ADMISSION GAMES INDUCED BY STABLE MATCHING RULES

Presentation at Brown University and University of Rochester, Department of Economics

MARILDA SOTOMAYOR

Universidade de São Paulo, Brown University

e-mail: marildas@usp.br

This paper deals with the cooperative model introduced in Sotomayor (1996), known nowadays as the SCHOOL CHOICE MODEL.

**GRADUATE CENTER ADMISSION MARKET**

There are a set of students and a set of graduate centers, that will be simply called centers. Each student can be enrolled in one center, at most. The centers have quotas, representing the maximum number of students it can admit. The students have ordered lists of preferences over the centers. **It is assumed that a given procedure to evaluate the students is available for the centers.** The students are then ranked by each center according to the result of this evaluation. This ranking defines the preference list of the center over individual students. The idea is then to investigate the strategic behavior of the students facing a stable matching mechanism which considers as fixed the preferences of the centers.
ADMISSION MARKET OF STUDENTS TO GRADUATE CENTERS IN ECONOMICS IN BRAZIL

In this market, the students are submitted to five tests: Mathematics, Statistics, Microeconomics, Macroeconomics and a test about Brazilian Economy. The schools attribute weights to each one of the five subjects. The quotas and weights of the centers and the students’ scores in the tests are common knowledge. The students are then evaluated by the centers in accordance to their weighted average score in the tests. These evaluations are used by the centers to determine a ranking of the students.

The school choice model, as it has been described in the literature, is a strategic model in which only the students have strategic behavior due to the fact that the schools rank the students according to some pre-determined criterion of priority.

Although it is not considered by the authors, in a real school choice market:

- **all the students must be admitted by some school**, so the number of students is not bigger than the number of seats in the schools and every student is acceptable to every school;

- **the priorities** of the schools over the students are determined according to the information obtained from the students, which can be falsified by these agents.
GRADUATE CENTER ADMISSION MODEL

• There is no information that a student can give to a graduate center that can change his/her position in the preference list of the given center.

• There is no restriction on the number of participants on each side of the market; any institution is free of keeping some unfilled positions and any student is free of remaining unmatched.

Therefore, the school choice model is treated in the literature as a center admission model.
GRADUATE CENTER ADMISSION MODEL

\[ M=(S,C,P_S,\{P_C,q_C\}). \]

\( S= \) set of students with \( m \) elements.
\( C= \) set of centers with \( n \) elements.
\( q(c)= \) quota of center \( c \)
\( P(s)= c_1, c_2, s, c_3 \ldots \), denotes de list of strict preferences of student \( s \).
\( P(c)= s_1, s_3, s_2, c, s_4, \ldots \), denotes de list of strict preferences of center \( c \).
**Matching $\mu$:** An allocation $\mu$ of the students to the centers which respects the quotas of these agents.

A student who is not matched to any center will be matched to himself and will be called **unmatched**.

A center that has some number of unfilled positions will be **matched to itself** in each of these positions.

**Notation:** $\mu(s)=c$; $\mu(s)=s$; $\mu(c)=\{s_2, s_4, c, c, c\}$. 
KEY NOTION: STABILITY

A matching that can be upset by a coalition of agents is not stable.
Intuitive idea: Suppose all participants are present at the same time in the market and that they can freely communicate to each other. Centers make offers to the students, who accept them or do not accept them. A matching will be a set of agreements between the centers and the students. An agreement between a center and a student will be achieved if a better agreement for both of them cannot be obtained elsewhere. The matching that will emerge from these pairwise-interactions is called a stable matching.
Example 1. Let \( S = \{s_1, s_2, s_3\} \), \( C = \{c_1, c_2, c_3\} \) and \( q(c_j) = 1 \), for all \( j = 1, 2, 3 \). Let \( P_S \) and \( P_C \) be given by:

\[
\begin{align*}
P(s_1) &= c_1, c_2, c_3, s_1 \\
P(s_2) &= c_3, c_2, c_1, s_2 \\
P(s_3) &= c_1, c_3, s_3, c_2 \\
P(c_1) &= s_2, s_1, s_3, c_1 \\
P(c_2) &= s_3, s_1, s_2, c_2 \\
P(c_3) &= s_1, s_3, s_2, c_3
\end{align*}
\]

\[
\mu_1: \\
\begin{array}{c}
S_1 \\
C_3 \\
S_2 \\
C_1 \\
S_3 \\
C_2
\end{array}
\]

\( c_2 \) is unacceptable to \( s_3 \), so \( s_3 \) will not accept an offer from \( c_2 \). This pair will never achieve an agreement, so this matching will not occur. \( \mu_1 \) is not a stable matching. It is not feasible.

A feasible matching only matches a student to a center if both agents are mutually acceptable.
\( P(s_1) = c_1, c_2, c_3, s_1 \)
\( P(c_1) = s_2, s_1, s_3, c_1 \)
\( P(s_2) = c_3, c_2, c_1, s_2 \)
\( P(c_2) = s_3, s_1, s_2, c_2 \)
\( P(s_3) = c_1, c_3, s_3, c_2 \)
\( P(c_3) = s_1, s_3, s_2, c_3 \)

\[ \mu_2 : \]

\begin{align*}
S_1 & \rightarrow S_2 \\
C_2 & \rightarrow C_3 \\
S_3 & \rightarrow C_1
\end{align*}

\( \mu_2 \) cannot be considered stable. It is **destabilized** by the pair \( (s_1, c_1) \).

Hence the stable matchings will be the feasible matchings that are not destabilized by any pair.

**Definition 1.** A matching \( \mu \) is **stable** for the market \( M = (S, C, P_S, \{P_C, q_C\}) \) if it is feasible for \( M \) and there are no student \( s \) and center \( c \), not matched to one another at \( \mu \), such that

a) \( c >_{P(s)} \mu(s) \) and b) \( s >_{P(c)} \sigma \), for some \( \sigma \in \mu(c) \).
Definition 2. For a given market $M=(S,C,P_S,\{P_C,q_C\})$ a stable matching $\mu_S$ is called the student-optimal stable matching if $\mu_S(s) \geq P(s)$ for every student $s$ and for every stable matching $\mu$. The center-optimal stable matching $\mu_C$ is analogously defined.


A CENTRALIZED PROCEDURE FOR THE ADMISSION OF THE STUDENTS TO THE CENTERS.

• The lists of preferences and quotas of the centers are known.

• The students inform their lists of preferences to a Central.

• An algorithm that finds a stable matching for the resulting market is used by the Central.

A stable mechanism for the graduate center admission market, restricted to a given market $M(P) = (S, C, P_s, \{P_C, q_C\})$, maintains fixed $S$, $C$ and $\{P_C, q_C\}$ and let each student $s \in S$ replace, if he wishes, his true preference list, $P(s)$, by any list of preference $Q(s)$.

The resulting market is then $M(Q) = (S, C, Q, \{P_C, q_C\})$, where $Q$ denotes the profile of lists of preference $Q(s)$, one list for each student $s \in S$, where $Q(s)$ may be $P(s)$. 
$M(Q) = (S, C, Q, \{P_C, q_C\})$.

$Q$ - a preference profile for the students.

$E(Q)$ - set of stable matchings for $M(Q)$.

$f$ - stable matching rule.

$f: \{Q\} \rightarrow \{E(Q)\}$

$Q = (Q(s_1), \ldots, Q(s_m))$  

$f(Q) \in E(Q)$

**Student-optimal stable matching rule** - the participants are assigned in accordance with the student-optimal stable matching for the revealed preferences.

**Center-optimal stable matching rule** - the participants are assigned in accordance with the center-optimal stable matching for the revealed preferences.
IMPOSSIBILITY THEOREM (ROTH, 1982) No stable matching mechanism exists for which stating the true preferences is a dominant strategy for every agent.

There is a particular college admission market such that, for every stable matching rule, some agent can be better off by misrepresenting his/her/its preferences while the other agents tell the truth.

"Roth’s impossibility theorem shows that there does not exist any stable matching mechanism that makes it a dominant strategy for all students and colleges to report their true preferences. This implies that when students and colleges are confronted with a game $\Gamma(\phi, P)$ induced by a stable mechanism $\phi$, the reported preferences may well be different from the true preferences." (Ma, J. (2002), “Stable matchings and the small core in Nash equilibrium in the College Admission Problem”, Review of Econometric Design, 7, 117-134)
Suppose the mechanism restricted to the market $M(P)$ produces the optimal stable matching for the students when these agents report their true preferences $P$.

\[
f: \{Q\}_{P_S} \rightarrow \{E(Q)\}
\]

\[f(P_S) = \mu_S\]

(Dubins and Freedman, 1981) No student $s$ can be better off by stating $Q(s) \neq P(s)$.

(Sotomayor, 1996) If $f(P_S) \neq \mu_S$ then there is some student who can be better off by misrepresenting his true preferences.
The remarkable fact is that such beliefs, although incorrectly concluded, have pointed directions that led to the main achievements of the theory of stable matching mechanisms for the college admission model. They imply that, if the agents select their sincere strategies, then at least one of them is not playing his/her/its best response. Then it makes sense to look for the existence of Nash equilibria and to address the problem of how the stability, under the true preferences, of the equilibrium outcomes could be affected. (See Romero Medina (1998), Ma (2002), Sotomayor (1996, 2008), among others). Under the practical point of view they have contributed to the organization of markets and to the design of stable matching mechanisms for several college admission and school choice markets (See Kesten (2006 a-b), Abdulkadiroglu and Sonmez (2003), Balinski and Sonmez (1999), among others).
GENERAL IMPOSSIBILITY THEOREM (SOTOMAYOR) When any stable mechanism is applied to a college admission market in which preferences over individual agents are strict and there is more than one stable matching, then at least one agent of this market can profitably misrepresent his/her/its preferences, assuming the others tell the truth.

THEOREM. If a stable mechanism is applied to market $M(P)$ and it does not produce the optimal stable matching for a given side of the market when $P$ is selected, then some agent of the opposite side can profitably misrepresent his/her/its preferences assuming the others report their true preferences.
More generally, suppose the stable matching rule produces the optimal stable matching for the students when these agents report their true preferences. If a set $T$ of students states $Q(T)$, instead of $P(T)$, at least one student in $T$ will not be better off. That is, if $Q=(Q(T), P_{-T})$ and $f(Q) = \mu$ then $\mu_S(s) \geq_{P(s)} \mu(s)$ for some student $s \in T$.

Formally,

It is implied by Sotomayor (1996) that if a mechanism is not the student-optimal stable matching mechanism then it is manipulable by at least one student.

Hence, returning to the center admission model, the mechanism that produces the optimal stable matching for the students is the only strategy-proof stable mechanism.
Admission game induced by $f$: $\Gamma(f) = (S, \{Q\}, f, P_S)$.

$S$ - set of players;
$\{Q(s)\}$ – set of strategies of student $s$;
$f$ - outcome function;
$P_S$ - profile of true preferences of the players (sincere strategies).

Note that this game is different from the game induced by the College admission mechanism because the colleges can behave strategically and the centers cannot.

By Proposition 1, it is a dominant strategy for each student to report his sincere strategy in the admission game induced by the student optimal stable mechanism.
According to the result of Sotomayor (1996), if \( f(P_S) \neq \mu_S \) and each student expects the others to play their sincere strategies, it will in general be possible for at least one student to profitably deviate from his strategy.

**QUESTION:** If honest revelation of preferences is not the best policy for the students, is there any profile of strategies which have the property that once they are adopted there will be no advantage to any student in changing his strategy?
Definition 3. Let $F(f)$ be an admission game. The profile of strategies $Q$ is a Nash equilibrium of $F(f)$ (or $Q$ is a Nash equilibrium for the matching rule $f$) if for every student $s$, $f(Q, Q') \geq P(s)$ for every strategy $Q'(s)$ of student $s$.

Proposition 2: Existence of Nash equilibria (Sotomayor, 1996). Let $\mu$ be a stable matching for the market $M(P) = (S, C, P_s, \{P_C, q_C\})$. Suppose each student $s \in \mu(C)$ chooses the strategy $Q(s)$ of listing only $\mu(s)$ as the only acceptable center; (i.e., $Q(s) = \mu(s)$ if $s$ is matched at $\mu$); $Q(s) = s$ if $s$ is unmatched at $\mu$. Then $Q$ is a Nash equilibrium of the game induced by every stable matching rule and $\mu$ is the equilibrium outcome. Furthermore, $E(Q) = \{\mu\}$. 
The results mentioned before imply that truth telling is a Nash equilibrium for the admission game induced by $f$ if and only if the outcome produced by $f$ for these preferences is the student-optimal stable matching.

This means that under any other stable matching rule, at equilibrium, when students behave strategically, at least one student will be misrepresenting his/her true preferences; produced by a Nash equilibrium. On the other hand, the adoption of a stable matching rule should be the simplest way to allocate the students to the market $M(P)=(S,C,P_S,\{P_C,q_C\})$. Therefore, the stability of the Nash equilibrium outcome is a relevant issue that was investigated in the paper of (2008).

Other stable matching mechanisms previously studied in the literature produce Nash equilibrium outcomes that are stable under the true preferences. (Roth (1984), Alcalde (1996), Alcalde and Medina (2000), Alcalde, Perez Castrillo and R. Medina (1998), Sotomayor (2002)).
Theorem*: Let \( Q=(Q_S,P_C) \) be a Nash equilibrium of some stable matching rule \( f \). Then \( f(Q) \) is stable under the true preferences.

Without this restriction, Theorem 5.18 of Roth and Sotomayor (1990) due to Roth (1986) implies that every \textbf{unstable} matching is a Nash equilibrium outcome for the game induced by any stable matching rule.

If the centers are not allowed to behave strategically then Example 1 of Sotomayor (2008) shows that we may have equilibrium outcomes that are \textbf{unstable} under the true preferences.
Example 2. \( S=\{s_1, s_2, s_3\} \), \( C=\{c_1, c_2, c_3\} \) and \( q(c_j)=1, \text{ for all } j=1,2,3 \).

\[
\begin{align*}
P(s_1) &= c_1, c_2, c_3 \\
P(s_2) &= c_3, c_2, c_1 \\
P(s_3) &= c_1, c_3 \\
Q(s_1) &= c_2, c_3 \\
Q(s_2) &= c_3, c_1 \\
Q(s_3) &= c_1, c_3 \\
\end{align*}
\]

\[\mu_S(Q): \]
\[
\begin{align*}
s_1 &\rightarrow c_2 \\
s_2 &\rightarrow c_3 \\
s_3 &\rightarrow c_1 \\
\end{align*}
\]

\[\mu_C(Q): \]
\[
\begin{align*}
s_1 &\rightarrow c_1 \\
s_2 &\rightarrow c_1 \\
s_3 &\rightarrow c_1 \\
\end{align*}
\]

\( Q \) is a Nash equilibrium for the matching rule \( f \) such that \( f(Q)=\mu_S(Q) \).

Suppose \( Q'=\langle Q'(s_1), Q(s_2), Q(s_3) \rangle \) and \( s_1 \) prefers \( f(Q') \) to \( \mu_S(Q) \). Set \( \mu'=f(Q') \).

\[\mu': \]
\[
\begin{align*}
s_1 &\rightarrow c_1 \\
s_2 &\rightarrow c_3 \\
s_3 &\rightarrow (s_3) \\
\end{align*}
\]

or

\[\mu': \]
\[
\begin{align*}
s_1 &\rightarrow c_1 \\
s_2 &\rightarrow (s_2) \\
s_3 &\rightarrow (s_3) \\
\end{align*}
\]

Any of the cases contradicts the stability of \( \mu' \) under \( Q' \). Therefore, \( Q \) is a Nash equilibrium of every game in which \( Q \) realizes \( \mu_S(Q) \). However \( (s_1, c_1) \) destabilizes \( \mu_S(Q) \) under the true preferences, so \( \mu_S(Q) \) is unstable under \( P \). \( \blacksquare \)
For the center-optimal stable matching rule every Nash equilibrium outcome is stable under the true preferences.

Theorem 1: (Sotomayor 2008) (Stability of the Nash equilibrium outcome for a particular stable matching rule) Let $Q$ be a preference profile. Let $f$ be a stable matching rule such that $f(Q) = \mu_c(Q)$. If $Q$ is a Nash equilibrium for $\Gamma(f)$ then $\mu_c(Q)$ is stable for the market $M(P)$.

Theorem 2. (Sotomayor, 2008) In the game induced by the center-optimal stable matching rule, the set of Nash equilibrium outcomes coincides with the set of stable matchings under the true preferences.
Definition 4. A strategy profile \( Q \) is a **Nash equilibrium in the strong sense** for the admission game induced by some stable matching rule if for every student \( s \), for every matching \( \mu \) in \( E(Q) \), for every \( Q'=(Q'(s),Q-s) \) and every matching \( \mu' \) in \( E(Q') \) we have that \( \mu(s) \geq_{P(s)} \mu'(s) \).

A strategy profile \( Q \) is a **Nash equilibrium in the strong sense** if it is a Nash equilibrium of the game induced by any stable matching rule.

**Proposition 2:** Existence of Nash equilibria in the strong sense. Let \( \mu \) be a stable matching for the market \( M(P)=(S,C,P_S,\{P_C,q_C\}) \). Suppose each student \( s \in \mu(C) \) chooses the strategy \( Q(s) = \mu(s) \) if \( s \) is matched at \( \mu \); \( Q(s)=s \) if \( s \) is unmatched at \( \mu \). Then \( Q \) is a Nash equilibrium in the strong sense and \( \mu \) is the equilibrium outcome. Furthermore, \( E(Q)=\{\mu\} \).
**Random stable matching rule.** This is the name we give here to the procedure that picks, at random (i.e., with equal probability), a stable matching for the selected preferences.

This kind of rule is justified in centralized settings where neither of the center-optimal and student-optimal stable matching rules is considered socially fair.

When the strategies are selected, no one knows which matching rule will be used. Then, *ex-ante*, the outcome of the game will be a lottery. Therefore, if the students play some profile of strategies $Q$ whose set of stable matchings is given by, say, $\mu_1$, $\mu_2$ and $\mu_3$, then they expect that each of these matchings will be selected with probability $1/3$. It might be that $Q$ is optimal for the students if $\mu_1$ is selected but is not optimal for them if $\mu_2$ or $\mu_3$ is chosen. Thus, it is natural to expect that the students would want to protect themselves by playing a profile of strategies such that *ex-post*, after randomization is done, no one could regret the choice made in the sense that, if one of the students had changed her strategy she would have had no chance of being better off.
The solution concept that captures this idea of strategic equilibrium is that of **Nash equilibrium in the strong sense**. This game is well defined since the students know the probability distribution on the set of stable matchings for the strategy profile selected. The assumption that any stable matching has equal probability of being selected is natural in our context but our results do not require this.

**Definition 5.** The profile of preferences $Q$ is a Nash equilibrium of the random stable matching rule $F$ if for every student $s$ and strategy profile $Q' = (Q'(s), Q_{-s})$ we have that

$$\text{Prob}(F(Q) = \mu) > \text{Prob}(F(Q') >_{P(s)} \mu)$$

for every $\mu$ in $E(Q)$.

That is, when $s$ deviates and selects $Q'(s)$, the probability that $F$ chooses a matching preferred to some $\mu$ in $E(Q)$ by $s$ is less than the probability that $F$ chooses $\mu$ when $Q$ is selected.
Nash equilibrium in the strong sense is not only a Nash equilibrium ex post under the random stable matching rule, but it is equivalent to the concept of Nash equilibrium for such rule.

Proposition 3. (Sotomayor, 2008) The profile of strategies $Q$ is a Nash equilibrium of the game induced by the random stable matching rule if and only if it is a Nash equilibrium in the strong sense.

Proposition 4. (Sotomayor, 2008) Let $Q$ be a Nash equilibrium of some admission game. Then $Q$ is a Nash equilibrium in the strong sense if and only if $|E(Q)|=1$. 


It turns out that if the set of stable matchings for a Nash equilibrium preference profile is a singleton, the Nash equilibrium outcome is the center-optimal stable matching for the selected profile. By Theorem 1 such Nash equilibrium outcome must be stable under the true preferences. On the other hand, for every stable matching under the true preferences, the Nash equilibrium profile given in the existence result is a Nash equilibrium in the strong sense. Then the set of Nash equilibrium outcomes in the strong sense coincides with the set of stable matchings with respect to the true preferences.

Theorem 3. (Sotomayor, 2008) Let $Q$ be a Nash equilibrium in the strong sense of the admission game induced by some stable matching rule $f$. Then $f(Q)$ is stable under the true preferences. Moreover, any stable matching under the true preferences is a Nash equilibrium outcome in the strong sense of the game induced by $f$, and $\mu_S$ is the equilibrium outcome most preferred by the students.
REMARK. The assumption that the centers do not have strategic behavior is not necessary to the conclusions of all of these results if we restrict the Nash equilibria to those profile of preferences in which the centers select their true preferences. These results can also be derived from Theorem*. The existence of Nash equilibrium in the strong sense in this general setting can be derived from Proposition 2.
Theorem 4. (Sotomayor, 2008) Let $\mu$ be a stable matching for the market $M(P)=(S,C,P_S,\{P_C,q_C\})$. Suppose each center $c$ chooses its true preference $P(c)$ and each student $s \in \mu(C)$ chooses the strategy $Q(s)$ of listing only $\mu(s)$ as the only acceptable center; $Q(s)=s$ in case $s$ is unmatched at $\mu$. The resulting profile of strategies, $(Q_S,P_C)$, is a Nash equilibrium in the strong sense and $\mu$ is the equilibrium outcome. Furthermore, $E(Q_S,P_C)=\{\mu\}$.

The argument to show that no center can profit by deviating from its strategy is that, since every student $s$ only lists $\mu(s)$, then a center $c$ can only be feasibly matched to students matched to it at $\mu$, whatever strategy it uses, so it cannot be better off.
Definition 8. (Aumann, 1959). A profile of strategies is a strong equilibrium point if no coalition of players can all gain by a simultaneous deviation (while the players outside the coalition maintain their strategies).

A special class of strategies used in practice by the participants consists of truncations of the true preferences.

\[ P(s) = c_6, c_2, c_3, c_1, c_5, c_4, s \qquad Q(s) = c_6, c_2, c_3, c_1, s \]

Definition 9. The list of preference \( Q(s) \) is a truncation of \( P(s) \) at agent \( v \) if \( Q(s) \) ranks the centers in the same order as \( P(s) \), but ranks as unacceptable all centers which are ranked below \( v \). We say that \( Q(s) \) is a truncation of \( P(s) \) if it is a truncation of \( P(s) \) at some agent \( v \).
Proposition 5. (Sotomayor 2008) For each student $s$ let $Q(s)$ be a truncation of $P(s)$. If the profile of strategies $Q$ is a Nash equilibrium in the strong sense then $E(Q) = \{\mu_s\}$.

Proposition 6. (Sotomayor 2008) For each student $s$ let $Q(s)$ be a truncation of $P(s)$. Then $Q$ is a strong equilibrium point in the strong sense if and only if $E(Q) = \{\mu_s\}$.

Consequently, every Nash equilibrium in the strong sense is a strong equilibrium in the strong sense.

Corollary 1. (Sotomayor 2008) Suppose the students only play, in equilibrium, truncations of the true preference lists. Let $f$ be any stable matching rule. Then $f$ implements the student-optimal stable matching correspondence in strong equilibrium in the strong sense concept.

- I prove that the stability of the Nash equilibrium outcome is not always reached.
- I introduce the concept of **Nash equilibrium in the strong sense** as a refinement of the Nash equilibrium concept, interpreted as the *ex-post* Nash equilibrium of a certain **random stable matching rule**.
- I characterize the Nash equilibria in the strong sense as those preference profiles for which the set of stable matchings is a singleton.
- I show that **the Nash equilibria in the strong sense yield stable matchings under the true preferences** and
- provide an implementation of the set of stable matchings by any stable matching rule via the Nash equilibrium in the strong sense concept.

_______________(2005) “Reaching the core through a random stable allocation mechanism” . Mimeo.


• I prove that the Nash equilibrium in the strong sense concept is indeed equivalent to the Nash equilibrium concept for the random stable matching rule.

• I found out that almost all results for the center admission model carry over the college admission model, without any restriction on the strategic behavior of the colleges.
Nash equilibria

\[ \{Q_1, Q_2, Q_3, Q, Q'\} \]

\[ \mu' = \mu \]

\[ E(P) = \{\mu_1, \mu_2, \mu_3\} \]