College Admissions Mechanisms: Student-optimality vs. College-optimality

Petcharat Viriyakattiyaporn

Department of Computer Science
University of British Columbia

1. Abstract

The college admissions problem is an example of a two-sided matching market. Even though it is closely-related to the stable marriage problem, it is in fact not equivalent. However, the problems are similar enough that the mechanisms used to solve them share some properties—the mechanisms are only guaranteed to produce an optimal stable outcome for one of the parties, but not both. In this paper, we will try to answer which of the two mechanisms is more preferable in the college admissions problem, at least in practice. While a men-optimal stable mechanism and a women-optimal stable mechanism in the marriage problem are symmetric of each other, a student-optimal stable mechanism and a college-optimal stable mechanism in the college admissions problem are not. Due to this asymmetry, it has been shown that a college-optimal stable mechanism has more undesirable properties than its counterpart. Moreover, the results from experiments that ran a student-optimal stable mechanism and a college-optimal stable mechanism over real data from the National Resident Matching Program that matches hospitals and medical interns (including medical specialties) in the US revealed that participating programs (colleges) would not have been much worse-off under a student-optimal stable matching (approximately 2% of the programs would have been worse off).

2. Introduction

Suppose there are *n* students who are applying for colleges and there are *k* colleges that these students can apply for. Each student has a strict preference ordering over all colleges, and each college also has a strict preference ordering over all students. By *strict* preference, it means a student is not indifferent between two colleges, and vice versa. In reality, it is impossible for a college to accept all the students who apply for it, due to limited resources. In fact, a college only accepts a specific number of students (quotas) in each academic year. So, every student cannot possibly get into their top choices. On the other hand, a student also can accept offer of admission from only one college. Thus, it is not guaranteed that all students whom a college has made offers of admission will accept the offers. This problem is called the *college admissions problem*.

The college admissions problem has been widely studied by economists and game theorists. It is well-known to be closely related to the *stable marriage problem*. Instead of trying to find matching between students and colleges, the stable marriage problem tries to find matching between males and females in the community, such that each couple cannot be better off by pairing with other people in the community. In 1960, Gale and Shapley proposed a mechanism that is guaranteed to produce such matching. However, they also showed that the mechanism came in two flavors: one that gives the best possible matches achievable by males and one that gives the best possible matches achievable by females. It turned out that these two optimal matchings are not always the same. Only in some cases that they are the same, such as when the preference orderings of males and females coincide.

Although the college admissions problem is slightly different from the stable marriage problem, which we shall see how in the next section, we can still categorize mechanisms used to solve this problem as *student-optimal stable mechanism* and *college-optimal stable mechanism*. As similar as in the stable marriage problem, these two types of mechanisms do not always yield the same results. Our question is which of the two flavors is more desirable.

Thesis: In practice, student-optimal stable mechanism is more preferable than college-optimal stable mechanism in the college admissions problem.

In the next section, we will describe in more details about the nature of the college admissions problem as a two-sided matching. In section 4, we will explore pros and cons of student-optimal stable mechanism and college-optimal stable mechanism from game theory perspectives. Section 5 will discuss about computational experiments that ran these two types of mechanisms over data from the National Resident Matching Program (NRMP) and the Specialty Matching in Thoracic Surgery, in the US, and section 6 concludes.

3. College Admissions Problem and Two-Sided Matching

3.1) Two-Sided Matching

Two-sided matching involves two disjoint sets of agents. Agents in one group are matched to agents in another group, based on their declared preference ordering. Examples of two-sided matching problems include hospital-interns matching in the US under the National Resident Matching Program (NRMP), kidney donors-patients matching, and the stable marriage problem. The college admissions problem is also one of them because it involves two disjoint sets of agents (colleges and students) and it tries to find "most preferable" matching between those two sets based on preference ordering given by all agents. Below, we give formal definitions of the terms that will be used in further discussion in this paper.

Definition 1 (Matching): Given two disjoint sets C and S. A matching $\mu : C \cup S \rightarrow C \cup S \cup \{\varnothing\}$ is an assignment of an agent in one set to one or more agents in another set or to an empty set. If an agent is matched to an empty set, we say that the agent is "unmatched".

Ideally, a mechanism designer would wish to find a mechanism that produces a matching that satisfies all of the following properties.

Definition 2 (Individual Rationality): A matching μ is individually rational if no agent i prefer to be unmatched to being matched to $\mu(i)$.

Definition 3 (Stable Matching): A matching μ is stable if and only if it is individually rational and there does not exist an agent i such that $\mu(i) \neq s$ but $s \succ_i \mu(i)$ and $i \succ_s \mu(s)$, according to the given preference orderings from all agents.

An *unstable matching* is a matching that is not stable.

Definition 4 (Optimal Matching): A stable matching is optimal if every agent is at least well off under it as under any other stable matching.

Definition 5 (Dominant-Strategy Truthful): A mechanism is dominant-strategy truthful if it is a dominant strategy for all agents to state their truth preferences.

However, as we shall see shortly, there does not exist a stable matching mechanism that is dominant-strategy truthful in the context of the college admissions problem and the stable marriage problem.

3.2) Stable Marriage Problem

Given a set of men $M = \{m_1, \ldots m_n\}$ and a set of women $W = \{w_1, .., w_k\}$. Each man has a complete strict preference ordering over the set $W \cup \{\varnothing\}$, and each woman has a complete strict preference ordering over the set $M \cup \{\varnothing\}$. The goal is to find an optimal *one-to-one* stable matching between as many men and women as possible.

Gale and Shapley (1962) have proven that there always exists a stable set of marriage (matching) and also proposed an algorithm that is guaranteed to find such set of marriage [1]. The mechanism works as follow:

Step 1: Each man proposes to their most favorite woman.
Repeat the following, until min{n, k} women have been proposed.
Step 2: Each woman only keeps the man that she most prefers on the string, and rejects the rest of the proposals.
Step 3: The men who are rejected propose to the next woman they most prefer.
When min{n, k} women have received proposals, the mechanism terminates and each woman is required to marry the man that she has on her string.

The resulting marriage is a stable matching. Moreover, Gale and Shapley showed that this matching is optimal for men, but not necessary for women [1]. Therefore, this mechanism is called *Gale-Shapley men-optimal stable mechanism (with deferred acceptance)*. To see why this matching is stable, let's imagine that, under this matching, John is married to Jane. Suppose that John prefers Anne to Jane and Anne prefers John to her husband, then at some stage of the mechanism, John must have proposed to Anne and Anne will keep him on string and rejected her current husband—a contradiction.

It is also not hard to see why this matching is optimal for men but not for women because men are the ones who propose, so they would be able to attain their most favorite women possible (who also prefers them than any other man). Note that if, instead, women are the ones who propose, they would be able to attain their most favorite men possible, which is called *women-optimal matching*. In the men-optimal matching, women are only allowed to accept proposals from men who propose to them, who have likely been rejected by other women. There can exist a case that the only person who proposes to, say Jane, is the one that she least prefers. Then, she would be better off under the women-optimal matching. Thus, the men-optimal matching is not necessary the same as the women-optimal matching.

Note that the mechanism above produces a stable optimal outcome for men, even though the number of men and women are not equal. In general, the followings are known theorems which apply to the stable marriage problem. Theorem 1 and 2 were found by Gale and Shapley [1]. Theorem 3, 4, and 5 were found by Roth [8].

Theorem 1: For any preference ordering from both parties, a set of stable marriage is always non-empty.

Theorem 2: In a set of stable outcomes of the marriage problem, there exists a men-optimal stable matching x^* with property that, for every man m in the matching, $x^*(m)$ is the most-preferred achievable assignment (i.e. $x^*(m)$ is at least as good as any other stable outcome $x(m)$). Similarly, there exists a women-optimal stable matching y^* such that, for every woman w, $y^*(w)$ is w's most-preferred achievable assignment.

Theorem 3: There does not exist any outcome y that every man prefers to the men-optimal stable matching. Similarly, there does not exist any outcome z that every woman prefers to the women-optimal stable matching.

Theorem 4: There does not exist a stable mechanism that is dominant-strategy truthful in the marriage problem.

Theorem 5: The matching mechanism that always yields the man-optimal stable outcome for any given preference ordering makes it a dominant strategy for every man to state his true preferences in the marriage problem. Similarly, a mechanism that always yield women-optimal stable outcome makes it a dominant strategy for every woman to state her true preferences.

In fact, Gale and Sotomayor have shown that under any men-optimal stable mechanism, women can be better off by falsifying their preferences assuming all men are truthful, and vice versa. [2]

3.3) How College Admissions Problem Is Not Equivalent to Marriage Problem

Given a set of colleges $C=\{c_1,\ldots,c_n\}$ and a set of students $S=\{s_1,\ldots,s_k\}$. Each college has a strict preference ordering over the set $S \cup \{\varnothing\}$. Each student has a strict preference ordering over the set $C \cup \{\varnothing\}$. Moreover, each college i has a quota q_i that is the number of students that it can accept in this academic year. However, each student can accept only *one* offer of admission. The goal is to find an optimal stable matching between colleges and students such that q_i students are matched to college i and only one college is matched to a student.

The main aspect that sets this problem apart from the stable marriage problem is its *one-to-many* matching from colleges to students. Although Gale and Sotomayor [6] tried to prove that the two problems were equivalent once we replace a college i with q_i identical copies. Each copy of a college has identical preferences ordering as college i. Further, each student s replaces college i in his preference ordering with q_i copies such that $c_{i,1} \succ_s \ldots \succ_s c_{i,q_i}$. However, with this model of the problem, Theorem 3 from the marriage problem above do not directly apply because the model is missing specification of how a college would compare the *outcomes* of the problem, which are *sets* of students.

In fact, later on, Roth showed that as long as colleges have responsive preferences over the outcomes, Theorems 3 and 5 above do not hold in the college admissions problem [3]. Roth defined that a college c has *responsive* preference over outcomes if, for any matching $x(c)$ and $y(c)$, $y(c) \succ_c x(c)$ if and only if $y(c)$ can be obtained from $x(c)$ by replacing a student in $x(c)$ with another student who is more preferable; i.e. $y(c) = x(c) \cup \{s_k\} \setminus \{\sigma\}$ for $\sigma \in x(c)$ and $s_k \notin x(c)$ such that $s_k \succ_c \sigma$.

More precisely, Roth has shown that the following theorems hold for the college admissions problem [3].

Theorem 1*: For any preference ordering from both parties, a set of stable outcomes of the college admissions problem is always non-empty.

Theorem 2*: In a set of stable outcomes of the marriage problem, there exists a college-optimal stable matching x^* with property that, for every college c with quota q in the matching, $x^*(c)$ contains q most-preferred achievable students for c if the number of students achievable for c is at least q. Otherwise, it contains all of c's most-preferred students (i.e. $x^*(c)$ is at least as good as any other stable outcome $x(c)$). Similarly, there exists a student-optimal stable matching y^* such that, for every student s, $y^*(s)$ is s's most-preferred achievable college.

Theorem 3*: In the college admissions problem, there does not exist any outcome z that every student prefers to the student-optimal stable outcome. However, there may exist an outcome y that every college strictly prefers to the college-optimal stable outcome when colleges have responsive preferences.

Theorem 4*: There does not exist a stable mechanism that is dominant-strategy truthful in the college admissions problem.

Theorem 5*: The mechanism that always yields the student-optimal stable outcome for any given preference ordering makes it a dominant strategy for every student to state his true preferences in the college admissions problem. However, when colleges have responsive preferences, no stable matching procedure makes it a dominant strategy for every college to state its true preferences.

Theorem 1*, 2*, and 4* follow directly from Theorem 1, 2, and 4 in the marriage problem. Roth proved Theorem 3* and 5* by using counterexamples to disprove Theorem 3 and 5 in the marriage problem, which we will omit from this paper. Also, even though Theorem 3* states that there exists an outcome that every college prefers to the college-optimal stable outcome, it does not say that such outcome is stable, so it is still consistent with Theorem 2* (where it says college-optimal stable outcome is at least as good as any other stable outcome).

4. Student-Optimal Stable Mechanism vs. College-Optimal Stable Mechanism

In this section, we will discuss pros and cons of student-optimal and college-optimal stable mechanisms in general, including but not limited to Gale-Shapley mechanism. Please recall that we want to answer the question: Which one of the two mechanisms should we use in the college admissions problem? If the college admissions problem were equivalent to the stable marriage problem, we would not be able to decide on the answer based on game theory alone because both student-optimal and college-optimal stable mechanisms would be symmetric to each other.

However, as we have seen in the previous section, the college admissions problem is not equivalent to the stable marriage problem and that the two types of mechanisms are no longer symmetric. This gives us some hope that we might be able to determine which of the two flavors should be used in practice. Based on the theorems discussed in the previous section, we can now summarize some pros and cons of each type of mechanisms.

The main advantage of a student-optimal stable mechanism is that it guarantees an outcome that is stable and optimal for students, with respect to the declared preferences of all agents. Another advantage is that this type of mechanism makes it a dominant strategy for every student to state his true preferences. This prevents students from declaring their preferences strategically. On the down side, a student-optimal stable mechanism does not produce optimal outcomes for colleges and that it does not prevent colleges from strategically declaring their preferences.

On the other hand, a college-optimal stable mechanism guarantees an outcome that is stable and optimal for colleges, according to the declared preferences of all agents. However, based on Theorem 3*, this outcome may not be the best outcome achievable by colleges (though that best outcome may not be stable), assuming that colleges have responsive preferences over the outcomes. Moreover, based on Theorem 5*, colleges have incentives to falsify their preference ordering to achieve better outcomes. In fact, under this type of mechanism, every student also has incentive to falsify their declared preferences based on Gale and Sotomayor [2]. In other words, every agent has incentive to lie under this mechanism.

So far, it seems like student-optimal stable mechanism has less disadvantages than the college-optimal one. However, the fact that student-optimal stable mechanism produces the less preferred outcomes for colleges than the outcomes produced by the college-optimal stable mechanism is still a problem because we want to maximize welfare of both parties as much as possible. One thing that theory has not been able to tell us is how different the outcomes produced by these two types of mechanisms are. In particular, we are interested to know, how much worse off colleges will become under a student-optimal stable mechanism. Thus, we now turn our head towards some experimental results, in real life setting, to answer this question.

5. Computational Experiments on the NRMP

In the present, newly graduated medical students or medical specialists and hospitals seeking interns and/or specialists in the US participate in the National Resident Matching Program (NRMP). This program centralizes procedure of hospitals-interns/specialists matching to alleviate chaos that have happened in the past when hospitals were competing with each other for interns and interns were competing for internships. Back in early 1900s, before NRMP was developed, hospitals usually made offers to medical students as early as in their junior years, which also forces medical students to make decision early without knowing what other offers might be forth coming. This led to a problem because hospitals then had very little information about students' performance, and students might regret accepting the offers when they could have received better offers afterward. [7]

Up to 1995, the NRMP used a matching procedure that yielded program-optimal (college-optimal) stable outcomes. However, in the fall of 1995, the Board of Directors of the NRMP commissioned the design of a new mechanism that was supposed to yield applicant-optimal (student-optimal) stable outcomes, along with a study to compare the new design with the pre-existing one. The main reasons that the new design was commissioned were:

(1) The pre-existing mechanism caused a lot of dissatisfaction among applicants, until there were questions raised whether the mechanism was unreasonably favorable to employers at the expense of applicants and whether applicants could strategically manipulate their declared preferences to get better positions;

(2) The pre-existing mechanism did not take into account any match variation that may be present, such as couples who seek 2 positions in the same vicinity [5]. In this situation, the problem is no longer equivalent to the college admissions problem that we have discussed so far.

Thus, a new applicant-optimal stable mechanism was designed for NRMP and Roth and Peranson conducted computational experiments on the Rank Order Lists (ROLs) received by NRMP from year 1987 and 1993-1996. They also ran the new mechanism on the ROLs for specialty matching in Thoracic Surgery from years 1991-1994 and 1996, in which case the matching is not affected by any match variation [5]. We will only focus on the results from the case of Thoracic Surgery specialty match because the college admissions problem of our interest does not include any match variation.

The results from the Thoracic Surgery case turns out to be minimal: in five years that were studied, only 4 applicants and 4 programs would have been affected by the new mechanism, which is approximately 2% of total positions filled (see Table 1). The effects were such that 4 applicants were better off and 4 programs were worse off, which is consistent to the theory. However, the results also showed that in 3 out of 5 years, the two mechanisms would have produced the same matching as shown in Table 2. It implies that, there is a unique optimal stable matching in those 3 years.

Table 1: Descriptive Statistics and Original Match Results in Thoracic Surgery Case

Category	1991	1992	1993	1994	1996
Applicant ROLs	127	183	200	197	176
Active programs	67	89	91	93	92
Program ROLs	62	86	90	93	92
Total quota	93	132	141	146	143
Positions filled	79	123	136	140	132

Table 2: Difference in Results between Pre-Existing Specialty Match and the New Mechanism

Year	Difference
1991	none
1992	2 applicants improve, 2 programs do worse
1993	2 applicants improve, 2 programs do worse
1994	none
1996	none

Roth and Peranson also provided insights about the results in [5]. They said that the reasons that the two mechanisms yielded such small difference were likely to be because:

(1) The preferences are highly correlated—similar programs tend to agree on which are the most desirable applicants and applicants tend to agree on which are the most desirable programs. When this happens, the set of optimal stable outcome is small.

(2) The number of positions an applicant can interview for before they design on their ROLs are relatively very small comparing to the market size (the total number of applicants and the total number of programs participated). If applicants can give a complete ROLs over all programs, then the percentage of applicants who could get

different stable matchings would grow as the total number of applicants and positions grow. However, since this is not possible in reality, the set of optimal stable outcomes from those two mechanisms is small (even if the preferences are uncorrelated), and hence, the minimal difference in resulting matches.

Note that this setting is also similar to the setting for the college admissions problem in real life.

(1) Colleges usually rank student applicants based on their academic records, standardized test scores, their essays, their resumes, and recommendation letters. Assuming no externalities, then the preference ordering of colleges should be highly correlated, as similar as the case of participating programs in the NRMP. On the other hand, most students tend to agree on which are the most desirable colleges (e.g. Ivy League universities are more desirable or some universities ranked on the top are more desirable and so on). So, their preferences should be highly correlated as well.

(2) Even if the preferences are not correlated, the number of colleges that a student can apply is relatively small, comparing to the total number of colleges available, due to his budget constraint. More specifically, in year 2007, there are 2,629 4-year post-secondary institutions in the US[1], while a student applicant applies for 5-7 colleges on average.

Thus, it is reasonable to believe that the magnitude of difference between the stable matchings that a student-optimal stable mechanism and a college-optimal stable mechanism produce will also be small, in the college admissions problem.

At this point, skeptical readers may question whether this small magnitude of difference can cause a huge drop in welfare on the programs (colleges) side. However, there was an argument made that a change in matches has a larger effects on the affected applicants than on the affected programs. Basically, the argument states that there is uncertainty in both programs' and applicants' rankings. However, while a program may not feel that much different between its 7th and 17th ranked candidates, an applicant can have clearer judgment between his 1st and 3rd choices.[5]

Based on the results of the experiments, the argument about effects of different matches on welfare of both parties, and the discussion in section 4, we can conclude that a student-optimal stable mechanism is more preferable than a college-optimal stable mechanism in practice.

6. Conclusion

In the context of the college admissions problem (when a match variation is not present), two main types of mechanisms are used to solve the problem: a student-optimal stable mechanism and a college-optimal stable mechanism. Even though it is tempting to say that this problem is similar to the stable marriage problem, Roth has shown us that it is not if we assume that colleges have responsive preferences over sets of students as outcomes [3]. Since this problem is not equivalent to the stable marriage problem, the two types of mechanisms are no longer symmetric. Due to this asymmetry, we want to find out which of these two types of mechanisms would be more preferable in practice.

Based on Theorem 1* and 2*, we have seen that a student-optimal stable mechanism only guarantees a stable and optimal outcome for students at the expenses of colleges,

---

[1] Source: National Center for Education Statistics, *Digest of Education Statistics 2007*. < http://nces.ed.gov/programs/digest/d07/tables/dt07_255.asp>

while the opposite is true for a college-optimal stable mechanism. However, based on Theorem 3*, there might exist an (unstable) outcome that is strictly preferred by every college to the college-optimal stable outcome. Moreover, while a student-optimal stable mechanism still makes it a dominant strategy for every student to state his true preferences, a college-optimal stable mechanism does not make it a dominant strategy for every college to state its true preferences (Theorem 5*). In other words, every agent has incentive to lie under a college-optimal stable mechanism. So, a college-optimal stable mechanism turns out to have more undesirable properties than the student-optimal one.

Furthermore, the computational experiments results of applying program-optimal (college-optimal) and student-optimal mechanisms on real Rank Order Lists submitted by applicants and participated programs in the Thoracic Surgery specialty match revealed that a minimal number of programs (around 2%) are worse off under the student-optimal mechanism. Roth and Peranson [5] reasoned that this phenomenon occurred because either (1) the preference orderings of agents were highly correlated, or (2) the number of programs that an applicant included in his ROLs was very small, relative to the total number of applicants and programs participated, which caused the set of stable matchings to be small. If this conjecture is true, then the difference between results of student-optimal and college-optimal mechanisms should also be small in the college admissions problem because, in reality, the preference orderings of students and colleges can be highly correlated. Even if they are not highly correlated, the number of colleges that an applicant can apply and rank is very small, relative to the total number of applicants and colleges. Also, regardless of the magnitude of the difference, it has been argued that a different match has greater effects on the welfare of the affected applicant than on the affected program.

Since a student-optimal stable mechanism has less undesirable properties than a college-optimal stable mechanism, and in reality setting, colleges would not be much worse off (while students would be a lot better off) under a student-optimal stable matching, then we can conclude that a student-optimal stable mechanism is more preferable than a college-optimal one in practice.

7. References

[1] Gale, D., Shapley, L. S., College Admissions and the Stability of Marriage. *American Mathematical Monthly*, January 1962, **69**(1), pp. 9-15.

[2] Gale, D., and Sotomayer, M., Ms. Machiavelli and the Stable Matching Problem. *American Mathematical Monthly*, April 1985, **9**(4), pp. 261-268.

[3] Roth, A. E., The College Admissions Problem Is Not Equivalent to the Marriage Problem. *Journal of Economic Theory*, 1985, **36**, pp. 277-288.

[4] Roth, A. E., The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Economic Theory*, 1984, **92**, pp. 991-1016.

[5] Roth, A. E. and Peranson, E., The Redesign of Matching Markets for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Revie*, September 1999, **89**(4), pp. 748-780.

[6] Gale, D. and Sotomayor, M., Some Remarks on the Stable Matching Problem, mimeo, 1983.

[7] Roth, A. E., The Origins, History, and Design of the Resident Match. *JAMA: Journal of the American Medical Association*, February 19, 2003, **289**(7), pp. 909-912.

[8] Roth, A. E., The Economics of Matching: Stability and Incentives. *Math. Oper. Res.* 7 (1982), pp. 47-57.