COLLEGE ADMISSIONS AND THE STABILITY OF MARRIAGE

By D. Gale and L. S. Shapley
Presented by Lucy Chang, Allison Jaros, Erin Lett, and Morgan Lucas (Coase Team)
February 28, 2012

I. Introduction

Gale and Shapley focused on a common problem faced by colleges based on their usual admissions procedure—namely, how to admit the ideal number of best-qualified applicants based on a specific quota without knowing precisely how many admitted applicants will accept. For example, the authors note the college may not know whether the applicant has applied elsewhere and, if so, how she ranked the colleges or whether the other colleges will admit her. As such, the college can likely only admit applicants of a number and quality that are "reasonably close" to their ideal.

The decentralized application process precludes coordination by colleges and students. Colleges cannot compare information on students to ascertain the likelihood that an applicant truly prefers their school to all others. Therefore, they must gamble on how many of their top applicants will ultimately decide to attend their college. To achieve their ideal class size, colleges are forced to estimate a certain level of over-admittance, which may result in class sizes that are too large. This creates a welfare loss to the college because they must expend resources to support a class that is too big. Further, there is a chance that, to comfortably estimate their ideal size, there will be a slight loss in student quality depending on the quality of the applicant pool. There is also a welfare loss to students: Had student preferences been perfectly known, some students may have been able to gain admission to a preferred college based on the fact that other students preferred to go elsewhere.

In response to this problem, colleges may use wait lists or ask applicants to list all colleges to which they have applied in order of preference, but the authors contend that these solutions simply create new problems. Consider the first option: Even if a student gains admission to their preferred school after having been placed on a waitlist, they may have already accepted admission to a less-preferred school; in which case the welfare loss is entirely borne by the student who still attends the less-preferred school. If a student, however, reneges on the less-preferred school, some of the welfare loss may be placed on that school (which must fill that spot to achieve their ideal class size). Under the second option, applicants may rationally feel that their chances of admission will be improved by listing every college as their first choice on each application. Assuming the colleges do not share information and colleges reward applicants for listing the college as their first choice, an applicant may generate a welfare loss: She will only attend one college, but has now taken valuable slots at a number of schools. Further, if she does not immediately reject these others, she may hurt the quality of the other colleges' classes, as students that should have been matched to their preferred college go to a less-preferred colleges to which they gained admission earlier (i.e., these students chose not to wait and see if they could get into a preferred college).

In this paper, the authors ultimately propose a new deferred acceptance procedure which removes uncertainties and, provided there are enough applicants, gives each college their ideal class size. They also consider the model using the example of marriages to demonstrate the availability of stable equilibria. Through these examples, the authors demonstrate how important it is to think about policy questions in the design of the matching process. Do we want students or colleges to achieve a better result from the application process? Should men benefit more than women from the marriage selection process? As the authors illuminate, how we answer these questions makes a difference in how we should design matching processes.

In this presentation, we add extensions to these ideas and consider new situations in which this model may prove insightful. Importantly, as noted in the authors' concluding remarks, this model is mathematical but not formula- or calculation-intensive: In describing the model as mathematical, the authors highlight that mathematics is simply an effort to argue with "sufficient precision" through a "moderately involved sequence of inferences." Thus, in offering their mathematical model and the accompanying illustrations, the authors indicate that, at its core, mathematics is simply a logical assessment of a problem engaged with sufficient precision to generate useful conclusions.

II. Definitions and Assumptions

In designing their model, the authors make a few assumptions that carry throughout their paper. The first of these is that there are no ties in how students and colleges or men and women rank their preferences. Additionally, every student and college, and man and woman, ranks every college and student, and woman and man, omitting only those which they would never consider under any circumstances.

A. <u>Definition: An assignment is unstable if there are two applicants (men) who are assigned to colleges (women), although an applicant (man) prefers the college (woman) to which the other applicant is assigned and the college (woman) he prefers also prefers him.</u>

If an assignment is unstable, it means that the assignment can be "upset by a college [man] and applicant [woman] acting in a manner which benefits both." In designing their model, the authors require that the end result not exhibit instability.

B. <u>Definition: An assignment is optimal if every applicant (man) is at least as well off under it as under any other stable assignment.</u>

By definition, if there is an optimal assignment it is unique. Therefore, the authors seek to design a method of assignment that will achieve a stable and optimal assignment.

III. The Model in the Context of Marriage

In their model, the authors demonstrate how the use of a deferred acceptance procedure will lead to a stable set of outcomes that is optimal for the "proposing parties" or applicants in a two-pool matching game. In the context of marriage, the deferred acceptance procedure works in the following manner. After each woman and each man ranks each member of the opposite sex, the men (or the women) propose to their favorite woman (or man). Since in most cases some men (or women) will have the same favorite, some women (or men) will receive multiple proposals. Therefore, after every round of proposals, each woman holds on to her best proposal and rejects the others. In the second round, the men who were rejected in the first round propose to their second choice. The process continues until every woman has received a proposal.

A. Theorem 1: There always exists a stable set of marriages.

Using the deferred acceptance procedure, as long as neither men nor women misrepresent their preferences, and as long as preferences remain stable throughout the process, there will always exist a stable set of marriages. Intuitively, this makes sense. If a man would rather be with a particular woman than his wife, it has to be a woman who has already rejected him. Similarly, if a woman would rather be with a particular man than her husband, the preferred man is currently with a woman to whom he would rather be married.

The process of reaching a stable outcome requires no more than n^2-2n+2 iterations, where n is the number of women and men when there is an equal number of women and men. Additionally, it can result in more than one stable outcome, depending on whether men or women propose first. For instance, **Example 1** on page 11 only requires one iteration whether men or women are proposing and it has three stable outcomes. It is as follows:

	A	**B**	**C**
α	1,3	*2,2*	3,1
β	3,1	1,3	*2,2*
γ	*2,2*	3,1	1,3

Where A, B, and C are women and α, β, and γ are men. The matrix lists man's ranking of women first and woman's raking of men second. If the game is played out with men proposing, all of the men get their first choice, which is indicated by the shaded cells:

	A	**B**	**C**
Round 1	α	β	γ

This is a stable solution because every man has his first choice and would have no reason to change. A similar result happens if the women propose, with the solution indicated by the cells in bold:

	α	**β**	**γ**
Round 1	C	A	B

Again, this is stable because none of the women have any reason to change husbands. Notice that when women propose, they get their first choice and men get their last choice and when men propose, the reverse occurs. In addition to these two schemes, however, there is another stable outcome that comes from everyone getting their second choice, denoted in italics. This is optimal because no woman (or man) could be made better off by proposing to a man (or woman) that would accept her (or him).

Example 2 in the article shows a situation in which there is only one stable solution and the solution yields a situation in which neither men nor women get their first choice. This solution is marked by the shaded boxes.

	A	**B**	**C**	**D**
α	1,3	2,3	3,2	4,3
β	1,4	4,1	3,3	2,2
γ	2,2	1,4	3,4	4,1
δ	4,1	2,2	3,1	1,4

If men propose to women, the game plays out as follows:

	A	**B**	**C**	**D**
Round 1	α β	γ		δ
Round 2	α	γ		δ β
Round 3	α	δ γ		β
Round 4	α γ	δ		β
Round 5	γ	α δ		β
Round 6	γ	δ	α	β

If women propose to men, the same result is achieved but the iterations are different:

	α	**β**	**γ**	**δ**
Round 1		B	D	A C
Round 2		B	D A	C
Round 3		D B	A	C
Round 4		D	A	C
Round 5		D	A	C B
Round 6	C	D	A	B

Note that neither of the previous two examples required the maximum number of iterations. Since n=4, the maximum number of iterations is (4*4) – 2(4) + 2 = 10. For the maximum number of iterations to occur, at least one person has to get their last choice and none of the others can get their first choice. **Example 3** on page 13 demonstrates this situation. In it, there is only one stable outcome, which is marked by the shaded boxes.

	A	**B**	**C**	**D**
α	1,3	2,2	3,1	4,3
β	1,4	2,3	3,2	4,4
γ	3,1	1,4	2,3	4,2
δ	2,2	3,1	1,4	4,1

Starting with the situation in which the men propose, the game plays out as follows:

	A	B	C	D
Round 1	α β	γ	δ	
Round 2	α	β γ	δ	
Round 3	α	β	γ δ	
Round 4	α δ	β	γ	
Round 5	δ	α β	γ	
Round 6	δ	α	γ β	
Round 7	δ γ	α	β	
Round 8	γ	α δ	β	
Round 9	γ	δ	α β	
Round 10	γ	δ	α	β

If the women propose, the game ends the same but the number of iterations is reduced to 4:

	α	β	γ	δ
Round 1	C		A	B D
Round 2	C		A D	B
Round 3	C D		A	B
Round 4	C	D	A	B

Until now, all of the examples that have been presented have had equal numbers of men and women. If there are an uneven number of men and women, when b>g, the deferred acceptance procedure ends when every man has either been rejected or is the only man on some woman's string. When b<g, the game ends when b women have been proposed to. To see this, consider a variation on **Example 2**. First, remove woman D so that there are 4 men and 3 women. The rankings table becomes:

	A	B	C
α	1,3	2,3	3,2
β	1,4	3,1	2,3
γ	2,1	1,4	3,4
δ	2,2	1,2	3,1

If men propose, α is left single, with the results indicated in the shaded cells.

	A	B	C
Round 1	α β	γ δ	
Round 2	α γ	δ	β
Round 3	γ	δ α	β
Round 4	γ	δ	β α
Round 5	γ	δ β	α
Round 6	γ	β	α δ (α is rejected)

If women propose, the results are the same but there is only one iteration.

	α	β	γ	δ
Round 1		B	A	C

It is easy to see that women are generally better off when there are fewer of them than there are men. In this case, every woman ends up with her first choice and there is only one iteration. A good demonstration of this principle is done by taking out woman D in the second example. Although we will leave this one for you to play out, the results are:

	A	**B**	**C**	**D**
α	*1,3*	2,3	3,2	4,3
β	1,4	4,1	3,3	2,2
γ	2,2	*1,4*	3,4	4,1
δ	4,1	2,2	3,1	1,4

Where the shaded cells again represent the only stable outcome when there are four men and four women. The cells in bold demonstrate the pairings if woman D is taken out. Similarly, if one of the men is removed, men are better off than they would be if there were an equal number of men and women. If δ is taken out, the result is denoted by the italicized cells.

From the previous examples, it is clear that it makes a difference whether men propose or women propose. In fact, if men propose, they will be strictly better off than if women propose, which leads to theorem 2.

B. <u>Theorem 2: Every proposer is at least as well off under the deferred acceptance procedure as he would be under any other stable assignment.</u>

To see this, consider **Example 4** (not in text) with 5 men and 5 women. There are two stable outcomes in this:

	A	**B**	**C**	**D**	**E**
α	1,5	5,1	2,3	4,5	3,1
β	4,3	5,2	1,4	2,4	3,2
γ	3,2	2,3	1,5	4,3	5,3
δ	2,4	1,5	5,2	3,2	4,4
ε	3,1	1,4	2,1	4,1	5,5

Under the deferred acceptance procedure, if men are proposing, the result is indicated by the shaded cells. When women propose, the result is indicated by the cells in bold. These results demonstrate that men are better off when they propose to women than they are under the other stable result where women propose.

This can be seen again from **Example 1** above, where there are three stable solutions:

	A	**B**	**C**
α	1,3	*2,2*	3,1
β	3,1	1,3	*2,2*
γ	*2,2*	3,1	1,3

The first solution, where men propose, is indicated by the shaded cells; the second solution, where women propose, is indicated by the cells in bold; and the third solution, where everyone gets their second choice, is indicated by italicized cells. Once again, men are better off when they are the ones proposing than they are under any other stable solution.

Intuitively, this makes sense. If every man has a different first choice, the men will get their first choice under the deferred acceptance procedure, in which case, they cannot be better off under any other solution. If some men have the same preferences, then there must be more than one iteration. In the end, however, each man ends up with his highest-attainable choice. It is impossible for him to do better under a stable solution because every woman whom he prefers to the woman he is with has already rejected him (meaning that she is unattainable).

C. Misrepresentation of Preferences

Until now, we have assumed that no one misrepresents their preferences. If, however, misrepresentation does occur, it is possible that theorems 1 and 2 will not hold. In the case of marriages, a man can be made better off at the end of the deferred acceptance procedure when he is proposing by misrepresenting his preferences. If he does, though, the result will be unstable. Therefore, couples will get divorced and in the long run, a man who misrepresents his preferences will not end up better off. On the other hand, when men are the ones proposing, women can achieve a better result by misrepresenting their preferences. Since the best stable outcome that they can achieve is the one that would occur if they are the ones proposing, this is also the best outcome that they can achieve through misrepresentation. This concept will be explored further in the context of the college admissions process below.

III. Limitations of the Deferred-Acceptance Procedure

In presenting their argument for the deferred-acceptance procedure, Gale and Shapley make simplifying assumptions about the preferences of the agents (the colleges and the applicants or the men and the women). In particular, they posit: (1) fixed and exogenous preferences, (2) strict preferences, and (3) college preferences that treat applicants akin to gross substitutes. Before considering the application of the deferred-acceptance procedure outside the "world of mathematical make-believe," it will be helpful to question these three features of the model.

A. Fixed and Exogenous Preferences

First, Gale and Shapley implicitly assume that the preferences of the agents are fixed prior to the matching process. If this were not the case, Theorem one would fail; the deferred-acceptance procedure would not necessarily lead to a stable outcome. For example, if woman D in Example 2 did not realize that she preferred δ as a husband to γ until after she rejected δ the end result would be unstable.

The assumption of fixed preferences are unrealistic for several reasons. Love is fickle. The matching process might also be an information sharing process during which preferences are formed. To give one example, the individualized cost of attending a particular college is not often known to an applicant until late in the application process (because of scholarships, grants, and financial aid few students pay full undergraduate tuition out of pocket). Therefore, for price sensitive students, preferences are likely indeterminate until late in the application process.

Moreover, by treating preferences as fixed, Gale and Shapley encourage the reader to think of each agent's preferences as exogenous to the other agents' preferences. In other words, the preferences of woman A are not influenced by the preferences of woman B or man α in the marriage problem. (This is not a necessary implication of Theorem 1 and 2 taken alone. It follows from the natural assumption that preferences are not transparent prior to the matching process. Gale and Shapely suggest this when they note that colleges are not likely to know how applicants rank other colleges or which colleges will admit the applicants. If agent preferences are not transparent and are dependent on each other, then the preferences are likely to change during the matching process undermining the stability of the result). Here are some examples of endogenous preferences:

- In a pool of men seeking to marry, there exists an "alpha" man whose preferences determine the preferences of the other men.
- Men and women in the marriage market prefer to marry someone who prefers to marry them. For example, a woman may be insulted to learn she was a man's third choice. This may cause her to rank him lower in her preferences.
- A student may want to attend a college where her friends are going.

B. Strict Preferences

Second, Gale and Shipley assume that agents have strict preferences, i.e. there are no ties. If an applicant is indifferent between two or more colleges, she nonetheless must rank them. This is a simple way of handling weak preferences. As long as each agent randomly "splits" any ties, a stable outcome will result from the deferred-acceptance procedure.

However, a more expansive definition of stability may be appropriate if weak preferences exist. To illustrate this point consider the following example: A and B are two colleges and α and β are the applicants. Each college has a quota of one so that the admissions problem is directly analogous to a marriage problem.

Actual Preferences (with ties denoted by the same number)

	A	**B**
α	1,1	1,1
β	1,1	1,2

A Possible Result under the Gale Shapely Deferred-Acceptance Procedure

	A	**B**
α	1,1	2,1
β	2,2	1,2

In this case, α, β, and A are all indifferent among the possible outcomes. However, if the applicants and college with weak preferences are forced to randomly rank these preferences, the deferred-acceptance procedure may result in college B ending up with its second choice. Although the outcome is not unstable according to the definition given by Gale and Shapley, the situation appears "unstable" because it is possible to make B better off without harming any other agent. In such a situation, B could bribe the other agents to switch outcomes.

To account for weak preferences it may be desirable to adjust the matching procedure to maximize each agent's outcome given that some agents are indifferent. To do this it is necessary to create a matching procedure that forces the proposing set of agents to reveal more information about their preferences. In the extreme case, in which all agents are indifferent to all outcomes, the matching problem looks like a problem of allocating a homogenous consumer good among purchasers (or a homogenous input among producers).

However, extending the model to acknowledge weak preferences undermines the existence of an optimal assignment. For example, if there are two men and two women in a marriage problem and both women prefer the same man, but he is indifferent between them. There will be two stable outcomes, but neither outcome will be optimal for both women. It is a zero sum game where only one woman can have the preferred husband.

C. Colleges Treat Applicants As Supplements

Finally, the Gale and Shapley model ignores any preferences the college may have about the composition of the incoming class. In other words, applicants are not treated as complementary inputs. The problem with this approach is clear if you consider the case of an adoption problem. Consider an agency that matches children with adoptive parents. Suppose there are parents who would like to adopt one boy and one girl at the same time. How should they rank the available children? To solve this problem, the agency could allow the parents to rank boys and girls separately and match them accordingly. This is analogous to a college using a quota system in order to obtain a desired ratio of women or minorities as part of its incoming class.

D. Favoring Applicants

In addition to making assumptions about agent preferences, Gale and Shapley state that it is fitting for applicants to receive more consideration than colleges. They argue that "colleges exist for the students." A policy maker who believes the essential function of colleges is to produce an educated workforce or further technological development may disagree. The students can be viewed as inputs rather than consumers. In this case it makes more sense to give greater consideration to the preferences of the colleges.

IV. The Deferred-Acceptance Procedure and the College Admissions Problem

As stated by Gale and Shapley, Theorem 1 and 2, which were discussed above in the context of the marriage problem, hold for the college admissions context. Using the deferred-acceptance procedure results in a stable assignment of students which is optimal for the proposing side. In a procedure in which the colleges bid on students, this means that each college receives its most preferred achievable students, where the achievable students are the set of students that are assigned to the college in the stable outcomes.

A. Roth Critique

The Gale and Shapley paper elicited critique for asserting that the marriage and college admissions problem are equivalent. Alvin Roth wrote a paper discussing this issue. The example used below is borrowed from him. He notes that in a marriage problem it is possible to show that there is no unstable outcome that all the proposing group prefers to the optimal outcome resulting from the deferred-acceptance procedure. Likewise, the proposing group does not have an incentive to lie about its preferences. However, neither of these claims hold in the case of college admissions.

Unlike a marriage problem, colleges admit multiple applicants. However, the deferred-acceptance model only accounts for a colleges preferences over applicants, not its preferences over outcomes. It is impossible to determine whether a college with a quota of three would prefer to receive (s_1, s_6 , s_7) or (s_3, s_4, s_5), where s_i indicates the student ranked in slot i. Roth argues that an understanding of the college admission problem is incomplete without a consideration of these outcome preferences. A brief example shows that a college with "responsive" outcome preferences might have an incentive to misrepresent its preferences during the deferred-acceptance process. A college has responsive preference if for any two outcomes that differ in only one student, the college prefers the outcome with the preferred student.

Let there be three colleges, A, B and C and four applicants α,β,γ, and δ.

	A	**B**	**C**
α	2,1	3,1	1,2
β	2,2	1,2	3,3
γ	1,3	3,3	2,1
δ	1,4	2,4	3,4

Quotas: A=2, B=1, C=1

In a deferred-acceptance procedure in which the colleges bid on the students, the outcome is shaded below.

	A	B	C
α	2,1	3,1	1,2
β	2,2	1,2	3,3
γ	1,3	3,3	2,1
δ	1,4	2,4	3,4

College A receives its third and fourth choice, while the other colleges receive their second choices. However, if college A were to lie about its preferences during the game by ranking them as follows all the colleges would be better off (assuming A has responsive preferences over outcomes).

	A	B	C
α	2,3	3,1	1,2
β	2,1	1,2	3,3
γ	1,4	3,3	2,1
δ	1,2	2,4	3,4

This example shows that a college may have an incentive to misrepresent its preference during a matching game. The resulting solution is not stable but it is preferable to all the colleges. This situation can be intuitively explained by an agreement among the colleges not to compete for the most desirable students. This takes control away from the students and allows the colleges to improve their situation.

B. Practical Implications

Gale and Shapley do not propose any practical applications of their model. However, the early admission process employed by some colleges may be a form of the deferred-acceptance procedure. Colleges use this process as a way of encourage students to apply to their top ranking school first. During the in-class presentation, we will consider whether the early admission process encourages a better matching outcome.

V. Extensions

A. The Employment Market: Medical Interns and Hospitals

One obvious, and extensively written about, application is to the matching of medical residents to hospitals. This is similar to the Gale–Shapley college-admissions problem in that hospitals can accept more than one medical resident (not a strict one-to-one matching like in the stable-marriage problem). However, various extensions of the Gale–Shapley model have been explored using the medical intern context, some of which are discussed below.

After encountering a variety of problems early in the 20th century (e.g., competition for interns advanced the appointment date to two full years before the beginning of an internship; timing of offers and acceptances), a centralized market mechanism was adopted for assigning medical interns to hospitals in the 1950s—the National Intern Matching Program. Graduating medicals students now rank hospitals in order of preference and hospitals ranks medical students in order of preference. These preferences are submitted to a central office which uses an algorithm to match students to hospitals.[1] Like the Gale–Shapley model, a given outcome is unstable if some student or hospital receives an unacceptable assignment or if there exists a hospital program and student who each prefer the other to their assignment.

i. *Extension #1: Group Ordering*[2]

Suppose that students give each individual hospital a ranking, but hospitals rank students by group: a group of *q* students are given ranking 1, a group of *q* students are given ranking 2, and so on. In this example, the first round of the algorithm matches first choices (that is students and hospitals that both ranked each other first), the second round matches hospital first choices with student second choices, the third round matches student first choices with hospital second choices, the fourth round matches hospital first choices with student third choices, and so on. Suppose student s_j, ranked first by hospital h_i, has ranked hospital h_i as his third choice. If the student is not ranked as a 1 by his first or second choice hospitals, he might miss out on his third-choice hospital, although it ranked him in its first-choice group, because all available positions are filled before the fourth round. This is an unstable matching (the hospital will end up with some second choice students and the student ends up with his forth-choice hospital). In addition, the student would have the incentive to not reveal his true preferences and rank the third-choice hospital as his first choice, in order to obtain a better match for himself.[3]

ii. *Extension #2: Couples*

If a couple is in the same labor market (e.g., two medical students looking for internship placements), this introduces a new element into the matching problem: the preferences of the couple (two individual preferences) must be taken into account. (The couple does not necessarily have to be assigned to the same hospital, but in the same metropolitan area.) Stability is still defined largely the same: For students, stability requires that it is always better for either or both student in the couple to accept the matching than to be unemployed. For the hospitals, it must be better to accept the assigned student than to leave the position unfilled. If one (or both) student(s) in a couple can improve the given matching by switching to another hospital such that

---

[1] A detailed description of the matching algorithm used by the NIMP can be found in 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. 92, No. 6 (Dec., 1984), pp. 991–1016.

[2] This extension is largely taken from 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. 92, No. 6 (Dec., 1984), pp. 991–1016.

[3] Note: this problem is not completely solved by individual ordering, but is greatly diminished.

this hospital is better off as well, then this mutually beneficial trade will be carried out and the given matching is unstable.

In 1984, Roth noted that many couples seek internship positions outside of the voluntary National Intern Matching Program, hypothesized that the algorithm must not produce stable outcomes for couples, and then develops an example showing this is true.
Klaus and Klijn,[4] however, show that by assuming couples have "weakly responsive" preferences—the unilateral improvement of one partner's job is beneficial for the couple as well—that stable matchings will exist. This proven by showing that, when there is no negative externality from one partner's job onto the other partner or the couple, the market of couples can be treated as a market of "singles." The singles market is constructed by accounting for the couple's preferences in the preferences of the individuals. Klaus and Klijin then refer back to Gale–Shapley to prove that this "singles" market is like college admissions and will have a stable outcome.

During the in-class presentation, we will consider other applications of the Gale–Shapley model, including:

A. Doubles Tennis Partners
B. Athletic Teams
C. Some School-Assignment Programs

[4] This extension is largely taken from Bettina Klaus and Flip Klijn, *Stable Matchings and Preferences of Couples*, Journal of Economic Theory 121 (2005) 75–106.