


The Stable Marriage Problem and School Choice

This column will present the game-theoretic results contained in the original Gale-Shapley paper along with Roth's subsequent analysis. Pathak calls the deferred acceptance algorithm "one of the great ideas in economics," and Roth and Shapley were awarded the 2012 Nobel Prize in economics for this work...

David Austin 
 Grand Valley State
 University
 Email David Austin (/cgi-bin/fmail/fmail.cgi?emailto=52616e646f6d4956c534b2c4a31e4e85261db2017b0Austin)

Introduction

Every year, 75,000 New York City eighth graders apply for admission to one of the city's 426 public high schools. Until recently, this process asked students to list five schools in order of preference. These lists were sent to the schools, who decided which applicants to accept, wait-list, or reject. The students were then notified of their status and allowed to accept only one offer and one position on a waiting list. After the students had responded to any offers received, schools with unfilled positions made a second round of offers, and this process continued through a concluding third round.

This process had several serious problems. At the end of the third round of offers, nearly half of the students, usually lower-performing students from poor families, had not been accepted into a school. Many of these students waited through the summer only to learn they had been matched with a school that was not on their list of five schools.

This process also encouraged students and their parents to think strategically about the list of schools they submitted. Students that were rejected by the school at the top of their list might find that their second-choice school had no vacancies in the second round of offers. This made it risky for many students to faithfully state their true preferences, a view encouraged by the Education Department's advice that students "determine what your competition is" before creating their lists of preferred schools.

Lastly, schools would often underrepresent their capacity hoping to save positions for students who were unhappy with their initial offerings.

In the end, the process couldn't place many students while it encouraged all parties, both students and schools, to strategically misrepresent themselves in an effort to obtain more desirable outcomes not possible otherwise. Widespread mistrust in the placement process was a natural consequence.

Using ideas described in this column, economists Atila Abdulkadiroglu, Parag Pathak, and Alvin Roth designed a clearinghouse for matching students with high schools, which was first implemented in 2004. This new computerized algorithm places all but about 3000 students each year and results in more students receiving offers from their first-choice schools. As a result, students now submit lists that reflect their true preferences, which provides school officials with public input into the determination of which schools to close or reform. For their part, schools have found that there is no longer an advantage to underrepresenting their capacity.

The key to this new algorithm is the notion of *stability*, first introduced in a 1962 paper by Gale and Shapley. We say that a matching of students to schools is *stable* if there is not a student and a school who would prefer to be matched with each other more than their current matches. Gale and Shapley introduced an algorithm, sometimes called *deferred acceptance*, which is guaranteed to produce a

stable matching. Later, Roth showed that when the deferred acceptance algorithm is applied, a student can not gain admittance into a more preferred school by strategically misrepresenting his or her preferences.

This column will present the game-theoretic results contained in the original Gale-Shapley paper along with Roth's subsequent analysis. Pathak calls the deferred acceptance algorithm "one of the great ideas in economics," and Roth and Shapley were awarded the 2012 Nobel Prize in economics for this work.

The stable marriage problem

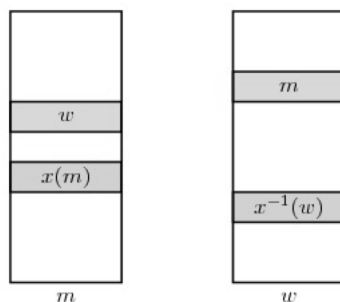
Besides matching students to schools, deferred acceptance has been applied in a wide variety of contexts, such as matching medical students to residency programs. In what follows, we will describe the algorithm within Gale-Shapley's original context, the stable marriage problem.

Suppose we have an equal number of men $M = \{m_1, m_2, \dots, m_n\}$ and women $W = \{w_1, w_2, \dots, w_n\}$. Every man lists the women in order of his preference, and every woman lists the men in order of her preference. We would like to arrange marriages between the men and women so that there is not a man and a woman who prefer one another to their spouses.

Before we go any further, let's acknowledge that our aim here is to model a mathematical problem. We will not, for instance, consider the realities of same-sex marriage, that individuals don't necessarily identify as either strictly male or female, and that women often propose to men. These issues all lead to mathematical situations that differ significantly from this one, which we hope to apply, more realistically, to the problem of matching students with schools.

Furthermore, it is relatively straightforward to extend this discussion to situations where there are an unequal number of men and women or where we allow polygamous matchings in which parties in one group may be matched with more than one party from the other group, which applies when, say, schools accept more than one student.

By a matching, we mean a one-to-one correspondence $x : M \rightarrow W$. A matching x is *unstable* if there is a man m and a woman w where m prefers w to $x(m)$ and w prefers m to $x^{-1}(w)$, as illustrated below. Otherwise, the matching is called *stable*.



In their 1962 paper, Gale and Shapely demonstrated that, given a set of preferences for every man and woman, there is always a stable matching; even better, they showed how to find a stable matching by applying the deferred acceptance algorithm, which we now describe.

- Step 1:** Every man proposes to the first woman on his list of preferences.
Every woman conditionally accepts the proposal from the man she most prefers out of those who have proposed. She rejects the other proposals.
- Step k:** Every man who is not conditionally engaged proposes to the woman he most prefers out of those who have not yet rejected him.
Every woman considers any new men who have proposed at this step and any man she had previously accepted and accepts the proposal from the man she most prefers, even if that means rejecting the man she had previously accepted.

End: This process continues until every woman has accepted a proposal at which time the conditional acceptances become final. At this step, the algorithm ends and $w = x(m)$ if w has accepted m .

Let's see how this works using an example provided by Gale and Shapley. Suppose there are four women $\{w_1, w_2, w_3, w_4\}$ and four men $\{m_1, m_2, m_3, m_4\}$ whose preferences are as shown below, in order from top to bottom.

w_1	w_1	w_2	w_4
w_2	w_4	w_1	w_2
w_3	w_3	w_3	w_3
w_4	w_2	w_4	w_1
m_1	m_2	m_3	m_4

m_4	m_2	m_4	m_3
m_3	m_4	m_1	m_2
m_1	m_1	m_2	m_1
m_2	m_3	m_3	m_4
w_1	w_2	w_3	w_4

Step 1: Each man proposes to the woman he most prefers:

- m_1 proposes to w_1
- m_2 proposes to w_1
- m_3 proposes to w_2
- m_4 proposes to w_4

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Notice that w_1 receives proposals from m_1 and m_2 . She chooses the proposal from m_1 since she prefers m_1 to m_2 .

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Step 2: Since m_2 has been rejected by w_1 , he proposes to his second choice w_4 .

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Now w_4 has proposals from m_2 and m_4 of which she chooses the one from m_2 .

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Step 3: m_4 proposes to w_2

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

who accepts the proposal and rejects m_3 .

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Step 4:

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Step 5:

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

Step 6:

	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

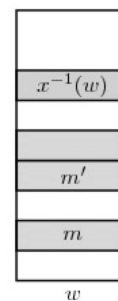
	w_1	w_2	w_3	w_4
m_1				
m_2				
m_3				
m_4				

We now arrive at the matching $x(m_1) = w_3, x(m_2) = w_4, x(m_3) = w_1$, and $x(m_4) = w_2$. Notice that if m proposes to w at one step of the algorithm and w' at a later step, then m must prefer w to w' . This means that m cannot propose to a woman twice, which implies that the algorithm will eventually terminate.

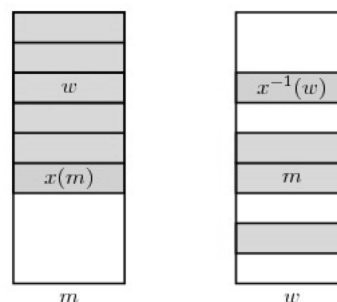
w
w'
$x(m)$
m

Also, we see that m proposes to every woman he prefers more than his match $x(m)$ before finally proposing to $x(m)$. That is, if m prefers w to $x(m)$, then w rejects m at some step of the algorithm.

Conversely, if w accepts m at one step of the algorithm and m' at a later step, then w prefers m' to m . This means that the men who propose to w and are rejected lie below $x^{-1}(w)$ on her list of preferences.



It is now easy to see that the matching x provided by the Gale-Shapley algorithm is stable. Suppose that a man m prefers a woman w more than his match $x(m)$. At some step of the Gale-Shapley algorithm, m proposed to w . Since w is not m 's ultimate match, however, she must have rejected m meaning she prefers $x^{-1}(w)$ to m . Therefore, it is not possible for m and w to prefer each other to their matches.



Comments?

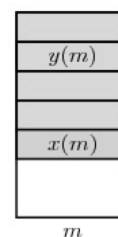
Optimal stable matches

How happy are the men and women with the matching x produced by the Gale-Shapley algorithm? For instance, is it possible to find a stable matching y where some men prefer their match in y to that in x ? Is there a stable matching y where some women prefer their match in y to that in x ?

We will see that no man will prefer his match in another stable match to his match in x . That is, if y is another stable match, then, for every man m , either $y(m) = x(m)$ or m prefers $x(m)$ to $y(m)$. We therefore call x an M -optimal stable matching. A moment's thought will convince you that there can be only one M -optimal stable matching.

Let's begin by assuming there is another matching y and a man m such that m prefers $y(m)$ to $x(m)$. We will see that this cannot happen if y is a stable matching.

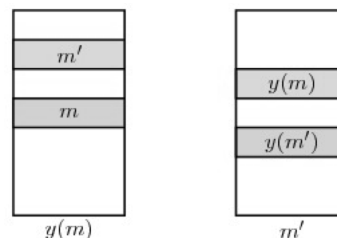
In this case, we will consider the steps of the Gale-Shapley algorithm that lead to the stable matching x . Since m prefers $y(m)$ to $x(m)$, then $y(m)$ must reject m at some step k of the algorithm. Of all such m , choose one such that no man m' has been rejected by $y(m')$ at an earlier step $k' < k$.



Since $y(m)$ rejects m , she must prefer a man m' who proposes to her at step k . Since m' is accepted by $y(m)$ at step k and has not been rejected by $y(m')$ at a step before k , m' must prefer $y(m)$ to $y(m')$.

Therefore, $y(m)$ prefers m' to m and m' prefers $y(m)$ to $y(m')$ meaning that y is not a stable matching.

In this way, we see that x is the M -optimal stable matching; every man is at least as happy with his match in x as he is in another stable match y .



Simple examples show that the matching x is not always optimal for the women. There is, however, a W -optimal stable matching: we simply apply the Gale-Shapley algorithm with the roles interchanged so that women propose to men.

Revealing one's true preferences

We have now seen that it is possible to create stable matchings with the Gale-Shapley algorithm and that these matchings are even the best possible for one set of participants. So far, we have assumed that each person represents his or her true preferences accurately. However, it seems possible that a man may attempt to obtain a more desirable match by misrepresenting his true preferences.

We will show that revealing one's true preferences is, in the language of game theory, a *dominant strategy* for men. This means that, assuming all other men and women keep the same preferences, a man cannot obtain a more desirable match by misrepresenting his preferences.

By P , we mean the set of preferences of all men and women, which we now assume to be their true preferences. We wish to consider the set of preferences P' , which are identical to P except for one man m_m , whom we call the "manipulator."

Also, we will use $x = GS(P)$ to denote the stable matching that results from the Gale-Shapley algorithm applied with the true preferences P while $y = GS(P')$ is the stable matching resulting from the manipulated preferences. Our goal is to show that m cannot improve his match; in other words, we will show that either $y(m) = x(m)$ or m prefers, according to his true preferences, $x(m)$ to $y(m)$.

We first show that we only need to consider a specific kind of misrepresentation. Remembering that $y = GS(P')$, we will consider *simple* equivalent misrepresentations P'' having the property that the manipulator m_m prefers $y(m_m)$ to all other women, according to his preference list in P'' . That is, the manipulator m_m obtains P'' from P' by simply moving his match in y to the top of his list. We will denote by $z = GS(P'')$, the stable matching that results from the simple misrepresentation P'' .

We make the following observations:

1. $y(m_m) = z(m_m)$. This means that the misrepresentation P'' leads to the same match for m_m as the misrepresentation P' . This is why we say that P' and P'' are equivalent misrepresentations. If m_m does not obtain a more desirable match in z , he won't in y either.

This is fairly straightforward to see. Remember that y is stable with respect to the preferences P' since y is obtained by the Gale-Shapley algorithm. This means that no man m and woman w prefer one another (in P') to their match in y . However, this implies that y is stable in P'' as well; the only party whose preferences have changed is m_m and, since he has moved $y(m_m)$ to the top of his list in P'' , he finds his match under y with respect to P'' at least as desirable as he does with respect to P' .

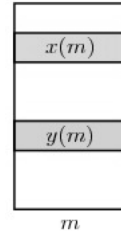
Remember now that z is the M -optimal stable matching with respect to P'' . Since y is also stable with respect to P'' , it follows that m_m prefers $z(m_m)$ at least as much as he does $y(m_m)$. However, he prefers $y(m_m)$ (in P'') more than any other woman so it must follow that $z(m_m) = y(m_m)$.

2. We will now assume that P' is a simple representation by a manipulator m_m and that $y = GS(P')$. We also assume that m_m misrepresents his true preferences to obtain a match $y(m_m)$ he prefers, in his true preferences P , at least as much as $x(m_m)$. With these assumptions, no man

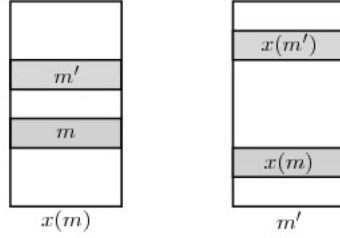
does worse in y than he does in x with respect to his true preferences. That is, for every man m , either $y(m) = x(m)$ or m prefers $y(m)$ to $x(m)$.

Notice that we are assuming that the manipulator fares no worse under y . In fact, he misrepresents his preferences hoping to improve his match.

Now suppose that there is a man m who fares worse in y than in x ; that is, suppose that m prefers $x(m)$ to $y(m)$ as shown. Since m is not the manipulator, m has the same preferences in both P and P' . Therefore, m is rejected by $x(m)$ at some step in $GS(P')$.



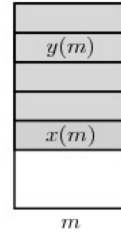
Let k be the first step in which some man m is rejected by $x(m)$ in $GS(P')$. More specifically, suppose that m is rejected by $x(m)$ in favor of m' in step k . This means that $x(m)$ prefers m' to m so it is not possible that m' proposed to $x(m)$ in $GS(P)$. Therefore, m' prefers $x(m')$ to $x(m)$.



Since m' proposes to $x(m)$ in step k of $GS(P')$, he must have been rejected by $x(m')$ at an earlier step of $GS(P')$. However, this is not possible since we assumed k is the first step in which a man m is rejected by $x(m)$ in $GS(P')$. It therefore follows that m prefers $y(m)$ at least as much as $x(m)$.

3. Since every man m is at least as happy with $y(m)$ as with $x(m)$, it follows that if m proposes to a woman w in $GS(P')$, then he must propose to her in $GS(P)$ also.

From this observation follows the useful fact that if w only receives one proposal in $GS(P)$, then she only receives one in $GS(P')$ as well. If that proposal comes from m , then $y(m) = x(m) = w$.



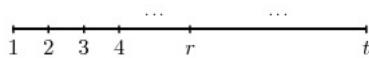
We have now assembled everything we need to explain why revealing one's true preferences is a dominant strategy for men. We assume that all the other men and women keep the same preferences, and we will show that the manipulator does not improve his match: either $y(m_m) = x(m_m)$ or m prefers $x(m_m)$ to $y(m_m)$.

We will begin by assuming that, for the manipulator, either $y(m_m) = x(m_m)$ or he prefers $y(m)$ to $x(m_m)$ since the result is certainly true if m_m prefers $x(m_m)$ to $y(m_m)$.

With this assumption, we know that no man does worse in y than in x , as we explained above. Also, if m proposes to a woman w who receives only one proposal, then $y(m) = x(m)$.

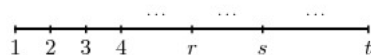
Looking at the steps of $GS(P)$, the manipulator will propose to the woman $x(m_m)$ he is matched with at some step r . Our strategy is to show that every man m who proposes to $x(m)$ at step r or later has $y(m) = x(m)$. This will then show that $y(m_m) = x(m_m)$.

We proceed by first looking at a man m who proposes to his match $x(m) = w$ in the last step t of $GS(P)$.



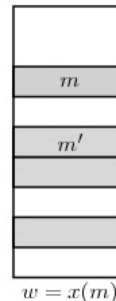
It follows that m is the only man who proposes to w . If not, she would have to reject a man whom she had previously accepted conditionally and that man would need to make another proposal in a subsequent step. Therefore, $y(m) = x(m)$.

We now consider a man who proposes to his match $x(m) = w$ at a step s where $r \leq s < t$, and we make the inductive assumption that any man who proposes to his match at step $s + 1$ or later satisfies $y(m) = x(m)$.

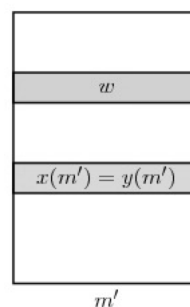


We then look at the set of men \bar{M} that w has rejected before accepting the proposal from her match m . If there are no men that w has rejected, then m is the only man that proposes to her and hence $y(m) = x(m)$.

If there are some men she has rejected, denote by m' the man she most prefers among that group. This man m' is the one she rejects at step s to accept m . Therefore m' proposes to his match at a step later than s , which means that $y(m') = x(m')$ by the inductive assumption.



Since m' proposes to his match at a later step, m' is not the manipulator so m' has the same preferences in both P and P' . This means that m' also proposes to w in $GS(P')$ and is ultimately rejected. Hence, w receives another proposal in $GS(P')$.



We have seen that if m makes a proposal to a woman in $GS(P')$, then he also does so in $GS(P)$. Since the only other proposal w receives in $GS(P)$ is from m , this means that the final proposal she receives in $GS(P')$ is also from m . From this, we conclude that $y(m) = x(m)$.

Therefore, any man who proposes to his match at step s must have $y(m) = x(m)$. By induction, this shows that every man who proposes to his match at step r or later must have $y(m) = x(m)$. Since the manipulator proposes to his match at step r , this means that $y(m_m) = x(m_m)$. In other words, the manipulator does not improve his match through his misrepresentation.

Curiously, examples show that it is possible for other men who propose to their match before step r to improve their match. The misrepresentation by the manipulator m_m does not improve his own match, but it might improve the match of others!

Comments?

Summary

Returning to the question of matching students to schools, we should ask which group will be allowed to make the proposals. Running the algorithm with students making the proposals gives the students no incentive to misrepresent their preferences though schools may have an incentive to do so. We would, however, expect that a school's misrepresentations would be generally applied to students and, hence, more easily detected. Schools may also be compelled, perhaps through transparency requirements or other legal means, to avoid certain types of misrepresentation, such as racial discrimination.

Indeed, one of Roth's collaborators, Atila Abdulkadiroglu, says he has received calls from parents looking for inside advice on how to better their student's match. His reply is simple: "Rank them in true preference order."

Stability is an intuitively appealing condition to impose on a matching system, such as the New York City school application process. In a stable match, no party has an incentive to seek a different match; for instance, any woman preferred by a man to his wife is not available to him since she prefers her husband to him.

One may ask, however, if evidence supports the importance of the role played by stability in matching systems. Indeed, Roth and his collaborators have studied many different markets with

this question in mind. For instance, the process by which American medical students are matched to residency programs was modified in the 1950s to one that closely resembles the deferred acceptance algorithm. This system has proved to be remarkably successful, and, as a result, a British Royal Commission recommended each region of the British National Health Service adopt a similar system.

As it turns out, each region adopted a slightly different matching algorithm since the details of the American system were not described in the American medical literature. The systems in some regions thrived while those in other regions failed. Roth and his collaborators determined that the stable systems were the ones that succeeded while the unstable ones failed.

About the application of game theory to economics, Roth writes that "the real test of our success will be not merely how well we understand the general principles which govern economic interactions, but how well we can bring this knowledge to bear on practical questions of microeconomic engineering." Indeed, examples like the New York City high school application process give testimony to this success.


Comments?

References

- D. Gale, L.S. Shapley. College Admissions and the Stability of Marriage. *American Mathematics Monthly*. **69**, 9-15. 1962.
The original paper introducing the deferred acceptance algorithm.
- It is a truth universally acknowledged that this Numberphile video presents an amusing example (<https://www.youtube.com/watch?v=Qcv1IqHWAzg>) of the stable marriage problem using characters from Jane Austen's *Pride and Prejudice*.
- Alvin E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operation Research*. Vol. 7 (4). 1982.
Here, Roth proves that representing one's true preferences is a dominant strategy for men.
- Alvin E. Roth. What Have We Learned from Market Design? *Innovation Policy and the Economy*. Vol 9, 79 - 112. 2009.
An interesting, well-written survey article that highlights several situations in which the ideas in this column have been applied.
- Alvin E. Roth. Misrepresentaiton and Stability in the Marriage Problem. *Journal of Economic Theory*, **34**, 383-387. 1984.
- Alvin E. Roth. The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy*. Vol. Y2 (6). 1984.
- Alvin E. Roth, Elliott Peranson. The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *The American Economic Review*. Vol. 89 (4). 1999.
- Stable matching: Theory, evidence, and practical design.
(http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2012/popular-economicsciences2012.pdf) *The Prize in Economic Sciences 2012*. The Royal Swedish Academy of Sciences.
A short survey of the impact of Roth and Shapley's work.
- How Game Theory Helped Improve New York City's High School Application Process.
(<http://www.nytimes.com/2014/12/07/nyregion/how-game-theory-helped-improve-new-york-city-high-school-application-process.html>) *New York Times*. December 5, 2014.

Comments (0)

This comment form is powered by **GentleSource**
Comment Script
(<http://www.gentlesource.com/comment-script/>). It can be included in PHP or HTML files and allows visitors to leave comments on the website.

David Austin 
Grand Valley State
University
Email David Austin (/cgi-bin/fmail/fmail.cgi?emailto=52616e646f6d4956c534b2c4a31e4e85261db2017bAustin)

