Introduction to the Theory of Two-Sided Matching Models

To see which results are robust and which are not, we'll look at some increasingly general models.

(Even before we look at complex design problems, we can get a headstart at figuring out which are our most applicable results by doing this sort of theoretical sensitivity analysis.)

- 1-1 matching: the "marriage" model
- Many-to-one matching (with simple preferences) : the "college admissions" model
- many to one matching with money and complex (gross substitutes) preferences

One to one matching: The marriage model

PLAYERS:

Men = $\{m_1,..., m_n\}$ Women = $\{w_1,..., w_p\}$

PREFERENCES (complete and transitive):

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

$P(w_j) = m_2, m_4, ... w_j ...$

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.

Strict preferences, and indifferences

If an agent is not indifferent between any two acceptable mates, or between being matched and unmatched, we'll say he/she has *strict* preferences. Some of the theorems we prove will only be true for strict preferences, and it will be useful to keep track of which ones, so I'm not assuming strict preferences unless I say so.

(This distinction has recently become much more important, as we've moved from labor markets to school choice problems)

An OUTCOME of the game is a *MATCHING*:

$\mu: M \cup W \rightarrow M \cup W$

such that $w = \mu(m)$ iff $\mu(w)=m$,
and for all m and w
either $\mu(w)$ is in M or $\mu(w) = w$, and
either $\mu(m)$ is in W or $\mu(m) = m$.

i.e. a man is matched to a woman only if she is also matched to him, and everyone is either matched to a person of the opposite gender or is single ("matched to him/herself)

Stable matchings

A matching is

BLOCKED BY AN INDIVIDUAL k if k prefers being single to being matched with $\mu(k)$, i.e. $k >_k \mu(k)$

BLOCKED BY A PAIR OF AGENTS (m,w) if they each prefer each other to μ, i.e.

$w >_m \mu(m)$ 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 the core, and in this simple model the set of (pairwise) stable matchings equals the core.

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

- 0. *If some preferences are not strict*, arbitrarily break ties
- 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 2.8 (Gale and Shapley): A stable matching exists for every marriage market.

(Theorems are numbered as in Roth and Sotomayor.)

Elements of the proof:

- the deferred acceptance algorithm always stops
- the matching it produces is always stable with respect to the strict preferences (i.e. after any arbitrary tie-breaking),
- and with respect to the original preferences.

Two-sidedness is important

(at least that's the way we thought about this before Ostrovsky 2008)

- Consider the one-sided "roommate problem" in which everyone can potentially be matched with anyone else.
- Example (GS) 4 roommates with preferences given by

1: 2,3,4 2: 3,1,4 3: 1,2,4 4: any prefs

No stable matching exists

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 μ_M 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.

Sketch of the proof:

Terminology: w is *achievable* for m if there is some stable μ such that $\mu(m) = w$.

Inductive step: 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 μ with $\mu(m) = w$, and $\mu(m')$ achievable for m'. Can't be stable: by the inductive step, (m',w) would be a blocking pair.

Common preferences

Let $\mu >_M \mu'$ denote that all men like μ at least as well as μ', with at least one man having strict preference. Then $>_M$ is a *partial* order on the set of matchings, representing the common preferences of the men. Similarly, define $>_W$ as the common preference of the women.

Theorem 2.13(Knuth)

When all agents have strict preferences, the common preferences of the two sides of the market are *opposed* on the set of stable matchings: if μ and μ' are stable matchings, then all men like μ at least as well as μ' if and only if all women like μ' at least as well as μ. That is, $\mu >_M \mu'$ if and only if $\mu' >_W \mu$.

Proof: immediate from definition of stability.

So the best outcome for one side of the market is the worst for the other.

Let's see what's going on.

For any two matchings μ and μ', and for all m and w, define
$\lambda = \mu \vee_M \mu'$ as the function that assigns each man his *more* preferred of the two matches, and each woman her *less* preferred:

- $\lambda(m) = \mu(m)$ if $\mu(m) >_m \mu'(m)$ and
- $\lambda(m) = \mu'(m)$ otherwise

- $\lambda(w) = \mu(w)$ if $\mu(m) <_w \mu'(w)$ and
- $\lambda(w) = \mu'(w)$ otherwise

Define $\nu = \mu \wedge_M \mu'$ analogously, by reversing the preferences.

Theorem 2.16 Lattice Theorem (Conway):

When all preferences are strict, if μ and μ' are stable matchings, then the functions

$\lambda = \mu \vee_M \mu'$ and $\nu = \mu \wedge_M \mu'$

- are both matchings.
- Furthermore, they are both stable.

So if we think of λ as asking men to point to their preferred mate from two stable matchings, and asking women to point to their *less* preferred mate, the theorem says that

No two men point to the same woman

- (this follows from the stability of m and m')

Every woman points back at the man pointing to her;

- the direction [λ(m) = w implies λ (w) = m] follows easily from stability, but the direction [λ(w) = m implies λ (m) = w] takes a bit more work. (We'll come back to this when we prove the Decomposition Lemma

- And the resulting matching is stable.
  - again, immediately from the stability of m and m'

Theorem 2.22: In a market (***M,W, P***) with strict preferences, the set of people who are **single** is the **same for all stable matchings**.

One strategy of proof: What can we say about the *number and identity* of men and women matched (and hence the number and identity unmatched) at μ_M and at μ_W?

i.e. denoting $M_{\mu M} = \mu_M(W \cap M)$, etc. what can we say about the relative sizes and containment relations of the sets $M_{\mu M}$, $W_{\mu M}$, $M_{\mu W}$, and $W_{\mu W}$?

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

Theorem 2.27 **Weak Pareto optimality for the men**: There is no individually rational matching μ (stable or not) such that $\mu >_m \mu_M$ for all m in M.

Proof (using the deferred acceptance algorithm…)

If μ were such a matching it would match every man m to some woman w who had rejected him in the algorithm in favor of some other man m' (i.e. even though m was acceptable to w).

Hence all of these women, μ(M), would have been matched under μ_M. That is, $\mu_M(\mu(M)) = M$.

Hence all of M would have been matched under μ_M and $\mu_M(M) = \mu(M)$.

But since all of M are matched under μ_M any woman who gets a proposal in the last step of the algorithm at which proposals were issued has not rejected any acceptable man, i.e. the algorithm stops as soon as every woman in $\mu_M(M)$ has an acceptable proposal.

So such a woman must be single at μ (since every man prefers μ to μ_M), which contradicts the fact that $\mu_M(M) = \mu(M)$.

Example 2.31 (μ_M not **strongly** Pareto optimal for the men)

M = {m1, m2, m3}, W = {w1, w2, w3}

P(m1) = w2,w1,w3 P(w1) = m1,m2,m3

P(m2) = w1,w2,w3 P(w2) = m3,m1,m2

P(m3) = w1,w2,w3 P(w3) = m1,m2,m3

μ_M = ([m1,w1], [m2,w3], [m3,w2]) = μ_W

But note that $\mu >_M \mu_M$ for

μ = ([m1,w2], [**m2**,w3], [m3,**w1**])

All m like μ at least as well as μ_M and some strictly prefer it.

One final lemma about stable and unstable matchings will help when we study strategic properties.

Lemma 3.5 Blocking Lemma (Gale and Sotomayor) : Let μ be any individually rational matching with respect to strict preferences ***P*** and let *M'* be all 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-M'* and *w* is in $\mu(M')$

Proof of blocking lemma

Case I: $\mu(M') \neq \mu_M(M')$. Choose w in $\mu(M')-\mu_M(M')$, say, $w=\mu(m')$. Then 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 II: $\mu_M(M')=\mu(M')=W'$.

Case II: $\mu_M(M')=\mu(M')=W'$. Let w be the woman in W' who receives the last proposal from an acceptable member of M' in the deferred acceptance algorithm. Since all w in W' have rejected acceptable men from M', w had some man m engaged when she received this last proposal. Claim: (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.

Strategic Behavior

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,**P**) whose outcome will be determined by a centralized clearinghouse, based on a list of preferences that players will state ("reveal"). If the vector of stated preferences is **Q**, the algorithm employed by the clearinghouse produces a matching h(**Q**). The *matching mechanism* h is defined for all (M,W,**Q**). If the matching produced is *always* a stable matching with respect to **Q**, we'll say that h is a *stable matching mechanism*.

Theorem 4.4 Impossibility Theorem (Roth)

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.

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 $\qquad$ w_1: m_2, m_1

m_2: w_2, w_1 $\qquad$ 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**)?

Of course, this kind of proof of the impossibility theorem leaves open the possibility that situations in which some participant can profitably manipulate his preferences are rare. The following result suggests otherwise.

Theorem 4.6

When any stable mechanism is applied to a marriage market in which preferences are strict and 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. (This agent can misrepresent in such a way as to be matched to his or her most preferred achievable mate under the true preferences at every stable matching under the mis-stated preferences.)

Incentives facing the men when the M-optimal stable mechanism is used

Theorem 4.7 (Dubins and Freedman, Roth)

The mechanism that yields the *M*-optimal stable matching (in terms of the stated preferences) makes it a dominant strategy for each man to state his true preferences.

Theorem 4.10 (Dubins and Freedman)

Let **P** be the true preferences of the agents, and let **P̃** differ from **P** in that some coalition **M̃** of the men mis-state their preferences. Then there is no matching μ, stable for **P̃**, which is strictly preferred to μ_M by all members of **M̃**.

Theorem 4.11 (Limits on successful manipulation.) (Demange, Gale, and Sotomayor).

Let **P** be the true preferences (not necessarily strict) of the agents, and let $\tilde{\mathbf{P}}$ differ from **P** in that some coalition C of men and women mis-state their preferences. Then there is no matching μ, stable for $\tilde{\mathbf{P}}$, which is preferred to *every* stable matching under the true preferences **P** by all members of C.

Note that Theorem 4.11 implies both Theorems 4.7 and 4.10.

Proof of Theorem 4.11:

Suppose some nonempty subset ~~**M**~~ $\cup$ ~~**W**~~ of men and women mis-state their preferences and are strictly better off under μ, stable w.r.t. ~~**P**~~ ,than under any stable matching w.r.t. **P**.

μ must be individually rational with respect to **P**, even though unstable.

Now construct strict preferences **P'**, so that if any agent x is indifferent under **P** between $\mu(x)$ and some other alternative, then under **P'** x prefers $\mu(x)$ (but otherwise make no change in the ordering of preferences **P**).

Then (m,w) blocks μ under **P'** only if (m,w) blocks μ under **P**.

Since every stable matching under **P'** is also stable under **P,**

$\mu(m) >_m \mu_M(m)$ for every m in ~~**M**~~, and (*)

$\mu(w) >_w \mu_W(w)$ for every w in ~~**W**~~

where μ_m and μ_W are the M- and W- optimal stable matchings for (M,W, **P'**).}

If **M** is not empty we can apply the Blocking Lemma (3.5) to the market (M,W, **P'**), since by (*) **M** is a subset of M'; thus there is a pair {m,w} which blocks μ under **P'** and so under **P** such that

$\mu_M(m) \geq_m \mu(m)$ and

$\mu_M(w) \geq_w \mu(w)$ (otherwise w and $\mu(w)$ would block μ_M, since w is in $\mu(M')$ by the blocking lemma).

Clearly m and w are not in $\mathbf{M} \cup \mathbf{W}$ and so are not mis-stating their preferences, so they will also block μ under **P**, contradicting that μ is stable under **P**.

If **M** is empty **W** is not, and the symmetrical argument applies.

Notice that the claim that individual men can't profitably manipulate their preferences is much more robust than the (knife-edge) claim that coalitions of men can't profitably manipulate, recall Example 2.31:

Example 2.31 (μ_M not strongly Pareto optimal for the men)

M = {m1, m2, m3}, W = {w1, w2, w3}

P(m1) = w2,w1,w3 P(w1) = m1,m2,m3

P(m2) = w1,w2,w3 P(w2) = m3,m1,m2

P(m3) = w1,w2,w3 P(w3) = m1,m2,m3

μ_M = ([m1,w1], [m2,w3], [m3,w2]) = μ_W But note that $\mu >_M \mu_M$ for μ= ([m1,w2], [**m2**,w3], [m3,w1])

Note that m2 can help the other men at no cost to himself…so a coaltion of all men can weakly help themselves (and strictly if there was some money)

What can we say about equilibrium?

Theorem 4.15 (Gale and Sotomayor): Pure strategy equilibria exist:

When all preferences are strict, let μ be any stable matching for (*M, W, **P***). Suppose each woman *w* in $\mu(M)$ chooses the strategy of listing only $\mu(w)$ on her stated preference list of acceptable men (and each man states his true preferences). This is an equilibrium in the game induced by the M-optimal stable matching mechanism (and μ is the matching that results).

(Think about how to prove this before looking in the book…)

Furthermore, every equilibrium mis-representation by the women must nevertheless yield a matching that is stable with respect to the true preferences. (But the proof of this—which is like the old joke about the dying financier—should raise doubts about it's applicability☺

Theorem 4.16 (Roth) Suppose each man chooses his dominant strategy and states his true preferences, and the women choose any set of strategies (preference lists) *P'(w)* that form an equilibrium for the matching game induced by the *M*-optimal stable mechanism. Then the corresponding *M*-optimal stable matching for (*M,W,* ***P'***) is one of the stable matchings of (*M,W,* ***P***).

Let's make all these results familiar by comparing them to a simple (auction) model of one seller of a discrete good and n buyers with money

Traders N={1, …n, n+1}

Seller s=n+1 owns the object, and has reservation price r_{n+1}

Buyers b=1,…n have reservation prices r_b

That is, each player i places a monetary value r_i on the object, and each buyer has sufficient cash to pay his reservation price.

If the seller sells the object to buyer b at a price p (and if no other monetary transfers are made), the seller earns p, buyer b earns r_b-p and all other buyers earn zero.

A feasible outcome is a non-negative vector of monetary payoffs $x=(x_1, x_2, \ldots x_n, x_{n+1})$ in $\mathbf{R}^{n+1}$ such that $\Sigma\, x_i \leq \max\{r_i\}$.

That is, it is feasible to have monetary transfers that aren't just between the buyer and seller. So the rules of the game are that a coalition S of players can distribute $\max\{r_i \mid i \text{ in } S\}$ if S contains the seller, and 0 otherwise.

A payoff vector is (pairwise) *stable* if it is individually rational and if there doesn't exist a buyer i and a price p such that $p>x_{n+1}$ and $r_i - p > x_i$.

The *core* of the game is the set of payoff vectors x such that no coalition of any size can afford to pay its members more than the sum of their payoffs at x.

Let r_{1*} be the highest reservation price, and r_{2*} the second highest (belonging to renumbered players 1* and 2* respectively)

Theorem 7.2:

- For any vector of reservation prices r, the core is nonempty.
- If the seller does not have the highest reservation price, then the core equals the set of feasible x such that $\mathbf{x_{n+1} = p}$ **for** $\mathbf{r_{2^*} \leq p \leq r_{1^*}}$, $\mathbf{x_{1^*} = r_{1^*} - p}$, and $x_i = 0$ for all players other than 1* and n+1.

  If the seller does have the highest reservation price (i.e. if 1* = n+1) then the core equals $\{(0,\ldots 0, r_{1^*})\}$
- The set of stable payoff vectors equals the core.

(NB: unlike the marriage model, this little model is asymmetric. In comparing most of the theorems we've just discussed, it will sometimes help to think of the bidders as the men…)

One more observation about the marriage model

Suppose μ and μ' are stable matchings, and for some m, $w = \mu(m) >_m \mu'(m) = w'$. Then the stability of μ' immediately implies that

$$\mu'(w) >_w \mu(w) = m.$$

But how about w'? (Recall the part of the Lattice Theorem, 2.16, that we deferred til now…)

The **Decomposition Lemma** (Corollary 2.21, Knuth):

Let μ and μ' be stable matchings in (M,W, **P**), with all preferences strict. Let $M(\mu)$ ($W(\mu)$) be the set of men (women) who prefer μ to μ', and let $M(\mu')$ ($W(\mu')$) be those who prefer μ'. Then μ and μ' map $M(\mu')$ onto $W(\mu)$ and $M(\mu)$ onto $W(\mu')$.

Proof: we've just observed above that $\mu(M(\mu))$ is contained in $W(\mu')$. So $|M(\mu)| \leq |W(\mu')|$.

Symmetrically, $\mu'(W(\mu'))$ is contained in $M(\mu)$, so $|M(\mu)| \geq |W(\mu')|$.

Since μ and μ' are one-to-one (and since $M(\mu)$ and $W(\mu')$ are finite), both μ and μ' are onto. So, to answer the question posed on the previous slide, a man or woman who prefers one stable matching to another is matched at *both* of them to a mate with the reverse preferences.

Median stable matchings

- Decomposability implies that, when preferences are strict and there are an odd number of stable matchings, there is a median, stable matching (and it is stable☺
- Proof: left to the reader…

Many-to-one matching: The college admissions model

PLAYERS: Firms = $\{f_1,..., f_n\}$ Workers= $\{w_1,...,w_p\}$

 # positions $q_1,...,q_n$

Synonyms (sorry:-): F=Firms = C=Colleges = H=Hospitals W=Workers = S=Students

PREFERENCES *over individuals* (complete and transitive), as in the marriage model:

$P(f_i) = w_3, w_2, ... f_i ...$ $[w_3 >_{fi} w_2]$

$P(w_j) = f_2, f_4, ... w_j ...$

An OUTCOME of the game is a *MATCHING*:

μ: $F \cup W \rightarrow F \cup W$

such that $\mu(f)$ contains w iff $\mu(w) = f$, and for all f and w

$|\mu(f)|$ is less than or equal to q_f

either $\mu(w)$ is in F or $\mu(w) = w$. so f is matched to the *set* of workers $\mu(f)$.

We need to specify how firms' preferences over matchings, are related to their preferences over individual workers, since they hire groups of workers. The simplest model is

Responsive preferences: for any set of workers $S \subset W$ with $|S|<q_i$, and any workers w and w' in W/S,

$S \cup w >_{fi} S \cup w'$ if and only if $w >_{fi} w'$, and

$S \cup w >_{fi} S$ if and only if w is acceptable to f_i.

A matching μ is *individually irrational* if $\mu(w) = f$ for some worker w and firm f such that either the worker is unacceptable to the firm or the firm is unacceptable to the student. An individually irrational matching is said to be blocked by the relevant individual. (**Note the modeling assumption here**.)

A matching μ is *BLOCKED BY A PAIR OF AGENTS* (f,w) if they each prefer each other to μ:

[$w >_f w'$ for some w' in $\mu(f)$ or $w >_f f$ if $|\mu(f)| < q_f$]

and $f >_w \mu(w)$

As in the marriage model, a matching is (pairwise) *stable* if it isn't blocked by any individual or pair of agents.

But now that firms employ multiple workers, it might not be enough to concentrate only on *pairwise* stability. The assumption of responsive preferences allows us to do this, however.

A matching μ is *blocked by a coalition* A of firms and workers if there exists another matching μ' such that for all workers w in A, and all firms f in A

1. $\mu'(w)$ is in A
2. $\mu'(w) >_w \mu(w)$
3. $\sigma \in \mu'(f)$ implies $\sigma \in A \cup \mu(f)$ (i.e. every firm in A is matched at μ' to new students only from A, although it may continue to be matched with some of its "old" students from μ. (THIS DIFFERS FROM THE STANDARD DEFINITION OF THE CORE…)
4. $\mu'(f) >_f \mu(f)$

A matching is *group stable* if it is not blocked by a coalition of any size.

Lemma 5.5: When preferences are responsive, a matching is group stable if and only if it is (pairwise) stable.

Proof: instability clearly implies group instability.

Now suppose μ is blocked via coalition A and outcome μ'. Then there must be a worker w and a firm f such that w is in $\mu'(f)$ but not in $\mu(f)$ such that w and f block μ. (Otherwise it couldn't be that $\mu'(f) >_f \mu(f)$, since f has responsive preferences.)

A related marriage market

Replace college *C* by q_C positions of *C* denoted by $c_1, c_2, \ldots, c_{(qC)}$. Each of these positions has C's preferences over individuals. Since each position c_i has a quota of 1, we do not need to consider preferences over groups of students.

Each student's preference list is modified by replacing *C*, wherever it appears on his list, by the string $c_1, c_2, \ldots, c_{(qC)}$, in that order.

A matching μ of the college admissions problem, corresponds to a matching μ' in the related marriage market in which the students in $\mu(C)$ are matched, in the order which they occur in the preferences P(C), with the ordered positions of C that appear in the related marriage market. (If preferences are not strict, there will be more than one such matching.)

Lemma 5.6: A matching of the college admissions problem is stable if and only if the corresponding matchings of the related marriage market are stable.

(NB: some results from the marriage model will translate immediately, but not those involving both stable and unstable matchings…)

Geographic distribution

Theorem 5.12: When all preferences over individuals are strict, the set of students employed and positions filled is the same at every stable matching.

The proof is immediate via the similar result for the marriage problem and the construction of the corresponding marriage problem (Lemma 5.6).

So any hospital 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. The next result shows that not only will such a hospital fill the same number of positions, but it will fill them with exactly the same interns at any other stable matching.

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.

(This will be easy to prove after Lemma 5.25, but my original proof, while correct, left me with a mistaken impression of what was going on. Try proving it as a generalization of the "pointing" in marriage markets.)

Comparison of stable matchings in the college admissions model

Overview: suppose one of the colleges, C, evaluates students by their scores on an exam, and evaluates entering classes according to their *average* score on the exam. (So even when we assume no two students have exactly the same score, so that college C's preferences over individuals are strict, it does not have strict preferences over entering classes, since it is indifferent between two entering classes with the same average score.) Then different stable matchings may give college C different entering classes. However *no two distinct entering classes that college C could have at stable matchings will have the same average exam score*. Furthermore, for any two distinct entering classes that college C could be assigned at stable matchings, we can make the following strong comparison. Aside from the students who are in both entering classes, *every* student in one of the entering classes will have a higher exam score than *any* student in the other entering class.

Lemma 5.25 (Roth and Sotomayor)

Suppose colleges and students have strict individual preferences, and let μ and μ' be stable matchings for ***(S, C, P),*** such that $\mu(C) \neq \mu'(C)$ for some C. Let $\underline{\mu}$ and $\underline{\mu}'$ be the stable matchings corresponding to μ and μ' in the related marriage market. If $\underline{\mu}(c_i) >_C \underline{\mu}'(c_i)$ for some position c_i of C then $\underline{\mu}(c_i) \geq_C \underline{\mu}'(c_i)$ for all positions c_i of C.

Proof

It is enough to show that $\mu(c_j) >_C \mu'(c_j)$ for all j > i. So suppose this is false. Then there exists an index *j* such that $\mu(c_j) >_C \mu'(c_j)$, but $\mu'(c_{j+1}) \geq_C \mu(c_{j+1})$. Theorem 5.12 (constant employment) implies $\mu'(c_j) \in \boldsymbol{S}$. Let $s' \equiv \mu'(c_j)$. By the *decomposition lemma* $c_j \equiv \mu'(s') >_{s'} \mu(s')$. Furthermore, $\mu(s') \neq c_{j+1}$, since

$s' >_C \mu'(c_{j+1}) \geq_C \mu(c_{j+1})$ (where the first of these preferences follows from the fact that for any stable matching μ' in the related marriage market, $\mu(c_j) >_C \mu'(c_{j+1})$ for all j). Therefore, since c_{j+1} comes right after c_j in the preferences of *s*' (or any *s*) in the related marriage problem, μ is blocked via s' and c_{j+1}, contradicting (via Lemma 5.6) the stability of μ.

(This proof also establishes the rural hospitals theorem).

Theorem 5.26: (Roth and Sotomayor)

If colleges and students have strict preferences over *individuals*, then colleges have strict preferences over those *groups* of students that they may be assigned at stable matchings. That is, if μ and μ' are stable matchings, then a college C is indifferent between $\mu(C)$ and $\mu'(C)$ only if $\mu(C) = \mu'(C)$.

Proof: via the lemma, and repeated application of responsive preferences.

Theorem 5.27: (Roth and Sotomayor)

Let preferences over individuals be strict, and let μ and μ' be stable matchings for (S***, C, P***). If $\mu(C) >_C \mu'(C)$ for some college *C*, then $s >_c s'$ for all *s* in $\mu(C)$ and *s*' in $\mu'(C)$-$\mu(C)$. That is, *C* prefers every student in its entering class at μ to every student who is in its entering class at μ' but not at μ.

Proof

Consider the related marriage market and the stable matchings $\mathbf{\mu}$ and $\mathbf{\mu}'$ corresponding to μ and μ'. Let $q_C=k$, so that the positions of C are $c_1,\ldots,c_k$.

First observe that C fills its quota under μ and μ', since, if not, Theorem 5.13 (Rural hospitals) would imply that $\mu(C) = \mu'(C)$. So $\mu'(C) - \mu(C)$ is a nonempty subset of S, since $\mu(C) \neq \mu'(C)$. Let $s' = \mathbf{\mu}'(c_j)$ for some position c_j such that s' is not in $\mu(C)$. Then $\mathbf{\mu}(c_j) \neq \mathbf{\mu}'(c_j)$.

By Lemma 5.25 $\mathbf{\mu}(c_j) >_C \mathbf{\mu}'(c_j) = s'$.

The Decomposition Lemma implies $c_j >_{s'} \mathbf{\mu}(s')$.

So the construction of the related marriage problem implies $C >_{s'} \mu(s')$, since $\mu(s') \neq C$.

Thus $s >_C s'$ for all s in $\mu(C)$ by the stability of μ, which completes the proof.

So, for considering stable matchings, we have some slack in how carefully we have to model preferences over groups. (This is lucky for design, since it reduces the complication of soliciting preferences from firms with responsive preferences…)

The results also have an unusual mathematical aspect, since they allow us to say quite a bit about stable matchings even without knowing all the preferences of potential blocking pairs.

Consider a College C with quota 2 and preferences over individuals P(C) = s1, s2, s3, s4. Suppose that at various matchings 1-4, C is matched to

1. {s1,s4},
2. {s2,s3},
3. {s1,s3}, and
4. {s2,s4}.

Which matchings can be simultaneously stable for some responsive preferences over individuals?

So long as all preferences over groups are responsive, matchings 1 and 2 cannot both be stable (Lemma 5.25), nor can matchings 3 and 4 (Theorem 5.27).

Strategic questions in the College Admissions model:

Theorem 5.16 (Roth)

A stable matching procedure which yields the student-optimal stable matching makes it a dominant strategy for all students to state their true preferences.

Proof: immediate from the related marriage market

Theorem 5.14 (Roth) No stable matching mechanism exists that makes it a dominant strategy for all hospitals to state their true preferences.

Proof: consider a market consisting of 3 hospitals and 4 students. H1 has a quota of 2, and both other hospitals have a quota of 1. The preferences are:

s1: H3, H1, H2

s2: H2, H1, H3

s3: H1, H3, H2

s4: H1, H2, H3

H1: s1, s2, s3, s4

H2: s1, s2, s3, s4

H3: s3, s1, s2, s4

The unique stable matching is

{[H1, s3,s4], [H2, s2], [H3, s1]}

But if H1 instead submitted the preferences s1,s4

the unique stable matching is

{[H1, s1,s4], [H2, s2], [H3, s3]}.