MANIPULABILITY OF STABLE MATCHING RULES
FOR THE COLLEGE ADMISSION MARKET

BY

MARILDA SOTOMAYOR

URL: www.marildasotomayor.com

Departamento de Economia	EPGE
Universidade de São Paulo/SP	Fundação Getúlio Vargas/RJ

THE COLLEGE ADMISSION MARKET: M= (S,C,P,q) (Gale and Shapley, 1962)

$S = \{s_1,\ldots,s_n\}$ = **set of students**

$C = \{c_1,\ldots,c_m\}$ = **set of colleges**

Each participant of one set has **strict preferences** over the participants of the other set, which are **transitive and complete**, and so they can be expressed by ordered lists of strict preferences.

$P(s_i) = c_2, c_3, s_i, c_1$

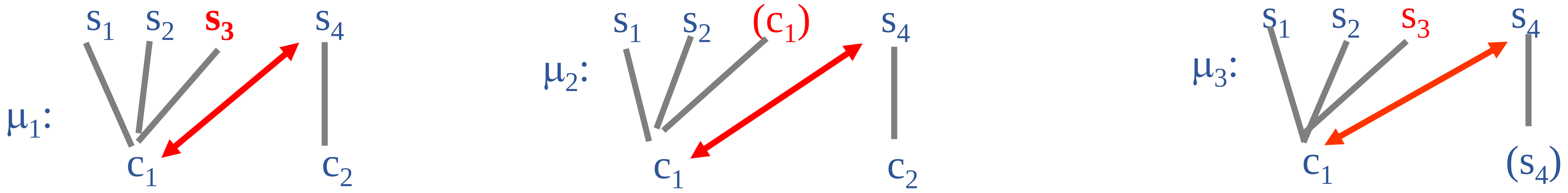
$P(c_j) = s_4, s_3, c_j, s_2, s_1$

q_j = **quota of college** c_j. Each student can be enrolled in one college at most.

ASSUMPTION: A college has preferences over groups of students, of size up to its quota, which are **responsive** to its preferences over individual students. That is, for any two groups of students that differ in only one student it prefers the one containing the more preferred student (Roth, 1985).

A student is *acceptable* to a college if the college prefers to admit the student to have one unfilled vacancy; a college is *acceptable* to a student if the student prefers to be enrolled at the college to be without school.

A matching is an assignment of the students to the colleges that respects the quotas of the colleges and such that no student is assigned to more than one college. It is *feasible* if all agents are acceptable to their partners.



A matching is *stable* if it is feasible and there is no pair, formed by a student and a college, such that the college prefers the student to some of the students assigned to it, or it did not fill its quota of students and it prefers the student to have one unfilled vacancy; the student prefers the college to the one to which he/she is assigned or he/she is not assigned to any college and prefers that college to be unassigned.

Gale and Shapley (1962) – *constructive proof.*

Algorithm: each student, following his/her ordered list of preferences, makes proposals to the colleges, which accept them temporarily or reject them. The algorithm ends when there are no more rejections.

Roughly speaking, the matching produced by the algorithm is the best stable matching for all students and the worst stable matching for all colleges. It is called *optimal stable matching for the students*.

By reverting the roles between students and colleges in the algorithm we get *the optimal stable matching for the colleges*, with symmetric properties. (Gale and Sotomayor, 1983, 1985).

Sotomayor (1996) – elementary *non-constructive* and shorter proof.

This talk concerns the way the college admission market operates, which is an issue of interest for Game Theory and Economics.

Suppose the admission process of students to colleges for a given college admission market, say *M(P)=(S,C,P,q)*, is in charge of a central planner who asks the participants to inform their lists of preferences.

Then, some specific stable matching for the profile of revealed preferences is selected.

Such process defines a ***stable matching rule H*** for *M(P)=(S,C,P,q).*

The domain of *H* is the set of all possible profiles of preferences *Q* of the participants. That is, *H(Q)* is a stable matching for the market *M(Q)= (S,C,Q,q).*

It turns out that, if some participant or group of participants is not honest in revealing preferences, we cannot expect the resulting matching be stable under the true preferences. The questions that naturally emerge are then:

A. Is it possible to design some stable matching rule for a given college admission market that elicits all students and colleges to reveal their true preferences?

A rule with such property is said to be ***non-manipulable*.**

B. Given a stable matching rule for a given college admission market, what is the prediction we could make about its manipulability or non-manipulability?

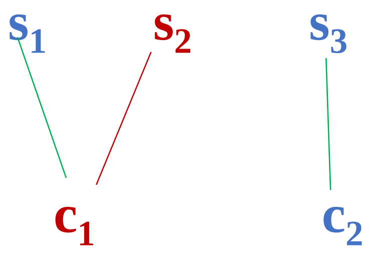
Example 1. $S=\{s_1, s_2, s_3\}$, $C=\{c_1,c_2\}$, $q_1=2$, $q_2=1$; H = stable matching rule; P= true preferences.

$P(s_1)= c_1, c_2$ $P(c_1)= s_1, s_2, s_3$

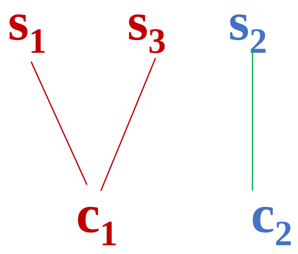
$P(s_2)= c_2, c_1$ $P(c_2)= s_3, s_2, s_1$

$P(s_3)= c_1, c_2$

$x_C =$



$x_S =$

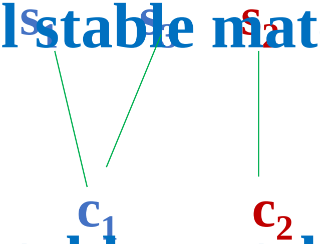


If $H(P)=x_C$ then **H is not manipulable by any college and it is manipulable by** $s_{2.}$

1. No student can manipulate the student optimal stable matching rule.

$P'(s_2)= c_2$

$x'_C=x_S=$



$c_2 >_{s2} c_1$

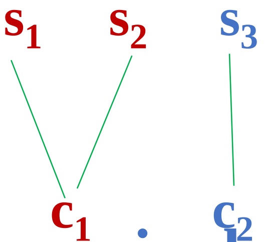
2. no college can manipulate the college optimal stable matching rule;

If $H(P)=x_S$ then **H is not manipulable by any student and it is manipulable by** c_1.

3. the optimal stable matching rule for one of the sides of the market is manipulable by some agent of the other side;

$P''(c_1)= s_1, s_2$

$x''_S= x_{C=}$



$\{s_1\ s_2\} >_{c1} \{s_1, s_3\}$

4. any stable matching rule for this market is manipulable.

1. No student can manipulate the student optimal stable matching rule.

The mathematical confirmation that phenomenon 1 holds for every college admission market is given by the Non-manipulability Theorem, due to Dubins and Freedman (1981). (The original proof of this result had about 20 pages. A shorter proof of it, with less than half a page, was my first contribution to the theory of incentives - Gale and Sotomayor, 1983, 1985 and Roth and Sotomayor, 1990).

This theorem implies that, in any college admission market, the best policy for any student, facing H_S, is to tell the truth. Indeed, the theorem is more general: if a group of students misrepresent their preferences, at least one of them will not be better off. A symmetric result holds for all colleges with a quota of one.

c_2 cannot manipulate the college optimal stable matching rule.

NON-MANIPULABILITY THEOREM (Dubins and Freedman, 1981): *For any college admission market, the student optimal stable matching rule is non-manipulable by any student and by any set of students; the college optimal stable matching rule is non-manipulable by any college and by any set of colleges, with a quota of one.*

2. No college can manipulate the college optimal stable matching rule.

In particular, **college c_1, which has a quota greater than one, cannot manipulate H_C in that example.** However, this phenomenon does not generalize to every college admission market. It is not hard to find situations in which **H_C is manipulated by a college with a quota greater than one** (Roth, 1985).

3. The optimal stable matching rule for one of the sides of the market is manipulable by some agent of the other side.

GENERAL MANIPULABILITY THEOREM (Sotomayor, 2012): *If some stable matching rule is used for a given college admission market, and it does not yield the optimal stable matching for one of the sides of the market, then honest revelation of preferences is not the best policy for some agent from this side.*

COROLLARY (Sotomayor, 2012): *If the college admission market has two or more stable matchings, then the student optimal stable matching rule is manipulable by some college and the college optimal stable matching rule is manipulable by some student.*

From the General Manipulability Theorem we can conclude the following:

GENERAL IMPOSSIBILITY THEOREM (Sotomayor, 2012): *If a college admission market has two or more stable matchings, then every stable matching rule for this market is manipulable.*

This theorem explains phenomenon 4:

4. Any stable matching rule for this market is manipulable.

When the market has **only one stable matching**, any stable matching rule associates the true preferences P to the college optimal stable matching, which coincides with the student optimal stable matching. Therefore, by the Non-manipulability Theorem, **only colleges with a quota greater than one may be able to manipulate**. We are able to construct examples where no college with a quota greater than one can manipulate any stable matching rule (Sotomayor, 2018). In such examples, the conclusions of these three results are false. Every stable matching rule is non-manipulable.

These theorems, together, provide the framework of the *Theory of Incentives for the college admission market*. This theory has helped to explain empirical economic phenomena and has contributed to the understanding of the operation of real markets. Along the years, they provided the bases for the implementation of allocation mechanisms in several real matching markets, as the **admission market of students to Universities in Turkey and Spain**, the **market of students and dormitories in Israel**, the **admission market to graduate centers of Economics in Brazil**, **school choice markets in Boston and New York**, as well as the reformulation proposed by Roth of **the market for medical residents and hospitals in the US** in 1998, etc.

Dubins and Freedman (1981), “Machiavelli and the Gale-Shapley algorithm”, *American Mathematical Monthly,* 88, 485-495.

Gale and Shapley (1962), “College admission and the stability of marriage”, *American Mathematical Monthly*, 69, 9- 15.

Gale and Sotomayor (1983, 1985), “Some remarks on the stable matching problem”, *Discrete Applied Mathematics*, 11, 223-232.

Roth (1985), “The college admissions problem is not equivalente to the marriage problem”, *Journal of Economic Theory,* 36, 277-288.

Roth and Sotomayor (1990), Two sided matching. A study in game-theoretic modeling and analysis”, *Econometric Society Monographs,* no. 18.

Sotomayor (1996), “A Non-constructive elementary proof of the existence of stable marriages”, *Games and Economic Behavior,* **13,** 135–137

_________(2012), “A further note on the college admission game”, *International Journal of Game Theory,* v. 41, p. 179-193.

_________(2018), “The college admission model is deficiente for the purpose of serving as vehicle for the manipulability analysis of a stable matching rule”, *working paper.*

MUITO OBRIGADA.

Albeit these results had never been proved until the paper of Sotomayor (2012), it was believed, for more than three decades, that **if the college optimal stable matching was used as an allocation rule for a given college admission market, then there would be, at least, one student and one college that could benefit themselves by falsifying their true preferences; if the student optimal stable matching was used, some college could benefit by misrepresenting preferences.** (Sotomayor (2018) shows that there are markets where the college admission rule is non-manipulable).

The main consequence of that belief was the preference of the organizers of real markets for the student optimal stable matching rule, instead of the college optimal stable matching rule. Roth, for example, reformulated the allocation procedure of the interns and hospitals in the United States, which previously used the Gale and Shapley algorithm with the hospitals making the proposals, and proposed a new procedure which involves the matching rule that yields the student optimal stable matching.

Example of Gale-Shapley algorithm. **$S=\{s_1, \ldots, s_4\}$, $C=\{c_1,c_2,c_3\}$, $q_1=2$, $q_2=q_3=1$.** The true preferences over acceptable agents are the following:

$P(s_1)= c_2, c_1, c_3$ $P(c_1)= s_1, s_2, s_3, s_4$

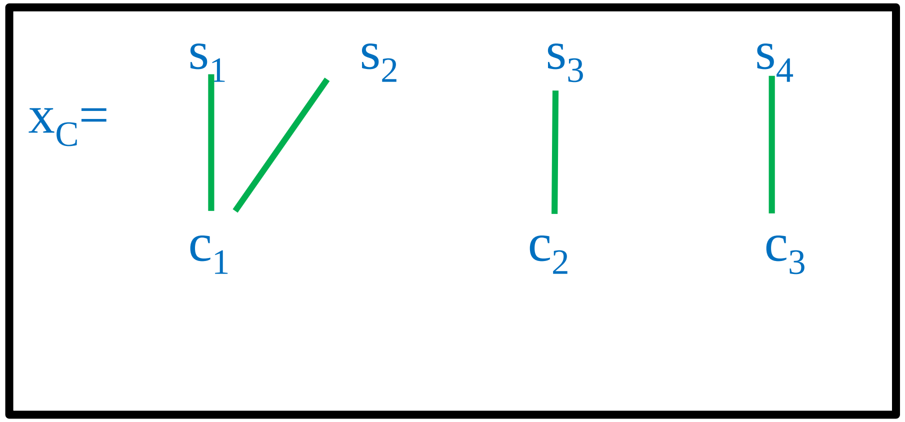
$P(s_2)= c_3, c_1, c_2$ $P(c_2)= s_2, s_3, s_4, s_1$

$P(s_3)= c_1, c_2, c_3$ $P(c_3)= s_3, s_4, s_2$

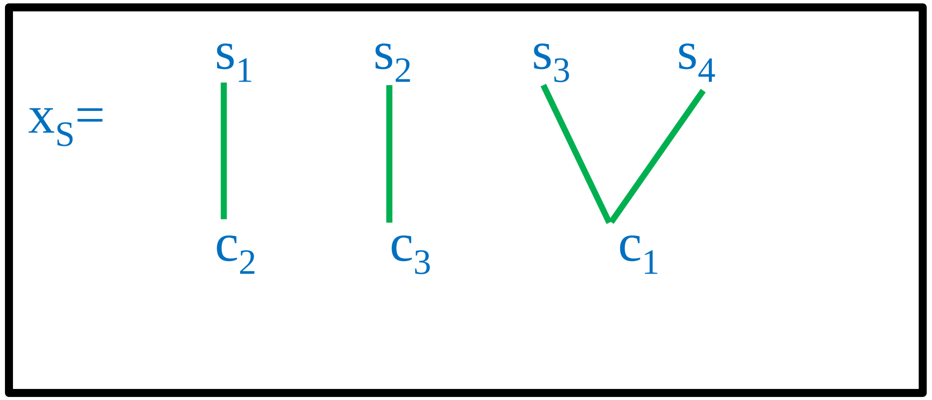
$P(s_4)=c_1, c_3$

	s_1	s_2	s_3	s_4
$x_C =$	c_1	c_1	c_3	c_3
		c_2	c_2	

College-optimal stable matching



Student-optimal stable matching



Example 4. **$S=\{s_1, \ldots, s_3\}$, $C=\{c_1,c_2\}$, $q_1=2$, $q_2=1$.** The true preferences over individuals are the following:

$P(s_1)= c_2, c_1$

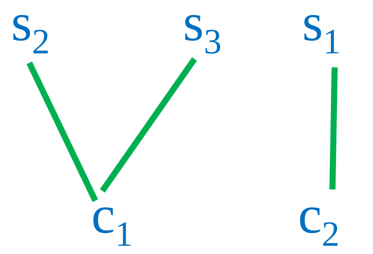
$P(s_2)= c_1, c_2$

$P(s_3)= c_1, c_2$

$P(c_1)= s_1, s_2, s_3$

$P(c_2)= s_1, s_2, s_3$

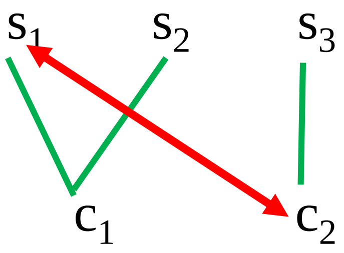
$x_C=$



$= x_S$

There is no manipulation for c_1 that benefits it under any stable matching rule.

$y=$

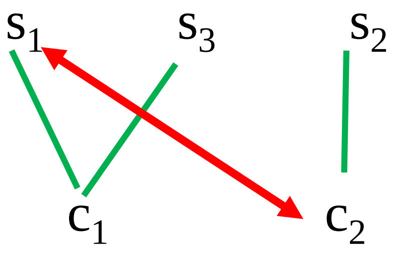


$c_2 >_{s1} c_1$

$s_1 >_{c2} s_3$

or

$y'=$



$c_2 >_{s1} c_1$

$s_1 >_{c2} s_2$

H_C is not manipulable by any student and by any college;
H_S is not manipulable by any student and by any college .

Any stable matching rule is non-manipulable.

Example 5. **$S=\{s_1, \ldots, s_3\}$, $C=\{c_1,c_2\}$, $q_1=2$, $q_2=1$.** The true preferences over individuals are the following:

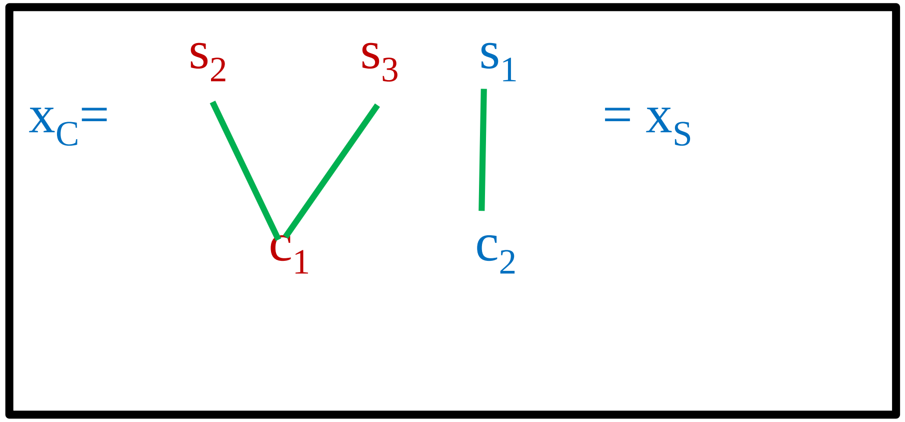
$P(s_1)= c_2, c_1$

$P(s_2)= c_1, c_2$

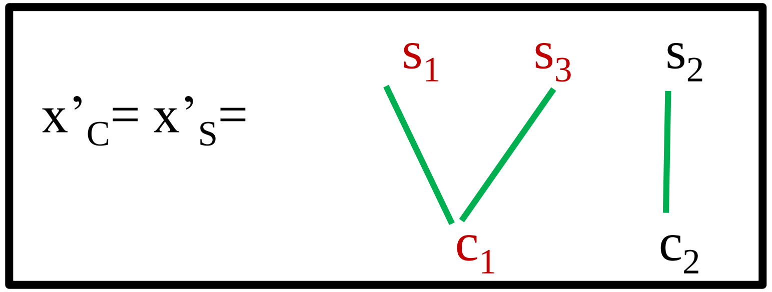
$P(s_3)= c_1, c_2$

$P(c_1)= s_1, s_2, s_3$

$P(c_2)= s_2, s_1, s_3$



$P'(c_1)= s_1, s_3$



$\boxed{\{s_1, s_3\} >_{c1} \{s_2, s_3\}}$

H_C is manipulable by c_1.
H_S is manipulable by c_1.

Any stable matching rule for this market is manipulable.

H_C is not manipulable by any student.
H_S is not manipulable by any student.