Affirmative Action through Minority Reserves: An Experimental Study on School Choice

Flip Klijn
Joana Pais
Marc Vorsatz

This version: September 2014

(February 2014)

Affirmative Action through Minority Reserves: An Experimental Study on School Choice*

Flip Klijn[†]   Joana Pais[‡]   Marc Vorsatz[§]

September 12, 2014

Abstract

Minority reserves are an affirmative action policy proposed by Hafalir *et al.* [20] in the context of school choice. We study in the laboratory the effect of minority reserves on the outcomes of two prominent matching mechanisms, the Gale-Shapley and the Top Trading Cycles mechanisms. Our first experimental result is that the introduction of minority reserves enhances truth-telling of some minority students under the Gale-Shapley but not under the Top Trading Cycles mechanism. Secondly, for the Gale-Shapley mechanism we also find that the stable matchings that are more beneficial to students are obtained more often relative to the other stable matchings when minority reserves are introduced. Finally, the overall expected payoff increases under the Gale-Shapley but decreases under the Top Trading Cycles mechanism if minority reserves are introduced. However, the minority group benefits and the majority group is harmed under both mechanisms.

Keywords: affirmative action, minority reserves, school choice, deferred acceptance, top trading cycles, truth-telling, stability, efficiency.

JEL–Numbers: C78, C91, C92, D78, I20.

1 Introduction

The concern for diversity has led many school districts in the U.S. to take affirmative action in promoting equal access for those groups that have been traditionally excluded, thus becoming more representative of their surrounding community. As such, school choice programs give students the opportunity to express their preferences over schools and thus provide choice to students, but the final assignment is also shaped by admission policies that aim at maintaining racial, ethnic, or socioeconomic balance.[1]

In the last decade, matching theory and mechanism design have been employed to accommodate affirmative action in school choice. For instance, Abdulkadiroğlu and Sönmez [2] and Abdulka-diroğlu [1] consider a cap or maximum quota on the number of students from the same group that a school can admit. Kojima [25] shows that affirmative action policies based on maximum quotas can be detrimental to the very minorities they are supposed to help. More precisely, for the widely studied Gale-Shapley (GS) and Top Trading Cycles (TTC) mechanisms there are environments in which every minority student is hurt by the introduction of maximum quotas. The intuition behind this result is as follows. If there is a school that is mostly wanted by majority students, it may end up with unfilled seats and unassigned majority students may create competition for seats at other schools, thus hurting minority students.

To overcome this problematic feature of placing bounds on the number of seats for majority students, Hafalir *et al.* [20] propose an affirmative action policy by reserving seats for minority students. Here, schools give higher priority to minority students up to the point when minorities fill the reserves, so that a school may assign some of its reserved seats to majority students provided that no minority student prefers that school to her assigned school. Hafalir *et al.* [20] adapt GS and TTC to minority reserves and explore the properties of the resulting mechanisms. They show that the GS and TTC mechanisms with minority reserves preserve the property of strategy-proofness, i.e., no student can ever benefit by misrepresenting her preferences. Moreover, when all students tell the truth there is a clear sense in which minority reserves present an improvement over majority quotas. Indeed, the GS with minority reserves (weakly) Pareto dominates the GS with majority quotas and, considering minority students only, is not strictly Pareto dominated by the standard GS. As for TTC, the advantage is less clear, but it is still the case that, for minority students, the TTC with minority reserves is not strictly Pareto dominated by the standard TTC.

Still, the theoretical results may easily break down in practice. It is well documented in the experimental literature on school choice that agents do not always realize it is in their best interest to reveal their true preferences when confronted with strategy-proof mechanisms.[2] Individual behavior that deviates from truthtelling can have serious consequences for all participants involved. Since distinct but strategy-proof mechanisms may be perceived differently, they could give rise to very different types and levels of non-truthful behavior. Therefore, whether or not affirmative

---

[1]See, for instance, Abdulkadiroğlu and Sönmez [2].

[2]See, for instance, Calsamiglia *et al.* [6], Chen and Kesten [7], Chen and Sönmez [8], Featherstone and Niederle [14], Guillen and Hing [19], Klijn *et al.* [24], and Pais and Pintér [27].

action policies actually benefit minority students may well depend on how agents perceive the different mechanisms.

This provided the motivation for our experimental study. We consider four mechanisms: the standard GS, the standard TTC, and the two counterparts proposed in Hafalir *et al.* [20], i.e., GS with minority reserves and TTC with minority reserves. We present a stylized environment where six students —four majority and two minority— are to be assigned to three schools with two positions each. The subjects take the roles of the students. As in real-life school choice applications, schools are not strategic as their priorities are fixed by law. When minority reserves are present, each school reserves exactly one seat for minority students. Preferences of students over schools are such that it is possible to assign each student to the school she ranks first. The (unique) Pareto-efficient outcome is obtained under any of the four mechanisms when all students rank their best school first, namely by reporting the true preferences. This common outcome under (rational) straightforward behavior helps us to compare and measure possibly different outcomes in the experiment. Schools' priorities over students are such that there is a strong opposition of interests between the two sides of the school choice problem. As a result, under the standard mechanisms as well as under GS with minority reserves, the Pareto-efficient outcome can only be reached when exactly all students coordinate and rank their most preferred school first. Under TTC with minority reserves, necessary conditions are less demanding, but still some coordination is needed to achieve the Pareto-efficient outcome.[3]

In our setting, minority reserves may have a positive effect on truth-telling. In fact, minorities may feel protected when minority reserves are used and thus compelled to rank their best school first and in particular tell the truth more often. Given the coordination element in this setup, as minority students tell the truth more often, the easier it is for other students to recognize that ranking the best school first or —since there is no reason to report a preference list that switches the ordering of the remaining schools— truth-telling is the best possible strategy. It follows that spillovers on the truth-telling rates of majority students may occur.

Our results show that truth-telling by minority students increases when minority reserves are introduced in GS, but not in TTC (Result 1). We might expect that increased truth-telling by minority students would positively affect the truth-telling rates of majority students, but this is not observed in the data.

We then analyze fairness of the final outcome using the standard notion of stability. In our setup, a matching is stable if it has no blocking pairs, i.e., for any student, all the schools she prefers to the one she is assigned to have exhausted their capacity with students that have higher priority. Results show that the proportion of stable matchings does not necessarily increase when minority reserves are introduced; however, under GS, the stable matchings that are more beneficial to students are obtained more often with respect to the other stable matchings when minority reserves are added. In fact, the probability distribution over stable matchings under GS with minority reserves first-order stochastically dominates the distribution obtained under the standard GS mechanism, but the converse holds for TTC (Result 2a). Following Hafalir *et al.* [20] we then

---

[3]See Appendix A for a formal statement and proof of these results.

analyze a weaker notion of stability, which disregards those blocking pairs containing majority students that do not represent an instance of justified envy when minority reserves are present. Ordering such "minority-stable" matchings according to average payoff, we obtain a similar dominance result: the probability distribution over minority-stable matchings under GS with minority reserves "almost" first-order stochastically dominates the distribution obtained under the standard GS mechanism, while the probability distribution over minority-stable matchings under the standard TTC first-order stochastically dominates the distribution obtained under TTC with minority reserves (Result 2b). Looking into the composition of blocking pairs, the mechanisms with minority reserves —particularly GS with minority reserves— appear to be fairer towards minority students as the results show that the probability of belonging to a blocking pair is significantly lower for the average minority student than for the average majority student under these mechanisms (Result 2c).

Finally, we have the following findings on efficiency. Despite the simplicity of the environment and the strong coordination element of the setup, expected payoffs do not necessarily increase with the introduction of minority reserves. In fact, the introduction of minority reserves in both GS and TTC harms majority students, even though it benefits minority students inasmuch as the expected payoff of minority students as a group increases (Result 3). This efficiency advantage is actually very clear for minority students in GS as the distribution of a minority student's payoffs under GS with minority reserves first-order stochastically dominates her payoff distribution under the standard GS (Result 3). Instead, the distribution of several majority students' payoffs under the standard TTC first-order stochastically dominates that under the TTC with minority reserves.

Our results therefore indicate that, while the benefits of introducing minority reserves in TTC are not clear, GS with minority reserves leads to higher rates of truth-telling by minority students and to outcomes that favor minority students both in what fairness measures and welfare are concerned.

These results contrast with the findings of Kawagoe *et al.* [23], who experimentally study minority reserves as well as maximum quotas. Unlike our study, they consider exclusively GS and analyze three mechanisms —the standard GS, GS with minority reserves, GS with majority quotas— and, using the same participants, test two markets that differ in students' preferences.[4] In the debate minority reserves *vs.* majority quotas, Kawagoe *et al.* [23]'s experimental results are in line with theory as GS with majority quotas never delivers higher average payoffs than GS with minority reserves. Nevertheless, contrary to our results, introducing minority reserves does not necessarily benefit minority students. The reason for this is that truth-telling by minority students is not necessarily higher under GS with minority reserves than under the standard GS. In what stability is concerned, conclusions are not so clear, but GS with minority reserves fares worse than the majority quotas counterpart in one of the markets.

Other types of affirmative action in school choice are studied in the literature as well. Kominers and Sönmez [26] generalize Hafalir *et al.* [20] to allow for slot-specific priorities. Westkamp [29]

---

[4]To be precise, they test the standard GS plus two different descriptions of GS with minority reserves and of GS with majority quotas.

studies the German university admissions system, where priorities may also vary across slots, and Braun *et al.* [4] and Braun *et al.* [5] complement his analysis by conducting a field experiment and a laboratory experiment, respectively. The laboratory experiment confirms the result in Westkamp [29] that the mechanism that is currently used for university admissions in Germany, designed to give top-grade students an advantage, actually harms them. Kamada and Kojima [22] study entry-level medical markets in Japan, where regional caps are used to overcome the shortage of doctors in rural areas. Moreover, Ehlers *et al.* [12], Fragiadakis *et al.* [16], and Fragiadakis and Troyan [17] combine lower and upper bounds on the number of students of each type. Finally, Echenique and Yenmez [11] and Erdil and Kumano [13] study generalizations of schools' priorities over sets of students with the aim of capturing diversity.

The remainder of the paper is organized as follows. We describe the stylized school choice problem and the mechanisms in Section 2. In Section 3, we present our experimental hypotheses and, in Section 4, we describe the obtained experimental results. Section 5 concludes.

2 The experiment

2.1 The school choice problem

Our experiment aims at analyzing the effect of minority reserves on the functioning of GS and TTC in school choice. We consider throughout the problem described in Table 1, which we explain next.

	Preferences						Priorities		
	M_1	M_2	M_3	M_4	m_1	m_2	s_1	s_2	s_3
Best match	s_1	s_1	s_3	s_3	s_2	s_2	m_1	M_3	M_1
Second best match	s_2	s_2	s_1	s_1	s_3	s_3	m_2	M_4	M_2
Third best match	s_3	s_3	s_2	s_2	s_1	s_1	M_4	M_2	m_2
Fourth best match							M_3	M_1	m_1
Fifth best match							M_1	m_1	M_3
Sixth best match							M_2	m_2	M_4

Table 1: Preferences of students over schools (left) and priority orderings of schools over students (right).

Six students look for a seat at one of three schools. Four students —indicated by M_1, M_2, M_3, and M_4— belong to the *majority group*, and two students —m_1 and m_2— form the *minority group*. The three schools are denoted by s_1, s_2, and s_3. Each school offers exactly two seats.

Table 1 shows that the students can be divided into three groups according to their preferences. Students M_1 and M_2 like s_1 most and s_3 least, students M_3 and M_4 like s_3 most and s_2 least, and students m_1 and m_2 like s_2 most and s_1 least. Since all schools offer two seats, it is feasible to assign all students to their most preferred school, which is therefore the unique Pareto-efficient

outcome.[5] The priority orderings of the schools are such that students M_1, M_3, and m_1 are ranked fifth in their most preferred, fourth in their second most preferred, and first in their least preferred school. Similarly, students M_2, M_4, and m_2 are ranked last in their most preferred, third in their second most preferred, and second in their least preferred school. We chose this particular problem for two main reasons. First, in the absence of any positive discrimination in favor of the minority students, m_1 is in exactly the same situation as M_1 and M_3 (and m_2 faces exactly the same decision problem as M_2 and M_4). This symmetry helps us to evaluate the effect of the minority reserves on the two mechanisms in a clear-cut way. Second, the setup exhibits a strong coordination element. In fact, if under any of the mechanisms all students rank their best schools first, namely by reporting the true preferences, each student will be admitted to her top school. However, by looking at the priority orderings only, it seems "easier" for a student to get a seat at less desirable schools. In fact, the structure of the priority orderings puts tension on obtaining the Pareto-efficient outcome to such an extent that under the standard mechanisms and under GS with minority reserves it is possible to reach the Pareto-efficient matching only if all students rank their best schools first. Under TTC with minority reserves the necessary conditions are less stringent, but still some coordination is needed to reach this outcome as the two minority students and the majority student M_3 have to rank their best schools first.[6]

During the experiment, subjects assume the role of students that seek to find a seat at one of the schools. Schools are not strategic players. Given the information in Table 1, the subjects' task is to submit a ranking over schools (not necessarily the true preferences) to be used by a central clearinghouse to assign students to schools. They receive 12 experimental currency units (ECU) in case they end up in their most preferred school, 9 ECU if they get a seat in their second most preferred school, and 6 ECU if they study in their least preferred school.

We consider four different matching mechanisms. Our two baseline mechanisms that treat all students equally are the standard GS student proposing deferred acceptance mechanism (GSs) and the standard TTC mechanism ($TTCs$). Following Hafalir *et al.* [20], the corresponding two modified mechanisms, denoted by GSm and $TTCm$, favor the minority group by obliging each school to reserve one seat to students from this group in case it is demanded; that is, each school has a minority reserve of 1. For the particular school choice setting at hand, the two GS mechanisms are as follows:

Gale-Shapley mechanisms (GSs and GSm)

Step 1: Each student sends an application to the school she ranked first.

Step 2: Each school that receives at least one application acts as follows.

(GSs) It temporarily accepts the applicant with the highest priority. It also temporarily accepts the applicant with the highest priority among all remaining applicants (if any). The rest of the applicants (if any) are rejected.

---

[5]Schools are considered mere objects and are thus not taken into account for welfare evaluation.

[6]See Appendix A for formal statements and proofs of these results.

(*GSm*) If the school receives no application from minority students, proceed as in Step 2 of *GSs*. If the school receives at least one application from minority students, then it temporarily accepts the minority applicant with the highest priority; it also temporarily accepts the applicant with the highest priority among all remaining (majority or minority) applicants (if any); the rest of the applicants (if any) are rejected.

Step 3: Whenever a student is rejected by a school, she applies to the next highest ranked school.

Step 4: Each school that receives at least one new application acts as follows.

(*GSs*) Among all new and retained applications, the school temporarily accepts the applicant with the highest priority. Among the remaining (new and retained) applications (if any), it also temporarily accepts the applicant with the highest priority (if any). The rest of the applicants (if any) are rejected.

(*GSm*) The school considers the new and the retained applications. If none of these applications is from a minority student, proceed as in Step 4 of *GSs*. If there is at least one application from a minority student, the school temporarily accepts the minority applicant with the highest priority; it also temporarily accepts the applicant with the highest priority among all remaining (majority or minority) applicants (if any); the rest of the applicants (if any) are rejected.

Step 5: Steps 3 and 4 are repeated until no more students are rejected, and the assignment is finalized. Each student is matched to the school that holds her application at the end of the process.

Similarly, for the particular school choice setting at hand, the two TTC mechanisms are as follows:

Top Trading Cycles mechanisms (TTCs and TTCm)

Step 1: Each student points to the school she ranked first.

(*TTCs*) Each school points to the student with the highest priority.

(*TTCm*) Each school points to the minority student with the highest priority.

There is at least one cycle of students and schools. Each student in any of the cycles is matched to the school she is pointing to and the school's number of available seats is reduced by one.

Step 2: Each unmatched student points to the school she ranks highest among all schools that still have available seats.

(*TTCs*) Each school with at least one available seat points to the student with the highest priority among all remaining students.

($TTCm$) A school that was not matched to a minority student before points to

- the unmatched minority student if there is one,[7] and otherwise points to
- the majority student with the highest priority among all remaining students.

A school that was already matched to a minority student points to the student (minority or majority) with the highest priority among all remaining students.

There is at least one cycle of students and schools. Each student in any of the cycles is matched to the school she is pointing to and the school's number of available seats is reduced by one.

Step 3: Repeat Step 2 until all students are matched.

2.2 Procedures

The experiment was programmed within the z–Tree toolbox provided by Fischbacher [15] and carried out in the computer laboratory at a local university. We used the ORSEE registration system by Greiner [18] to invite students from a wide range of faculties. In total, 175 undergraduates from various disciplines participated in the experiment.[8]

We considered two matching mechanisms in each session. At the beginning of the session, subjects were anonymously matched into groups of six. Each subject received instructions for one of the four mechanisms together with an official payment receipt and was told that she would play two games under this mechanism: once in her true and once in a fictitious role.[9]

Subjects could study the instructions at their own pace and any doubts were privately clarified. Participants were also informed that after this first phase of two games they would take part in a second phase with a different matching mechanism. Subjects knew that their decisions in the first phase would not affect their payoffs in the second phase (to avoid possible hedging across phases) and that they would not receive any information regarding the decisions of any other player (so that they could not condition their actions in the second phase on the behavior of other

---

[7]Since in Step 1 all schools point to minority students and since there is at least one cycle, at least one minority student is matched in Step 1. Hence, in Step 2 there is at most one minority student.

[8]Since not all sessions had a number of participants that is a multiple of 6, members of filled groups were randomly added to the extra participants that could not form an additional group of 6 students on their own in order to be able to determine the payoff of the extra students (this insures that there is no deception for the extra subjects).

[9]This procedure allows us to collect the same number of observations for both minority and majority students. Since subjects did not know which of their roles was true and which was fictitious, incentives are not affected. The true (fictitious) role of the six subjects in a group were: M_1 (m_1) for the first, M_2 (m_2) for the second, M_3 (m_1) for the third, M_4 (m_2) for the fourth, m_1 (M_1) for the fifth, and m_2 (M_2) for the sixth subject. We distinguish between true and fictitious roles in the instructions because it is not possible to calculate the outcome when subjects act in their fictitious roles as there are four subjects who play the game in the role of a minority student and two subjects in the role of a majority student. So, the actions in the fictitious roles cannot be used for payment.

participants under the first matching mechanism). Also, no feedback whatsoever was provided during the entire experiment.

After completing the two games under the first mechanism, subjects received instructions for the second mechanism. The group assignment did not change and subjects took decisions in the same roles (neither the true nor the fictitious role changed). To prevent income effects, either the first or the second mechanism in the true roles was payoff relevant, which was randomly determined by the central computer at the end of the experiment. The whole procedure was known by the subjects from the beginning.

We ran a total of eight sessions in such a way that (i) each mechanism was played four times, (ii) each mechanism was played the same number of times in phase 1 and phase 2, and (iii) either the mechanics (GS or TTC) or the setting (minority reserves or not) was fixed in each session.[10] Subjects received 1 Euro per ECU earned during the experiment. A typical session lasted about 75 minutes and subjects earned on average 12 Euro (including a 3 Euro show-up fee) for their participation.

3   Hypotheses

We derive the null hypothesis of the experiment by analyzing the economic incentives of the students. It is well-known that in the setting without minority reserves, the Gale-Shapley and the Top Trading Cycles mechanisms are strategy-proof (see Dubins and Freedman [10], Roth [28], and Abdulkadiroğlu and Sönmez [2]); that is, no student can gain from misrepresenting her preferences. Strategy-proofness of the corresponding mechanisms with minority reserves is established in Hafalir *et al.* [20]. If students indeed report their preferences truthfully, all four mechanisms assign all students to their most preferred school.[11] It follows that the resulting matching is the student-optimal stable matching and the unique Pareto-efficient outcome.

Null Hypothesis: *In all four mechanisms, preferences are revealed truthfully. Hence, all four mechanisms generate the student-optimal stable matching and are Pareto-efficient.*

The construction of the alternative hypotheses starts from the general experimental observation that not all subjects reveal preferences truthfully when the standard mechanisms are employed; see, for instance, Calsamiglia *et al.* [6], Chen and Kesten [7], Chen and Sönmez [8], Featherstone and Niederle [14], Guillen and Hing [19], Klijn *et al.* [24], and Pais and Pintér [27]. To say it differently, we presume that there is a considerable number of subjects who do not realize that

---

[10]More precisely, the combinations of mechanisms in the eight sessions were as follows: session $1 = GSs - GSm$, session $2 = GSm - GSs$, session $3 = GSm - TTCm$, session $4 = TTCm - GSm$, session $5 = TTCm - TTCs$, session $6 = TTCs - TTCm$, session $7 = TTCs - GSs$, and session $8 = GSs - TTCs$. We have 93 observations for GSs, 83 observations for GSm, 92 observations for $TTCs$, and 82 observations for $TTCm$.

[11]We focus on truth-telling, as it is interesting on its own right. Nevertheless, to be precise, having all students ranking their preferred school first is sufficient for such an assignment to be obtained.

it is in their best interest to report their true preferences under GSs and $TTCs$. Then, since M_1, M_3, and m_1 (and similarly, M_2, M_4, and m_2) all face the same situation when there are no minority reserves and the mechanisms do not distinguish between majority and minority students, it is natural to hypothesize that *at least* the level of truthfully reported preferences is the same for these students.

ALTERNATIVE HYPOTHESIS 1: *In the absence of minority reserves, that is for both GSs and TTCs, there is no difference in the level of truthfully reported preferences between M_1, M_3, and m_1. Similarly, the degree of truth-telling among M_2, M_4, and m_2 is the same.*

On the other hand, the introduction of minority reserves has the potential to positively affect the level of truth-telling among minority students, now facing reduced competition for their top schools. This effect is clear in the case of student m_1. In the presence of minority reserves, m_1 will be assigned to her most preferred school, s_2, in case she puts that school on the top of her reported preferences independently of the behavior of the other students. Having realized this, m_1 should rank her best school s_2 first. Furthermore, as there is no reason why m_1 would invert the true order of the other schools on her list, ultimately we may expect her to tell the truth. So, even though m_1 may not be aware of the strategy-proofness of GSm and $TTCm$, she may be tempted to reveal her true preferences because minority reserves render the strategic uncertainty of this student irrelevant.[12]

The case for m_2 is less obvious as she may still face competition for her top school, whose reserved seat might be taken by m_1. Nonetheless, the strategic uncertainty of m_2 is reduced by the introduction of minority reserves as they (i) should help her to obtain a better idea about the behavior of m_1 and (ii) provide her with an advantage over majority students. These two factors should help m_2 to make better decisions; that is, rank the best school first more often. Since there is no reason why m_2 would invert the order of her second best and least preferred schools, this implies (a) telling the truth more often in the presence than in the absence of minority reserves, and (b) telling the truth more often than the corresponding majority students when there are minority reserves.

ALTERNATIVE HYPOTHESIS 2: *Minority students report preferences truthfully more often in the presence than in the absence of minority reserves. In both GSm and TTCm, the level of truth-telling of student m_1 (m_2) is higher than the level of truth-telling of M_1 and M_3 (M_2 and M_4).*

The higher level of truth-telling of minority students in the presence of minority reserves might be foreseen by the other students. Since having all students ranking their best school first is sufficient to achieve the unique Pareto-efficient outcome in this setup, majority students may

---

[12]A player has strategic uncertainty if she is unsure about the actions or beliefs (or beliefs of beliefs, *etc.*) of others; see Brandenburger [3] for a formal definition and Heinemann *et al.* [21] for an experimental application to coordination games.

recognize truth-telling as (one of) the best possible strategies and may also be inclined to report preferences truthfully under the mechanisms with minority reserves.

ALTERNATIVE HYPOTHESIS 3: *The mechanisms with minority reserves generate more truth-telling among all students than the corresponding mechanisms without minority reserves.*

So far, we have discussed how minority reserves might affect the level of truth-telling. Next, we analyze the implications of these hypotheses on important measures of fairness (stability and stability with minority reserves) and on welfare (measured by average payoffs).

In our setup, a matching is stable if it has no "blocking pairs." This means that, for any majority or minority student, all the schools she prefers to the one she is assigned to have exhausted their capacity with students that have higher priority. Therefore, in a stable matching no student justifiably envies any other student, i.e., she does not form a "blocking pair" with any school. Using standard techniques it can be easily verified that there are a total of five stable matchings in our school choice problem (see Table 2). In the stable matching that is least attractive to the students, denoted by μ_1, each student is assigned to her worst school. So, the expected payoff from this matching (including the show-up fee of 3 ECU) is equal to 9 ECU. In the stable matching μ_2, three students (M_2, M_4, and m_2) are assigned to their second best school and the other three students (M_1, M_3, and m_1) are at their worst school. The expected payoff of μ_2 is thus 10.5 ECU. The stable matching μ_3 assigns all students to their second best school, which leads to an expected payoff of 12 ECU. In the stable matching μ_4, the students M_1, M_3, and m_1 are assigned to their best and the students M_2, M_4, and m_2 to their second best school. Hence, the expected payoff of μ_4 is 13.5 ECU. Finally, all students get a seat at their best school in the student-optimal stable matching μ_5. Therefore, this matching is Pareto-efficient (and the expected payoff per student is 15 ECU). Note that the students unanimously weakly prefer μ_5 to μ_4, μ_4 to μ_3, μ_3 to μ_2, and μ_2 to μ_1.

Student	Stable matching				
	μ_1	μ_2	μ_3	μ_4	μ_5
M_1	s_3	s_3	s_2	s_1	s_1
M_2	s_3	s_2	s_2	s_2	s_1
M_3	s_2	s_2	s_1	s_3	s_3
M_4	s_2	s_1	s_1	s_1	s_3
m_1	s_1	s_1	s_3	s_2	s_2
m_2	s_1	s_3	s_3	s_3	s_2
Expected payoff (in ECU)	9	10.5	12	13.5	15

Table 2: Stable matchings.

Since implementing minority reserves gives some advantage to minority students, Hafalir *et al.*

[20] introduce a different stability concept that gives less blocking ability to majority students. In our setup, a matching is stable with minority reserves —let us call it minority-stable— if it has no blocking pairs or if all of its blocking pairs are composed of a majority student M and a school s that M prefers to the one she is assigned to, such that two conditions are met: (i) s has exhausted its capacity with majority students that have higher priority than M and with at least one minority student that has lower priority than M, but still (ii) s admitted a number of minority students smaller or equal to its minimum reserve. It follows that, besides the five stable matchings, two other matchings, which we denote by μ_6 and μ_7, emerge as minority-stable in our school choice problem (see Table 3).[13] In the minority-stable matching μ_6, majority students M_1, M_2, and M_3 are matched to their worst schools, majority student M_4 is matched to her second best school, minority student m_1 is matched to her best school, and minority student m_2 is matched to her worst school. Hence, the expected payoff of μ_6 is 10.5 ECU. The minority-stable matching μ_7 assigns majority student M_1 to her worst school, majority students M_2, M_3, and M_4 to their second best schools, minority student m_1 to her best school, and minority student m_2 to her second best school. The expected payoff of μ_7 is therefore 12 ECU.

Student	Minority-stable matching						
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7
M_1	s_3	s_3	s_2	s_1	s_1	s_3	s_3
M_2	s_3	s_2	s_2	s_2	s_1	s_3	s_2
M_3	s_2	s_2	s_1	s_3	s_3	s_2	s_1
M_4	s_2	s_1	s_1	s_1	s_3	s_1	s_1
m_1	s_1	s_1	s_3	s_2	s_2	s_2	s_2
m_2	s_1	s_3	s_3	s_3	s_2	s_1	s_3
Expected payoff (in ECU)	9	10.5	12	13.5	15	10.5	12

Table 3: Minority-stable matchings.

Under Alternative Hypothesis 2, minority students are more likely to report preferences truthfully when the mechanisms with minority reserves are employed. For m_1, this implies that she directly obtains a seat at her top school s_2. In terms of stable matchings, m_1 is assigned to s_2 only in μ_4 and μ_5, and in terms of minority-stable matchings, m_1 is assigned to s_2 in μ_4, μ_5, μ_6, and μ_7. For m_2 the effects are less clear, since even though this student sees her strategic uncertainty reduced, her final assignment when telling the truth still depends on the behavior of majority students. Nevertheless, when both minority students submit their true preferences, m_2 faces no competition for her second best school s_3, so that she should obtain a seat in either her best school s_2 or in school s_3.

It follows that if stability is obtained when minority reserves are introduced, we should expect the matchings μ_4 and μ_5 to be reached more often. Similarly, if minority-stable matchings are

---

[13]See Appendix B for a proof of this claim.

reached, the matchings μ_4, μ_5, and μ_7 should be reached more often. However, it is clear that the behavior of majority students determines whether a stable or a minority-stable matching is reached, so that we are not able to derive implications from Alternative Hypothesis 2 on how the proportion of stable or minority-stable matchings changes when minority reserves are introduced. This leads to the following alternative hypothesis.

ALTERNATIVE HYPOTHESIS 4: *The probability of obtaining μ_4 or μ_5 relative to the other three stable matchings is higher in the presence than in the absence of minority reserves. Similarly, the probability of obtaining μ_4, μ_5, or μ_7 relative to the other four minority-stable matchings is higher in the presence than in the absence of minority reserves.*

The affirmative action targets the minority group, and we have already seen that the mechanisms with minority reserves help one of these students to ensure herself a seat at her most preferred school in an easy way. In fact, any successful discriminatory policy should increase the payoffs (or at least the average payoff) of the minority students when minority reserves are introduced. In our particular school choice problem, the effect of the minority reserves on the expected payoff of the majority group could be positive as well. Indeed, if the majority students see their own strategic uncertainty reduced and become more likely to reveal their true preferences (Alternative Hypothesis 3) or at least to rank their preferred schools first, the frequency of the Pareto-efficient outcome and their own payoffs could increase as well.

ALTERNATIVE HYPOTHESIS 5: *No student is harmed by the presence of minority reserves.*

The hypothesis that minority reserves may benefit all students pertains to this particular environment, where if all students tell the truth (or at least rank their best schools first), the same outcome is obtained under the four mechanisms. In other, more general contexts, this is not the case as the mechanisms with and without minority reserves may produce different outcomes for the same preferences and the effects of affirmative action on efficiency may actually be quite disappointing under full truth-telling.[14]

4   Results

We now present the aggregate results of our experiment. We are interested in how the different mechanisms perform in terms of truth-telling (Section 4.1). Afterwards, we study the implications of individual behavior on fairness (Section 4.2) and welfare (Section 4.3).

---

[14]In fact, it can only be guaranteed that if the affirmative action benefits no minority student, then at least one minority student is not worse off (see Hafalir *et al.* [20]). Moreover, when no minority student benefits from affirmative action, then also no majority student benefits (see Doğan [9]).

4.1 Truth-telling

Table 4 shows the probabilities with which each of the six possible strategies is played. We use the notation (2,3,1) for the ranking where a student lists her second most preferred school first, her least preferred school second, and her most preferred school last. The other five strategies (1,2,3), (1,3,2), (2,1,3), (3,1,2), and (3,2,1) have similar interpretations. There are no order effects (for a given pair of mechanisms, the distribution of reported preferences does not depend on which of the two mechanisms students face first and which second) and, therefore, we pool the data.

Mechanism	Submitted ranking					
	(1,2,3)	(1,3,2)	(2,1,3)	(2,3,1)	(3,1,2)	(3,2,1)
GSs	0.39	0.02	0.34	0.11	0.02	0.12
GSm	0.39	0.07	0.26	0.09	0.06	0.13
$TTCs$	0.48	0.05	0.31	0.07	0.03	0.06
$TTCm$	0.34	0.06	0.33	0.11	0.05	0.11

Table 4: Probability distribution of submitted rankings (aggregated over all six roles).

It can be seen from the second column that truth-telling is the most prominent strategy in all four mechanisms —ranging from 34% in treatment $TTCm$ to 48% in treatment $TTCs$—, in line with many experimental studies on matching,[15] yet considerably lower than what is predicted by theory (the Null Hypothesis). Consequently, there are many subjects who do not realize that it is in their best interest to report preferences truthfully. The effects the different mechanisms and roles have on truth-telling can be identified with the help of Table 5.

Mechanism	Student					
	$\underline{M_1}$	M_2	$\underline{M_3}$	M_4	$\underline{m_1}$	m_2
GSs	$\underline{0.47}$	0.37	$\underline{0.37}$	0.27	$\underline{0.44}$	0.37
GSm	$\underline{0.25}$	0.30	$\underline{0.43}$	0.14	$\underline{0.76}$	0.44
$TTCs$	$\underline{0.53}$	0.55	$\underline{0.50}$	0.33	$\underline{0.58}$	0.39
$TTCm$	$\underline{0.50}$	0.42	$\underline{0.21}$	0.07	$\underline{0.55}$	0.30

Table 5: Probability of truth-telling.

First, we know from Table 1 that students M_1, M_3, and m_1 face the same situation. In particular, they are ranked fifth in their most preferred school, fourth in their second most preferred school, and first in their least preferred school. A similar kind of symmetry holds true for students M_2, M_4, and m_2. Consequently, one would expect that the proportion of truth-telling of the

---

[15]See the examples mentioned in Section 3.

minority students is the same as that of their majority counterparts in the absence of minority reserves when there is no positive discrimination and the matching mechanisms treat subjects equally (Alternative Hypothesis 1). Mann-Whitney U tests confirm this intuition in all cases.[16]

Second, the introduction of minority reserves has one important effect. Independently of the behavior of the other students, m_1 can now directly obtain a seat in her most preferred school by ranking this school first and m_2 faces reduced competition as minority reserves give her an advantage over majority students. Hence, Alternative Hypothesis 2 stated that minority students tell the truth more often than their majority counterparts in the presence of minority reserves and increase their own level of truth-telling with respect to the situation when there are no minority reserves. In fact, we find that student m_1 tells the truth more often than both M_1 and M_3 in GSm and more often than M_3 in $TTCm$.[17] Student m_1 also tells the truth more frequently in GSm than in GSs (two-sided $p=0.0019$). On the other hand, m_1 tells the truth more often in $TTCs$ than in $TTCm$. So, minority reserves do not positively affect the level of truth-telling of this student under TTC. This interpretation is further strengthened when one compares the behavior of m_1 in GSm with that in $TTCm$. The level of truth-telling reaches 76% in GSm but "only" 55% in $TTCm$ (two-sided $p=0.0406$).[18] Furthermore, there is no statistical difference between m_2 and M_2 nor M_4 and no evidence that m_2 tells the truth more often under minority reserves.[19]

Finally, we would also like to know whether the possible higher level of truth-telling of minority students in the presence of minority reserves is anticipated by the other participants, and whether this change in their beliefs induces them to tell the truth more often as well (Alternative Hypothesis 3). However, we do not find any significant spillover effect of this kind.[20] So, either the other subjects are not aware of the different situation of the minority students or they do not respond to this change in their beliefs. We summarize as follows.

Result 1. *There is a positive effect of minority reserves on the level of truth-telling of m_1 under the Gale-Shapley mechanism. There are no spillover effects on other students.*

4.2 Fairness

We now analyze the effects of minority reserves on fairness. We base our measures of fairness on blocking pairs since a blocking pair represents a possible instance of justified envy. First,

---

[16]The two-sided p-values under GSs are 0.4840 for the comparison between m_1 and M_1, 0.3458 for m_1 and M_3, 0.4840 for m_2 and M_2, and 0.3458 for m_2 and M_4. The two-sided p-value under $TTCs$ are 0.4840 for the comparison between m_1 and M_1, 0.7728 for m_1 and M_3, 0.5653 for m_2 and M_2, and 0.4237 for m_2 and M_4.

[17]The two p-values of the Mann Whitney U tests are 0.0015 (0.0107) for the comparison with M_1 (M_3) in GSm and 0.0368 for the comparison with M_3 in $TTCm$.

[18]And this occurs while it follows from Table 5 that the baseline level of truth-telling for m_1 in the absence of minority reserves is higher for the Top Trading Cycles than for the Gale-Shapley mechanism (58% *vs.* 44%).

[19]The two-sided p-values of the Mann Whitney U tests range from 0.1294 (when comparing m_2 with M_4 under GSm) to 0.5297 (for the comparison with M_2 under $TTCm$). The two-sided p-values associated to the introduction of minority reserves are 0.8798 for GS and 0.4116 for TTC.

[20]The two-sided p-values of the Mann Whitney U tests range from 0.0835 to 0.8173.

we consider the standard measure and look for matchings characterized by the complete absence of blocking pairs (stable matchings). Second, we consider a measure adapted to the presence of minority reserves where certain blocking pairs that involve majority students are not permitted (minority-stable matchings). Finally, we provide a more detailed analysis by presenting the expected number of blocking pairs, split out according to different types.

4.2.1 Stable matchings

We first analyze the effects of minority reserves on the probability of obtaining stable matchings.

Mechanism	Matchings						
	μ_1	μ_2	μ_3	μ_4	μ_5	$\mu_4 + \mu_5$	Stable
GSs	0.0043	0.0602	0.2062	0.0702	0.0043	0.0763	0.3471
	(0.0125)	*(0.1744)*	*(0.5973)*	*(0.2034)*	*(0.0125)*	*(0.2158)*	*(1.0000)*
GSm	0.0013	0.0180	0.0025	0.1062	0.0038	0.1100	0.1317
	(0.0099)	*(0.1366)*	*(0.0190)*	*(0.8058)*	*(0.0288)*	*(0.8346)*	*(1.0000)*
$TTCs$	0.0007	0.0000	0.0024	0.0216	0.0212	0.0428	0.0460
	(0.0152)	*(0.0022)*	*(0.0522)*	*(0.4696)*	*(0.4609)*	*(0.9304)*	*(1.0000)*
$TTCm$	0.0037	0.0317	0.0141	0.0525	0.0137	0.0662	0.1157
	(0.0320)	*(0.2740)*	*(0.1219)*	*(0.4538)*	*(0.1184)*	*(0.5722)*	*(1.0000)*

Table 6: Proportion of stable matchings. In parentheses we present relative values.

Our first observation is that under GSs the stable matching μ_3, where each student is assigned to her second best school, is reached in almost 21% of the cases, corresponding to 60% of the times a stable matching is obtained (see Table 6). This is due to the fact that, instead of revealing their true preferences, many students rank their second best school first under this mechanism. By contrast, under GSm, m_1 tells the truth significantly more, resulting in her immediate assignment to her top school. This change in behavior prevents the outcome from being μ_3 and, combined with the fact that there are no spillovers on truth-telling of other students, has two implications. First, the probability of reaching a stable matching under GSm is significantly lower than under GSs (35% vs. merely 13%). Second, within the set of stable matchings, μ_4 is reached more often, guaranteeing that the probability of obtaining μ_4 and μ_5 relative to the other three stable matchings is higher in the presence than in the absence of minority reserves, so that the relevant part of Alternative Hypothesis 4 cannot be rejected.[21] In fact, the relative numbers in Table 6 allow us to go further as a closer inspection unveils a sharper result: ordering stable matchings from the least favored stable matching μ_1 to the favorite stable matching μ_5, the probability distribution over stable matchings obtained under GSm first-order stochastically dominates the one obtained under GSs. It is clearly

---

[21]Due to the very large number of recombinations, all pairwise comparisons are significant at $p = 0.01$. This is also true for the remaining analysis on fairness and the next subsection on payoffs.

the case that, conditional upon stability, introducing minority reserves in Gale-Shapley actually shifts the probability mass over matchings towards those that are more beneficial to students.

In contrast, introducing minority reserves under Top Trading Cycles increases the chances of reaching a stable matching from 5% to 12%. Still, since the probability of obtaining the matchings μ_4 and μ_5 increases less than proportionally from 4% to 7%, the chances of reaching μ_4 and μ_5 within the set of stable matchings actually decrease from roughly 93% to 57%. This ensures that, in what the Top Trading Cycles is concerned, the corresponding part of Alternative Hypothesis 4 must be rejected. As noted above, there is no increase in m_1's truth-telling rate when minority reserves are introduced. In fact, no minority or majority student ranks the best school first more often in the presence than in the absence of minority reserves. Note however that contrary to $TTCs$ (where ranking the best school first by all students is necessary to reach μ_5), under $TTCm$ several strategy profiles lead to μ_5 (i.e., not only the best-school-first profile). Majority students M_1 and M_2 actually use the associated strategies under $TTCm$ more often than the best-school-first strategy under $TTCs$. On the other hand, M_3, M_4, and m_2 use the associated strategies under $TTCm$ less often than the best-school-first strategy under $TTCs$.[22] Even though we have seen before that these differences in behavior are not significant statistically, they actually have an impact on outcomes, rendering all stable matchings more frequent except for μ_5. This happens to such an extent that the distribution over stable matchings under $TTCs$ first-order stochastically dominates the distribution over stable matchings under $TTCm$.

Result 2a. *The probability distribution over stable matchings obtained under GSm first-order stochastically dominates the one obtained under GSs. On the other hand, the probability distribution over stable matchings under TTCs first-order stochastically dominates the one obtained under TTCm.*

4.2.2 Minority-stable matchings

We now analyze the effects of the minority reserves on the probability of obtaining minority-stable matchings using the numbers in Table 7.

To start, note that even if under GSm the matching μ_4 is the most frequent stable matching, it only represents 11% of all outcomes. In fact, most importantly, introducing minority reserves in Gale-Shapley increases the frequency of the minority-stable (but not stable) matching μ_7 to 21% of all outcomes, thus becoming the most frequent minority-stable matching, reached in 60%

---

[22]Majority student M_1 ranks the best first or switches the two best schools 89% of the time under $TTCm$, but ranks the best first only 63% of the time under $TTCs$ and M_2 submits the lists that lead to μ_5 under $TTCm$ (any list) 100% of the time, while ranking the best first under $TTCs$ only 55% of the time. However, M_3 ranks her best school first 56% of the time under $TTCs$, but only 35% of the time under $TTCm$; M_4 uses the strategies that lead to the Pareto-efficient matching 40% of the time under $TTCs$, but only 21% of the time under $TTCm$; m_1 ranks the best school first 67% of the time under both mechanisms; and m_2 ranks her best school first 41% of the time under $TTCs$, but only 30% of the time under $TTCm$. Please see Appendix A for the necessary and sufficient conditions to reach the Pareto-efficient matching under each mechanism.

of the minority-stable outcomes. All of this ensures that, in what Gale-Shapley is concerned, even if the proportion of stable matchings is sharply reduced from 35% to 13% of all outcomes, the proportion of minority-stable matchings decreases only slightly from 39% to 36%. And once more, in what Gale-Shapley is concerned, the corresponding part of Alternative Hypothesis 4 cannot be rejected. Moreover, ordering minority-stable matchings according to their average payoffs, we can again go a bit further in comparing the probability distributions over minority-stable matchings obtained under the two versions of Gale-Shapley. Ordering the minority stable matchings from the lowest average payoff matching μ_1 to the highest average payoff matching μ_5, it turns out that the probability distribution over minority-stable matchings obtained under GSm "almost" first-order stochastically dominates the one obtained under GSm.[23]

Mechanism	Matchings								
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	$\mu_4 + \mu_5 + \mu_7$	Minority–stable
GSs	0.0043	0.0602	0.2062	0.0702	0.0043	0.0047	0.0410	0.1173	0.3928
	(0.0109)	(0.1533)	(0.5249)	(0.1833)	(0.0109)	(0.0120)	(0.1044)	(0.2986)	(1.0000)
GSm	0.0013	0.0180	0.0025	0.1062	0.0038	0.0098	0.2137	0.3237	0.3553
	(0.0037)	(0.0507)	(0.0070)	(0.2989)	(0.0107)	(0.0276)	(0.6015)	(0.9111)	(1.0000)
$TTCs$	0.0007	0.0000	0.0024	0.0216	0.0212	0.0058	0.0012	0.0440	0.0530
	(0.0141)	(0.0009)	(0.045)	(0.4075)	(0.4000)	(0.1094)	(0.0226)	(0.8302)	(1.0000)
$TTCm$	0.0037	0.0317	0.0141	0.0525	0.0137	0.0179	0.1069	0.1731	0.2403
	(0.0154)	(0.1319)	(0.0587)	(0.2185)	(0.0570)	(0.0745)	(0.4449)	(0.7203)	(1.0000)

Table 7: Proportion of minority-stable matchings. In parentheses we present relative values.

Introducing minority reserves under Top Trading Cycles increases the chances of reaching a minority-stable matching from 5% to 24%. On the other hand, even though the introduction of minority reserves leads to achieving μ_7 significantly more to almost 11% of all outcomes, the probability of reaching μ_4, μ_5, and μ_7 within the set of minority-stable matchings decreases from roughly 83% to 72%. This ensures that the corresponding part of Alternative Hypothesis 4 must be rejected. In fact, as it happens when we consider stable matchings, slight differences in behavior under the two versions of the Top Trading Cycles increase the frequency of every minority-stable matching except for μ_5. Finally, the distribution over minority-stable matchings under $TTCs$ first-order stochastically dominates the distribution over minority-stable matchings under $TTCm$.

---

[23]In fact, the probability of obtaining μ_1, the least favored minority-stable matching is higher under GSs than under GSm (1.1% *vs.* 0.3%), the probability of obtaining μ_1, μ_2, or μ_6 is again higher under GSs (17.6% *vs.* 8.2%), the probability of obtaining these three matchings or μ_3 or μ_7 is higher under GSs (80.6% *vs.* 69.0%), but the probability of reaching any of the mentioned matchings or μ_4 is slightly higher under GSm (98.88% under GSs *vs.* 98.93% under GSm).

Result 2b. *The probability of obtaining μ_4, μ_5, or μ_7 relative to the other four minority-stable matchings is higher in GSm than in GSs. Also, the probability distribution over minority-stable matchings under TTCs first-order stochastically dominates the one obtained under TTCm.*

4.2.3 Blocking pairs

To complete this section on fairness, we now look into the expected probability of belonging to a blocking pair, distinguishing between majority and minority students. The numbers in Table 8 reveal that in the standard mechanisms the average majority and the average minority student have roughly the same probability of being part of a blocking pair (both have 20% probability of being part of a blocking pair under GSs, while under $TTCs$ the average majority student has 36% probability and the average minority student has 38% probability of being part of a blocking pair). This is to be expected because GSs and $TTCs$ treat all students equally.

Mechanism	Students							
	M_1	M_2	M_3	M_4	m_1	m_2	Majority	Minority
GSs	0.1938	0.1494	0.3878	0.0543	0.2301	0.1988	0.1963	0.2024
	(0.1106)	(0.0884)	(0.3544)	(0.0472)			(0.1502)	
GSm	0.5882	0.4245	0.2392	0.3063	0.0960	0.0328	0.3896	0.0644
	(0.1972)	(0.1357)	(0.0941)	(0.0197)			(0.1117)	
$TTCs$	0.3471	0.4674	0.3716	0.2721	0.2695	0.4913	0.3646	0.3804
	(0.2926)	(0.2962)	(0.3006)	(0.1004)			(0.2474)	
$TTCm$	0.4859	0.6090	0.3799	0.1565	0.2162	0.0852	0.4078	0.1507
	(0.2813)	(0.3419)	(0.1931)	(0.0507)			(0.2168)	

Table 8: Average probability of belonging to a blocking pair. In parentheses we present the average probability with which majority students belong to a valid blocking pair according to the minority-stable matching concept.

This symmetry breaks down once affirmative action policies through minority reserves are implemented as the average probability with which a minority student belongs to a blocking pair is significantly reduced, while the reverse holds for majority students. It follows that the frequency with which a minority student is part of a blocking pair is significantly lower than that of a majority student under GSm and $TTCm$ (6% *vs.* 39% under GSm and 15% *vs.* 41% under $TTCm$, respectively). The same result holds even if we exclude those blocking pairs containing majority students that are ruled out according to the minority-stable matching concept (6% *vs.* 11% under GSm and 15% *vs.* 22% under $TTCm$, respectively).

As a final remark, the numbers in Table 8 suggest that GSm protects minority students better than $TTCm$ since the average probability of being part of a blocking pair is significantly lower under GSm than under $TTCm$ (6% *vs.* 15%). Moreover, it seems that it is this effect on minority

students that is mainly driving the fact that GSm reaches stable outcomes more often than $TTCm$.

Result 2c. *Under both GSm and $TTCm$, the average probability of belonging to a blocking pair is significantly lower for minority students than for majority students. The probability with which the average minority student belongs to a blocking pair is significantly lower under GSm than under $TTCm$.*

4.3 Efficiency

Next, we turn our attention to the expected payoff. Any successful discriminatory policy should increase the payoffs (or at least the average payoff) of the minority students when minority reserves are introduced. In our particular setup, we could additionally expect majority students not to be worse off with the introduction of minority reserves (Alternative Hypothesis 5).

Mechanism	Overall	Students						
		M_1	M_2	M_3	M_4	m_1	m_2	Minority
GSs	12.06	12.33	12.02	11.98	12.07	12.06	11.92	11.99
	(3.38)	(4.59)	(1.99)	(5.55)	(1.58)	(4.25)	(2.20)	(6.65)
GSm	12.27	11.97	11.67	12.27	11.26	14.34	12.12	13.23
	(5.05)	(8.09)	(5.05)	(3.93)	(2.94)	(3.19)	(1.31)	(3.20)
$TTCs$	12.70	13.44	12.24	13.10	12.01	13.52	11.99	12.71
	(5.64)	(4.34)	(6.11)	(5.19)	(6.13)	(4.13)	(5.23)	(9.61)
$TTCm$	12.25	12.29	11.33	12.16	11.80	13.44	12.47	12.96
	(5.26)	(6.88)	(5.18)	(5.80)	(2.48)	(5.69)	(2.99)	(8.27)

**Table 9:** Expected payoffs. In parentheses we present variances.

The numbers in the first column in Table 9 indicate that, on average, expected payoffs are lower under $TTCm$ than under $TTCs$ and, even though they are slightly higher under GSm than under GSs, the corresponding variances indicate that there are instances in which GSm delivers considerably lower expected payoffs than GSs. The rejection of Alternative Hypothesis 5 is due to the change in payoffs of majority students. In fact, a closer inspection of the above table reveals that all majority students (except M_3 when we compare GSm with GSs) obtain lower expected payoffs once minority reserves are introduced. On the contrary, when we focus on minority students, they appear to benefit from the introduction of minority reserves. In fact, expected individual payoffs are higher under the mechanisms with minority reserves than under the corresponding standard mechanisms in most of the instances. There is one exception: minority student m_1 obtains a slightly lower payoff under $TTCm$ than under $TTCs$. Despite this fact, when we consider minority students as a group (i.e., focus on the joint distribution of their payoffs), we clearly see that expected rewards increase and risk is reduced when minority reserves are added to each of the two standard mechanisms.
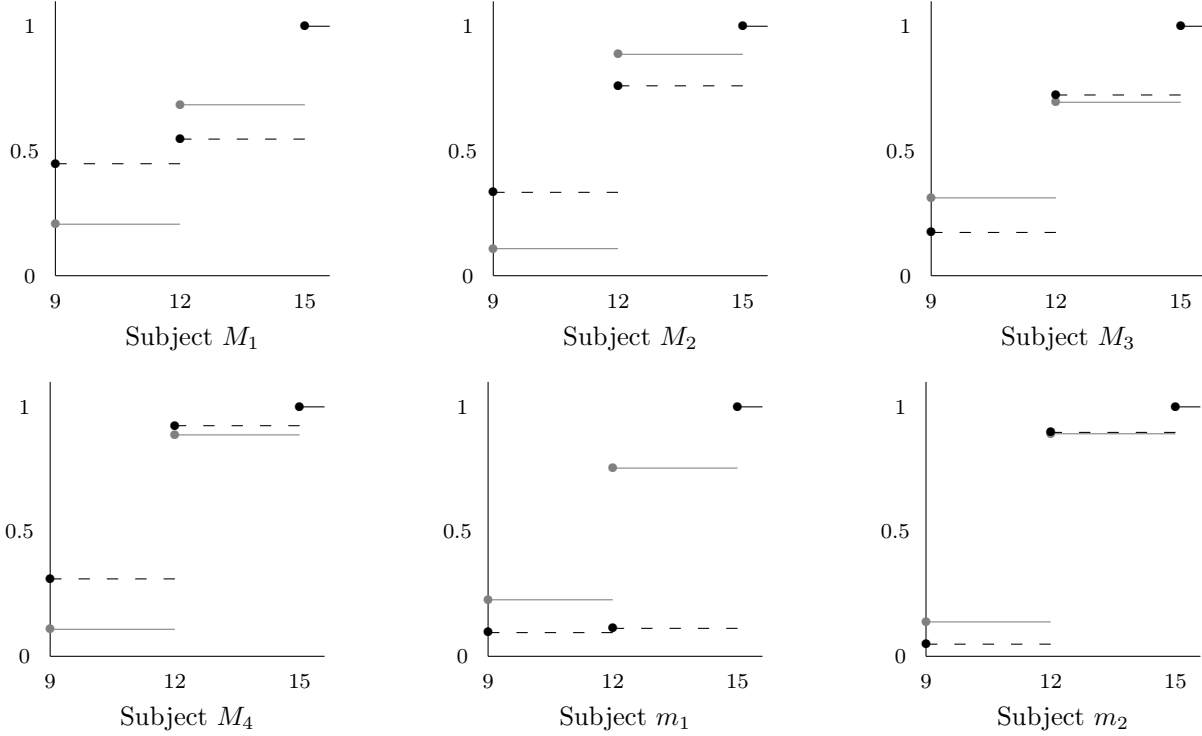
Figure 1: Cumulative distributions of payoffs in treatment GSs (solid gray) and GSm (dashed black).

Even though, on average, minority students benefit from the introduction of minority reserves, there is one fundamental difference between the two standard mechanisms. In fact, the comparison of the full distributions of individual payoffs reveals that while introducing minority reserves in GSs benefits one minority student in a very clear-cut way, without harming the other minority student, this does not happen under $TTCs$. In fact, in Figure 1 we can see that the distribution of minority student m_1's payoffs under GSm has first-order stochastic dominance over his payoff distribution under GSs. The same conclusion cannot be drawn for TTC. As Figure 2 shows, there is no first-order stochastic dominance of $TTCm$ over $TTCs$ with respect to any individual student. Instead, the distribution of payoffs under $TTCs$ dominates the distribution under $TTCm$ for M_1, M_2, and M_3. We summarize these findings in Result 3.

Result 3. *Minority reserves harm majority students, but benefit minority students inasmuch as the average payoff of this group increases. Moreover, the payoff distribution of m_1 under GSm first-order stochastically dominates her payoff distribution under GSs, whereas the payoff distribution under $TTCs$ first-order stochastically dominates the distribution under $TTCm$ for several majority students.*
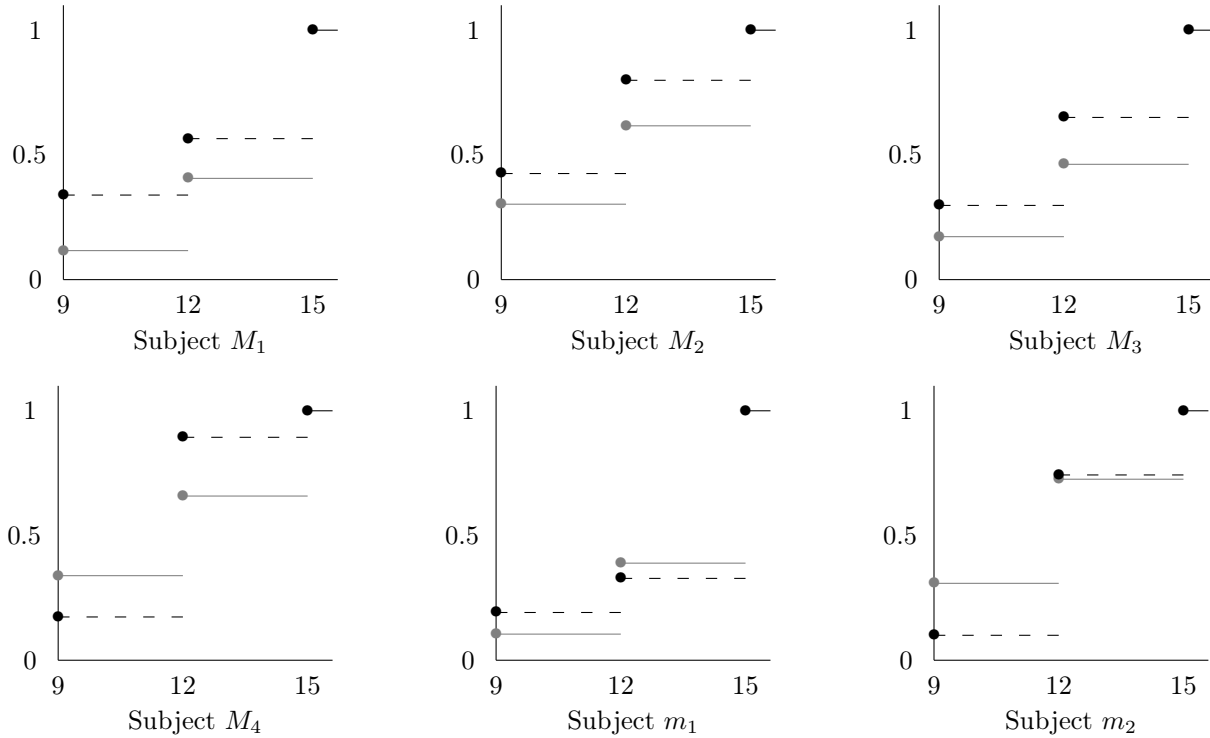
Figure 2: Cumulative distributions of payoffs in treatment $TTCs$ (solid gray) and $TTCm$ (dashed black).

5 Conclusion

We have analyzed in this experimental paper the effects of minority reserves on the Gale-Shapley and the Top Trading Cycles mechanisms in a school choice problem. Our main experimental finding highlights that adding minority reserves increases the level of truth-telling of some of the minority students when the Gale-Shapley mechanism is employed but not when the Top Trading Cycles mechanism is used. This result deserves some more detailed discussion.

Strategy-proofness of the four mechanisms implies that it is in the best interest of all students to reveal their preferences truthfully. Since we find that subjects tell the truth in the standard setting without minority reserves in about 39% of the cases under the Gale-Shapley and with 48% under the Top Trading Cycles mechanism, it is clear that many subjects are not aware of the strategy-proofness property. The corresponding percentages for the mechanisms with minority reserves are 39% for the Gale-Shapley and 34% for the Top Trading Cycles mechanism. This suggests that adding minority reserves increases the difficulty to understand the induced game when the Top Trading Cycles mechanism is employed (in fact, truth-telling decreases for all six types of students under this mechanism). The picture for the Gale-Shapley mechanism is more positive. In particular, we find that the minority student m_1 who can ensure herself a seat at her most preferred school independently of the behavior of the others increases her level of truth-telling from 44% for the mechanism without minority reserves to 76% for the mechanism with minority reserves. Yet, the affirmative action policy does not affect the level of truth-telling of the other

22

students.

The discussion above suggests that in our particular school choice problem, the Gale-Shapley mechanism performs better than the Top Trading Cycles mechanism when minority reserves are present. This intuition is confirmed when one looks at different measures of fairness and the level of welfare. While there is almost no difference in the overall efficiency between the two mechanisms with minority reserves, the Gale-Shapley mechanism tends to be more stable and more minority-stable than the Top Trading Cycles mechanism independently of whether minority reserves are employed or not. And when we restrict attention to the obtained (minority-)stable matchings, the best (minority-)stable matchings are reached more often under the Gale-Shapley mechanism than under the Top Trading Cycles mechanism. Moreover, the Gale-Shapley with minority reserves benefits minority students inasmuch as their expected payoff as a group is the highest under this mechanism and the expected probability of being part of a blocking pair is the lowest.

Nevertheless, two facts should be taken into account when evaluating the introduction of minority reserves. First, even if the Gale-Shapley mechanism appears to deliver better results than the Top Trading Cycles, the numbers obtained reveal that there may be a price to pay, as even though the proportion of minority-stable matchings decreases slightly, stability is actually severely affected by introducing minority reserves in Gale-Shapley: the proportion of stable matchings obtained decreases from 35% to 13% of all matchings. Second, these results were obtained using a simple experimental setup with the flavor of a coordination game that puts pressure on the level of truth-telling but where minimum reserves could actually bring a clear improvement for all students in terms of efficiency. Yet, we observed that several majority students were still harmed by this policy when the Gale-Shapley mechanism was applied.

References

[1] A Abdulkadiroğlu (2005). College admissions with affirmative action. *International Journal of Game Theory* 33(4): 535–549.

[2] A Abdulkadiroğlu and T Sönmez (2003). School choice: A mechanism design approach. *American Economic Review* 93(3): 729–747.

[3] A Brandenburger (1996). Strategic and structural uncertainty in games. In R Zeckhauser, R Keeney and J Sibenius (eds.) *Wise choices: Games, decisions, and negotiations*, Harvard Business School Press.

[4] S Braun, N Dwenger, and D Kübler (2010). Telling the truth may not pay off: An empirical study of centralized university admissions in Germany. *The B.E. Journal of Economic Analysis & Policy (Advances)* 10(1): 1–38.

[5] S Braun, N Dwenger, D Kübler, and A Westkamp (2014). Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior* 85(1): 232–251.

[6] C Calsamiglia, G Haeringer, and F Klijn (2010). Constrained school choice: An experimental study. *American Economic Review* 100(4): 1860–1874.

[7] Y Chen and O Kesten (2013). From Boston to Chinese parallel to deferred acceptance: Theory and experiments on a family of school choice mechanisms. Discussion Paper, Social Science Research Center Berlin (WZB), Research Area Markets and Politics, Research Unit Market Behavior, No. SP II 2013-205.

[8] Y Chen and T Sönmez (2006). School choice: An experimental study. *Journal of Economic Theory* 127(1): 202–231.

[9] B Doğan (2013). Responsive affirmative action in school choice. Mimeo, Department of Economics, University of Rochester.

[10] L Dubins and D Freedman (1981). Machiavelli and the Gale–Shapley algorithm. *American Mathematical Monthly* 88(7): 485–494.

[11] F Echenique and B Yenmez (2013). How to control controlled school choice. Working paper, Caltech SS.

[12] L Ehlers, I Hafalir, M Yenmez, and M Yildirim (2014). School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory* 153(1): 648–683.

[13] A Erdil and T Kumano (2013). Prioritizing diversity in school choice. Working paper, Washington University.

[14] C Featherstone and M Niederle (2013). Improving on strategy-proof school choice mechanisms: An experimental investigation. Working Paper, Stanford University.

[15] U Fischbacher (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2): 171–178.

[16] D Fragiadakis, A Iwasaki, P Troyan, S Ueda, and M Yokoo (2013). Strategyproof matching with minimum quotas. Working paper, Stanford University.

[17] D Fragiadakis and P Troyan (2014). Market design with distributional constraints: School choice and other applications. Working paper, Stanford University.

[18] B Greiner (2004). The online recruitment system ORSEE 2.0 - A guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne.

[19] P Guillen and Alexander Hing (2014). Lying through their teeth: Third party advice and truth telling in a strategy proof mechanism. *European Economic Review* 70(1): 178–185.

[20] I Hafalir, M Yenmez, and M Yildirim (2013). Effective affirmative action in school choice. *Theoretical Economics* 8(2): 325–362.

[21] F Heinemann, R Nagel, and P Ockenfels (2009). Measuring strategic uncertainty in coordination games. *Review of Economic Studies* 76(1): 181–221.

[22] Y Kamada and F Kojima (2013). Efficient matching under distributional concerns: Theory and applications. Working paper, Stanford University.

[23] T Kawagoe, T Matsubae, and H Takizawa (2013). Affirmative action in school choice problem: An experiment. Working paper, Future University Hakodate, Hokkaido.

[24] F Klijn, J Pais, and M Vorsatz (2013). Preference intensities and risk aversion in school choice: A laboratory experiment. *Experimental Economics* 16(1): 1–22.

[25] F Kojima (2012). School Choice: Impossibilities for Affirmative Action. *Games and Economic Behavior* 75(2): 685–693.

[26] S Kominers and T Sönmez (2013). Designing for diversity: Matching with slot-specific priorities. Working paper, University of Chicago.

[27] J Pais and Á Pintér (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior* 64(1): 303–328.

[28] A Roth (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research* 7(4): 617–628.

[29] A Westkamp (2013). An analysis of the German university admissions system. *Economic Theory* 53(3): 561–589.

Appendix A

For each of the four mechanisms we provide a necessary and sufficient condition on the submitted rankings of the students to obtain the Pareto-efficient outcome μ_5.

Proposition 1. *Under GSs, the Pareto-efficient outcome μ_5 is reached if and only if all students rank their best school first.*

Proof. The "if"-part is immediate and hence we only prove the "only if"-part. Consider student M_1 and assume she ranks school s_3 first. Since M_1 is s_3's highest priority student, she is tentatively accepted and ends up matched to s_3, so that μ_5 is not reached. Now assume M_1 ranks s_2 first. Then, either M_1 is tentatively accepted by and ends up assigned to s_2 (so that s_2 never ends up with m_1 and m_2) or M_1 is rejected, in which case s_2 tentatively accepts two other students with higher priority than M_1, and s_2 does not end up admitting m_1 and m_2. In either case, μ_5 is not reached. It follows that, in order to reach μ_5, student M_1 has to rank her best school first. The same reasoning applies to M_3 and m_1, who face exactly the same decision problem.

Now consider student M_2 and assume she ranks school s_3 first. Since M_2 is s_3's second highest priority student (and s_3 has two seats to fill), she is tentatively accepted and ends matched to this school, so that μ_5 is not reached. Now assume M_2 ranks school s_2 first. Then, either M_2 is tentatively accepted by and ends up assigned to s_2 (in which case s_2 never ends up with m_1 and m_2, who have lower priority than M_2) or M_2 gets rejected. If M_2 gets rejected, s_2 tentatively accepts two other students with higher priority and, again, s_2 does not end up admitting m_1 and m_2. In both cases, μ_5 is not reached. It follows that, in order to reach μ_5, student M_2 has to rank her best school first. The same reasoning applies to M_4 and m_2, who face exactly the same decision problem. $\qquad\square$

Proposition 2. *Under TTCs, the Pareto-efficient outcome μ_5 is reached if and only if all students rank their best school first.*

Proof. The "if"-part is immediate and hence we only prove the "only if"-part. As it happens in the standard Gale-Shapley mechanism, in the standard Top Trading Cycles mechanism, students M_1, M_3, and m_1 face the same decision problem, while M_2, M_4, and m_2 also face the same problem (but different from that of students M_1, M_3, and m_1).

Draw the graph with each school pointing at the highest priority student, i.e., s_1 pointing at m_1, s_2 pointing at M_3, and s_3 pointing at M_1. If M_1 does not rank s_1 first, she can rank s_3 first, in which case she will be assigned to s_3 (and we do not reach μ_5) or she can rank s_2 first. In the latter case, M_3 cannot rank s_3 first (otherwise we have a cycle, M_1 gets s_2, and we do not reach μ_5), nor s_2 (or M_3 is assigned to s_2), so that she can only point at s_1. As a consequence, if m_1 points at s_1, she is assigned to s_1 (and again we do not reach μ_5), if m_1 points at s_2, we have a cycle and M_3 is assigned to s_1 (and again we do not reach μ_5), if she points at s_3, we have a cycle and M_1, M_3, and m_1 are assigned to their second best schools (and we do not reach μ_5). As a consequence, M_1 ranks her best school first. The same reasoning holds for M_3 and m_1 who face the same decision problem. It follows that, in the first round of the algorithm, one cycle forms where s_1 points at m_1, m_1 points at s_2, s_2 points at M_3, M_3 points at s_3, s_3 points at M_1, and finally M_1 points at s_1. So, m_1 is assigned to s_2, M_3 is assigned to s_3, and finally M_1 is admitted

at s_1. This holds independently of the lists submitted by students M_2, M_4, and m_2, who remain in the market for the second round of the algorithm.

In the second round, the students in the market are M_2, M_4, and m_2 and a similar reasoning applies. We now have s_1 pointing at m_2, s_2 pointing at M_4, and s_3 pointing at M_2. Now, if M_2 ranked s_3 first, she is now pointing at s_3 (which has one seat to fill) and she will be assigned to s_3 (and we do not reach μ_5). If M_2 ranked s_2 first, she is now pointing at this school, in which case M_4 cannot rank s_3 first (otherwise we have a cycle, M_2 gets s_2, and we do not reach μ_5), nor s_2 (or M_4 is assigned to s_2), so that she can only point at s_1. It follows that, if m_2 points at s_1, she is assigned to s_1 (and again we do not reach μ_5), if m_2 points at s_2, we have a cycle and M_4 is assigned to s_1 (and again we do not reach μ_5), if she points at s_3, we have a cycle and M_2, M_4, and m_2 are assigned to their second best schools (and we do not reach μ_5). So, M_2 has to rank her best school first. The same holds for M_4 and m_2, who face the same decision problem. $\qquad\square$

Proposition 3. *Under GSm, the Pareto-efficient outcome μ_5 is reached if and only if all students rank their best school first.*

Proof. The "if"-part is immediate and hence we only prove the "only if"-part. Consider minority student m_1 and assume she ranks school s_1 first. In this case, being the student with highest priority for school s_1, she is tentatively held by and ultimately assigned to s_1, and we do not reach μ_5. Now assume m_1 ranks s_3 first. We can have m_1 being tentatively accepted by and finally matched to s_3 (in which case s_3 does not end up with M_3 and M_4) or m_1 being rejected. In the latter case, since one of the seats is reserved for minority students, s_3 receives an offer from m_2 and another offer from either M_1 or M_2; here, again, s_3 does not end up admitting M_3 and M_4. So, independently of what the other students do, if m_1 does not rank s_2 first, matching μ_5 is not reached. Hence, m_1 does rank s_2 first.

Consider minority student m_2. Suppose m_2 ranks s_1 first. Then, she is tentatively assigned and ends up matched to s_1, since m_2 is a minority student and she is the second highest priority student for s_1 (and s_1 has two seats to fill). As a consequence, matching μ_5 is not reached. Now suppose m_2 ranks s_3 first. Then, m_2 is tentatively held by and ultimately matched to s_3, since m_2 is the highest ranked minority student for s_3. As a consequence, matching μ_5 is not reached. Hence, m_2 ranks s_2 first.

Consider majority student M_1. Suppose M_1 ranks s_3 first. Then, she is tentatively assigned and ends up matched to s_3 as she is the highest priority student for s_3 (and only one of s_3's seats is reserved for minority students), so that μ_5 is not reached. Now suppose M_1 ranks s_2 first. In this case, either M_1 is tentatively accepted by and ends up matched to s_2 (in which case s_2 never ends up with m_1 and m_2) or M_1 gets rejected. If M_1 gets rejected, s_2 accepts two other students, at least one of them with higher priority than M_1 (since only one seat of s_2 is reserved for minority students and M_1 has higher priority than both m_1 and m_2); in this case, again, s_2 does not end up admitting both m_1 and m_2, i.e., matching μ_5 is not reached. Hence, M_1 ranks s_1 first.

Consider majority student M_2 and suppose she ranks s_3 first. Then, she is tentatively assigned and ends up matched to s_3, unless this school receives proposals from both M_1 and a minority student. In either case, s_3 does not end up matched to M_3 and M_4. Now suppose M_2 ranks s_2 first. Then, M_2 is either tentatively accepted by and matched to s_2 (in which case s_2 never ends up with m_1 and m_2) or M_2 gets rejected. If M_2 gets rejected, s_2 accepts two other students, one

of whom must higher priority than M_2 (since only one seat of s_2 is reserved for minority students and M_2 has higher priority than both m_1 and m_2). So, again, s_2 does not end up admitting both m_1 and m_2, i.e., matching μ_5 is not reached. Hence, M_2 ranks s_1 first.

Consider majority student M_3 and suppose she ranks s_2 first. Then, she is tentatively assigned and ends up matched to s_2 as she is the highest priority student for s_2 (and only one of s_2's seats is reserved for minority students), so that μ_5 is not reached. Now suppose M_3 ranks s_1 first. Then, M_3 is either tentatively accepted by and ends up matched to s_1 (in which case s_1 never ends up with M_1 and M_2) or gets rejected by s_1. If M_3 gets rejected, s_1 admits two students with higher priority than M_3. In either case, we do not reach μ_5. Hence, M_3 ranks s_3 first.

Consider majority student M_4 and suppose she ranks s_2 first. Then, she is either tentatively accepted and finally matched to s_2 or s_2 receives offers from both M_3 and a minority student. In either case we do not reach μ_5. Now suppose M_4 ranks s_1 first. Then, M_4 is either tentatively accepted by and ends up matched to s_1 or is rejected by s_1 (implying that s_1 admits both m_1 and m_2). In either case, we do not reach μ_5. Hence, M_4 ranks s_3 first. $\qquad\square$

Proposition 4. *Under TTCm, the Pareto-efficient outcome μ_5 is reached if and only if*

(i) m_1, m_2, and M_3 rank their best schools first;

(ii) M_1 ranks her first school first or switches her first two schools;

(iii) M_4 ranks her first school first or [ranks her first school second and her third school first].

Proof. The "if"-part is easy to check. We now prove the "only if"-part. Draw the graph with each school pointing at the minority student with highest priority, i.e., s_1 and s_2 pointing at m_1 and s_3 pointing at m_2.

Assume m_1 ranks s_1 first. Then, a cycle forms and m_1 is assigned to s_1. Hence, and μ_5 is not reached. Now assume that m_1 ranks s_3 first and consider m_2. If m_2 ranks s_1 first, a cycle forms and μ_5 is not reached (m_2 is assigned to s_1 and m_1 is assigned to s_3). If m_2 ranks s_2 first, a cycle forms and μ_5 is again not reached (since m_1 is assigned to s_3). Finally, if m_2 ranks s_3 first, a cycle forms and μ_5 is again not reached (m_2 is assigned to s_3). It follows that m_1 ranks s_2 first and, independently of the lists submitted by the other students, in the first round of the algorithm, a cycle forms and m_1 is assigned to s_2. This is the only cycle that forms, unless m_2 ranks s_3 first, in which case we would not reach μ_5.

In the second round of the algorithm, we have s_1 and s_3 pointing at m_2, while s_2 points at M_3. Now consider m_2. We know m_2 does not rank s_3 first. Suppose m_2 ranks s_1 first. Then, a cycle forms and m_2 is assigned to s_1 in the second round of the algorithm, so that μ_5 is not reached. Hence, m_2 ranks s_2 first.

Consider M_3 and suppose she ranks s_2 first. Then, a cycle forms and she is assigned to s_2. Suppose M_3 ranks s_1 first. Then, a cycle forms where M_3 is assigned to s_1 and m_2 is assigned to s_2. In either case, μ_5 is not reached. So, M_3 ranks s_3 first and, independently of the lists submitted by M_1, M_2, and M_4, a cycle forms in the second round of the algorithm where m_2 is admitted at s_2 and M_3 is admitted at s_3.

In the third round of the algorithm, we therefore have s_1 pointing at M_4 and s_3 pointing at M_1. As a consequence, if M_1 is pointing at s_3, she is admitted at s_3 (so that μ_5 is not reached).

So, either M_1 ranks s_1 first or ranks s_2 first and s_1 second. In either case (since both seats in s_2 are taken), M_1 is now pointing at s_1.

As for M_4, if she is pointing at s_1, she is admitted at s_1 (so that μ_5 is not reached). So, either M_4 ranks s_3 first or ranks s_2 first and s_3 second if μ_5. In either case (again since both seats in s_2 are taken), she is now pointing at s_3. Hence, in this third round, a cycle forms and M_1 is admitted at s_1 and M_4 is admitted at s_3.

The only student left is M_2 and she is assigned (in the fourth round) to the only seat left, in school s_1, independently of the list of schools she submitted. $\qquad\square$

Appendix B

In this Appendix we show that the matchings μ_1 to μ_7 in Table 3 constitute the set of minority-stable matchings.

Proposition 5. *The set of minority-stable matchings is $\{\mu_1, \ldots, \mu_7\}$.*

Proof. A matching is minority-stable if it has no blocking pairs or if all of its blocking pairs are composed of a majority student M and a school s that M prefers to the one she is assigned to, such that two conditions are met: (i) s has exhausted its capacity with majority students that have higher priority than M and with at least one minority student that has lower priority than M, but still (ii) s admitted a number of minority students smaller or equal to its minimum reserve.

From this definition it immediately follows that the stable matchings μ_1 to μ_5 are minority-stable. Moreover, the definition implies that, in our setup, every minority-stable (but not stable) matching is a matching blocked only by the pair (M_1, s_2) or by (M_2, s_2) or by both. This blocking occurs because M_1 or M_2 or both justifiably envy one of the minority students m_1 or m_2, so that either m_1 or m_2 but not both (or condition (ii) of the definition would be violated) are matched to s_2.

Let us start with the case of the matchings blocked only by (M_1, s_2). Such matchings must have (i) M_1 matched to s_3, (ii) M_2 matched to s_1 or s_2, and (iii) either m_1 or m_2 matched to s_2. As for the matchings that can be blocked only by (M_2, s_2), they must have (i) M_2 matched to s_3, (ii) M_1 matched to s_1 or s_2, and (iii) either m_1 or m_2 matched to s_2. Finally, the matchings that can be blocked by both (M_1, s_2) and (M_2, s_2) have (i) M_1 and M_2 matched to s_3 and (ii) either m_1 or m_2 matched to s_2.

In Table 10 we present all matchings that fulfill these conditions. Lines 1 to 13 correspond to matchings that are candidates to be blocked by (M_1, s_2) but not by (M_2, s_2), lines 14 to 27 are candidates for blocking by (M_2, s_2) but not by (M_1, s_2), and lines 28 to 31 are candidates for blocking by the two pairs. For each candidate matching we present, whenever possible, a blocking pair different from (M_1, s_2) and (M_2, s_2), which immediately shows that the candidate matching is not minority-stable.

		Matching						Blocked by
		M_1	M_2	M_3	M_4	m_1	m_2	
1.		s_3	s_1	s_1	s_2	s_2	s_3	(M_1, s_1)
2.		s_3	s_1	s_2	s_1	s_2	s_3	(M_1, s_1)
3.		s_3	s_1	s_2	s_3	s_2	s_1	(M_1, s_1)
4.	$[\mu_7]$	s_3	s_2	s_1	s_1	s_2	s_3	—
5.		s_3	s_2	s_1	s_3	s_2	s_1	(m_2, s_3)
6.		s_3	s_2	s_3	s_1	s_2	s_1	(m_2, s_3)
7.		s_3	s_1	s_1	s_2	s_3	s_2	(M_1, s_1)
8.		s_3	s_1	s_2	s_1	s_3	s_2	(M_1, s_1)
9.		s_3	s_1	s_2	s_3	s_1	s_2	(M_1, s_1)
10.		s_3	s_1	s_3	s_2	s_1	s_2	(M_1, s_1)
11.		s_3	s_2	s_1	s_1	s_3	s_2	(m_1, s_2)
12.		s_3	s_2	s_3	s_1	s_1	s_2	(m_1, s_2)
13.		s_3	s_2	s_1	s_3	s_1	s_2	(m_1, s_2)
14.		s_1	s_3	s_1	s_2	s_2	s_3	(M_4, s_1)
15.		s_1	s_3	s_2	s_1	s_2	s_3	(M_3, s_1)
16.		s_1	s_3	s_2	s_3	s_2	s_1	(M_3, s_3)
17.		s_1	s_3	s_3	s_2	s_2	s_1	(M_4, s_1)
18.		s_2	s_3	s_1	s_1	s_2	s_3	—
19.		s_2	s_3	s_1	s_3	s_2	s_1	—
20.		s_2	s_3	s_3	s_1	s_2	s_1	—
21.		s_1	s_3	s_1	s_2	s_3	s_2	(M_4, s_1)
22.		s_1	s_3	s_2	s_1	s_3	s_2	(M_3, s_1)
23.		s_1	s_3	s_2	s_3	s_1	s_2	(M_3, s_1)
24.		s_1	s_3	s_3	s_2	s_1	s_2	(M_4, s_1)
25.		s_2	s_3	s_1	s_1	s_3	s_2	—
26.		s_2	s_3	s_1	s_3	s_1	s_2	—
27.		s_2	s_3	s_3	s_1	s_1	s_2	—
28.	$[\mu_6]$	s_3	s_3	s_2	s_1	s_2	s_1	—
29.		s_3	s_3	s_1	s_2	s_2	s_1	(M_4, s_1)
30.		s_3	s_3	s_1	s_2	s_1	s_2	(m_1, s_2)
31.		s_3	s_3	s_2	s_1	s_1	s_2	(m_1, s_2)

Table 10: Candidate matchings to be blocked by (M_1, s_2), (M_2, s_2), or both.

Matchings in lines 18 to 20 and 25 to 27 are only blocked by (M_2, s_2). Nevertheless, in all these matchings M_2 justifiably envies the majority student M_1, so that none of these matchings is minority-stable. Matching μ_6 is only blocked by both (M_1, s_2) and (M_2, s_2) because M_1 and M_2 justifiably envy m_1, whereas μ_7 is only blocked by (M_1, s_2) because M_1 justifiably envies m_1. It follows that only these two candidate matchings emerge as minority-stable. □

Instructions (translated from Spanish)

Welcome

Thank you for participating in the experiment. The objective of this session is to study how individuals make decisions in a particular situation. The session is going to last about 2 hours. In addition to the 3 Euros show up fee you receive for your participation, you can earn additional money depending on the decisions made during the experiment. In order to ensure that the experiment takes place in an optimal environment, we ask you to respect the following rules:

1. Do not speak with other participants.

2. Turn off your mobile phone.

3. If you have a question, raise your hand.

   If you do not follow these rules, it is impossible for us to make use of the data, and we have to exclude you from the session. In that case, you will not receive any compensation. During the experiment, payoffs are expressed in ECU (experimental currency units). You will receive 1 Euro for each ECU gained during the experiment. Since your final payoff depends on your decisions, it is of utmost importance that you read the instructions very carefully. If you are not sure to fully understand the functioning of the experiment at any point in time, please, do not hesitate to raise your hand and ask.

Procedures

In this session, you are going to make decisions in an economic environment. In this environment, there are a total of six different roles. At the beginning of the experiment, the central computer divides all participants into groups of six. Within each group of six participants, each participant is assigned to TWO of the six roles. One of the assigned roles is the TRUE role, the other is FICTITIOUS. You will only learn your true role at the end of the experiment. Hence, you will have to make decisions in both roles you get assigned.

We are going to consider two variations of the basic environment. Thus, you will have to make a total of four decisions (2 roles × 2 variations) in this experiment. At the end of the experiment, the central computer randomly selects one of the two variations. The outcome of the randomly selected variation (together with the true roles of all group participants) is going to determine your final payoff.

It is important to note that the group assignment does not change during the experiment. Neither you nor the other participants in your group know or are going to learn the identity of the group participants. Moreover, your are not going to receive any information regarding the behavior of the other participants or the possible monetary implications of your own decisions until the end of the experiment.

The decision environment

The basic decision environment in the experiment is as follows: There are six students (**the six roles**) —let us call them M_1, M_2, M_3, M_4, m_1, and m_2— to be assigned to a school. The students m_1 and m_2 are called minority students. The students M_1, M_2, M_3, and M_4 are called majority students. There are three schools —denoted s_1, s_2, and s_3— and every school has two available seats. Since the schools differ in their location and quality, students have different opinions of which school they want to attend. The desirability of schools in terms of location and quality is expressed in the following table:

	M_1	M_2	M_3	M_4	m_1	m_2
Most preferred school	s_1	s_1	s_3	s_3	s_2	s_2
Second most preferred school	s_2	s_2	s_1	s_1	s_3	s_3
Least preferred school	s_3	s_3	s_2	s_2	s_1	s_1

Table 11: Preferences of students over schools.

Schools have a priority ordering that is predetermined independently for each school and that, among other things, depends on the distance from the student to the school and whether a sibling already attends the school. The following table summarizes the priority ordering of each school.

	School s_1	School s_2	School s_3
Best candidate	m_1	M_3	M_1
Second best candidate	m_2	M_4	M_2
Third best candidate	M_4	M_2	m_2
Fourth best candidate	M_3	M_1	m_1
Fifth best candidate	M_1	m_1	M_3
Worst candidate	M_2	m_2	M_4

Table 12: Priority orderings of schools over students.

To decide which students are assigned to which school, each student is asked to submit a ranking over schools; that is, each student has to order the three schools (**the decision you have to make during the experiment**). Observe that students can submit whatever ranking they like, it does not have to coincide with the actual preferences. In other words, each students has to select (and hand in) one of the following six lists (each column represents a possible ranking):

	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5	Ranking 6
1st position	s_1	s_1	s_2	s_2	s_3	s_3
2nd position	s_2	s_3	s_1	s_3	s_1	s_2
3rd position	s_3	s_2	s_3	s_1	s_2	s_1

Table 13: Each student has to select one of the six rankings.

Given the submitted rankings of the six students, the final assignment of the students to the schools is determined by a procedure in the central computer. We remind you that there are **two variations** of this procedure. The first procedure is as follows. We will present the second variation once the first variation is completed.

The first variation

[The participants receive one of the following four instructions depending on the session. The names of the mechanisms are not included in the instructions.]

Mechanism 1: Gale-Shapley standard (GSs)

Step 1 Every student applies to the school she ranked first.

Step 2 Every school that receives **at least one application** acts as follows.

- It temporarily accepts the applicant with highest priority.
- It also temporarily accepts the applicant with highest priority among all remaining applicants (if any).
- The rest of the applicants are rejected (if any).

Step 3 Whenever a student is rejected by a school, she applies to the next highest ranked school.

Step 4 Every school that receives **at least one <u>new</u> application** mixes the retained and the new applications. These applications are then processed as follows:

- It temporarily accepts the applicant with highest priority.
- It also temporarily accepts the applicant with highest priority among all remaining applicants (if any).
- The rest of the applicants are rejected (if any).

Step 5 Steps 3 and 4 are repeated until all students are matched, and the assignment is completed. Each student is assigned to the school that holds her application at the end of the process.

Mechanism 2: Gale-Shapley with minority reserves (GSm)

Step 1 Every student applies to the school she ranked first.

Step 2 Every school that receives **at least one application** acts as follows.

- If the school receives one or more **applications from minority students**, then it temporarily accepts the minority applicant with highest priority.

  It also temporarily accepts the applicant with highest priority among all remaining (majority or minority) applicants (if any).

  The rest of the applicants are rejected (if any).

- If the school receives **no applications from minority students**, then it temporarily accepts the majority applicant with highest priority.

  It also temporarily accepts the applicant with highest priority among all remaining majority applicants (if any).

  The rest of the applicants are rejected (if any).

Step 3 Whenever a student is rejected by a school, she applies to the next highest ranked school.

Step 4 Every school that receives **at least one new application** mixes the retained and the new applications. These applications are then processed as follows:

- If there are one or more **applications from minority students**, then it temporarily accepts the minority applicant with highest priority.

  It also temporarily accepts the applicant with highest priority among all its remaining (majority or minority) applicants (if any).

  The rest of the applicants are rejected (if any).

- If there is **no application from minority students**, then it temporarily accepts the majority applicant with highest priority.

  It also temporarily accepts the applicant with highest priority among all its remaining majority applicants (if any).

  The rest of the applicants are rejected (if any).

Step 5 Steps 3 and 4 are repeated until all students are matched, and the assignment is completed. Each student is assigned to the school that holds her application at the end of the process.

Mechanism 3: Top Trading Cycles standard ($TTCs$)

Step 1 Each student points (with her index finger) to the school she ranked first.

Each school points (with its index finger) to the student with highest priority.

There always is some cycle of students and schools. In each cycle each element (i.e., student or school) points to the next element and the last element points to the first element.

*For **example**, it can happen that some submitted rankings are such that the cycle (Ana, Santa Fé, Jorge, Los Pinos) forms in the first step (it is possible to have cycles with more or less students and schools):*
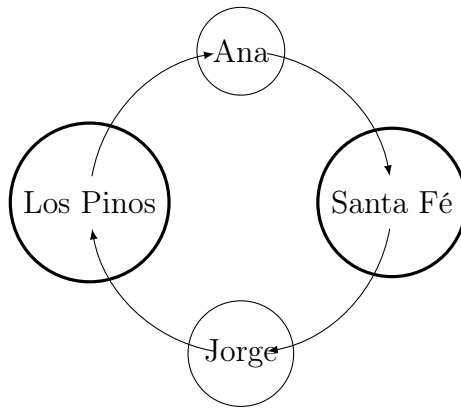
Each student in any of the cycles is matched to the school she is pointing to and that school's number of available seats is reduced by one. Each student in any cycle is sent home with a letter that confirms her obtained seat.

*In the **example**, Ana obtains a seat at "Santa Fé" and Jorge obtains a seat at "Los Pinos". The two schools forming part of the cycle reduce their number of seats by 1.*

Step 2 Each student that still has no school seat points to the school she ranked highest among all schools that still have available seats.

Each school with available seats points to the student with highest priority among all remaining students.

There always is some cycle of students and schools. In each cycle each element (i.e., student or school) points to the next element and the last element points to the first element.

Each student in any of the cycles is matched to the school she is pointing to and that school's number of available seats is reduced by one. Each student in any cycle is sent home with a letter that confirms her obtained seat.

**Step 3** Step 2 is repeated until all students are matched.

Mechanism 4: Top Trading Cycles with minority reserves ($TTCm$)

Step 1 Each <u>minority student</u> points (with her index finger) to the school she ranked first.

Each school points (with its index finger) to the <u>minority student</u> with highest priority.

There always is some cycle of minority students and schools. In each cycle each element (i.e., minority student or school) points to the next element and the last element points to the first element.

*For **example**, it can happen that some submitted rankings are such that the cycle (Pedro, La Fuente) forms in the first step (it is possible to have cycles with more students and schools):*
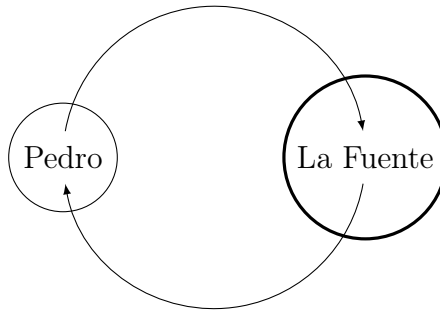
Each student in any of the cycles is matched to the school she is pointing to and that school's number of available seats is reduced by one. Each student in any cycle is sent home with a letter that confirms her obtained seat.

*In the **example**, Pedro obtains a seat at "La Fuente". The school "La Fuente" reduces its number of seats by 1.*

Step 2 All students (minority or majority) that still have no school seat point to the highest ranked school (according to the submitted ranking) that still has at least one seat available.

Each school with available seats acts as follows.

- A school that was **not matched** to a minority student in Step 1 points to the unmatched minority student (if any). In case all minority students are already matched, it points to the majority student with highest priority.

- A school that was **matched** to a minority student in Step 1 points to the (minority or majority) student with highest priority.

There always its some cycle of students and schools. In each cycle each element (i.e., student or school) points to the next element and the last element points to the first element.

*For **example**, it can happen that some submitted rankings are such that the cycle (Pablo, Cielo Azul, María, La Fuente, Juan, Dos Torres) forms in the second step (it is possible to have cycles with less students and schools):*
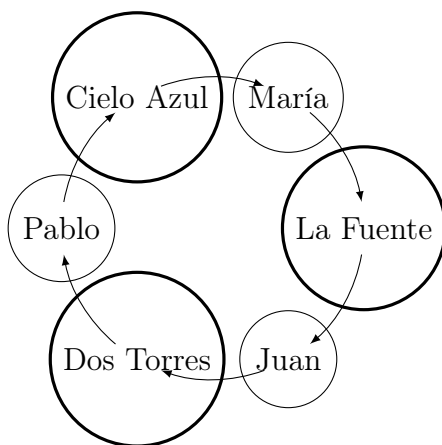
Each student in any of the cycles is matched to the school she is pointing to and that school's number of available seats is reduced by one. Each student in any cycle is sent home with a letter that confirms her obtained seat.

*In the **example**, Pablo obtains a seat at "Cielo Azul", María obtains a set at "La Fuente", and Juan obtains a seat at "Dos Torres". Each school of the cycle reduces its number of seats by 1.*

Step 3 Step 2 is repeated until all students are matched.

Procedures - continuation

You are going to play the two variations with the following monetary payoffs. You will receive 12 ECU in case you end up in your most preferred school, 9 ECU if you manage to get a seat in your second most preferred school, and 6 ECU if you study in your least preferred school.

The second variation

[Depending on the session, the subjects get the instructions for a different mechanism once the first variation is completed.]