House Allocation Problems (One-Sided Matching)

Fuhito Kojima[1]

April 12, 2018

[1]Stanford University.

House Allocation Problem

The house allocation problem: introduced by Hylland and Zeckhauser (1979). Suggested reading is Sonmez and Unver (2009).[2]

A house allocation problem is tuple $(A, H, \succ)$ where

1. A is a set of agents
2. H is a set of goods ("houses")
    - Assume (for now) goods are mutually distinct and $|H| = |A|$; almost everything generalizes easily.
3. Each agent $a \in A$ has strict preferences over houses, $\succ_a$ (weak preferences are denoted $\succeq_a$). $\succ = (\succ_a)_{a \in A}$ is the preference profile.

Possible applications include on-campus housing, organ allocation, office allocation and (some) student placement problems.

[2] *Matching, Allocation, and Exchange of Discrete Resources*, forthcoming, Handbook of Social Economics.

Matching μ is a function specifying who gets what good.

$\mu(a)$ is the house that agent a receives in μ.

A matching μ is **Pareto-efficient** if there is no other matching ν such that

1. $\nu(a) \succeq_a \mu(a)$ for every agent a, and
2. $\nu(a) \succ_a \mu(a)$ for some agent a.

(Terminology: If such an alternative matching ν exists, then we say ν Pareto-dominates μ.)

Mechanism

A (deterministic) mechanism is a rule that assigns a matching for each preference profile. $\varphi(\succ)$ is the matching when agents report $\succ$ under mechanism φ.

A mechanism φ is **strategy-proof** if telling the true preferences is a weakly dominant strategy (a best action no matter what others do) for everyone.

A mechanism is **Pareto-efficient** if $\varphi(\succ)$ is Pareto-efficient for every preference profile $\succ$.

Serial Dictatorship

A serial dictatorship mechanism (priority mechanism) specifies an order and lets the first agent receive her favorite good, the next agent his favorite good among the remaining ones, and so on.

Formally, mechanism π^f first specifies a priority ordering function f, where $f(i)$ is the agent with the i^{th} priority. Then the first agent $f(1)$ receives her favorite good, the next agent $f(2)$ his favorite good among the remaining ones, and so on.

The serial dictatorship mechanism is very easy to implement: decide the order (randomly, or using some existing priority such as seniority) and let applicants choose according to the order.

Serial dictatorship is used in many applications (with some variations: discussed later): office allocation for professors, NYC school choice system and Columbia, Harvard, and Stanford housing allocation (Pathak 2008, Kojima and Manea 2010, Che and Kojima 2010), just to name a few.

In addition, serial dictatorship has several good properties (perhaps, these properties are part of reason for wide use of this mechanism).

Theorem

Serial dictatorship is Pareto-efficient.

Proof sketch: Prove the claim by contradiction: Suppose there is matching ν that Pareto-dominates $\mu := \pi^f(\succ)$. Consider the agent $a = f(i)$ with the highest priority who obtains a strictly better good in ν than in $\mu = \pi^f(\succ)$. Then we know

1. $\nu(a) = \mu(b)$ for some $b = f(j)$ with $j < i$: because that is why a cannot get $\nu(a)$ in π^f, and
2. $\nu(b) = \mu(b)$: because (i) $\nu(b) \succeq_b \mu(b)$ by assumption and (ii) $\nu(b) \succ_b \mu(b)$ cannot be true (why?)

This is a contradiction.

Theorem

Serial dictatorship is strategy-proof.

Proof sketch: Let f be the priority order.

The first agent $f(1)$ of the priority order obtains the favorite good for her when she tells the truth, so she has no incentives to lie.

The second agent $f(2)$ of the order gets her favorite good among the remaining goods, so she has no incentives to lie.

... and so on.

Group Strategy-Proofness

Actually, serial dictatorship has even a stronger incentive property.

Consider the possibility that a group of agents collude and misreport preferences jointly. Can we assure a mechanism to be immune to such joint manipulations?

Formally, let $\succ_B = (\succ_a)_{a \in B}$ and $\succ_{-B} = (\succ_a)_{a \in A \setminus B}$.
Mechanism φ is **group strategy-proof** if there is no group of agents $B \subset A$ and preferences $\succ'_B$ such that

1. $\varphi(\succ'_B, \succ_{-B}) \succeq_a \varphi(\succ_B, \succ_{-B})$ for all $a \in B$ and
2. $\varphi(\succ'_B, \succ_{-B}) \succ_a \varphi(\succ_B, \succ_{-B})$ for at least one $a \in B$.

In words, a mechanism is group strategy-proof if no group of agents can jointly misreport preferences in such a way to make some member strictly better off while no one in the group is made worse off.

Theorem

Serial dictatorship is group strategy-proof.

Proof is omitted (Exercise): An intuition is that the mechanism only uses preference information of an agent when it is her turn to choose, so the best she can do is to report her true favorite remaining good as her favorite choice. Whenever she does so, the subsequent part of the mechanism proceeds exactly as when she reports true preferences.

Housing Market (Shapley and Scarf 1974)

A **housing market** is $((a_k, h_k)_{k \in \{1,\ldots,n\}}, \succ)$ such that

1. $\{a_1, \ldots, a_n\}$ is a set of agents and $\{h_1, \ldots, h_n\}$ is a set of houses, where agent a_k owns house h_k.
2. Each agent a has strict preferences $\succ_a$ over houses.

Mechanism

Most terminologies are defined in the same manner as in house allocation problems. Just to review some of them:

Matching μ is a function specifying who gets what good: $\mu(a)$ is the house that agent a receives in μ.

As before, a **mechanism** is a rule φ that prescribes matching $\varphi(\succ)$ when agents report preference profile $\succ$.

A mechanism φ is **strategy-proof** if telling the true preferences is a dominant strategy (a best action no matter what others do) for everyone.

A mechanism is **Pareto-efficient** if $\varphi(\succ)$ is Pareto-efficient for every preference profile φ.

Stability Concept: The Core

As in the case of two-sided matching, a prescribed matching is sustainable only if no (groups of) agents can profitably deviate from it. The concept of the **core** is central.

Matching μ is in the **core** if there is no coalition of agents B and a matching ν such that

1. For any $a \in B$, $\nu(a)$ is the initial house of some $b \in B$, and
2. $\nu(a) \succeq_a \mu(a)$ for all $a \in B$ and $\nu(a) \succ_a \mu(a)$ for some $a \in B$.

Remark: Core is a central solution concept in a branch of game theory, called cooperative game theory. In many-to-one two-sided matching, stability is equivalent to core.

A matching is **individually rational** if every agent obtains a house that is at least as good as her initial house.

It is immediate to see that

1. Any core matching is individually rational (consider a one-person "coalition" $B = \{a\}$),
2. Any core matching is Pareto efficient (consider $B = A$).

There always exists a core matching

Theorem (Shapley and Scarf 1974)

There exists a core matching for any housing market.

This is a fundamental result in the housing market, and is similar in spirit to the existence theorem by Gale and Shapley (1962) for the two-sided matching.

Gale's Top Trading Cycles (TTC) algorithm

The following proof is attributed to Gale, based on **Gale's Top Trading Cycles (TTC) algorithm**.

Step 1: Each agent points to the owner of his/her first choice house. There exists at least one cycle and no cycles intersect (why?). Remove all the cycles and assign each agent in a cycle the house whose owner he or she is pointing to.

Step t: Each agent points to the owner of his/her first choice house among the remaining ones. There exists at least one cycle and no cycles intersect. Remove all the cycles and assign each agent in a cycle the house whose owner he or she is pointing to.

Example of Gale's TTC algorithm

Let $A = \{a_1, a_2, a_3, a_4\}$, and

a_1	a_2	a_3	a_4
h_2	h_4	h_2	h_4
$\vdots$	h_1	h_1	$\vdots$
	$\vdots$	h_3	$\vdots$
		$\vdots$	

A more elaborate example can be found in the survey by Sonmez and Unver (Example 1).

Equivalent Representation of the TTC algorithm

For future purposes, the following equivalent description of TTC is useful; for now, use whichever you like better!

Step 1: Each agent points to his/her first choice house and each house points to its initial owner. There exists at least one cycle and no cycles intersect. Remove all the cycles and assign each agent in a cycle the house he or she is pointing to.

Step t: Each agent points to his/her first choice house among the remaining ones and each house points to its initial owner. There exists at least one cycle and no cycles intersect. Remove all the cycles and assign each agent in a cycle the house he or she is pointing to.

Proof of the Theorem

Recall we wanted to show:

Theorem (Shapley and Scarf 1974)

There exists a core matching for any housing market.

Let μ be the resulting matching from TTC. Suppose there is a coalition B that deviates profitably by matching ν.

Consider the subset of agents in B who strictly prefer their allocation under ν to those in μ, and let a be an agent who is matched first among this subset in the TTC algorithm.

Then $\nu(a)$ is owned by an agent $b \in B$ who is removed by the TTC algorithm in a strictly earlier step (say cycle C_l).

Then, b obtains a house of $b' \in B \cap C_l$ both in ν and μ, . . . , and $b^* \in B \cap C_l$ obtains $\nu(a)$ both at ν and μ (why?). So contradiction.

Review

- House Allocation problem
  - Pareto efficiency, strategy-proofness
  - Serial dictatorship
- Housing market (a.k.a. house exchange) problem
  - Pareto efficiency, individual rationality, strategy-proofness, core
  - Gale's top trading cycles (TTC) mechanism

The Uniqueness of the core matching

Is there any other matching in the core?

Theorem (Roth and Postlewaite 1977)

The matching produced by Gale's TTC algorithm is the unique core matching.

Proof of The Theorem

We have already seen that the TTC algorithm finds a core matching, μ, so we will show there is no other core matching.

Consider an arbitrary matching $\nu \neq \mu$, and fix agent a to be one of the first agents with $\nu(a) \neq \mu(a)$ (according to the order of being matched in TTC).

Let C_l be the set of agents that form a cycle that includes a. Then, any b who are matched before C_l satisfies $\nu(b) = \mu(b)$.

By construction of TTC, $\mu(b) \succeq_b \nu(b)$ for all $b \in C_l$ (because, in ν, all preferred goods are allocated to those who are matched before C_l).

Moreover, since $\nu(a) \neq \mu(a)$ and $a \in C_l$, we have $\mu(a) \succ_a \nu(a)$.

Since, for any $b \in C_l$, $\mu(b)$ is an initial house owned by some other agent in C_l, these facts imply that C_l can profitably deviate from ν by μ, so ν is not in the core.

TTC as a mechanism

As a mechanism, does TTC have good incentive properties?

Theorem (Roth 1982)

The TTC algorithm is strategy-proof.

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

Axiomatic characterization

TTC has good properties. Are there other mechanisms satisfying nice properties like TTC?

Such an approach is called **(axiomatic) characterization**: Find the set of properties (also called axioms) that are exactly necessary and sufficient for the mechanism to be in a certain class of mechanisms.

Theorem (Ma 1994)

A mechanism is strategy-proof, Pareto-efficient and individually rational if and only if it is TTC.

House Allocation with Existing Tenants

Based on Abdulkadiroglu and Sonmez (1999), motivated by on-campus housing practices.

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.

Each agent has strict preferences over houses (and being unmatched).

There are houses owned by existing tenants and vacant houses.

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

1. Housing market when all agents are existing tenants and there is no vacant house.
2. House allocation problem when all agents are newcomers.

Random Serial Dictatorship with Squatting Rights

Used in undergrad housing at Carnegie-Mellon, Duke, Michigan, Northwestern and Pennsylvania, etc.

1. Each existing tenant decides whether they want to participate in the housing lottery or keep the current house. Those who decides to keep their houses are assigned the current houses. All other houses become available for assignment in later steps.

2. An ordering of agents is decided. The ordering may be uniformly random or may favor some subgroup of agents (for example, seniors over juniors).

3. Serial dictatorship is applied to all available houses and agents (except for existing tenants already assigned their current houses).

Something is wrong with random serial dictatorship with squatting rights.

Existing tenants are not guaranteed to get at least as good a house as their current house. Individually irrational!
$\rightarrow$ Some existing tenants may not want to enter the lottery even if they want to move.

This may result in loss of gains from trade, and the resulting matching may not be Pareto efficient.

Some good properties we want for house allocation mechanisms:

1. Pareto efficiency
2. Strategy-proofness
3. Individual rationality

Also, recall that there were good mechanisms in special cases:

1. Serial dictatorship (house allocation problem)
2. Gale's top trading cycles (housing markets)

Can we find mechanisms with the above good properties? Does our knowledge on SD and TTC help?

The "YRMH-IGYT" mechanism

The **you request my house - I get your turn (YRMH-IGYT)** mechanism with ordering f is a generalization of SD.

1. Let the agent with the top priority receive her first choice good, the second agent his top choice among the remaining goods and so on, until someone requests the house of an existing tenant.
2. If the existing tenant whose house is requested has already received a house, then proceed the assignment to the next agent. Otherwise, insert the existing tenant at the top of the priority order and proceed with the procedure.
3. If at any step a cycle forms, the cycle is formed by existing tenants $(a_1, \dots, a_k)$ where a_1 points to a house of agent a_2, who points to the house of a_3, and so on. In such a case assign these houses by letting them exchange (a bit like TTC!), and then proceed with the algorithm.

Example of YRMH-IGYT mechanism

Let $A = \{a_1, a_2, a_3, a_4\}$, and

a_1	a_2	a_3	a_4
h_2	h_4	h_2	h_4
$\vdots$	h_1	h_1	$\vdots$
	$\vdots$	h_3	$\vdots$
		$\vdots$	

Let $f = (a_1, a_2, a_3, a_4)$ be the ordering of agents and,

1. a_1 and a_4 are existing tenants (with h_1 and h_4 occupied).
2. a_1, a_2, a_4 are existing tenants.
3. a_1, a_2 are existing tenants.

Example 2 from the survey by Sonmez and Unver is a more elaborate example.

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 simply does not happen, so the mechanism coincides with serial dictatorship straightforwardly.
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.
3. Indeed, we can think of YRMH-IGYT 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. So we sometimes call the mechanism TTC as well.

Properties of YRMH-IGYT mechasnism

Theorem

Any TTC (YGMH-IGYT) mechanism is individually rational, strategy-proof, and Pareto-efficient.

Proof sketch: As TTC (YGMH-IGYT) is a common generalization of serial dictatorship and Gale's TTC, Pareto efficiency and strategy-proofness are "inherited" from these mechanisms (the proof is quite similar).

Also, individual rationality is inherited from Gale's TTC, and can be understood as follows: Whenever some agent points to a house of an existing tenant, she is promoted to the top of the priority. Whenever her top choice at this stage is her own house she can keep it by forming a "cycle" composed of herself and her house.

Characterization

Based on Sonmez and Unver, "House Allocation with Existing Tenants: A Characterization."

They consider variable population, that is, how assignment changes when some agents (and possibly their houses) are removed from the problem.

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

Theorem

A mechanism is Pareto efficient, individually rational, strategy-proof, weakly neutral, and consistent if and only if it is a TTC (YRMH-IGYT) mechanism.

So, there is a sense in which TTC (YRMH-IGYT) are the "right" mechanisms.

Experimental Evidence

Based on Chen and Sonmez (2002).
They conduct laboratory experiments, comparing random serial dictatorship with squatting rights and YRMH-IGYT algorithm.

1676 THE AMERICAN ECONOMIC REVIEW DECEMBER 2002

TABLE 4—EFFICIENCY—ORIGINAL ENVIRONMENT
(Standard Errors in Parentheses)

Mechanisms	Sessions	Observed efficiency	Expected efficiency with 1 million lotteries	Recombinant estimation of mean efficiency
RSD (original)	R_o1	0.673	0.693 (0.031)	$\hat{\mu}_{rsd} = 0.754$ (0.020)
	R_o2	0.737	0.741 (0.023)	$\sigma^2 = 0.00358$
	R_o3	0.836	0.849 (0.027)	$\varphi = 0.000203$
	R_o4	0.661	0.698 (0.027)	
	R_o5	0.750	0.802 (0.033)	
RSD (large)	R_l1	0.743	0.742 (0.007)	$\hat{\mu}_{rsd} = 0.742$ (0.001)
	R_l2	0.737	0.746 (0.013)	$\sigma^2 = 0.000331$
				$\varphi = 0.0000000542$
TTC (original)	T_o1	0.924	0.934 (0.061)	$\hat{\mu}_{ttc} = 0.889$ (0.020)
	T_o2	0.743	0.802 (0.044)	$\sigma^2 = 0.00332$
	T_o3	0.901	0.871 (0.050)	$\varphi = 0.000157$
	T_o4	0.930	0.911 (0.021)	
	T_o5	0.877	0.890 (0.025)	
TTC (large)	T_l1	0.913	0.903 (0.010)	$\hat{\mu}_{ttc} = 0.875$ (0.006)
	T_l2	0.837	0.830 (0.012)	$\sigma^2 = 0.000692$
				$\varphi = 0.00000107$

Guillen and Kesten (2010) show that a house allocation mechanism used in MIT is a version of DA. Then they conduct laboratory experiments.

TABLE 2. EFFICIENCY (Standard Errors in Brackets)

Mechanisms	Group	Observed efficiency	Expected efficiency	Recombinant est of mean efficienc
NH4	NH4-1	1	1 (0)	
	NH4-2	.8701299	.8738139 (0.025)	$\hat{\mu} = 0.893$ (0.092
	NH4-3	.9047619	.8874434 (0.028)	$\sigma^2 = 0.0045$
	NH4-4	.8528138	.8445917 (0.028)	$\varphi = 0.0036$
	NH4-5	.8138528	.8050089 (0.033)	
TTC	TTC-1	.8095238	.8166418 (0.025)	$\hat{\mu} = 0.813$ (0.054
	TTC-2	.8398268	.8742235 (0.017)	$\sigma^2 = 0.0018$
	TTC-3	.7705628	.7692916 (0.022)	$\varphi = 0.0012$
	TTC-4	.8354979	.8287112 (0.021)	
	TTC-5	.7532467	.7902112 (0.020)	

TTC in School Choice

Serial dictatorship and Gale's TTC are good solutions in house allocation and housing markets, respectively.

TTC by AS is a common generalization of SD and Gale's TTC to house allocation with existing tenants, with good properties.

The school choice problem can be regarded as a further generalization, and so a generalization of TTC proves to be a promising market design.

Summary

We have looked at

1. House Allocation Problem,
2. Housing Market,
3. House Allocation with Existing Tenants

For each problem, we want the mechanism to satisfy good properties, such as Pareto efficiency, strategy-proofness, individual rationality, core, and so on.

Serial dictatorship, Gale's TTC, and TTC (YRMH-IGYT) are good mechanisms in each of the problems.

Methodologically, we have learned axiomatic characterization.

A preview of the next topic; Kidney Exchange

We will learn kidney exchange next, using the theory we have learned so far.

Mathematically, the problem of kidney exchange is quite similar to house allocation with existing tenants.

1. Kidney patients want to obtain a kidney for transplantation.
2. There are kidneys from diseased donors as well as “good Samaritan donors” (similar to “vacant houses”).
3. Some kidney patients have willing but incompatible donors (similar to “existing tenants”).

However, there are some medical and logistical constraints that may make a direct application of existing theories impossible. This fact motivates new theories to be explored.

Reading for the next topic

We will learn kidney exchange next, using the theory we have learned so far.

Look at chapter 3 of the survey,
Sonmez and Unver, "Matching, Allocation, and Exchange of Discrete Resources," http://www2.bc.edu/ unver/

This is a good introduction to kidney exchange. If you want more detail, read original papers by Roth, Sonmez and Unver such as
"Kidney Exchange," Quarterly Journal of Economics 2004,
"Pairwise Kidney Exchange," Journal of Economic Theory, 2005,
"Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatibility-Based Preferences," American Economic Review, 2007

Kidney Exchange

Fuhito Kojima

April 12, 2018

[1]Stanford University.

Kidney Exchange

Transplant is an important treatment of serious kidney diseases.

Over 90,000 patients are on waiting lists for kidney in the U.S.

In 2011, there were

1. 11,043 transplants from diseased donors,
2. 5,771 transplants from living donors, while
3. 4,697 patients died while on the waiting list (and 2,466 others were removed because they were "too sick to transplant").

Kidneys cannot be bought and sold

Buying and selling kidneys is illegal in the U.S. as well as many other countries.

Section 301 of the National Organ Transplant Act states:

"it shall be unlawful for any person to knowingly acquire, receive or otherwise transfer any human organ for valuable consideration for use in human transplantation."

Given that constraint, donation is the most important source of kidneys.

There are two sources of donation:

1. Deceased donors: A centralized mechanism has been used for allocation of deceased donor kidneys.
2. Living donors: Living donors usually come from friends or relatives of a patient (because the monetary transaction is prohibited). Live donation has been increasing recently.

Donor Types	2008	1998	1988
All donors	10,920	9,761	5,693
Deceased donors	5,992	5,339	3,876
Live donors	4,928	4,422	1,817

Table: Number of donors by donor types. Data obtained at http://www.optn.org/

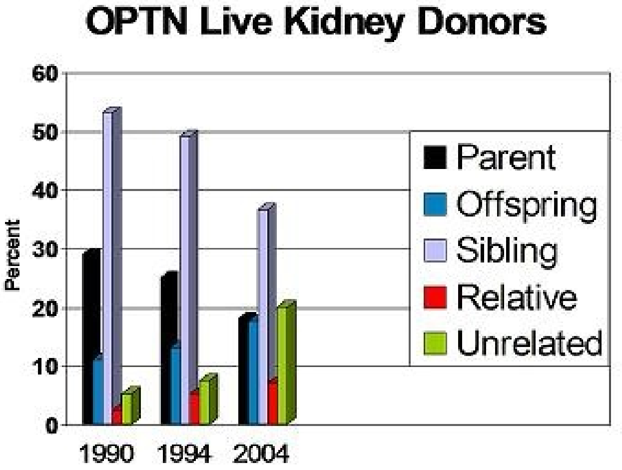

Figure: Live donors by relationship to patients (thanks to Al Roth for providing the graph).

For a successful transplant, the donor kidney needs to be **compatible** with the patient.

1. Blood type compatibility: There are four blood types, O, A, B and AB.
   - O type patients can receive kidneys from O type donors
   - A type patients can receive kidneys from O or A type donors
   - B type patients can receive kidneys from O or B type donors
   - AB type patients can receive kidneys from donors of any blood type (that is, O, A, B or AB)
2. There is another compatibility issue around some proteins called HLA Tissue Compatibility.

A problem with transplant from live donors: transplant is carried out if the donor kidney is compatible with the patient. Otherwise the willing donor goes home and the patient cannot get transplant.

Is there any way to increase the number and quality of transplant?

A **paired exchange** (aka paired donation): Take a look at the web page of Alliance for Paired Donation at
http://www.paireddonation.org/anim.htm

A **list exchange**: "Match" one incompatible patient-donor pair and the deceased donor waiting list. That is,

- 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, Sonmez and Unver (economists) as well as doctors design the clearinghouse.

Potential issues include

1. Efficiency (Pareto efficiency; maximizing number of transplantation)
2. Fairness
3. Incentives (Strategy-proofness)

Do patients and doctors behaves strategically? Here is one example indicating they do.

A news report by Reuters (2003-7-29)

> *Three Chicago hospitals were accused of fraud by prosecutors on Monday for manipulating diagnoses of transplant patients to get them new livers.*
> *Two of the institutions paid fines to settle the charges.*
> *"By falsely diagnosing patients and placing them in intensive care to make them appear more sick than they were, these three highly regarded medical centers made patients eligible for liver transplants ahead of others who were waiting for organs in the transplant region," said Patrick Fitzgerald, the U.S. attorney for the Northern District of Illinois.*

The model

A kidney exchange model is composed of

1. A set of donor-patient (kidney-transplant) pairs,
2. A preference over all kidneys and "high priority in the waitlist" (in exchange of donating a kidney.)

A **matching** is a function that specifies which patient obtains which kidney (or waitlist). We assume that the wait list can be matched with any number of patients.

Design 1 (RSU 2004)

Roth, Sonmez and Unver (2004) assume that

1. There is no limit on the number of pairs participating in one exchange.
2. Patients have strict preferences over compatible kidneys and the waitlist: Some justification by Opelz (1997). He shows that, in his data, increase in the number of HLA mismatch decreases the likelihood of kidney survival. Other characteristics such as body size and donor age affect kidney survival.

Connection with House allocation with Existing Tenants

With the assumption of RSU (2004), the kidney exchange problem is mathematically very similar (almost identical!) to **house allocation with existing tenants** (Abdulkadiroglu and Sonmez 1999) :

Kidney Exchange	House allocation with existing tenants
patient	agent (tenant)
donor	occupied house
waitlist	vacant houses

Also from the correspondence, we could include other features:

1. Good Samaritan donors (donors who give kidneys although they are not paired with an incompatible patient) can be treated as vacant houses,
2. Patients without a paired donor can be treated as "newcomers."

One difference is that the waitlist w can be matched to multiple patients, but this can be accommodated straightforwardly.

YRMH-IGYT (TTC)

Because the mathematical structure is very similar to house allocation with existing tenants, a promising solution is

YRMH-IGYT mechanism (a.k.a. TTC mechanism).

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.
2. 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.
3. If at any step a cycle forms, assign these kidneys by letting them exchange, and then proceed with the algorithm.

In this environment, TTC is a big winner:

Theorem

The TTC mechanism is

1. *Pareto efficient*
2. *strategy-proof,*
3. *individually rational.*

Remarks: (1) In RSU they consider many variants of the TTC mechanism, which they call TTCC (Top Trading Cycles and Chains) mechanisms. The one which they pick as the winner corresponds to TTC (this point is formally pointed out by Krishna and Wang 2007).

(2) Sonmez and Unver (2008) give axiomatic characterization of the TTC mechanism, thus adding one more justification of using this mechanism.

Design 2 (RSU 2005)

RSU discussed the design with doctors, who say

1. Only pairwise exchanges (or at least short ones) may be possible (at least initially)
    - because 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 of a market with

1. N: the set of incompatible donor-patient pairs
2. A list of who are "mutually compatible" with whom.
3. An (individually rational) matching μ is bilateral, i.e., $\mu(i) = j$ if and only if $\mu(j) = i$.
4. The model is like a "roommate problem", but importantly, agents have "dichotomous" preferences, that is, a kidney is just "good" or "bad" for the agent.
5. Desiderate such as Pareto efficiency and strategy-proofness are defined as before.
6. Additional policy question: Does Pareto efficient matching maximize the number of exchanges?

Efficiency Property

Theorem (Lemma 1 of RSU (2005))

All Pareto optimal matchings match the same number of pairs.

The set of matchings forms a mathematical object called "matroid."

This claim does not hold if larger exchanges are possible.

The priority mechanism

Consider the following **priority mechanism (serial dictatorship)**:

1. Order pairs in some way (ordering could be random or favor waiting time, etc.),
2. If there is any matching in which the top priority pair is matched, then match that pair. Otherwise, skip that pair.
3. 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.

   ⋮

4. Match the k^{th} 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.

The priority mechanism

Theorem (RSU 2005)

The priority mechanism is Pareto efficient and strategy-proof.

The reason for this claim is very intuitive (why?).

RSU (2005) consider stochastic mechanisms as well, to get a "fair" solution.

Design 3 (RSU 2007)

An exchange involving more than two pairs may be difficult, but may not be infeasible.

Still, the logistical constraints are likely to matter: two-way (pairwise) exchanges are easier than three way exchanges, and three-way exchanges are easier than four-way exchanges, and so on.

How much efficiency gain can we obtain through larger exchanges?

Three-way Exchanges Can Add A Lot of Transplants

Example: A pair is denoted as type x-y if the patient and donor are ABO blood-types x and y, respectively. Consider a population composed of

1. O-B, O-A, A-B, A-B, B-A (blood-type incompatible),
2. A-A, A-A, A-A, B-O (positive crossmatch).

Assume there is is no tissue rejection between patients and other patients' donors.

1. If only two-way exchanges are possible:
   (A-B,B-A), (A-A,A-A), (O-B,B-O).
2. 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.

Four-way Exchanges Can Add Only A Little

Example: Consider a population composed of

1. O-A, A-B, B-AB (blood-type incompatible),
2. AB-O (positive crossmatch).

Assume there is no tissue rejection between patients and other patients' donors.

1. If only two-way and three-way exchanges are possible: (O-A, A-B, AB-O).
2. If four-way exchanges are also feasible: (AB-O, O-A, A-B, A-AB).

However, a situation like the above example is rare, because

1. AB-type is rare (only 3.85 percent in U.S. population),

Patient ABO Blood Type	Frequency
O	48.14 %
A	33.73 %
B	14.28 %
AB	3.85 %

2. AB-O incompatible pair above should come from tissue incompatibility, not blood-type incompatibility.

Given that four-way exchanges are even more difficult than three-way, it may not be the first priority ...

Question: What about 5-way or larger exchanges? Are large exchanges worth the trouble?

For theoretical analysis, RSU make a few assumptions.

1. No patient is tissue-type incompatible with another patient's donor.
2. Patient-donor pairs of types O-A, O-B, O-AB, A-AB, and B-AB are on the "long side" of the exchange in the sense that at least one pair of each type remains unmatched in each feasible set of exchanges.
3. $\#(\text{A-B}) > \#(\text{B-A})$.
4. There is either no type A-A pair or there are at least two of them. The same is also true for each of the types B-B, AB-AB, and O-O.

Theorem (RSU 2007)

Consider a patient population for which Assumptions 1, 2, 3 and 4 hold and let μ be any maximal matching (when there is no restriction on the size of the exchanges). Then there exists a maximal matching ν that consists only of two-way, three-way, and four-way exchanges, under which the same set of patients get transplant as in matching μ.

The Theorem means that four-way exchanges suffice: any maximal matching can be achieved just using two-way, three-way and four-way exchanges.

Simulation

TABLE 2

Pop. size	Method	Type of exchange: Two-way	Two-way, three-way	Two-way, three-way, four-way	No constraint
	Simulation	8.86 (3.4866)	11.272 (4.0003)	11.824 (3.9886)	11.992 (3.9536)
$n = 25$	Upperbound 1	12.5 (3.6847)	14.634 (3.9552)	14.702 (3.9896)	
	Upperbound 2	9.812 (3.8599)	12.66 (4.3144)	12.892 (4.3417)	
	Simulation	21.792 (5.0063)	27.266 (5.5133)	27.986 (5.4296)	28.09 (5.3658)
$n = 50$	Upperbound 1	27.1 (5.205)	30.47 (5.424)	30.574 (5.4073)	
	Upperbound 2	23.932 (5.5093)	29.136 (5.734)	29.458 (5.6724)	
	Simulation	49.708 (7.3353)	59.714 (7.432)	60.354 (7.3078)	60.39 (7.29)
$n = 100$	Upperbound 1	56.816 (7.2972)	62.048 (7.3508)	62.194 (7.3127)	
	Upperbound 2	53.496 (7.6214)	61.418 (7.5523)	61.648 (7.4897)	

Notes: Simulation results about average number of patients actually matched and predicted by the formulae to be matched. The standard errors of the population are reported in parentheses. The standard errors of the averages are obtained by dividing population standard errors by square root of the simulation number, 22.36.

But typical patient pool has lots of highly sensitized patients... Ashlagi and Roth (2013)

Incentives

Is a maximal matching mechanism incentive compatible?

Hatfield (2005) shows that the answer is yes: if a kidney exchange mechanism satisfies a property he calls "consistency," then it is strategy-proof.

Many mechanisms satisfy consistency, and hence are strategy-proof.

Computational Issues

Is it easy to find a maximal matchings?

Finding a maximal two-way matching is relatively easy.

Finding a maximal matching with three way and up is known to be computationally difficult (NP-complete).

Abraham, Blum, and Sandholm (2007) present an algorithm to find the maximal matchings. According to them, their algorithm is fast enough to use for 10000 pairs or so (reasonably large size, consistent with U.S. case).

Dynamic Kidney Exchange

Unver (2010 REStud).

So far we have considered a one-shot problem of matching patients and donors, but in reality patients and donors arrive and leave the pool over time.

Unver considers how the transplantation center should decide who to match, when to match, etc.

Unver studies how to organize the dynamic kidney exchange mechanism. He shows

1. When only two-way exchanges are feasible, it is optimal to conduct all exchanges as soon as they become available.
2. When there is no limit on size of the exchange, sometimes it is optimal not to conduct all the currently available exchanges and wait until more more patients can be matched.

More recent; Akbarpour, Li, and Oveis Grahan (2014), Anderson, Ashlagi, Gamarnik, and Kanoria (2015).

Ultruistic ("Good Samaritan") Donors

Simultaneous surgeries constraints may be indispensable for matching incompatible pairs with each other, so two-way exchange may be the only option.

But that may not be the case if an exchange is initiated by a good Samaritan donor (non-directed altruistic donor), because even if someone reneges on the plan, no patient ends up getting no kidney while losing her willing donor.

Take a look at the web page of Alliance for Paired Donation at http://www.paireddonation.org/anim2.htm

Sonmez and Unver (2014, Journal of Economic Theory), Sonmez, Unver, and Yenmez (2017)

More issues

How to incorporate compatible pairs?

Weighting different transplants (how good the match is, transportation cost, etc).

How to organize a transplantation network when there are many transplant centers? (Ashlagi and Roth 2011).

Stochastic mechanisms and fairness (RSU 2005, Yilmaz 2008).

Immunosupressant procedures: Sonmez, Unver, and Yilmaz (2018), Chum and Heo (2017), Andersson and Kratz (2017).

Other organs: Ergin, Sonmez, and Unver (2017) on livers and lung.

Nikzad, Akbarpour, Rees, and Roth (2017) on "global kidney exchange"

Summary

In most countries, organ allocation cannot use monetary transfers, resulting in difficulty in efficiently allocating the organs.

Matching theory can improve efficiency of organ allocation.

Designs involve such issues as efficiency, incentives and fairness.

Many theories are motivated by details of the model (dichotomous preferences, logistical constraints, blood types, etc.).

A lot of unresolved issues.