Decentralized vs Centralized Matching[1]

Date: June 5, 2013

SL

This file first argues that the fundamental difference between decentralized and centralized matching systems is one about timing. In centralized systems, the market clears at once while in decentralized systems waitlists clear sequentially, possibly necessitating multiple rounds. The crucial decision for applicants in decentralized systems is accordingly an acceptance strategy that specifies which offers they should accept at which point in time. In centralized systems, the crucial decision is the application strategy (i.e. should one submit preferences truthfully or is there a need to game the system). The need to clear sequentially is inevitable in decentralized systems because applicants may and in equilibrium will have to reject offers because all applicants optimally put their names down on multiple waitlists.[2] Second, the file uses a simple example to demonstrate that decentralized systems are prone to two sources of inefficiency. (1) Unfilled slots. Even with aggregate excess demand they may fail to fill slots. (2) Inefficient matching of slots. Slots may not be allocated in the most efficient manner.

Setup. Assume that there are two centres (or colleges), called A an d B, each with a capacity of one slot. There are three individuals (parents or students) $i = 1, 2, 3$, each with demand for one slot.

Preferences. There are four outcomes for each individual. Individual i can be matched to A, to B, to an outside option O or not matched at all. We will assume that the prior probability that an individual prefers A to B is α, and we refer to such an individual as a "type-A" individual. Individuals who prefer B to A are called "type-B" and have prior probability of $1 - \alpha$. The value of the outside option, which is available in unlimited supply (and which in the context of child care may be thought of as grandmother looking after the child, one parent working less so the other parent can work a bit etc, or accepting a slot in an under-demanded centre that is far off the parent's commuting route) is $v_O = 1$. The value of not being matched is $v_N = 0$ (which can be interpreted as staying home instead of working). The value of the preferred slot is $v_{(1)} = 3$ for every individual while $v_{(2)} = 2$ is the value of the second-best outcome for every individual. Table 1 displays these match values.

Priorities. Each centre has a priority list over individuals. There being three individuals, there are 6 possible permutations (i.e. six different priority lists) at each centre and therefore 36 possibilities across the two centres. We assume that each of these has the same probability (i.e. uniform random priority). Uncertainty about priority is meant to capture the idea that even though someone may know

[1] Filename: decentralized.tex

[2] This suggests that decentralized systems may work better than centralized ones if deadlines are heterogenous across participants on both sides of the market (say, workers and firms), which may be the case for the typical labor market where vacancies and job seekers arise continuously over time. It is also consistent with the notion that centralized matching systems are popular for entry-level job markets, each of which has the same (or a similar) starting date.

Match	type-A	type-B
A	3	2
B	2	3
O	1	1
N	0	0

Table 1: Match values.

their rank on a waitlist at a given centre it is impossible to say with perfect accuracy whether one gets an offer from this centre because all parents put their names onto multiple waitlists. The assumption that the probability is uniform is made for analytical simplicity. All randomization is independent and all probabilities and values are common knowledge. The only private information individuals have is their type.

Timing. The timing is as follows.

At time $t = 1$, the first round (R1) offers are made to applicants. That is, every centre makes an offer to the individual with highest priority at this centre. Individuals who receive offer(s) accept definitely or reject definitely.

At time $t = 2$, second round (R2) offers are made to the individual with second highest priority by centre(s) whose first round offer(s) have been rejected. Individuals who receive offer(s) accept or reject definitely.

At time $t = 3$, third round offers (R3) may be made if necessary (i.e. if there are still unfilled slots).

Between $t = 1$ and $t = 2$, individuals' outside option expires if not accepted definitely. This captures the idea that arrangements must be made ahead of time (with grandmother, the spouse's work etc). If the outside option were always valid (i.e. even at date $t = 3$), inefficient matching would still occur but there would be no unfilled slots.

Analysis. In R1, an individual may get either of the following:

1. an offer, denoted a, from A,
2. an offer, denoted b, from B,
3. an offer from both, denoted ab,
4. no offer, denoted n.

If the offer includes the first-best option, e.g. offer a for type-A's, or offer b for type-B's or offer ab, an individual obviously optimally accepts the best offer. In case the offer is only second-best (i.e. b for type-A and a for type-B) or if no offer is received, the optimal strategy is less trivial and depends on what the other individuals are expected to do.

It is next shown that the following "rushing behavior" is a symmetric (Bayes' Nash) equilibrium (BNE): Every R1 offer is accepted; individuals who receive no offer accept the outside option.[3]

To see that this is an equilibrium, recall that accepting a first-best offer in R1 is obviously optimal. So we only need to determine what to do if one receives a second-best offer in R1 or no offer. To see that immediately accepting even a second-best offer is optimal given the stipulated behavior by all others, notice that receiving only the second-best offer means that someone else has received the other offer, which this someone else will accept given the stipulated equilibrium behavior. Thus, there is simply no point waiting for the other, better offer (as it will never come). So we are left to determine the optimal behavior upon no offer.

Interestingly, this is a little bit more complicated because if i receives no offer there is a chance someone else receives no offer because someone is lucky enough to have highest priority everywhere and accordingly receives two offers. This would then induce R2 offers. To determine the probability of this happening, we need to determine the probability that someone else has received no R1 offer conditional on i not having received one, denoted $\Pr(h \quad \text{or} \quad j \quad \text{received} \quad ab|i \quad \text{received} \quad n)$ with $j, h \neq i$. This probability is given by Bayes' rule as

$$\mu \equiv \Pr(h \quad \text{or} \quad j \quad \text{received} \quad ab|i \quad \text{received} \quad n) = \frac{\beta\eta_{both}}{\beta\eta_{both} + (1-\beta)\eta_{none}},$$

where β is the prior probability that h or j get *both* offers, η_{both} is the probability that i gets none when either h or j gets both (of course, $\eta_{both} = 1$) and η_{none} is the probability that i gets none when neither h nor j get both.

The prior probability that h gets two offers is simply $1/9$. The same being true for j, we have $\beta = 2/9$.

There are 8 possibilities for i not to receive an offer at A and B and the identities of the individuals who receive offers a and b to differ: The priority list at A can be: (h, j, i) or (h, i, j), in which cases the priority list at B must be: (j, h, i) or (j, i, h). This gives 2^2 possibilities. And then there are two permutations (the first ranked at A could be j instead), which gives us 8 possibilities. There being 36 possibilities a priori, this gives us $\eta_{none} = \frac{8}{36} = \frac{2}{9}$. Hence we get

$$\mu = \frac{\frac{2}{9}1}{\frac{2}{9}1 + \left(1-\frac{2}{9}\right)\frac{2}{9}} = \frac{9}{16}.$$

However, the probability of receiving an offer in R2 is only $\mu/2$ because with probability 1/2 the other individual who has not received an offer in R1 will have higher priority at the centre that will make an offer. Notice that whoever receives an offer in R2 will accept it if he has not already accepted the outside option. For a type-A individual the expected continuation payoff when rejecting the outside

[3] The conditions for this "rushing" equilibrium to be unique remain to be determined.

option after having received no offer in $t = 1$ is therefore

$$U_A = \frac{\mu}{2}[2\alpha + 3(1-\alpha)] = \frac{9}{32}[3-\alpha],$$

where α is the probability that the individual who received ab accepted a, so that our individual in question will only receive the offer b. Similarly,

$$U_B = \frac{\mu}{2}[3\alpha + 2(1-\alpha)] = \frac{9}{32}[2+\alpha],$$

is the expected payoff of going to R2 for an individual of type-B who received no offer in $t = 1$. Notice that both U_A and U_B are no more than $\frac{27}{32}$, which is less than 1, the value of the outside option. Thus, all individuals who receive no offer in $t = 1$ optimally accept the outside option. Thus, the stipulated behavior is an equilibrium.

Evaluation and Comparison. An important property of the equilibrium under the decentralized matching system is that only first priority slots are ever allocated.

Under a centralized matching system using the applicant proposing Gale-Shapley DA algorithm, every applicant has a dominant strategy to submit preferences truthfully.[4] Accordingly, slots are allocated efficiently (i.e. in the sense of maximizing ex post surplus) subject to the constraints imposed by centres' priorities.

Consider, for example, the case where 1 and 2 are type-A's and 3 is type-B (when they're all the same types the allocation does not matter for efficiency as long as all slots are allocated and so the focus on this case is without loss of generality except that one will have to replace α's by $1-\alpha$'s to account for the case where two prefer B and one prefers A). Under DA, the ultimate allocation will be that one of the type-A's is matched to A unless 3 has highest priority at A but not at B. If 3 has highest priority at B, he will be matched to B and we have the ex post surplus maximizing allocation.

If all are the same types, DA allocates always ex post efficiently while the decentralized system allocates inefficiently with probability 1/9 because that's the probability of leaving one slot open (because the same agent has highest priority at both centre).

Tables 2 and 3 provide the summary for the matchings and parental surplus created under, respectively, a centralized matching system running the (parental proposing) DA algorithm and in the BNE under the decentralized system derived above. The top left entry in table 2 says that under the priority lists $p_A = (1,2,3)$ (meaning 1 has highest, 2 second-highest and 3 lowest priority at A and $p_B = (1,2,3)$ parent 1 will be matched to A, parent 2 to B and the parental surplus will be $S = 5(= 3+2)$. Analogously, the top left entry in table 3 specifies that, under the priorities $p_A = (1,2,3)$ and $p_B = (1,2,3)$, 1 is matched to A in the BNE and B's slot remains empty, giving a parental surplus of $S = 3$.

[4] This is true for the domain of ordinal preferences and is expected to extend to the preferences imposed here.

Priorities, Allocations, and Surplus (without v_O)						
priorities	$p_B = (1,2,3)$	$p_B = (1,3,2)$	$p_B = (2,1,3)$	$p_B = (2,3,1)$	$p_B = (3,1,2)$	$p_B = (3,2,1)$
$p_A = (1,2,3)$	A:1; B:2; S=5	A:1; B:3; S=6	A:1; B:2; S:5	A:1; B:2; S:5	A:1; B:3; S:6	A:1; B:3; S:6
$p_A = (1,3,2)$	A:1; B:2; S=5	A:1; B:3; S=6	A:1; B:2; S:5	A:1; B:2; S:5	A:1; B:3; S:6	A:1; B:3; S:6
$p_A = (2,1,3)$	A:2; B:1; S=5	A:2; B:1; S=5	A:2; B:1; S:4	A:2; B:3; S:6	A:2; B:3; S:6	A:2; B:3; S:6
$p_A = (2,3,1)$	A:2; B:1; S=5	A:2; B:1; S=5	A:2; B:1; S:4	A:2; B:3; S:6	A:2; B:3; S:6	A:2; B:3; S:6
$p_A = (3,1,2)$	A:3; B:1; S=4	A:3; B:1; S=4	A:3; B:2; S:4	A:3; B:2; S:4	A:1; B:3; S:6	A:1; B:3; S:6
$p_A = (3,2,1)$	A:3; B:1; S=4	A:3; B:1; S=4	A:3; B:2; S:4	A:3; B:2; S:4	A:2; B:3; S:6	A:2; B:3; S:6

Table 2: DA: AAB.

Priorities, Allocations, and Surplus						
priorities	$p_B = (1,2,3)$	$p_B = (1,3,2)$	$p_B = (2,1,3)$	$p_B = (2,3,1)$	$p_B = (3,1,2)$	$p_B = (3,2,1)$
$p_A = (1,2,3)$	A:1; B:∅; S=3	A:1; B:∅; S=3	A:1; B:2; S=5	A:1; B:2; S=5	A:1; B:3; S=6	A:1; B:3; S=6
$p_A = (1,3,2)$	A:1; B:∅; S=3	A:1; B:∅; S=3	A:1; B:2; S=5	A:1; B:2; S=5	A:1; B:3; S=6	A:1; B:3; S=6
$p_A = (2,1,3)$	A:2; B:1; S=5	A:2; B:1; S=5	A:2; B:∅; S=3	A:2; B:∅; S=3	A:2; B:3; S=6	A:1; B:3; S=6
$p_A = (2,3,1)$	A:2; B:1; S=5	A:2; B:1; S=5	A:2; B:∅; S=3	A:2; B:∅; S=3	A:2; B:3; S=6	A:1; B:3; S=6
$p_A = (3,1,2)$	A:3; B:1; S=4	A:3; B:1; S=4	A:3; B:2; S=4	A:3; B:2; S=4	A: ∅; B:3; S=3	A: ∅; B:3; S=3
$p_A = (3,2,1)$	A:3; B:1; S=4	A:3; B:1; S=4	A:3; B:2; S=4	A:3; B:2; S=4	A: ∅; B:3; S=3	A: ∅; B:3; S=3

Table 3: BNE: AAB.

Priorities, Allocations, and Surplus (without v_O)						
priorities	$p_B = (1,2,3)$	$p_B = (1,3,2)$	$p_B = (2,1,3)$	$p_B = (2,3,1)$	$p_B = (3,1,2)$	$p_B = (3,2,1)$
$p_A = (1,2,3)$	2	3	0	0	0	0
$p_A = (1,3,2)$	2	3	0	0	0	0
$p_A = (2,1,3)$	0	0	1	3	0	0
$p_A = (2,3,1)$	0	0	1	3	0	0
$p_A = (3,1,2)$	0	0	0	0	3	3
$p_A = (3,2,1)$	0	0	0	0	3	3

Table 4: Surplus Difference $S_{DA} - S_{BNE}$. Differences of 1 and 2 are due to inefficient allocations (given priority constraints) of slots in the BNE under the decentralized system; differences of 3 are due to the failure to allocate all slots.

Table 4 then takes the difference in S between tables 2 and 3. Notice that entries in table 4 are never negative and sometimes positive. This illustrates the superiority of the well-designed centralized matching system. Observe also that whenever there is a 3 in table 4, this means that a slot remains empty.[5] This was alluded to as inefficiency (1) above. Whenever there is a 1 or a 2 in the table, all slots are filled but the allocation of slots is inefficient given the constraints imposed by priorities: The slots are not allocated to the parents who value them the most. This is inefficiency (2).

[5]The probability of this happening was computed to be 2/9, and consistent with that we find a 3 in eight of the 36 cells.

Discussion of (and Motivation for) the Assumptions and their Implications A number of questions arise naturally in regards to the above model, which are addressed next.

- Uniform probability: This assumption is imposed for analytical ease as it allows us to analyze the model without carrying around additional parameters. It should be completely innocuous in the sense that the key insights – two sources of inefficiency under decentralized matching – should arise much more generally in equilibrium. (There may be a theorem out there.)

- Consumer surplus and welfare. Implicitly the model assumed that parents perceive different centres as heterogenous but that centres do not care about which parent (or child) fills their slots. In other words, centres care primarily about filling slots. This seems a sensible set of assumptions. Accordingly, the surplus in tables 2, 3, and 4 captures "consumer surplus" with consumers being defined as parents. If all centres get the same benefit $\pi > 0$ per slot filled, then there is an additional welfare loss of π whenever there is an empty slot in the decentralized system.

- The assumption that centres do not care about which parents they're matched to (they have priorities, much like public schools are typically assumed tohave) but parents care about which centres they are matched to, centres do not suffer "much" from an inefficient matching system (disregarding the transaction costs of running and clearing independent wait lists, which may be substantial, of course, but is not part of the model) but parents do. Accordingly, centres in high demand have little or no incentives to change a decentralized matching system. (Notice that high-demand could be modeled in the present setup by assuming that the outside option expires only after $t = 2$.) Parents are the ones that suffer from the inefficient system, but because of an enormous collective action problem (parents are transitory players, and each individual parent's benefit from changing the system is dwarfed by the cost of doing so), parental initiative is very unlikely to change the system.

- Multiple offers per round: The assumption that in each (or at least the first round) multiple offers are made may seem questionable at first; however, it captures the idea that fortunate enough parents may hold different exploding offers at some points in time. If they reject both, they will have to wait and hope for the next round offer, which is exactly what happens (off the equilibrium path) in the present setup. (However, nothing of substance is expected to change if only one offer were made each round; it is a little bit hard to see though how such coordination could occur within a decentralized system.)