Notes on Stable Matching

1 Introduction

Imagine a group of N boys and another group of N girls. Everyone wants to be matched with (one) member of the opposite sex. The problem is how to do it. If people didn't care who their partner is, then it is not hard to come up with a matching. For example, you could arrange the boys and girls in order of age and pair the oldest boy with the oldest girl, the second oldest boy with the second oldest girl, and so on. As long as no two members of the same sex have exactly the same birthday, this procedure assigns one (and only one) girl to every boy and one (and only one) boy to every girl. Of course, you could order people by their names, the length of their hair, their grade point average, or their wealth and also come up with matches.

One trouble with these approaches (the trouble that these notes will focus on) is that they ignore the fact that people have preferences. Boys like some girls better than others. Girls also care who they are matched with. One would like to tailor the matches so that they are somehow consistent with individual preferences. In this section, I am going to explain a way to do this that has nice properties.

I need to be precise about several things before I start. First, a **matching** is just an assignment of one (and only one) boy to each girl. When there are equal numbers of boys and girls, there are a lot ($N!$) of different assignments. I will say that a pair is **mated** if they are matched together. I will talk about a boy and girl being mated and so on.

Second, I must describe what preferences are. Here I assume that every individual has a preference ordering over the members of the opposite sex. That is, if I denote the boys by $B_1, B_2, \ldots, B_N$ and the girls $G_1, G_2, \ldots, G_N$, then the preferences of each individual are represented by an ordered list of members of the opposite sex. For example, B_1's preferences may be: $G_1 \succ G_2 \succ \ldots \succ G_N$, meaning that the first boy likes G_1 better than any other girl, likes G_2 second best, and so on. I will assume that everyone's preferences are strict (so that if you ask a girl: "Who do you prefer: B_i or B_j?" She'll give you a definite answer. She won't say: "I don't care."), but otherwise I will make no other assumption. If there is a universally accepted notion of attractiveness, then all of the boys will have the same preferences. I allow this possibility, but I do not require it.

Finally, I must come up with some condition that describes what it means to have a good match. Some ideas have superficial appeal, but clearly are too much to ask for. For example, you could say that a match is good if everyone is matched to his or her favorite choice. It should be clear that in general this is not possible. Even if there are only two boys and two girls, if both boys favor the same girl, then one of them will be disappointed. Another possibility is to say that a match is good if someone is happy. This condition rules out silly matches (for example, matching everyone to his or her least favorite, if that is possible), but in some sense is asking too little. Here is the condition, called

stability, that I propose. A match between boys and girls is stable if there does not exist a boy-girl pair (all them Ben and Jen) such that:

1. Ben is not paired with Jen.
2. Ben prefers Jen to his mate.
3. Jen prefers Ben to her mate.

Suppose that you could find a pair that satisfied the three properties. The matching is unstable in the following sense. Ben can approach Jen and suggest that she dump her current partner in favor of him. Ben hopes Jen will accept (because he likes her better than his current partner). Jen will accept (because she likes Ben better than her current partner). So, if a matching isn't stable, you will expect divorces. Notice that stability does not prevent someone from being stuck with his least preferred choice: Suppose that I am every girl's least favorite boy. If I am matched to my least favorite girl, then the first two conditions above hold between me and any girl I'm not matched with, but the third condition fails. When I request that someone break up with their mate to hook up with me, she'll turn me down.

With this introduction, the questions that I want to ask are: "Do stable matchings exist?" "How do I find them?"

2 An Algorithm

In this section I constructive show that stable matchings exist by describing an algorithm that arrives at a stable matching after a finite number of steps. The algorithm, as well as the formulation of the problem, is due to David Gale and Lloyd Shapley, mathematicians who work at UCB and UCLA respectively.

Gale-Shapley Algorithm:

Step 1. Each boy "proposes" to the favorite girl on his list.

Step 2. Each girl who receives at least one proposal, "dates" the boy she prefers among those who propose; "rejects" the rest. Girls with no proposals do nothing.

Step 3. If no boy is rejected, stop. You have a stable matching between girls and their current dates. Otherwise, rejected boys cross the name of the girl who rejected them off their list and then propose to the favorite among those remaining. Boys who are "dating" repeat their proposal.

Step 4. Return to Step 2.

I must show that the algorithm is well defined, that is, it is possible for everyone to follow the steps; that when it stops it provides a matching; that it is guaranteed to stop; and that when it stops the matching it provides is stable.

If a girl ever receives a proposal, then she has a date in each subsequent period (because she never rejects all of the boys who propose and the boy she is dating cannot propose to anyone else). The algorithm stops when each girl is dating exactly one boy (so that no boy is rejected). It follows that the algorithm must stop before any boy is rejected by every girl. (If a boy is rejected by all but one girl, then he proposes to the last girl on his list. At that time, the other girls must all be dating someone – the boy they preferred to the rejected boy or someone even better. So when the last girl gets a proposal, the algorithm stops.)

The algorithm is well defined because no boy is rejected by all of the girls (so there is always someone left to propose to) and because preferences are strict (so that there is always a favorite).

The algorithm ends at a matching because after the final round of proposals, no girl could have received more than one proposal (because that would lead to a rejection) and each boy must be dating at least one girl.

The algorithm ends in a finite number of steps because in each round (until the last one) at least one boy is rejected. No boy can be rejected more than $N-1$ times. Since there are N boys, in no more than $N(N-1)$ rounds the process must stop.

Finally, the algorithm arrives at a stable match. Suppose Ben is not married to Jen, but Ben prefers Jen to his mate Gwen. I must show that Jen prefers her mate to Ben (otherwise, the match will not be stable). If Ben prefers Jen to Gwen, then according to the algorithm, he proposed to Jen before he proposed to Gwen. So Jen must have rejected him. She would only have done that if some boy she preferred had also proposed to her. Gwen must end up mated with someone at least as good (according to her preferences) as the guy who was better than Ben. Hence she prefers her mate to Ben and the match must be stable.

Interestingly, not only does the algorithm come up with a stable matching, it comes up with a stable matching that is **boy optimal**: among all other stable matchings, it is the one that is unanimously preferred by the boys. That any matching should have this property is surprising: Different boys certainly have different preferences over **all** matchings. (For example, if two boys have the same favorite girl, then they disagree about which is the best matching.) On the other hand, if you limit attention to stable matchings, the boys all have the same interest. The second reason why the result is surprising is that the algorithm seems to give a lot of power to the girl. They can, after all, select who to reject. The result shows that it is the power to make offers, however, that is more valuable.

Proposition 1 *The Gale-Shapley Algorithm supplies a boy-optimal stable match.*

Proof For this argument, say that boy B_i is an eligible mate for G_j if there exists some stable matching in which they are married. We want to show that the algorithm assigns to each boy his most preferred eligible partner. Suppose, in order to reach a contradiction, some boy is paired with someone other than his best eligible partner. Since boys propose in decreasing order of preference,

some boy is rejected by an eligible partner. Let Lloyd be the first such boy, and let Amy be first valid partner that rejects him. When Lloyd is rejected, Amy must have available a boy, say David, whom she prefers to Lloyd. Let M be the stable matching in which Amy and Lloyd are mates. Let Beth be David's partner in the matching M. Now we can obtain a contradiction by showing that M is not a stable matching. By assumption, David was not rejected by any eligible partner at the point when Lloyd is rejected by Amy (since Lloyd is first to be rejected by an eligible partner). Thus, David prefers Amy to Beth. But Amy prefers David to Lloyd. The matching cannot M cannot be stable. This contradiction establishes the result. *Q.E.D.*

Proposition 2 *The Gale-Shapley Algorithm finds the girl-pessimal stable matching. (Each girl is married to worst eligible partner.)*

Proof Suppose in the Gale-Shapley Algorithm Amy is matched to David, but David is not the worst eligible partner for Amy. There exists stable matching M in which Amy is paired with Lloyd, whom she likes less than David. Let Beth be David's partner in M. David prefers Amy to Beth (by Proposition 1). Hence Amy and David prefer each other to their mate in M. M therefore cannot be a stable matching. This contradiction establishes the proposition. *Q.E.D.*

Proposition 2 is less surprising in view of Proposition 1. These results suggest perhaps the following question. Suppose that boys and girls know that matches will be made according to the algorithm. Is it in their best interest to made and respond to proposals honestly? It is possible to prove that if all of the girls behave honestly (rejecting boys they don't prefer in favor of those that they do prefer), then it is in the best interest of boys to behave honestly (after all, this leads to their most preferred stable match). Girls, on the other hand, might gain by dating someone who is not their best current option with the expectation that they'll eventually get someone they like even better. For example, if

BOY	1	2	3
Adam	Amy	Beth	Cara
Bill	Beth	Amy	Cara
Carl	Amy	Beth	Cara

Boys' Preferences

GIRL	1	2	3
Amy	Bill	Adam	Carl
Beth	Adam	Bill	Carl
Cara	Adam	Bill	Carl

Girls' Preferences

It is easy to check that the Gale-Shapley Algorithm arrives at the match: Adam-Amy; Bill-Beth; Carl-Cara if girls are honest. But look at what happen when Amy is dishonest in the first round. In the first round she gets proposals from Adam and Carl. Suppose that she rejects Adam (who she likes better than

Carl). Then Adam will propose to Beth, who will accept him and reject Bill. Bill then approaches Amy, who is now able to dump Carl. Eventually, Cara gets stuck with Carl and the match: Adam-Beth; Bill-Amy; and Carl-Cara results. This match is stable, but it is better for Amy and Beth (and worse for Adam and Bill) than the Gale-Shapley match.

3 Comments

You might view the discussion so far as an exercise in recreational mathematics. Good fun, perhaps, but with no insight into anything practical.

It is entertaining to try to draw sociological conclusions from the model. I am not prepared to make the claim that the results demonstrate that a match-making institution in which boys make proposals is better for boys and creates an environment where it pays for girls to misrepresent their true feelings. Nor does the mathematics provide serious foundation for the claim that heterosexual matching is more stable than homosexual matching. Pursue these lines of thought if they amuse you.

It turns out that the algorithm is truly useful in situations where people and matched to jobs (labor markets) or objects are matched to people (allocation processes like auctions). Before the problem was posed and solved mathematically, the national resident matching program used an equivalent procedure to match medical school graduates to hospital residency programs. In this application, the hospitals are the boys and have preferences over their candidates for residency appointments. The medical students were the girls. They have preferences over residency programs. The object is to match doctors to hospitals. The matching process was complicated and highly unstable with individual doctors and hospitals making and breaking deals outside of a structured environment until some one figured out a way to generate stable matchings. The environment is different from the boy-girl problem in several ways. Some of these are not important (the algorithm can be generalized to take them into account). For example, residents work in one hospital, but hospitals hire more than one resident (so the mating is polygamous). Some of the differences are harder to handle (for example, some medical students are married to medical students and they care not just about where they work, but whether their partner also has a nearby job). These considerations may make it impossible to find a stable matching.

Current research tries to take realistic considerations into account and find out when it is possible to arrive at stable assignments and when it is possible to come up with algorithms to compute them.

4 Postscript

This discussion is based on the article "College Admissions and the Stability of Marriage," by David Gale and Lloyd Shaply published in *The American*

Mathematical Monthly in 1962. Lloyd Shapley and Alvin Roth received the 2012 Nobel Memorial Prize in Economics for their work in this area. (David Gale died in 2007.) The article is seven pages long. This is how it ends:

> Most mathematicians at one time or another have probably found them- selves in the position of trying to refute the notion that they are people with "a head for figures," or that they "know a lot of formulas." At such times it may be convenient to have an illustration at hand to show that mathematics need not be concerned with figures, either numerical or geometrical. For this purpose we recommend the statement and proof of our Theorem 1. 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 because of unfamiliarity with the subject matter.
>
> What, then, to raise the old question once more, is mathematics? The answer, it appears, is that any argument which is carried out with sufficient precision is mathematical, and the reason that your friends and ours cannot understand mathematics is not because they have no head for figures, but because they are unable to achieve the degree of concentration required to follow a moderately involved sequence of inferences. This observation will hardly be news to those engaged in the teaching of mathematics, but it may not be so readily accepted by people outside of the profession. For them the foregoing may serve as a useful illustration.