable* to k.

For simplicity I'm going to assume *all preferences are strict*…

An OUTCOME of the game is a *MATCHING*:

i.e. it is a function that matches men and women to one another (μ:M$\cup$W $\rightarrow$M$\cup$W)

such that w is matched to m iff m is matched to w

(w = μ(m) iff μ(w)=m),

and for all m and w each is matched either to a member of the opposite set or is single (“matched to him/herself”)

(either μ(w) is in M or μ(w) = w, and either μ(m) is in W or μ(m) = m.

Everyone’s preferences over matchings is determined entirely by their preferences over who they are matched to.

Stable matchings

A matching is

BLOCKED BY AN INDIVIDUAL k if k prefers being single to being matched with $\mu(k)$,

$(k >_k \mu(k))$

BLOCKED BY A PAIR OF AGENTS (m,w) if they each prefer each other to who they are matched to

$(w >_m \mu(m)$ and $m >_w \mu(w))$

- A matching μ is *STABLE* if it isn't blocked by any individual or pair of agents.

- NB: A stable matching is efficient, and in this simple 1-1 matching model the set of (pairwise) stable matchings equals the core (i.e. the set of outcomes not blocked by a coalition of *any* size).

Deferred Acceptance Algorithm, with men proposing (Gale-Shapley 1962)

- 1 a. Each man m proposes to his 1st choice (if he has any acceptable choices).
- b. Each woman rejects any unacceptable proposals and, if more than one acceptable proposal is received, "holds" the most preferred and rejects all others.
- k a. Any man rejected at step k-1 makes a new proposal to its most preferred acceptable mate who hasn’t yet rejected him. (If no acceptable choices remain, he makes no proposal.)
- b. Each woman holds her most preferred acceptable offer to date, and rejects the rest.
- STOP: when no further proposals are made, and match each woman to the man (if any) whose proposal she is holding.

Theorem (Gale and Shapley): A stable matching exists for every marriage market.

Proof:

1. the algorithm always stops (since no man proposes twice to the same woman)
2. the matching produced is stable.
    a. No man ever proposes to an unacceptable woman, and no woman ever holds the offer of an unacceptable man, so no individual blocks the match
    b. if some man would prefer to be matched to a woman other than his assigned mate, he must, have already proposed to her, and she has rejected him, meaning she has a man she strictly prefers, hence they cannot form a blocking pair

Theorem 2.12 (Gale and Shapley)

When all men and women have **strict** preferences, **there always exists an M-optimal stable matching** (that every man likes at least as well as any other stable matching), **and a W-optimal stable matching**. Furthermore, the matching produced by the deferred acceptance algorithm with men proposing is the M-optimal stable matching. The W-optimal stable matching is the matching μ_W produced by the algorithm when the women propose.

Proof:

Let's call w *achievable* for m if there is some stable matching (μ) at which m and w are matched (μ(m) = w).

It will be sufficient to show that no man is rejected by an achievable woman in the man-proposing deferred acceptance algorithm. (So when the algorithm stops every man is matched to an achievable mate, hence to his most preferred achievable mate.)

suppose that up to step k of the algorithm, no m has been rejected by an achievable w, and that, at step k, w rejects m (who is acceptable to w) and (therefore) holds on to some m'.

Then w is not achievable for m.

Consider a matching at which m and w are matched (μ with μ(m) = w), and at which m' is matched to some w' who is achievable for him (μ(m') achievable for m'). This matching can't be stable,: (m',w) would be a blocking pair.

So there's no first step k at which a man is rejected by an achievable woman.

Converging to a stable matching by *satisfying* blocking pairs

- if (m',w') is a blocking pair for a matching μ, a new matching ν can be obtained from μ by *satisfying* the blocking pair if m' and w' are matched to one another at ν, their mates at μ (if any) are unmatched at ν, and all other agents are matched to the same mates at ν as at μ.
- Knuth (1976) found an example with a cycle, but in which there did exist a choice of blocking pairs did converge to a stable matching. He posed an open problem: Does such a path always exist?
- Obstacle to proof: satisfying blocking pairs isn't monotone in participant welfare…

Roth and Vande Vate: finding a stable matching by satisfying one blocking pair at a time

Theorem (Roth and Vande Vate, 1990): Let μ be an arbitrary matching for a marriage problem (*M, W,* ***P***). Then there exists a finite sequence of matchings $\mu_1,\ldots,\mu_k$, such that $\mu = \mu_1,\ldots \mu_k$ is stable, and for each $i = 1,\ldots,k$-1, there is a blocking pair (m_i, w_i) for μ_i such that μ_{i+1} follows from μ_i by satisfying the pair (m_i, w_i).

μ_1 with blocking pair (m1,w1)

μ_2 satisfies (m1,w1)

A is a set of agents such that there are no blocking pairs for μ_2 contained in A, and such that μ_2 does not match any agent in A to any agent outside of A.

.

m1,w1
m2,w2
…
mk,wk
mj
wl

μ_q with blocking pair (m',w'):
at most one of m' and w' is contained in A(q) (say m'). **Add w'** matched to her most preferred such m' in A, and operate DA within the set A to rematch within the set A. Every time a proposal is 'held' a blocking pair has been satisfied.

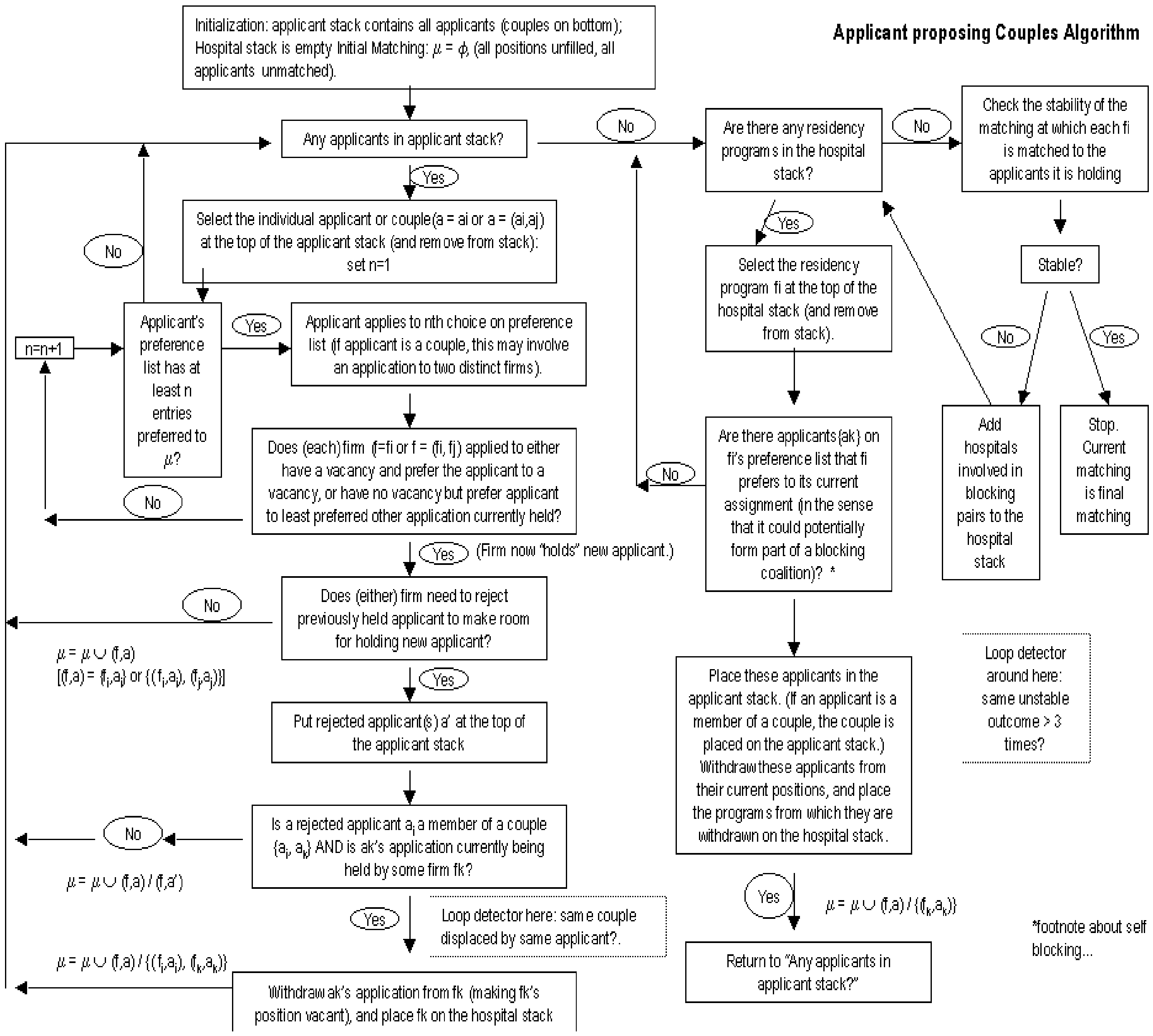
μ_{q+1}

there are no blocking pairs for μ_{q+1} contained in A, and μ_{q+1} does not match any agent in A to any agent outside of A.

Continue until all m and w are contained in A…

NRMP Match, Roth-Peranson algorithm

1. Deferred acceptance (with couples, etc.)



2. Satisfying one blocking pair at a time

An *especially* simple existence proof: Sotomayor, Marilda. (1996). "A non-constructive elementary proof of the existence of stable marriages." Games and Economic Behavior

DEFINITION (1-1 matching). The matching μ I s *simple* if, in the case a blocking pair (m,w) exists, w is single. (NB: the everyone-unmatched matching is simple).

Theorem: the set of *stable matchings* is non-empty.

Proof: the set of simple matchings is non-empty and finite. So it contains at least one element (μ*) that is Pareto optimal for the men. This matching must be stable: if not, there would be a blocking pair (m,w) with the woman unmatched .(If w is involved in more than one blocking pair, choose (m,w) with her most preferred mate.) The new matching at which that blocking pair is satisfied is <u>also simple</u> (w is no longer involved in any blocking pair), and <u>Pareto superior for men</u>. (i.e. man m is better off at the new outcome, and only his mate if any at μ*, a woman, is worse off).

Properties of the set of stable matchings (other than non-emptiness)

Constant employment theorem: In a market (*M,W, P*) with strict preferences, the set of people who are **single** is the **same for all stable matchings**.

Proof: What can we say about the *number and identity* of **men and women matched** (and hence the number and identity unmatched) at the M- and W-optimal stable matchings (μ_M and at $\mu_{W)}$?

(i.e. denoting $M_{\mu M}$ = $\mu_M(W) \cap M$,)

μ_M	$\lvert M_{\mu M} \rvert$	$\lvert W_{\mu M} \rvert$
μ_W	$\lvert M_{\mu W} \rvert$	$\lvert W_{\mu W} \rvert$

Many to one matching: “college admissions”

- a set of colleges and a set of students, identical to the marriage model except each college C_i wishes to be matched to $q_i \geq 1$ students, while each student is interested in being matched to only 1 college.
- Each student has a preference over colleges, each college has a preference over individual students, and a matching assigns each student to no more than one college, and each college C_i to no more than q_i students.
- Essentially the same deferred acceptance algorithm (with college C_i proposing at each point to its q_i most preferred students who hadn't yet rejected it in the college-proposing version, or rejecting all but the q_i most preferred applications it had received at any point of the student-proposing version) produces a stable matching defined as before, i.e. with no student-college blocking pairs defined precisely as for the marriage model.
- For each college admissions problem there’s a closely related marriage problem in which each college C with q positions is replaced with q colleges c each with 1 position…

Theorem 5.13 Rural hospitals theorem (Roth ‘86):

When preferences over individuals are strict, any hospital that does not fill its quota at some stable matching is assigned *precisely the same set of students* at every stable matching.

proof: through the constant employment/unemployment result together with side-optimal stable matchings (shown to me by Scott Kominers): Consider the related marriage problem, and suppose hospital h, with n positions only fills k of them, so the Doctor-optimal stable match is

$$\mu_D(h) = (d1, d2,\ldots\mathbf{di}\ldots dk, h,\ldots h)$$

$$\text{Let } \nu(h) = (d1, d2,\ldots\mathbf{dj}\ldots dk, h,\ldots h).$$

Suppose ν is stable. Contradiction: di, who is matched to some h’ at ν, forms a blocking pair with h at ν (involving one of it’s empty positions) so ν isn’t stable.

Matching with wages and contracts (1-to-n)

- M firms (i)
- N workers (j)
- X possible (monetary or non-monetary) contractual terms (x)
- Each worker has a strict preference ordering over pairs (i, x)
- Each firm has a strict preference ordering over sets of pairs {(j, x)} such that no worker j appears more than once in the set
- A *matching* is a set μ of triples {(i, j, x)} such that no worker j appears more than once in the set

Firms' preferences are substitutable

- Each firm i has a strict preference ordering over sets of pairs {(j, x)} such that no worker j appears more than once in the set
- The preference ordering gives rise to a choice function C_i: for any set Y of pairs {(j, x)}, choice function $C_i(Y)$ returns the firm's most preferred subset of Y
- Firm i's preferences are *substitutable* if the following condition holds:

  For any set Y of pairs {(j, x)}, for any two distinct pairs (j, x) and (j', x'), if $(j,x) \in C_i(Y)$, then $(j,x) \in C_i(Y \setminus (j',x'))$.

Stability

- Matching μ is stable if
  - No worker wants to unilaterally drop his assigned match
  - No firm wants to unilaterally drop any subset of its assigned matches
  - There is no combination of firm i, worker j, and contractual terms x such that
    - Worker j strictly prefers (i, x) to its match (possibly empty) under μ
    - Firm i strictly prefers adding (j, x) to its set of assigned matches (possibly dropping some of the assigned matches at the same time)

- Theorem (Crawford-Knoer, Kelso-Crawford, Roth, Hatfield-Milgrom)
  - A stable matching exists

New proof

- A matching μ is *simple* if it is individually rational and any blocking triple (i, j, x) involves worker j who is unmatched under μ.
- Consider a simple matching μ^* that is Pareto optimal for the firms
- It must be stable.
- Why? Suppose there is a blocking triple involving an (unassigned) worker j. Of all such triples, pick the one, (i, j, x), that is most preferred by worker j. Consider a new matching μ' that augments μ^* by adding the triple (i, j, x) to it and dropping any matches of firm i that it doesn't want to keep after this addition. Matching μ' is simple (by substitutability the firm doesn't want to add any new workers) and Pareto dominates μ^* (for the firms). Contradiction.

Strategic behavior/dominant strategies

Impossibility Theorem (Roth '82)

No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.

Remark on proof: for an impossibility theorem, one example for which *no* stable matching mechanism induces a dominant strategy is sufficient.

(A stable matching mechanism is a mechanism that produces a stable matching for *every* input of preferences)

Consider an example with 2 agents on each side

with true preferences $\mathbf{P} = (\mathbf{P}_{m1}, P_{m2}, P_{w1}, P_{w2})$ as follows:

m_1: w_1, w_2 $\qquad\qquad$ w_1: m_2, m_1

m_2: w_2, w_1 $\qquad\qquad$ w_2: m_1, m_2

In this example, what must an arbitrary stable mechanism do? I.e. what is the range of h(**P**) if h is a stable mechanism?

Given h(**P**), and the restriction that h is a stable mechanism, can one of the players x engage in a profitable manipulation by stating some $P_x' \neq P_x$ such that x prefers h(**P’**) to h(**P**)?

Shapley, Lloyd, and Herbert Scarf. "On cores and indivisibility." *Journal of mathematical economics* 1,1, 1974

- each agent initially possesses a single unit of an indivisible good, a "house." Agents have preferences over all the houses, which can be traded. But no money can be used: trades have to be house swaps, among cycles of any length.
- S&S modeled the game as a game V in characteristic function form without sidepayments:
- Theorem (Shapley and Scarf 1974): ***V is a balanced game; hence the market in question has a nonempty core.***
- And then, literally as an afterthought:
- "After the proof in sect. 4 had been discovered, David Gale pointed out to the authors a simple constructive method for finding competitive prices in this market, and hence a point in its core."

Gale’s top trading cycles (TTC) algorithm

- Each agent points to her most preferred house (and each house points to its owner). There is at least one cycle in the resulting directed graph (a cycle may consist of an agent pointing to her own house.) In each such cycle, the corresponding trades are carried out and these agents are removed from the market together with their assignments.
- The process continues (with each agent pointing to her most preferred house that remains on the market) until no agents and houses remain.
- Theorem (Shapley and Scarf 1974): the allocation x produced by the top trading cycle algorithm is in the core (no set of agents can all do better than to participate)
- Proof: try to construct a coalition of agents that could have done better by trading among themselves…

Let $R \subseteq N$, and define a *top trading cycle for R* to be any set S, $\phi \subset S \subseteq R$, whose s members can be indexed in a cyclic order:

$$S = \{i_1, i_2, \ldots, i_s = i_0\},$$

in such a way that each trader i_ν likes the $i_{\nu+1}$'st good at least as well as any other good in R. It is evident that every nonempty $R \subseteq N$ has at least one top trading cycle, for we may start with any trader in R and construct a chain of best-liked goods that eventually must come back to some earlier element.[8]

Using this idea, we can partition N into a sequence of one or more disjoint sets:

$$N = S^1 \cup S^2 \cup \ldots \cup S^p,$$

by taking S^1 to be any top trading cycle for N, then taking S^2 to be any top trading cycle for $N - S^1$, then taking S^3 to be any top trading cycle for $N - (S^1 \cup S^2)$, and so on until N has been exhausted. We can now construct a payoff vector x by carrying out the indicated trades within each cycle. That is, if $i = i_\nu^j \in S^j$, then x_i is i's utility for the good of trader $i_{\nu+1}^j$. We assert (1) that x, so constructed, is in the core, and (2) that a set of competitive prices exists for x.

To establish (1), let S be any coalition. Consider the first j such that $S \cap S^j \neq \phi$. Then we have

$$S \subseteq S^j \cup S^{j+1} \cup \ldots \cup S^p = N - (S^1 \cup \ldots \cup S^{j-1}).$$

Let $i \in S \cap S^j$. Then i is already getting in x the highest possible payoff available to him in S. No improvement is possible for him, unless he deals outside of S. Hence S cannot strictly improve, and it follows that x is in the core.

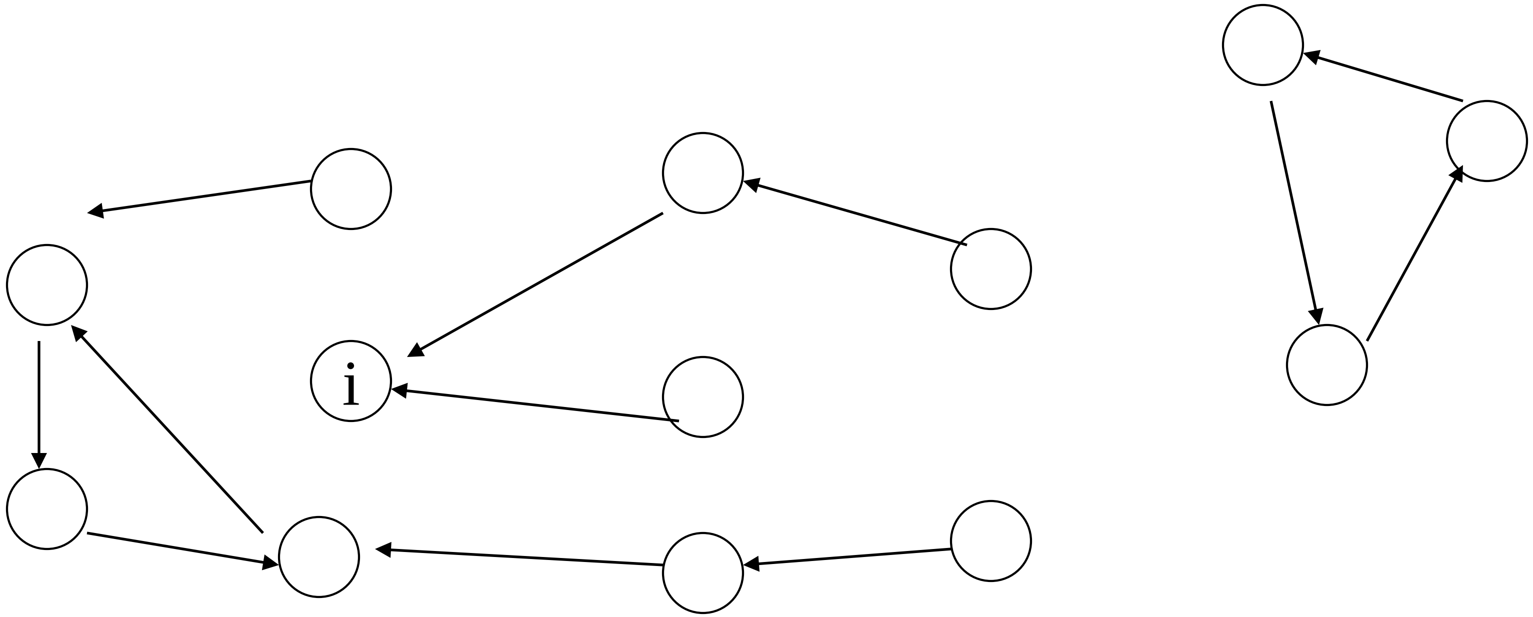
Theorem (Roth and Postlewaite 1977): When preferences are strict, the *order* in which cycles are removed doesn’t change the outcome.

- Theorem (Roth ’82): if the top trading cycle procedure is used, it is a dominant strategy for every agent to state his true preferences.

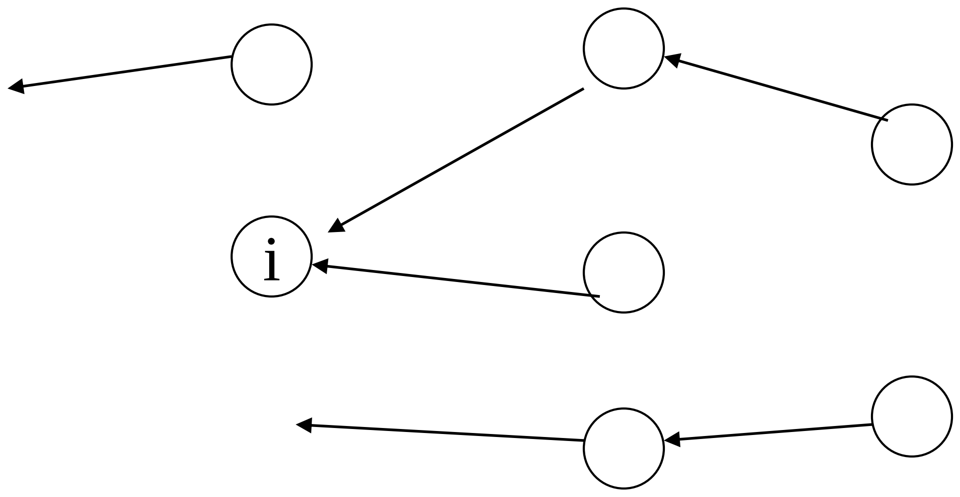
Strategyproofness-Exposing Mechanism Descriptions by Yannai A. Gonczarowski, Ori Heffetz, Clayton Thomas, 27 Sep 2022

- Description of TTC to each participant:
- TTC operates as if we will consult YOUR preferences only after removing all cycles that do not contain you (which contain options that you can never have).
- So, when we look at your preferences, you will form a cycle with whatever item you point to as your first choice among the items that are available to you.

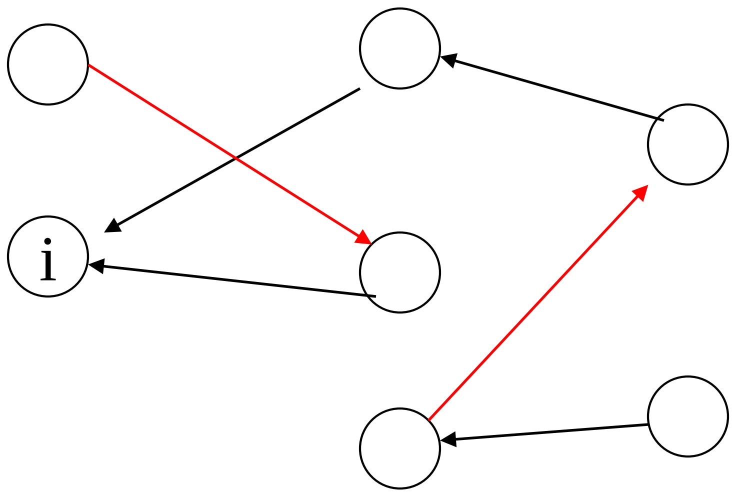
Theorem (Roth ’82): if the top trading cycle procedure is used, it is a dominant strategy for every agent to state his true preferences.



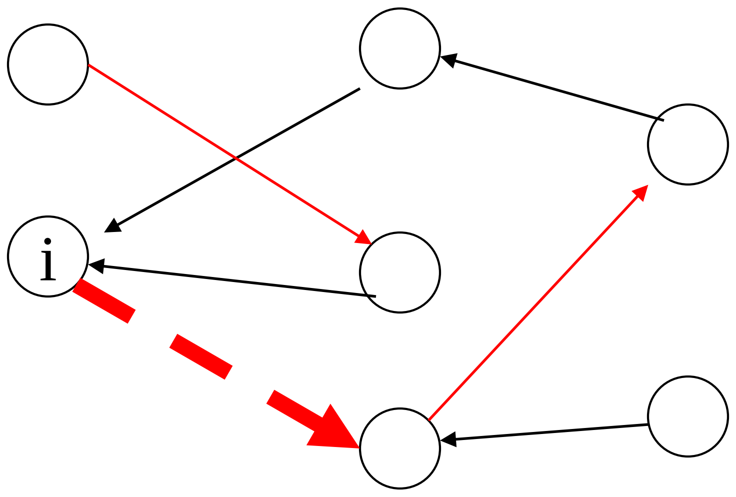
The cycles that don’t include i leave the market (regardless of i’s preferences), but i’s choice set (the chains pointing to i) remains, and can only grow



When there are no more cycles without i, all chains end at i. So i's preferences will be used simply to determine which of the items that are available to him he likes best (and will receive)



When there are no more cycles without i, all chains end at i. So i's preferences will be used simply to determine which of the remaining items he likes best (and will receive)



i points to the most preferred of the objects he could possibly get, and receives it, (as part of a cycle of length 4).

John von Neumann Theory Prize (INFORMS)

2022 Winner(s)

- **Vijay Vazirani**, University of California, Irvine

1983

WINNER(S)

- **Herbert E. Scarf**, Yale University

1981

WINNER(S)

- **Lloyd S. Shapley** , University of California

1980

WINNER(S)

- **David Gale**
- **Harold W. Kuhn**,
- **Albert W. Tucker**