Let's (Not) Talk about Sex:
Gender Awareness and Stereotype-Threat on Performance under Competition[*]

Nagore Iriberri[+] and Pedro Rey-Biel[**]

First version: September 2011

This version: September 2012

Abstract

We use a laboratory experiment to study how gender differences in performance under competition are affected by the provision of information regarding rivals' gender (strong manipulation) and by priming own gender (subtle manipulation). We use two tasks which differ regarding perceptions about which gender outperforms the other. We observe women's underperformance only under two conditions: 1) the task is perceived as favoring men and 2) rivals' gender is explicitly mentioned. This result is consistent with *stereotype-threat*. Blind competition (omitting information about gender) is a safe alternative to avoid women's underperformance in competition.

Keywords: gender differences, competition, gender perception, stereotype-threat.

JEL classification: C72; C91; D81.

[*] We thank Carlos J. Gómez-Ariza for his valuable guidance in choosing the tasks, Pablo López Aguilar for programming the experimental software and Kurt Schmidheiny for valuable advice. We also thank Stefano Gagliarducci and seminar audiences for insightful comments. Nagore Iriberri acknowledges financial support from Fundación Rafael del Pino, Ministerio de Educación y Ciencia (SEJ2007-64340 and ECO2009-11213) and the support of the Barcelona GSE Research Network and the Government of Catalonia. Pedro Rey-Biel acknowledges financial support from Ministerio de Educación (ECO2009-07616), Barcelona GSE Research Network and Generalitat de Catalunya.

[+] Universitat Pompeu Fabra and Barcelona Graduate School of Economics. Ramón Trías Fargas 25-27, 08005 Barcelona (Spain). Tel: (+34) 935422690. E-mail: nagore.iriberri@upf.edu.

[**]Universitat Autònoma de Barcelona and Barcelona Graduate School of Economics. Department d´Economia i d´Historia Econòmica, 08193 Bellaterra. Barcelona (Spain). Tel: (+34) 935812113. E-mail: pedro.rey@uab.es.

1. Introduction

Gender differences in labor market outcomes persist, being a continuous object of study among economists. In addition to the classical explanations based on gender differences in human capital and preferences, or statistical discrimination, recently, two seminal papers have proposed gender differences in competitiveness as a complementary explanation. Gneezy et al. (2003) showed that women underperform compared to men in competitive environments, referred to as gender differences in performance under competition, while Niederle and Vesterlund (2007) showed that women are more likely to avoid competitive environments than men, referred to as gender differences in tournament entry. Since labor markets are inherently competitive, these results would imply a gender gap in wages either because women may be less effective in performing in certain competitive environments or because they may be less likely to seek promotions. The importance of these results has led to abundant follow-up studies which have been reviewed in Croson and Gneezy (2009) and in Niederle and Vesterlund (2011). They point out the importance of understanding which institutions and policies mitigate the presence of gender differences in competitiveness.

This paper studies the effect of one potential institution, the provision of information to individuals before they participate in head-to-head competitions. Our aim is to understand which pieces of information activate "stereotype-threat" (Steele, 1997), defined as the concern arising from a situation where a person confirms a negative stereotype about their social group. Steele and Aronson (1995) and Ryan and Ryan (2005) argue that very subtle manipulations can activate stereotype-threat and affect performance. In particular, we combine tasks differing in perceptions regarding which gender outperforms the other with the provision of information regarding rivals' gender or priming own gender. Our goal is to understand under which circumstances women underperform compared to men in competitive environments and whether this underperformance can be enhanced or reduced by providing or omitting gender information which may activate or deactivate stereotype-threat.

We propose a laboratory setting to test for the effects of information. In our experiment individuals perform two real effort tasks, first under a piece-rate incentive scheme and then under a competitive scheme. Although our focus is on performance in competitive environments, we control for inherent individual differences in ability using individual performance under piece-rate.

There are three different treatments in our experiment. First, in the control treatment, participants are not provided with any information before they compete. Second, in the “Rival’s Gender” treatment, we use a strong manipulation consisting in telling participants the gender of their competing rival, before they compete. Third, in the “Own Gender” treatment, we use a subtler manipulation consisting in asking participants to fill in their gender before competing, which is used to prime their own gender.

Our experiment proposes two new cognitive tasks in which perceptions regarding which gender outperforms the other are opposed. Previous research has shown that individual perceptions about whether the task favors one gender over the other are an important determinant in the emergence of gender differences in competitiveness (Günther et al., 2010; Shurchkov, 2011; Cárdenas et al., 2012).

We elicited incentivized measures of individuals’ confidence and their perceptions regarding each of our tasks favors one gender over the other. We also obtain demographic variables as well as non incentivized measures on individuals’ general attitudes toward competition. All these variables will be used as additional controls in our analysis.

We find that women’s relative underperformance in competitive environments not only depends on the task but also on the information provided (or omitted). Consistent with previous research, women underperform compared to men in competitive environments only when the task is perceived as a male task but not when it is perceived as a female task. That is, when competing in the task that is perceived to favor men women perform 20% lower than men, but show no underperformance when competing in the task that is perceived to favor women. More importantly we show that the information or its omission is crucial and that it interacts in non-trivial ways with perceptions about the tasks. Even when competing in the task that is perceived to favor men, the omission of information regarding gender can mitigate the underperformance of women.

First, information regarding the rival’s gender is the determinant piece of information in our experiment, making women underperform compared to men. When no information is provided we do not find evidence that women underperform compared to men. Second, we find that information on rival’s gender affects women and men very differently. In the task that is perceived to favor men, it has a positive effect on men’s performance under competition, increasing their performance by almost

60%, but a negative effect on women's performance when competing, reducing their performance in about 40%. On the other hand, in the task that is perceived to favor women the effect is positive for both men and women. Third, there are no differential effects depending on whether individuals are told they are competing against a male or a female rival, such that the content of the information per se does not make a differential effect, not even when we interact rivals' gender with subjects' own gender. What is determinant is talking or (not talking) about gender but not the actual rivals' gender. Finally, when priming own gender we do not observe that women significantly underperform with respect to men even in the task that is perceived to favor men, although it has a positive effect for both male and female subjects in the task that is perceived to favor women. In fact, we find that women react even more positively than men in the task where they are expected to outperform men and their own gender is primed, showing that women perform even better than men in competitive environment (see section 3.3).

In order to explain where our results come from, we use subjects' individual characteristics, perceptions about the tasks and attitudes toward competition as additional controls in our analysis. While there are no significant differences between men and women in standard demographics in our sample, we do find that women clearly show a significantly lower competitive attitude and stronger beliefs that the tasks used in the experiment favor men. When we add these variables as additional controls, while gender differences in performance persist, they become weaker as they are partially explained by these variables.

Most previous studies have focused on policies and institutions that can change gender differences in tournament entry.[1] However, women underperformance in competitive environments deserves equal attention since it might actually explain why women avoid entering into competitive environments in the first place. Gneezy et al. (2003) manipulate the gender composition of the competing group, which is always visible to the subjects, and they find that women underperform in mixed gender groups but not in all female groups. Follow up studies also make the gender composition of the group visible to the participating subjects, and in some experiments subjects even

[1] For example, Cason et al. (2010), Wozniak (2010), and Ertac and Szentes (2010) show that information on relative performance differences reduces the gender gap in tournament entry. Dargnies (2011) show that gender differences in tournament entry also depend on whether competition is at the individual level or at the team level. Other papers directly manipulate the gender composition of the competing group, making it visible to participating subjects, which results in women being more likely to enter competitions among groups with a higher proportion of women (Sutter and Rützler, 2010; Booth and Nolen, 2009; Gupta et al., 2011; and Grosse and Riener, 2010).

compete face to face (Gneezy and Rustichini, 2004; Antonovics et al., 2009; Dreber et al. 2011). The contribution of our paper is that our treatment design allows us precisely to study how different degrees of manipulations regarding the information on gender in combination with perceptions about which gender outperforms the other within a task, can explain gender differences in competitive environments. We show that the provision of gender information explains women's undeperformance in competitive environments and that it interacts in non-trivial way with the perception of the task, activating stereotype-threat.

Our paper implies the following lessons regarding the institutions that could mitigate women's underperformance compared to men. Rival's gender should not be highlighted unless the task is perceived as a female favoring task. In other words, mentioning gender is highly detrimental for women when the task is perceived as one favoring men. As long as the settings allows for it, omitting explicit information regarding rival's gender, or ultimately, making competitions gender blind, is a safe alternative as it prevents women from underperforming in competitive environments.[2]

The paper is organized as follows. Section 2 describes the experimental design and procedures, giving detailed information regarding the tasks and individuals' perceptions about them. Section 3 contains the results. We first analyze the aggregate data in order to see when women underperform compared to men in competitive environments. We then compare each informational treatment with the control, where no information is provided, focusing on whether women and men react differently to the informational treatments. Finally, we extend the analysis using individual characteristics and perceptions as additional controls. Section 5 concludes with a discussion. The Appendix contains translations of the instructions and the post-experiment questionnaire.

2. Experimental Design and Procedures

Twelve experimental sessions were conducted in the Laboratori d'Economia Experimental (LEEX) at Universitat Pompeu Fabra using z-Tree experimental software (Fischbacher, 2007). A total of 240 subjects, 20 per session, were recruited using the ORSEE recruiting system (Greiner, 2004), ensuring that subjects had not participated in

[2] Following the Lucas critique (1976), providing or omitting information might become meaningful in itself. If information regarding rivals' gender is omitted only when the task is perceived as one that favors men, female competitors will learn information omission is bad news. Omitting this information always might be a safer recommendation.

similar experiments in our laboratory in the past.[3] Our recruiting method ensured that half of the subjects were men and half were women, without subjects noticing there was a gender aspect involved in the experiment. Upon arrival, subjects were called into the lab in random order and were seated in individual cubicles separated by screens. Subjects could observe that individuals of both genders were participating in the experiment but there was no special emphasis on the gender composition of the subject pool.

Each experimental session lasted one hour, including assignment of subjects to their seats and payment. Throughout the experiment we ensured anonymity and effective separation between subjects. They were paid individually and in private. All instructions appeared on screen and were read aloud to all subjects. Once the experiment had concluded, subjects filled in a voluntary questionnaire while they waited to be paid.

The experiment consisted of two tasks, which subjects performed in a sequence of four-minute periods each, first under piece-rate incentives and then under a pair-wise tournament. For piece-rate incentives, one of the two tasks was selected randomly for payment and subjects were paid 15 euro cents for each correct solution they gave in such task. For the pair-wise tournaments, subjects needed to be matched. Participants were ranked according to their performance in each of the tasks under piece-rate and then the top performer was matched with the second highest performer, the third with the fourth and so on until the participant ranked nineteenth is matched with the one ranked twentieth. This matching protocol was public knowledge (see instructions in the Appendix). Based on the tournament literature, Lazear and Rosen (1981), in order to study the pure effect of competition on all participants is important that the competition is similarly tight, which is ensured by our matching protocol. Also, ensuring it is public knowledge that individuals will be competing with individuals of similar ability makes the informational treatment effects harder to observe. For tournament payment, one of the two tasks was randomly chosen and the subject who performed best in each pair in such task earned 30 euro cents per correct answer, while the other subject earned

[3] One subject actually repeated the experiment using a different username but was identified and thus, the data from the second participation was eliminated from the analysis. One subject suffered small computer glitches during the experiment which prevented her from having the full time to perform one of the tasks. Finally, one subject submitted 0 correct responses the first time she undertook one of the tasks. For these subjects and tasks, data are omitted from the analysis, which explains the small sample size variation across treatments in the data analysis.

nothing.[4] Additionally, once all tasks have concluded, subjects could earn 10 euro cents per each of 16 questions rewarding predictive accuracy. Finally, subjects also earned a 3 euro participation fee. Average total payments were 13.68 euro with a large standard deviation, 6.35, due to the competitive environment.

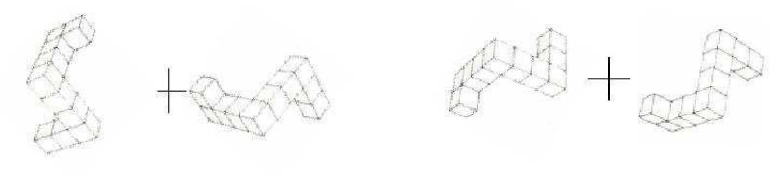
The two tasks were chosen based on the extensive literature in Psychology regarding inherent gender differences in cognitive abilities. Our objective was to find two tasks in which under piece-rate incentives each gender would perform better than the other and, at the same time, where there existed common perceptions that one gender would outperform the other. Kimura (2004) argues that consistent gender differences in abilities are hard to find and that observed differences greatly depend on the specific details of the tasks. Nevertheless, there exists a degree of consensus that men are better than women at tasks involving spatial skills, while women outperform men in tasks involving certain verbal and memory skills (Kimura, 1999). In particular, for a male favoring task we chose a mental rotation task, see Shepard and Metzler (1971) for a description of the task and Maccoby and Jacklin (1974) for a review of gender differences in this task. For a female favoring task we chose a symbol digit substitution task, see Wechsler (1958) for a description of the task and see Majeres (1983) for evidence on gender differences in this task. A non-incentivized pilot experiment conducted using paper and pencil at a different university with 184 subjects of the same age prior to our main experiment confirmed that not only men outperformed women in the mental rotation task we chose and that women outperformed men at the symbol digit substitution task, but also that subjects on average expected these results when asked which gender would on average do better at each task.[5]

We adapted both the mental rotation and symbol digit substitution tasks to our computerized setting, which facilitated the provision of information and matching protocol. The mental rotation task (MRT) in our experiment consisted of showing pairs of three-dimensional figures to subjects who had to answer whether such figures were "identical" or "mirror" figures. Identical figures are those for which, after rotating one of them, you would get the other one. Mirror figures are those for which, no matter the

[4] Since subjects did not know until the end of the experiment if, for each task, they would be paid according to their performance under piece-rate incentives or under the tournament, independent of their attitude towards risk, they always had incentives to perform the best they could.

[5] In our non-incentivized pilot, we gave subjects two minutes to perform each task. For the mental rotation task men on average solved 15.56 figures correctly and women did 12.21. For symbol digit substitution task men on average gave 27.65 correct answers while women gave 30.48. Both differences are significant at the 1% level. On average, subjects assigned highest frequency to the expectation that men outperform women in the mental rotation task (43%), while the opposite beliefs are obtained for the codification task (42% of subjects expected that women outperformed men).

number of rotations, one figure is never identical to the other one, and furthermore one is the reflection of the other figure. Figure 1 shows a pair of identical, shown in (a), and a pair of mirror images, shown in (b), from the experiment.



(a) Identical Figures (b) Mirror Figures

Figure 1. Mental Rotation Task (MRT)

The symbol digit substitution task (SDST) in our experiment consisted of showing subjects codes, which associated nine numbers to nine letters and subjects had to de-codify sequences of three letters into numbers. Codes were changed every nine three letter sequences, so that the task would involve both memory and codification abilities. Our SDST adapts the original symbol digit substitution task to the computerized setting by changing two dimensions.[6] Figure 2 shows an example of one of the codes used in the experiment as well as one three letter sequence and the corresponding correct answer.

B	H	U	T	R	W	C	A	K
1	2	3	4	5	6	7	8	9

Three letter sequence: KHR

Correct answer: 925

Figure 2. Code used in symbol digit substitution task (SDST) and a three letter sequence with its solution

Our main measure of subjects' performance is the number of correct answers subjects give for each of the tasks. We also use the number of submitted answers and

[6] Notice that when adapting this task to the computer we modified two elements. First, our codes associate numbers to letters, while in the original task codes associate numbers to symbols, and thus, subjects were asked to fill in numbers instead of symbols, since the z-tree software would only read numbers as variables. Second, sequences were presented in three letter strings instead of much longer strings commonly used. Shorter sequences allow us more precise performance measures.

the quality of submitted answers, which is calculated by the proportion of correct answers out of the submitted ones. Finally, to measure performance in competitive environments, we also look at the probability of winning each respective tournament.

The experiment had three treatments, with 40 male and 40 female subjects in each treatment. In all treatments subjects perform both tasks, MRT and SDST, under a piece-rate scheme (called Tasks 1 and 2 in the experiment) and then they repeat both tasks under a tournament scheme (called Tasks 3 and 4 in the experiment). In the "Control" treatment subjects received no information regarding their rival's gender and their gender was not primed. In the "Rival's Gender" treatment, subjects were told the gender of their rival ("Your matched participant is a boy/girl") before competing in each task. In the "Own Gender" treatment subjects were asked to fill in their own gender "for administrative purposes" before competing in each task.

Once the four tasks concluded, subjects were given an incentivized questionnaire. This included questions regarding their number of correct answers, their relative ranking and whether women or men outperformed or not the other gender, for each task. For each correct answer subjects earned 10 euro cents. Additionally, in treatments where the information contained in such questions had not been provided in the past, subjects were asked questions regarding the gender of the rival and/or whether they believed they had outperformed or not their rival under piece-rate incentives. Finally, subjects filled in a questionnaire regarding standard demographics (gender, age, nationality, mother language and studies), and questions regarding their attitude toward competition. All these variables will be used as controls when analyzing the results. See Figure 3 for the timeline of the experiment and see the Appendix for experimental instructions and for the questionnaire.

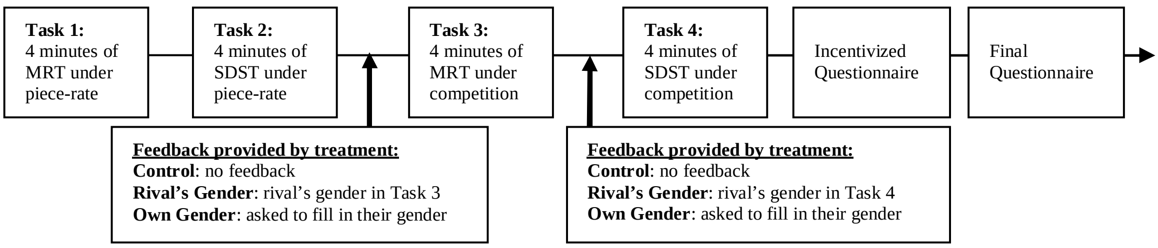


Figure 3. Timeline of the Experiment

We can now check whether our choice of tasks satisfies the inherent gender differences we had aimed for with our design. Notice that under piece-rate incentives, all four treatments are exactly the same, thus we can aggregate data in order to look for gender differences in each of the tasks. Figure 4 shows the cumulative distribution function (CDF) of the number of correct answers by gender in each of the tasks. For MRT, the performance by males statistically dominates the one by females (two-sample Kolmogorov-Smirnov test for equality of distribution functions yields a *p*-value of 0.02). However, for SDST, we cannot reject that the two cumulative distributions are equal (two-sample Kolmogorov-Smirnov test for equality of distribution functions yields a *p*-value of 0.12). This result differs from the results from our pilot, in which we observed that MRT was a male favoring task while SDST was a female favoring task. However, when we adapted SDST to our computerized setting this is no longer the case (see footnote 6). Canada and Brusca (1991) find that there is a technological gender gap favoring men when tasks are computerized, which might explain the differences we find between the paper and pencil and computerized versions of this task.

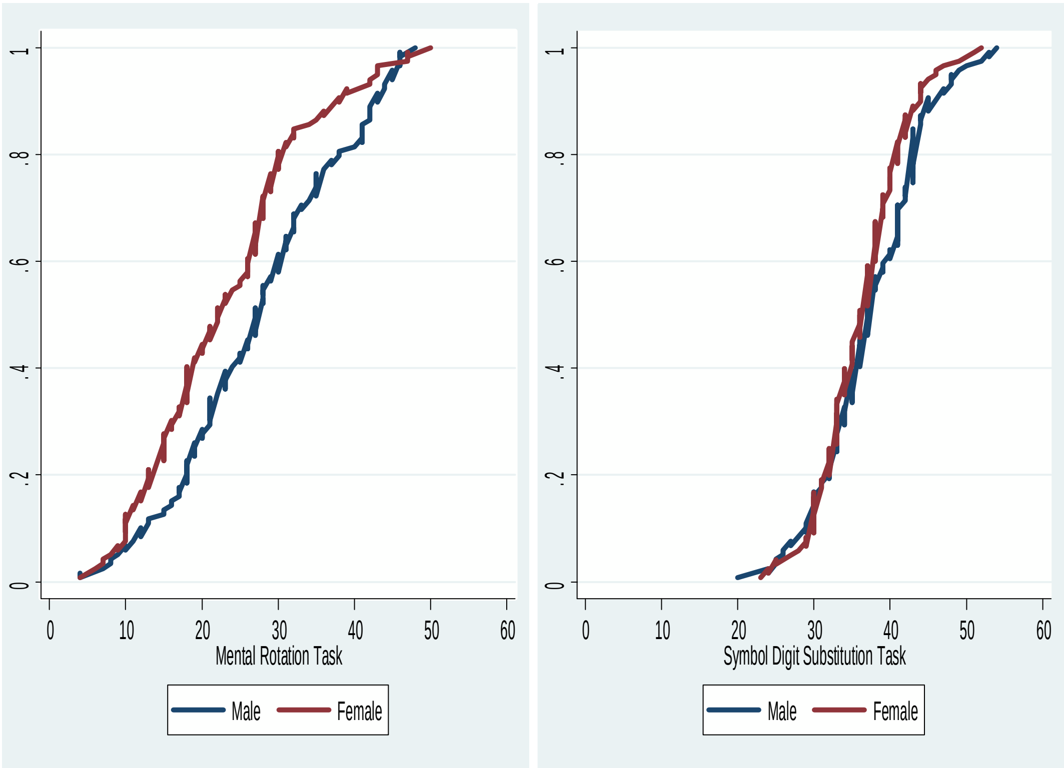


Figure 4. CDF of Number of Correct Answers in MRT and SDST by Gender under Piece-Rate

More importantly, perceptions regarding which gender is favored by each task do not change when the tasks are adapted to a computerized setting. Figure 5 uses answers from the questionnaire administered after subjects concluded the experiment to graph the average frequency assigned by all subjects to each gender outperforming the other at each task under piece-rate incentives (see last question in Screen 11 of the instructions). Clearly, on average MRT is perceived to be a male favoring task while SDST is perceived to be a female favoring task, as they were perceived in the pilot.

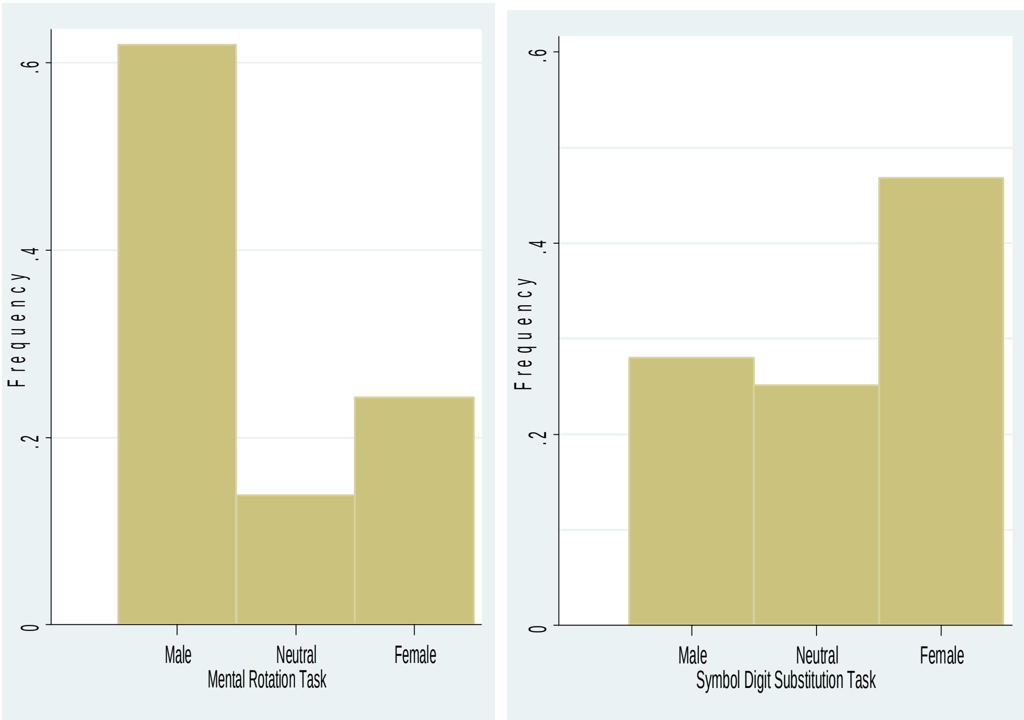


Figure 5. Histograms of Perceptions in MRT and SDST under Piece-Rate

The design of our experiment leaves us with two interesting cases. On the one hand, we have MRT, in which men not only outperform women but there exists a consensus that this is the case and, on the other, we have SDST, where perceptions regarding a female advantage are not confirmed by performance data. This will allow us in section 3.3 to further explore the role of perceptions in explaining gender differences in performance under competition.[7]

3. Results

3.1. Do Women Underperform in Competition?

[7] Our measures of gender perceptions about the tasks are obtained once subjects have performed both tasks under piece-rate and tournament schemes, such that they could be interpreted as some type of ex-post justification of their individual experiences. First, notice that subjects had monetary incentives to express their true perceptions. Second, perceptions were elicited before subjects were provided with their performance results and thus only in treatments "Performance" and "Gender and Performance" subjects could have a partial indication of whether they had an ex-ante advantage with respect to their rival in each tournament. Finally, the correlation between the number of submitted responses to each task and perceiving the task as favoring the opposite gender, although negative, it is always below 16%.

We start by exploring whether women perform worse than men in competitive environments. First, we consider raw data, looking at averages and distributions. Second, we carry out regression analysis using performance measures under piece-rate as controls for individual ability. For both analyses, first we use the aggregate data over all treatments, where we pool the control as well as the three treatment groups, and then, we look at each treatment separately.

First, we look at simple averages and cumulative distribution functions (CDFs), in order to evaluate whether there exist gender differences in performance across tasks and incentive-schemes, as shown in Figure 6. We draw the CDFs of the individual improvement from performing under piece-rate to competitive environment, that is, the difference between the number of correct answers in the competitive environment and the number of correct answers under the piece-rate incentive scheme. Thus, the way we measure changes in performance under competition uses an imperfect control for individual differences in ability. Figure 6 shows that men improve more than women in MRT, showing gender differences when performing under competitive environments, while there is no significant difference in improvement in SDST.[8]

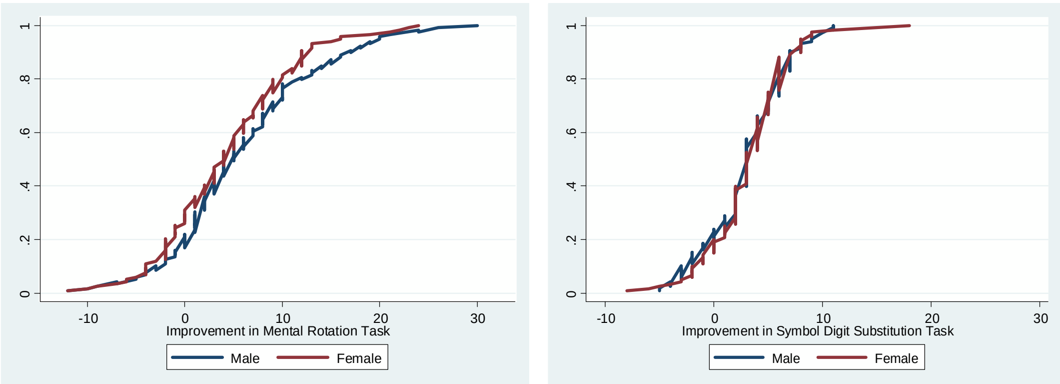
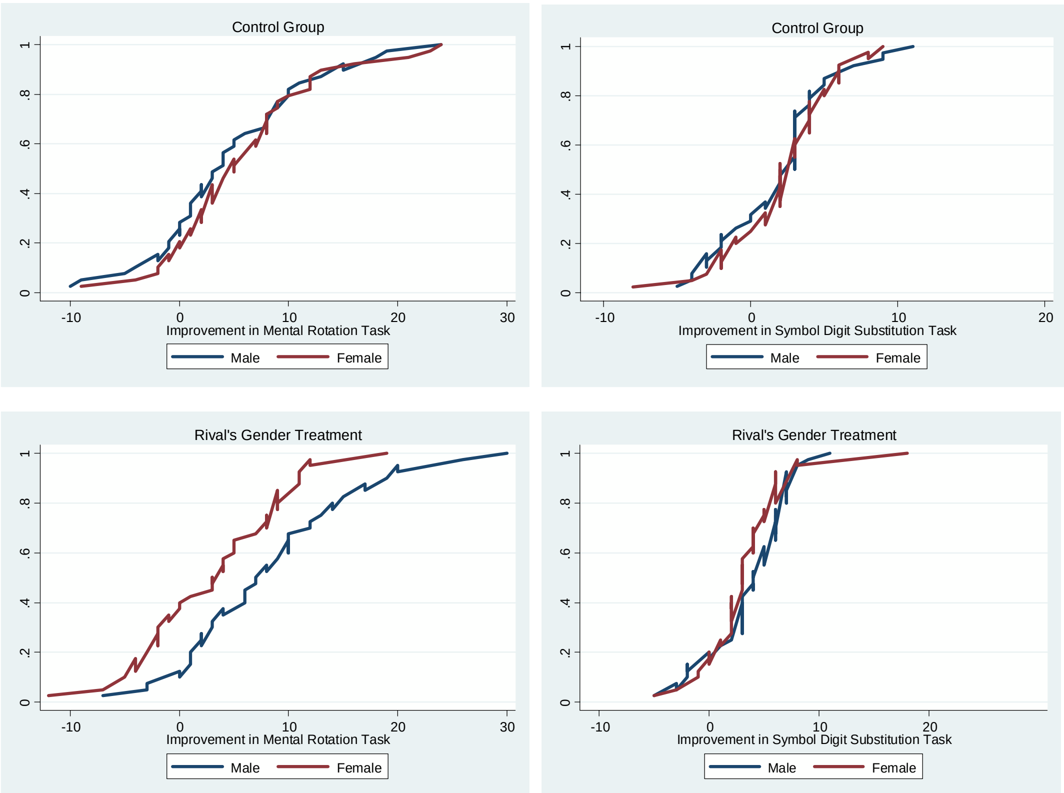


Figure 6. CDF of Improvement in the Number of Correct Answers from Piece-Rate to Competition Overall by Tasks

Moreover, when we look at improvement by gender separated by different treatments, shown in Figure 7, two results are noteworthy. First, for MRT (shown in 7(a)), the gender differences in improvement from piece-rate to competition are only observable in the treatments in which the rivals' gender is provided, that is, in "Rival's Gender". However, we do not observe gender differences in performance under

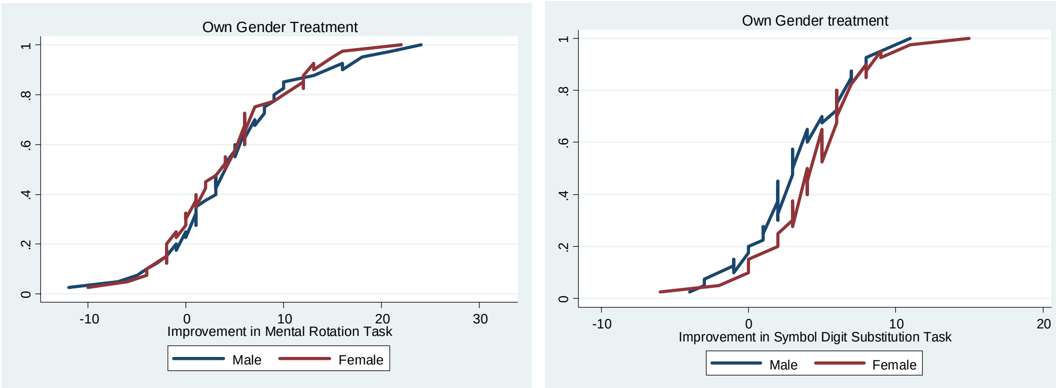
[8] One-way ANOVA test for equality of means can be rejected for MRT but not for SDST with *p*-values of 0.0959 and 0.979, respectively. The Mann-Whitney test for the null hypothesis that the median of the distributions is the same across gender cannot be rejected for MRT and SDST with *p*-values of 0.1392 and 0.7744, respectively. Finally, two-sample Kolmogorov-Smirnov test for equality of distribution functions cannot reject that male and female improvement is the same for both MRT and SDST with *p*-values of 0.581 and 0.979 respectively.

competition in the "Control" and "Own Gender" treatments.[9] Second, for SDST (shown in 7(b)), even when we disaggregate the improvement for different treatments, we do not find statistically significant evidence for any gender difference.[10] Notice in any case, that in Figure 7(b), females' CDF lies above men's in the "Own Gender" treatment, while, if anything, it lies below men's in the "Rival's Gender" treatment, possibly indicating a stronger performance enhancing effect on women than men when reinforcing own gender in the task expected to favour women. This result will be further confirmed in section 3.3, where we add additional controls.



[9] One-way ANOVA test for equality of means across gender can be rejected for "Rival's Gender" treatment but not for the "Control" and "Own Gender" treatments with *p*-values of 0.0019, 0.4642 and 0.7046, respectively. The Mann-Whitney test for the null hypothesis that the median of the distributions is the same across gender rejects that the male and female improvement is the same for the treatments "Rival's Gender" but cannot reject it for the "Control" and "Own Gender" treatments with *p*-values of 0.0048, 0.4650, and 0.6963, respectively. In a similar way, two-sample Kolmogorov-Smirnov test for equality of distribution functions rejects that male and female improvement is the same for "Rival's Gender" treatment but cannot reject for the other "Control" and "Own Gender" treatments with *p*-values of 0.097, 0.986 and 1.000, respectively.

[10] One-way ANOVA test for equality of means across gender cannot be rejected for the "Control", "Rival's Gender" and "Own Gender" treatments with *p*-values of 0.7608, 0.5049 and 0.2009, respectively. The Mann-Whitney test for the null hypothesis that the median of the distributions is the same across gender cannot reject that the male and female improvement is the same for all treatments with *p*-values of 0.6399, 0.1506 and 0.1626, respectively. In a similar way, two-sample Kolmogorov-Smirnov test for equality of distribution functions cannot reject that male and female improvement is the same for all treatments with *p*-values of 0.968, 0.573 and 0.400, respectively.



(a) Mental Rotation Task (b) Symbol Digit Substitution Task

Figure 7. CDF of Improvement in the Number of Correct Answers from Piece-Rate to Competition by Tasks and Treatments

Note that the results in Figures 6 and 7 only control for individual ability taking the difference between the performance in competition and the performance under piece-rate. Thus, we turn to regression analysis using performance under piece-rate incentives as control for individual ability, which we believe is the proper control. Table 1 presents results for the aggregate data. The first four columns, (1) to (4), refer to the MRT and the second four columns, (5) to (8), refer to the SDST. The table includes four performance measures as outcome variables: probability of winning, number of correctly submitted answers, number of submitted answers, and finally, the quality of submitted answers, defined as the number of correct divided by the number of submitted answers. As mentioned, we control for individual ability in all regressions, which is measured by the corresponding performance variables in the piece-rate environment. For the probability of winning, the relevant control for ability is the difference between the subject's number of correct answers and the rival's number of correct answers when they performed each task under piece-rate. In addition to individual ability, when we pool the control and three treatment groups, we include dummy variables for each of the treatments, where the omitted group is the control group.

Table 1
Women Performance in Competition

	MENTAL ROTATION (MRT)				SYMBOL DIGIT SUBSTITUTION TASK (SDST)			
	Prob. Of Winning	Correct	Submitted	Quality	Prob. Of Winning	Correct	Submitted	Quality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Correct_i-Correct_j)_PR	0.115***				0.151***			
	(0.0335)				(0.0445)			
Correct_PR		0.849***				0.757***		
		(0.0437)				(0.0345)		
Submitted_PR			0.666***				0.852***	
			(0.0417)				(0.0352)	
Quality_PR				0.728***				0.0349
				(0.0577)				(0.0405)
Female	**-0.341****	**-2.242****	**-2.691*****	**-3.009***	**-0.0901**	**-0.0218**	**-0.120**	**0.331**
	(0.166)	**(0.945)**	**(0.983)**	**(1.640)**	**(0.166)**	**(0.432)**	**(0.430)**	**(0.366)**
Rival's_Gender	-0.0256	0.147	-1.829	2.119	-0.107	1.234**	1.323**	0.152
	(0.204)	(1.138)	(1.185)	(2.004)	(0.204)	(0.528)	(0.526)	(0.450)
Own_Gender	0.00439	-0.842	-0.121	0.411	-0.103	1.652***	1.983***	-0.376
	(0.204)	(1.136)	(1.185)	(2.026)	(0.204)	(0.528)	(0.526)	(0.450)
Constant	0.262	10.40***	21.48***	15.77***	0.248	11.14***	7.716***	93.48***
	(0.168)	(1.538)	(1.678)	(5.194)	(0.169)	(1.370)	(1.442)	(3.925)
Observations	238	238	238	238	238	238	238	238
R-squared		0.643	0.567	0.429		0.679	0.721	0.013

Notes: The outcome variable in columns (1) and (5) is *Prob. of Winning*, which takes the value of 1 when the subject won or tied in the competition and 0 otherwise; in columns (2) and (6) is *Correct*, which measures the number of correctly solved answers under competition; in columns (3) and (7) is *Submitted*, which measures the number of submitted answers under competition; and finally in columns (4) and (8), is *Quality*, which is the ratio between *Correct* and *Submitted*. *(Correct_i-Correct_j)_PR*, measures the difference between the subject's correctly solved answers and her rival's correctly solved answers under piece rate. *Correct_PR*, *Submitted_PR* and *Quality_PR* measure the number of correctly submitted answers, correctly solved answers and the quality of answers when performing under piece-rate. *Female* is a dummy variable that takes the value of 1 when the subject is a woman and 0 otherwise, *Rival'_Gender* is a dummy variable that takes the value of 1 when the subject is informed about her rival's gender and *Own_Gender* is a dummy variable that takes the value of 1 when the subject is asked to fill in their own gender. All estimations are the result of Ordinary Least Square regressions, except for columns (1) and (5), which are estimated using Probit estimator. Standard errors in parentheses, where *** denotes significant at 1%, ** denotes significant at 5%, and * denotes significant at 10%.

Table 1 shows that women clearly underperform when competing in the MRT but not in the SDST, such that, women's underperformance is highly dependent on the task. This result is robust when looking at all four outcome variables. When competing in the MRT, women not only submit fewer answers but get fewer correct answers, the quality of their performance is significantly lower and as a consequence, they have a lower probability of winning compared to men. On average, after controlling for individual ability differences, women get about 20% fewer correct answers than men under competition. However, there is no evidence that women perform worse than men when competing in the SDST. Also and as expected, individual ability controls are highly significant for all performance measures with the exception of quality for SDST.[11] From now on, we will focus on the number of correctly submitted answers, as the main outcome variable.[12] Another interesting finding is that for SDST both informational treatments, "Rival's Gender" and "Own Gender", show higher performance than in the control group, which suggests that these treatments increase individuals' performance.

Table 2
Women Performance in Competition by Treatment

	MENTAL ROTATION (MRT)			SYMBOL DIGIT SUBSTITUTION TASK (SDST)		
	Control	Rival's Gender	Own Gender	Control	Rival's Gender	Own Gender
	Correct	Correct	Correct	Correct	Correct	Correct
	(1)	(2)	(3)	(4)	(5)	(6)
Correct_PR	0.899***	0.917***	0.758***	0.736***	0.694***	0.835***
	(0.0787)	(0.0751)	(0.0715)	(0.0566)	(0.0577)	(0.0646)
Female	**0.581**	**-5.403*** **	**-1.921**	**0.351**	**-1.194**	**0.787**
	(1.702)	**(1.639)**	**(1.531)**	**(0.729)**	**(0.717)**	**(0.793)**
Constant	7.690***	10.47***	11.74***	11.71***	15.28***	9.496***
	(2.552)	(2.205)	(2.286)	(2.154)	(2.233)	(2.495)
Observations	78	80	80	78	80	80
R-squared	0.651	0.684	0.626	0.694	0.673	0.685

Notes: The outcome variable in all regressions is *Correct*, which measures the number of correctly solved answers under competition. *Correct_PR* measures the number of correctly submitted answers when performing under piece-rate. *Female* is a dummy variable that takes the value of 1 when the subject is a woman and 0 otherwise. All estimations are the result of Ordinary Least Square regressions. Standard errors in parentheses, where *** denotes significant at 1%, ** denotes significant at 5%, and * denotes significant at 10%.

Table 2 reports estimation results separately for each of the three treatments. As before, and as columns (4) to (6) show, women do not perform worse than men when competing in the SDST. Interestingly, as columns (1) to (3) show, we can see that the underperformance of women under competition previously documented in the MRT, is now only significant in the "Rival's Gender" treatment. In the "Control" and "Own Gender" treatments, women do not perform significantly worse compared to men in competitive environments. This shows that women's underperformance is not only

[11] It turns out that individuals make very few errors in SDST. This explains that the quality measure under piece-rate has no power in explaining the quality under competition.
[12] When choosing the main outcome variable among the four possible variables, we discard quality because it is a combination of correct and submitted answers. Between correct and submitted, a principal will always be more interested in the number of correct than in the number of submitted answers. Finally, probability of winning is only relevant under competition and it also depends on the matching protocol.

dependent on the task itself, but that it is also highly dependent on the information provided.

Notice that a crucial difference between MRT and SDST is the perception about which gender is favoured by the task, such that, MRT is perceived as a task favouring men, while SDST is perceived as a task favouring women. Our results show that women underperform compared to men in competition only when both the task is perceived as favouring men *and* the information about rival's gender (strong manipulation) is provided, such that the gender information brings subjects' attention to their perceptions, reinforcing them. When own gender is primed (weak manipulation), we see that the effect is negative but not significant. In the case of SDST, the former condition is not satisfied given that the task is perceived as favouring women.

In the following sections we will further analyze these two results, the importance of the task (in relation to perceptions) and of the provided information. In section 3.2 we will study the effect of each piece of information, separating the effect of information about rival's gender and information on own gender, further exploring the importance of the content of such information. In Section 3.3 we will study the impact of other control variables, such as the role of perceptions.

3.2. Explaining Gender Differences in Performance through Informational Treatments

In this section we compare each informational treatment with the control. The goal is to learn if information and in particular its content regarding rival's gender and priming of own gender is affecting women and men differently when they perform under competition.

In each subsection, we will be looking at informational treatment effects in four different steps. We start identifying the overall treatment effect. We then proceed to see if the overall treatment effect is significantly different for women, that is, we will look for differential treatment effects based on gender. Additionally, in the case of "Rival's Gender" treatment, we further look at two more aspects. First, we control for the content of the information, such that, there might be differential treatment effects based on the content of the information. For example, when rival's gender is provided, knowing one is competing against a man or a woman might have a different effect. Second, we will look for differential treatment effects that depend both on subjects' gender and the content of the provided information. In all regressions we will control for individual ability using the performance variable under piece-rate.

3.2.1. Does Information about Rival's Gender affect Men and Women Differently?

In the gender informational treatment, "Rival's Gender", after subjects have performed under piece-rate and right before performing under competition, they were provided with information about the rival's gender, such that they were told that either they were competing against a man or a woman. In this section we will study if this information is having any impact on subjects' performance and whether it is affecting men and women differently.

Table 3
Informing about Rival's Gender

	MENTAL ROTATION (MRT)				SYMBOL DIGIT SUBSTITUTION TASK (SDST)			
	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Correct_PR	0.893***	0.909***	0.896***	0.923***	0.722***	0.716***	0.723***	0.716***
	(0.0548)	(0.0542)	(0.0568)	(0.0569)	(0.0403)	(0.0403)	(0.0406)	(0.0409)
Female	-2.485**	0.641	-2.452**	-0.799	-0.399	0.359	-0.416	-0.321
	(1.192)	(1.672)	(1.210)	(2.385)	(0.510)	(0.721)	(0.513)	(1.015)
Rival_Male			-0.325	-2.409			0.373	-0.357
			(1.733)	(2.430)			(0.728)	(1.045)
*Female*Rival_Male*				3.147				1.378
				(3.352)				(1.458)
Rival's_Gender	**0.219**	**3.273****	**-0.0146**	**2.966**	**1.213****	**1.973*****	**1.538****	**1.943***
	(1.176)	**(1.642)**	**(1.687)**	**(2.342)**	**(0.509)**	**(0.722)**	**(0.721)**	**(1.043)**
Female*Rival's_Gender		**-6.058*****		**-5.781***		**-1.506**		**-0.765**
		(2.316)		**(3.323)**		**(1.018)**		**(1.441)**
Rival's_Gender*Rival_Male			**0.467**	**1.046**			**-0.658**	**0.104**
			(2.398)	**(3.341)**			**(1.027)**	**(1.462)**
Female*Rival's_Gender*Rival_Male				**-0.815**				**-1.570**
				(4.685)				**(2.050)**
Constant	9.368***	7.409***	9.459***	8.108***	12.64***	12.47***	12.40***	12.64***
	(1.832)	(1.948)	(1.905)	(2.084)	(1.577)	(1.575)	(1.634)	(1.697)
Observations	158	158	158	158	158	158	158	158
R-squared	0.655	0.669	0.655	0.673	0.681	0.685	0.681	0.688

Notes: The outcome variable in all these regressions is the correct number of submitted answers under competition. *Correct_PR* measures the number of correctly submitted answers when performing under piece-rate. *Female* is a dummy variable that takes the value of 1 when the subject is a woman and 0 otherwise, *Rival_Male* is a dummy variable that takes the value of 1 when the subject is a man and 0 otherwise, and *Rival's_Gender* is a dummy variable that takes the value of 1 when the subject is informed about her rival's gender. All estimations are the result of Ordinary Least Square regressions. Standard errors in parentheses, where *** denotes significant at 1%, ** denotes significant at 5%, and * denotes significant at 10%.

Table 3 reports the results. Columns (1) to (4) report the results for MRT, while columns (5) to (8) report the results for SDST.

We start with the results regarding MRT. In all four regressions, the ability control is positive and highly significant, as expected. Column (1) provides the estimation results for the overall treatment effect. The treatment variable, *Rival's_Gender*, is not significant, that is, when rival's gender is provided performance is not different from when this information is absent. However, notice that *Female* is negative and significant at 5%, showing women perform significantly worse in competition than men. Now, when we look for differential treatment effects that depend on gender, shown in column (2), we clearly see that women and men are very differently affected by the information about their rival's gender. While the effect of information regarding rivals' gender is positive for men, it is highly negative for women, as the variable *Female*Rival's_Gender*, is negative and significant at 1%. Note that the variable *Female* now becomes non-significant which suggests that the underperformance of women is only significant when the information about rival's gender is provided. When no information is provided and controlling for their initial ability, a woman on average correctly solves about 8 figures while a man does 7, this difference not being significant. However, when information about rival's gender is provided and controlling for their initial ability, a woman on average correctly solves about 5 figures while a man solves 11, this difference being highly significant. In other words, the information about rival's gender increases men's performance when competing by 58%, while it decreases women's performance by almost 43%.

Columns (3) and (4) study the effect of the content of the information regarding rival's gender. We define the dummy variable, *Rival_Male*, which takes the value of 1 when the rival is a man and takes the value of 0 when the rival is a woman. Column (3) reports the estimates when we control for the content of the information. The content of the information per se is not having a differential effect, as it is insignificant, while women show to underperform in competition. Finally, in column (4) we interact the content of the information with subjects' gender. The key variable in this regression is the triple interaction, that is, *Female*Rival's_Gender*Rival_Male*, which is insignificant, meaning that the content of the gender information is not affecting male and female subjects differently. This is important as it shows that it is the provision of information about rival's gender what affects negatively to women and positively to men, but not the actual content of the information.

We now look at SDST. Overall, the treatment variable is positive and significant at 5%, as shown in column (5), which suggests that when the information about rival's gender is provided performance is increased. However, there is no differential treatment effect based on gender, as the interaction term between *Female* and *Gender_Info*, although negative as in MRT, is now insignificant. Moreover, as it can be observed in columns (7) and (8), the treatment effect is not dependent on the content of the information, neither there are differential treatment effects based on gender and the content of the information.

To summarize, the effect of gender information affects women and men differently when performing in MRT but not in SDST. In MRT women are negatively affected by gender information while men are positively affected, the results being very strong. On the other hand, in SDST, although the sign of estimated coefficients point in the same direction, the differential negative effect for women is not significant, suggesting that both women and men are positively affected by the gender information.

3.2.2. Does Priming Own Gender affect Men and Women Differently?

In the "Own Gender" informational treatment, after subjects had performed under piece-rate and right before performing under competition, they were asked to fill in their gender "for administrative purposes" (see the Appendix for the exact instructions). In this section we study whether this subtle information manipulation has an impact on subjects' performance and whether it affects men and women differently.

Table 4
Priming Own Gender

	MENTAL ROTATION TASK (MRT)		SYMBOL DIGIT SUBSTITUTION TASK (SDST)	
	Correct	Correct	Correct	Correct
	(1)	(2)	(3)	(4)
Correct_PR	0.827***	0.828***	0.782***	0.784***
	(0.0532)	(0.0532)	(0.0427)	(0.0429)
Female	-0.715	0.139	0.520	0.333
	(1.144)	(1.604)	(0.537)	(0.765)
Own_Gender	**-0.845**	**-0.00514**	**1.666*****	**1.479***
	(1.102)	**(1.560)**	**(0.536)**	**(0.766)**
Female*Own_Gender		**-1.679**		**0.370**
		(2.206)		**(1.078)**
Constant	10.20***	9.758***	9.910***	9.963***
	(1.803)	(1.898)	(1.665)	(1.676)
Observations	158	158	158	158
R-squared	0.635	0.636	0.691	0.691

Notes: The outcome variable in all these regressions is the correct number of submitted answers under competition. *Correct_PR* measures the number of correctly submitted answers when performing under piece-rate. *Female* is a dummy variable that takes the value of 1 when the subject is a woman and 0 otherwise and *Own_Gender* is a dummy variable that takes the value of 1 when the subject is asked to fill in her gender. All estimations are the result of Ordinary Least Square regressions. Standard errors in parentheses, where *** denotes significant at 1%, ** denotes significant at 5%, and * denotes significant at 10%.

Table 4 reports the results. Columns (1) and (2) report results for MRT while columns (3) and (4) report results for SDST.

We start with results regarding MRT. In the two regressions, the ability control is positive and highly significant, as expected. Also, in all regressions *Female* is insignificant, showing that women do not underperform compared to men in competition in the “Control” and “Own Gender” treatments. Column (1) provides the estimation results for the overall treatment effect, *Own_Gender*, which is insignificant. When their own gender is made salient individuals do not perform differently. Now, when we look for differential treatment effects that depend on gender, shown in column (2), we see that although the differential effect for women is higher and negative it is nonetheless insignificant. Contrary to the information on rival’s gender, we cannot conclude that priming own gender affects differently men and women.

When we look at SDST, in columns (3) and (4) consistent with the findings when rival’s gender was provided, we find that the overall treatment effect is positive and significant at the 1% level. When the task is perceived as favoring female subjects, both male and female participants increase their performance in about 16% when their gender is made salient (weak manipulation). Furthermore, the positive effect is not significantly different for male and female subjects. However, in the next section we will show that when using additional controls, the treatment effect of priming own gender in the task perceived to favor women is actually significantly stronger for women than men, making women perform even better than men in competitive environments.

To summarize, the effect of rival's gender information on performance under competition is only partially replicated when looking at information that makes own gender salient. In MRT, despite the suggestive evidence which points in the direction that women are affected negatively by the weak informational manipulation, the negative effect is not significant. In SDST on the other hand, priming own gender affects both men and women positively.

3.3. Explaining Differences in Performance with Other Controls: How do Women Differ from Men?

We have shown that when information regarding rival's gender is provided women underperform under competition in MRT but not in SDST. Given we obtained data on individual characteristics during and after the experiment, we can further explore whether differences in these variables explain the effect of information on women's underperformance.

Table 5
Control Variables Separated by Male and Female Subjects

	All Subjects			Male Subjects			Female Subjects			
Variables	Obs	**Mean**	Std. Dev.	Obs	**Mean**	Std. Dev.	Obs	**Mean**	Std. Dev.	***P*-Value**
Demographics:										
Age	239	**21.28**	3.22	119	**21.54**	3.58	120	**21.03**	2.81	**0.17**
Foreign	239	**0.06**	0.24	119	**0.08**	0.27	120	**0.05**	0.22	**0.42**
Fields of Study:										
Social_Sciences	239	**0.64**	0.48	119	**0.64**	0.48	120	**0.63**	0.48	**0.93**
Humanities	239	**0.23**	0.42	119	**0.21**	0.41	120	**0.26**	0.44	**0.38**
Applied_Sciences	239	**0.05**	0.21	119	**0.08**	0.28	120	**0.01**	0.09	**0.01**
Natural_Sciences	239	**0.04**	0.19	119	**0.01**	0.09	120	**0.07**	0.25	**0.02**
Other_Fields	239	**0.05**	0.22	119	**0.07**	0.25	120	**0.03**	0.18	**0.23**
Attitudes toward Competition:										
Experience_Competing	239	**0.36**	0.48	119	**0.54**	0.50	120	**0.18**	0.38	**0.00**
Ability_Competing	239	**4.90**	1.35	119	**5.14**	1.40	120	**4.67**	1.27	**0.00**
Enjoy_Competing	239	**4.86**	1.66	119	**5.27**	1.66	120	**4.45**	1.56	**0.00**
Gender Perception about Tasks:										
MRT: Favors_Opposite_Gender	239	**0.38**	0.49	119	**0.20**	0.40	120	**0.56**	0.50	**0.00**
SDST: Favors_Opposite_Gender	239	**0.34**	0.48	119	**0.45**	0.50	120	**0.24**	0.43	**0.00**
Confidence:										
MRT: Guessed_Rank	238	**9.85**	4.67	119	**8.51**	4.32	119	**11.19**	4.63	**0.00**
SDST: Guessed_Rank	239	**9.99**	3.93	119	**9.20**	3.90	120	**10.77**	3.82	**0.00**
MRT: Confidence_Rank	238	**0.60**	5.36	119	**0.81**	5.25	119	**0.39**	5.49	**0.40**
SDST: Confidence_Rank	239	**0.54**	5.57	119	**0.82**	5.70	120	**0.27**	5.44	**0.58**

Notes: *Foreign* is a dummy variable that takes the value of 1 when the subject is non-Spanish. There are five fields of studies. Each of them measures the proportion of subjects studying each of the fields. *Experience_Competing* is a dummy variable that takes the value of 1 if the subject revealed she has actively participated in comeptitive activities. *Ability_Competing* and *Enjoy_Competing* measure the degree of agreement in a scale between 1 (total disagreement) and 7 Itotal agreement) to subjects reveal to the following statement: "I am good at/enjoy competing". *Favors_Opposite_Gender* is a dummy variable that takes the value of 1 when the subject is male/female and reveals that he thinks the task favors females/males and 0 otherwise. *Guessed_Rank* is a variable that measures subjects' guesses about their rank (between 1 and 20, 1 representing the best rank among 20 subjects) when performing under piece-rate. *Confidence_Rank* is represents the difference between the actual rank and the guessed rank when performing under piece-rate. The final column represents the *p*-value for the Kruskal-Wallis equality-of-populations rank test with ties.

We start by looking at whether there exist differences between men and women in the control variables we obtained in the experiment. Table 5 summarizes the individual characteristics grouped in four categories for all subjects, as well as, separated by gender. The last column includes the *p*-values for the Kruskal-Wallis equality-of-populations rank test. The variables in the first category, demographics, were elicited in the ex-post questionnaire. They include subjects' age, whether they are foreign or not, and their field of study, classified in five categories.[13] We do not observe significant differences between female and male subjects, except in the proportions of

[13] Social Sciences include fields such as Economics and Business, Humanities include fields such as Law, Applied Sciences include fields such as Engineering, and finally Natural Sciences include fields such as Biology.

subjects studying Applied and Natural Sciences, which have a low frequency in the sample. These differences go in the expected directions as we have more women studying Natural Sciences (mostly Biology) and fewer women studying Applied Sciences. The second category, attitudes toward competition, also elicited in the ex-post questionnaire, includes a dummy variable indicating whether subjects regularly participate in competitive activities (*Experience_Competing*), and two variables ranging from 1 to 7 (where 1 indicates subjects' total disagreement and 7 indicates total agreement) regarding whether subjects consider they are good at competing and whether they enjoy competing. In all three variables male subjects clearly show a significantly more competitive attitude. The variables in the third category, gender perception about tasks, were elicited with monetary incentives right after subjects concluded the tasks but before they could observe any result. In particular, we define *Favors_Opposite_Gender* as the proportion of subjects of each gender who thinks each task favors the opposite gender. Two observations are worth noting. First, on average both genders perceive MRT as a male favoring task while SDST is perceived as a female favoring task, which can be clearly observed in Figure 8. Second, a higher proportion of female subjects think MRT is a male favoring task (56%) than the proportion of male subjects thinking SDST is a female favoring task (42%). The final category, confidence, includes two types of variables. *Guessed_Rank* is defined as subjects' incentivized beliefs about their rank in each of the tasks, elicited after they concluded all tasks but before they observed any result. *Confidence_Rank*, used as a control variable, is defined as the difference between subjects' actual rank in each of the tasks under piece-rate and *Guessed_Rank*. In both tasks women expect to be ranked significantly lower than men and both women and men on average believe to be ranked worse than they actually are as the confidence measures are positive. Finally, despite men showing higher average levels of confidence than women, given the high standard deviations, these differences are not significant.

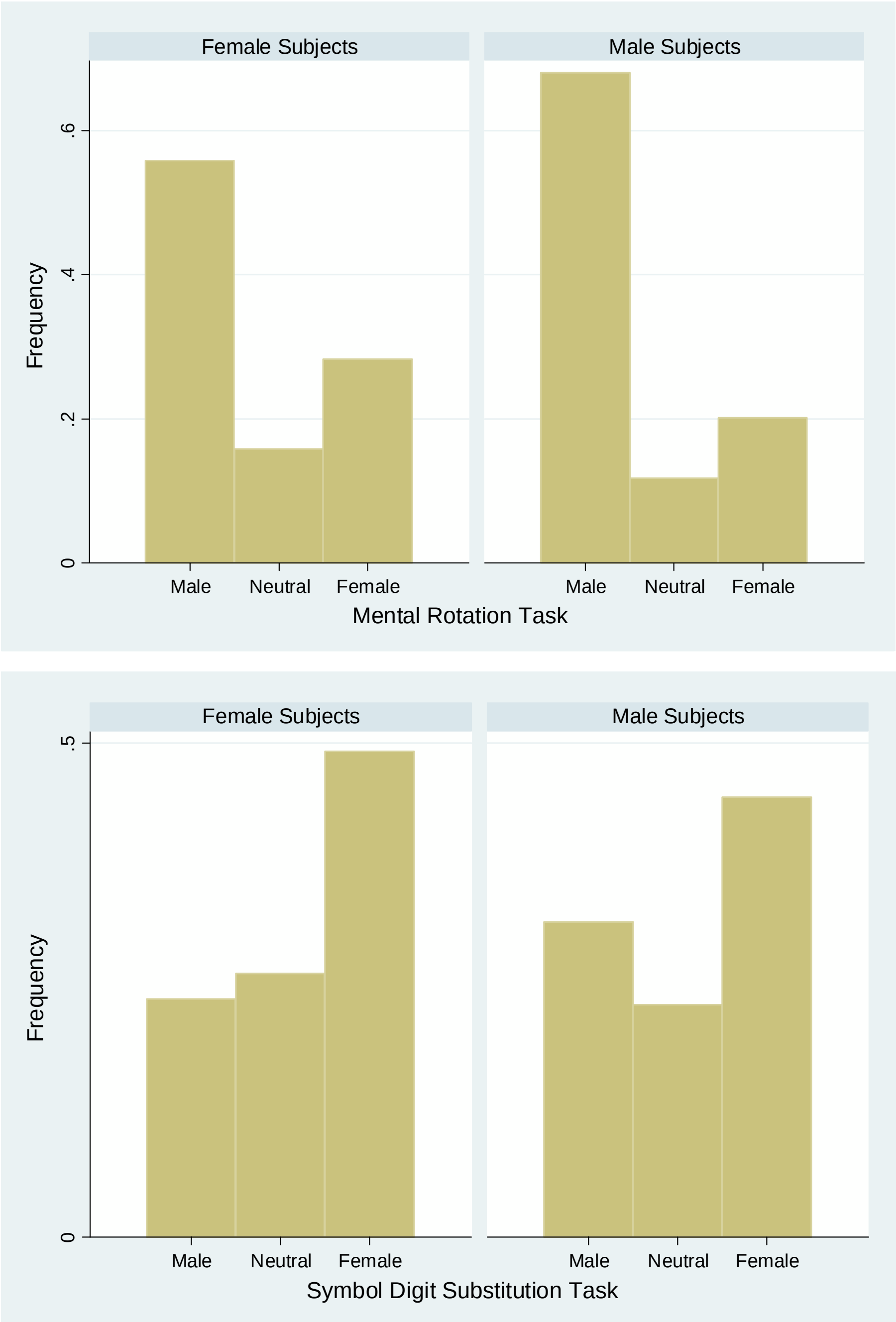


Figure 8. Histograms of Perceptions in MRT and SDST under Piece-Rate by Gender

In order to further explore gender differences under competition, we include all the presented variables as controls in our main regressions in Tables 1 and 2. Columns (1) and (5) of Table 6 replicate the regressions presented in columns (2) and (6) in Table 1 adding the additional controls. Columns (2) to (4) and columns (6) to (8) replicate the regressions in columns (1) to (3) and columns (4) to (6) in Table 2 adding the additional controls.

Table 6
Women Performance in Competition with Controls

	MENTAL ROTATION TASK (MRT)				SYMBOL DIGIT SUBSTITUTION TASK (SDST)			
	All	**Control**	**Rival's Gender**	**Own Gender**	**All**	**Control**	**Rival's Gender**	**Own Gender**
	Correct	Correct	Correct	Correct	Correct	Correct	Correct	Correct
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Correct_PR	0.890***	0.859***	0.906***	0.869***	0.788***	0.734***	0.716***	0.823***
	(0.0578)	(0.110)	(0.112)	(0.105)	(0.0492)	(0.0976)	(0.0796)	(0.0998)
Female	-1.079	0.293	-2.975*	-1.093	0.152	0.666	-1.057	2.035**
	(1.082)	(2.191)	(1.776)	(2.091)	(0.498)	(0.943)	(0.776)	(1.005)
Rival's_Gender	0.791				1.193**			
	(1.152)				(0.539)			
Own_Gender	0.328				1.484***			
	(1.171)				(0.548)			
Age	0.00302	0.174	0.0277	-0.0407	-0.0471	-0.00302	-0.219	-0.176
	(0.153)	(0.253)	(0.417)	(0.254)	(0.0714)	(0.111)	(0.182)	(0.131)
Foreign	-0.811	-2.189	-0.267	-9.232	-2.023**	-0.908	-2.099*	-0.678
	(2.033)	(4.424)	(2.702)	(7.647)	(0.940)	(1.990)	(1.180)	(3.724)
Social_Sciences	-2.373	-2.813	-3.427	-0.857	-0.0875	0.583	0.641	-0.347
	(2.071)	(4.222)	(4.157)	(3.212)	(0.962)	(1.850)	(1.839)	(1.665)
Humanities	-0.679	-1.290	-1.386	-0.966	1.088	1.325	1.908	1.524
	(2.197)	(4.177)	(4.307)	(3.663)	(1.025)	(1.844)	(1.880)	(1.860)
Applied_Sciences	-0.600	3.637	-6.167	-0.301	-0.103	2.420	-0.197	-0.735
	(2.795)	(5.536)	(6.982)	(4.277)	(1.321)	(2.452)	(2.929)	(2.223)
Natural_Sciences	-1.446	1.003	0.0196	-6.992	-1.004	0.831	0.843	-4.680
	(3.177)	(5.319)	(8.283)	(5.945)	(1.460)	(2.232)	(3.665)	(3.069)
Experience_Comp	-0.199	-1.129	0.501	0.00190	0.862*	-0.488	0.605	2.037**
	(1.093)	(2.459)	(1.805)	(1.871)	(0.515)	(1.068)	(0.786)	(0.958)
Ability_Competing	1.457***	0.957	2.903***	0.593	0.0901	0.364	0.899**	-0.0864
	(0.465)	(0.907)	(0.988)	(0.737)	(0.217)	(0.400)	(0.445)	(0.387)
Enjoy_Competing	-0.655*	-0.847	-0.670	-0.660	-0.0582	0.300	-0.837**	0.0329
	(0.380)	(0.705)	(0.797)	(0.652)	(0.182)	(0.328)	(0.354)	(0.337)
Favors_Opposite_Gender	-2.086*	-2.737	-3.158*	-0.412	-0.574	-1.249	-0.304	0.789
	(1.083)	(2.242)	(1.848)	(1.927)	(0.474)	(0.890)	(0.773)	(0.982)
Confidence_Ranks	0.258**	-0.00513	0.160	0.432**	0.0450	0.131	0.0677	-0.0501
	(0.113)	(0.221)	(0.207)	(0.195)	(0.0543)	(0.105)	(0.0849)	(0.102)
Constant	6.656	7.119	1.666	11.12	10.64***	8.162	17.82***	12.25**
	(4.794)	(8.790)	(10.09)	(7.925)	(2.800)	(4.988)	(5.138)	(5.425)
Observations	238	78	80	80	238	78	80	80
R-squared	0.679	0.692	0.771	0.676	0.705	0.739	0.753	0.739

Notes: The outcome variable in all these regressions is the correct number of submitted answers under competition. *Correct_PR* measures the number of correctly submitted answers. *Female* is a dummy variable that takes the value of 1 when the subject is a woman and 0 otherwise, *Rival's_Gender* is a dummy variable that takes the value of 1 when the subject is informed about her rival's gender and *Own_Gender* is a dummy variable that takes the value of 1 when the subject is asked to fill in her own gender. The rest of the variables are control variables that have been defined in the notes of Table 5. All estimations are the result of Ordinary Least Square regressions. Standard errors in parentheses, where *** denotes significant at 1%, ** denotes significant at 5%, and * denotes significant at 10%.

We start commenting on the results regarding MRT. In column (1) of Table 6, the dummy variable for female subjects is no longer significant while in Table 1 this variable was negative and significant at 1%. This indicates that when adding controls women do not underperform in competition compared to men. Four control variables show significant effects and explain the lack of significance of the female dummy. Subjects who believe they are good at competing (*Ability_Competing*) as well as those who are overconfident (*Confidence_Rank*) perform better, which is shown by the positive and significant coefficients. We have seen in Table 5 that overall women think they are worse at competing than men, and that they are less overconfident than men although not significantly, such that these two variables are in part explaining the underperformance of women compared to men when competing. More importantly, subjects who think that the MRT favors the opposite gender perform significantly worse, which is shown by the negative and significant coefficient of *Favors_Opposite_Gender*. In Table 5, we also observed that a higher proportion of women than men thinks that they are facing a task that favors the opposite gender, which again partly explains the underperformance of women compared to men. Finally, participants who reveal that they enjoy competing do significantly worse, which is a priori counterintuitive,[14] but this is only significant at the 10% level.

In columns (2) to (4) of Table 6 we show the regressions with additional controls separated by treatment. As in Table 2, the female dummy is significant in the treatment in which information about rivals' gender has been provided. Compared to the regressions without controls, shown in column (2) in Table 2, the female dummy shows lower significance as shown in column (2) of Table 6, which is explained by the inclusion of the control variables. This indicates that the information regarding rivals' gender is crucial to explain women's underperformance in competition, while also provides an explanation to why the female dummy became insignificant when pooling all the treatments, shown in column (1) of Table 6.

Regarding SDST, column (5) in Table 6 shows that women still do not underperform in competition when adding further controls, consistent with the results in

[14] We are not aware of empirical evidence showing a positive correlation between enjoyment and performance in competitions.

Table 1. The four control variables we mentioned above, although they have the same sign as in the regressions for MRT, are not significant. Columns (6) to (8) replicate the regressions for each treatment, as in Table 2, but they now include the control variables. As opposed to previous regressions without controls, we now observe in column (8) that the female dummy is positive and significant in the treatment where own gender is primed. Thus, we conclude that in tasks perceived to favor women, priming own gender does not result in women's underperformance under competition with respect to men but it makes women perform better than men. This final result reinforces the idea that information, combined with a priori perceptions about tasks, can have important effects on performance under competition.

4. Discussion

We find that the provision or omission of information regarding gender is crucial in understanding observed gender differences in performance under competition. In particular, gender information interacts in no trivial ways with previously held (true or false) perceptions regarding whether one gender is more skilled at the particular task studied. We find that women's underperformance under competition only occurs in tasks in which men are expected to outperform women and when women are explicitly told their rivals' gender. On the contrary, we find that women enhanced their performance as least as much as men when they believe they are better at the task than men and own gender is primed.

Our results are compatible with the provision (omission) of information activating (deactivating) stereotype-threat, which has been shown to influence women's performance in a great manner. Therefore, manipulations such as omitting or providing information can reinforce or weaken previous perceptions about gender differences at the task as well as perceptions of competitive abilities and therefore affect performance. For example, as long as the setting allows for them, blind competitions seem to be a safe alternative against gender differences in competitive environments. On the contrary, competitive environments that make any gender information salient (strong or subtle manipulations) can translate into gender differences in competitive environments. In that respect, affirmative action policies based on gender may in fact have counterproductive effects, since while creating advantageous conditions for women they also make gender information salient. Further research about how to design competitive institutions such that women's performance is not affected by stereotype-threat should follow. In particular, we should study the effect of correcting false preconceptions about women's relative lower ability at jobs traditionally considered as male, when such perceptions are in fact not true.

5. **References**

Antonovics, K., Arcidiacono, P., Walsh R. (2009), "The Effects of Gender Interactions in the Lab and in the Field," *Review of Economics and Statistics* 91(1):152-163.

Booth, A., Nolen, P. (2009), "Choosing to compete: How different are girls and boys?," IZA Working Papers 4027 Bonn, Germany.

Canada, K., Brusca, F. (1991), "The Technological Gender Gap: Evidence and Recommendations for Educators and Computer-based Instruction Designers," *Educational Technology Research and Development* 39:43-51.

Cárdenas, J. C., Dreber, A., von Essen, E., Ranehill, E. (2012), "Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden," *Journal of Economic Behavior and Organization* 83(1): 11-23.

Cason, T. N., Masters, W. A., Sheremeta, R. M. (2010), "Entry into Winner-take-all and Proportional Prize Contests: an Experimental Study," *Journal of Public Economics* 94:604–11.

Croson, R., Gneezy, U. (2009), "Gender Differences in Preferences," *Journal of Economic Literature* 47(2):1–27.

Dargnies, M.-P. (2011), "Men Too Sometimes Shy Away from Competition: The Case of Team Competition," conditionally accepted for publication at *Management Science*.

Dreber, A., von Essen, E., Ranehill, E. (2011), "Outrunning the Gender Gap – Boys and Girls Compete Equally," *Experimental Economics* 14(4): 567-582.

Ertac, S., Szentes, B. (2010), "The Effect of Performance Feedback on Gender Differences in Competitiveness: Experimental Evidence," Working Paper Koc University, Turkey.

Fischbacher, U. (2007), "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics* 10(2): 171-178.

Gneezy, U., Niederle, M., Rustichini, A. (2003), "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics* 118:1049–74.

Gneezy, U., Rustichini, A. (2004), "Gender and Competition at a Young Age," *American Economic Review* 94:377– 81.

Greiner, B. (2004), "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics," University of Cologne WP Series in Economics 10.

Grosse, N., Riener, G. (2010), “Explaining gender differences in competitiveness: gender-task stereotypes,” Working Paper Friedrich Schiller University Jena, Germany.

Günther, C., Ekinci, N., Schwieren, C., Strobel, M. (2010), “Women Can’t Jump? An Experiment on Competitive Attitudes and Stereotype Threat,” *Journal of Economic Behavior and Organization* 75:395–401.

Gupta, N. D., Poulsen, A., Villeval M. (2011), “Gender Matching and Competitiveness: Experimental Evidence,” *Economic Inquiry*.

Kimura, D. (1999), *Sex and Cognition*, Cambridge, MA: MIT Press.

Kimura, D. (2004), “Human Sex Differences in Cognition: Fact, Not Predicament,” *Sexualities, Evolution & Gender* 6:45-53.

Lazear, E. P., Rosen, S. (1981), “Rank-Order Tournaments as Optimum Labor Contracts,” *The Journal of Political Economy* 89(5):841-864.

Lucas, R. (1976), "Econometric Policy Evaluation: A Critique", in Brunner, K.; Meltzer, A., *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, 1, New York: American Elsevier, pp. 19–46, ISBN 0444110070.

Majeres, R. L. (1983), “Sex Differences in Symbol Digit Substitution and Speeded Matching,” *Intelligence* 7:313-327.

Maccoby, E., Jacklin, C.N. (1974), *The Psychology of Sex Differences* Stanford, CA: Stanford University Press.

Niederle, M., Vesterlund, L. (2007), “Do Women Shy Away from Competition? Do Men Compete too Much?,” *Quarterly Journal of Economics* 122:1067–101.

Niederle, M., Vesterlund, L. (2011), “Gender and Competition,” *Annual Review of Economics* 3.

Ryan, K. E., Ryan, A. M. (2005), “Psychological Processes Underlying Stereotype Threat and Standardized Math test Performance,” *Educational Psychologist* 40:53-63.

Shurchkov, O., “Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints,” *Journal of the European Economic Association,* in press.

Shepard, R. N., Metzler, J. (1971), “Mental Rotation of Three Dimentional Objects,” *Science* 171:701-703.

Steele, C.M. (1997), “A Threat in the Air. How Stereotypes Shape Intellectual Identity and Performance,” *American Psychologist* 52:613-629.

Steele, C.M., Aronson, J. (1995), “Stereotype Threat and the Intellectual Test Performance of African Americans,” *Journal of Personality and Social Psychology* 89:797-811.

Sutter, M., Rützler, D. (2010), "Gender Differences in Competition Emerge Early in Life," IZA Discussion Paper 5015.

Wechsler, D. (1958), "The measurement and appraisal of adult intelligence," 4th ed. Baltimore, MD: Williams & Wilkins.

Wozniak, D., Harbaugh, W. T., Mayr, U. (2010), "Choices about Competition: Differences by Gender and Hormonal Fluctuations, and the Role of Relative Performance Feedback," Working Paper University of Oregon.

6. Appendix

Experimental Instructions

Below you can find a translation of the experimental instructions (originally in Spanish) which appeared sequentially on computer screens and were read aloud by the same experimentalist in all sessions. Variations for each treatment are indicated in parenthesis.

Instructions read to all subjects.

SCREEN 1

THANK YOU FOR PARTICIPATING IN OUR EXPERIMENT!

This is an experiment and thus, no talking, looking-around or walking is allowed. If you have any question or need help please raise your hand and one of the researchers will assist you. If you do not follow the indicated rules, WE WILL ASK YOU TO LEAVE THE EXPERIMENT AND YOU WILL NOT RECEIVE ANY PAYMENT. Thank you.

Both Pompeu Fabra and Autònoma de Barcelona universities have provided funds used in this experiment. You will receive 3 euros for having arrived on time. Additionally, if you follow the instructions correctly you may earn more money.

Each participant has an "Experiment Code" determined by the number which appears on each computer terminal. As you could observe when you arrived, your number has been assigned randomly. Participants will not be able to identify each other by their decisions nor their earnings. Researchers will observe each participant's earnings at the end of the experiment but we will not associate your decisions with any participants' names.

The experiment consists of 4 tasks. Before each task, we will inform you about the type of decisions you will have to take and about how your decisions will affect your earnings. Everything you earn will be paid in cash and in a strictly private manner at the end of the experimental session.

Your final earnings will be the sum of the 3 euros you receive for participating plus whatever you earn in 2 of the 4 tasks of the experiment. The computer will randomly determine if you will be paid for task 1 or task 2 of the experiment. Similarly, the computer will randomly determine whether you will be paid for task 3 or task 4 of the experiment.

Press OK to continue with instructions.

SCREEN 2

Let us see two examples:

- If the computer determines that you will be paid for tasks 1 and 4 of the experiment, your earnings will be: 3 euros for your participation + your earnings in task 1 + your earnings in task 4.

- If, for example, the computer determines that you will be paid for tasks 2 and 4 of the experiment, your earnings will be: 3 euros for your participation + your earnings in task 2 + your earnings in task 4.

At the end of the experiment, the program will inform you about your results in each of the tasks, which tasks have been randomly chosen for realizing the payments and what your final earnings are.

Press the OK button in order to start with the instructions for Task 1 of the experiment.

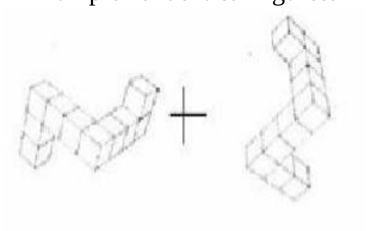
SCREEN 3

Task 1 Instructions

In task 1 of the experiment, you will see two geometric figures, one next to each other. These figures can either be “identical” or “mirror”. Your task consists of indicating, for each pair of figures, which is the case.

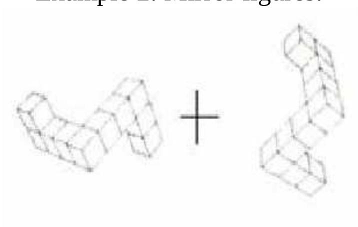
1. Identical: The two geometric figures are the same, although one of them may be rotated a certain number of degrees with respect to an axis. I.e., if we rotated one of the figures we would get the other one.

Example 1: Identical figures:



2. Mirror: The two geometric figures are different and, in fact, if we rotated one of them we would obtain a reflection of the other. I.e., if we rotated one of the figures we would never get two identical figures, because one would be the reflection of the other.

Example 2: Mirror figures:



SCREEN 4

The computer will show you, for the next 4 minutes, pairs of figures and your task will consist of identifying whether such pairs are identical or mirror figures. All participants in the experiment will see the same pairs of figures in exactly the same order. If at the end of the experiment the computer randomly chooses task 1, you will earn 15 euro cents for each correct answer.

Press OK to start with Task 1 of the experiment.

SCREEN 5

Task 2 Instructions

In task 2 you will be given some codes. Each code shows which letter of the alphabet corresponds to a number from 1 to 9. Your task consists of decoding sequences of letters, i.e., in associating numbers to letters following the given code.

For example, if the code is:

B	H	U	T	R	W	C	A	K
1	2	3	4	5	6	7	8	9

And the sequence of letters we give you is: TWK
The correct answer would be: 469

During the next 4 minutes, the computer will show codes and sequences of letters in order for you to write the corresponding numbers. All participants in the experiment will see the exact same sequences of letters and exactly in the same order. If at the end of the experiment the computer randomly chooses task 2, you will earn 15 euro cents for each correct sequence of letters.

Press OK in order to start with task 2.

SCREEN 6

Task 3 Instructions
In task 3, you will have to do the same as in task 1; i.e., the computer will show you for the next 4 minutes pairs of figures and your task consists of identifying whether they are identical or mirror figures.

All participants in the experiment will see the same pair of figures in exactly the same order.

In this task you are matched with another participant in the experiment. Matching has been determined by the number of correct answers in task 1 of the experiment. The computer will order participants from larger to smaller number of correct answers in task 1.

Task 1

1 Participant with the highest number of correctly identified figures
2
3
4
5
6
…
18
19
20 Participant with the lowest number of correctly identified figures

Using this order, the computer will match participants in the following manner. The first with the second, the third with the fourth, the fifth with the sixth and similarly until the ninetieth participant is matched with the twentieth participant in the ranking. You will not know your position in the ranking, i.e., you will not know whether you are the 1^{st}, 2^{nd} … or 20^{th} but, using this matching mechanism, it is guaranteed that the participant matched with you gave a similar number of correct answers in task 1 as you did.

When task 3 is finished, the computer will compare your number of correct answers in task 3 with the number of correct answers in task 3 of your matched participant, and earnings will depend on this comparison.

- If at the end of the experiment the computer determines you will be paid for task 3, you will earn double what you earned in task 1 for each correct answer; i.e., 30 euro cents per correct answer, whenever your number of correct answers is larger than the number of correct answers of your matched participant.
- You will earn nothing if your number of correct answers is lower than the number of correct answers of your matched participant.
- In the case of ties, each participant will earn 15 euro cents per correct answer.

Press OK in order to start with task 3 of the experiment.

(In treatment "Rival's Gender", before showing pairs of figures the following message appeared: "Your matched participant is a boy/girl". In treatment "Own Gender", before showing pairs of figures the following message appeared: "Please, fill in your gender for administrative purposes").

SCREEN 7

Task 4 Instructions

In task 4, you will have to do the same as in task 2; i.e., the computer will show you for the next 4 minutes different codes and sequences of letters and your task consists of decoding sequences of letters; i.e., in associating numbers to letters following the given code.

All participants in the experiment will see the same codes and sequences of letters in exactly the same order.

In this task you are matched with another participant in the experiment. Matching has been determined by the number of correct answers in task 2 of the experiment. The computer will order participants from larger to smaller number of correct answers in task 2.

Task 2

1 Participant with the highest number of correctly decoded sequences
2
3
4
5
6
…
18
19
20 Participant with the lowest number of correctly decoded sequences

Using this order, the computer will match participants in the following manner. The first with the second, the third with the fourth, the fifth with the sixth and similarly until the ninetieth participant is matched with the twentieth participant in the ranking. You will not know your position in the ranking, i.e., you will not know whether you are the 1st, 2nd … or 20th but, using this matching mechanism, it is guaranteed that the participant matched with you gave a similar number of correct answers in task 2 as you did.

When task 4 is finished, the computer will compare your number of correct answers in task 4 with the number of correct answers in task 4 of your matched participant, and earnings will depend on this comparison.

- If at the end of the experiment the computer determines you will be paid for task 4, you will earn double what you earned in task 2 for each correct answer; i.e., 30 euro cents per correct answer, whenever your number of correct answers is larger than the number of correct answers of your matched participant.
- You will earn nothing if your number of correct answers is lower than the number of correct answers of your matched participant.
- In the case of ties, each participant will earn 15 euro cents per correct answer.

Press OK in order to start with task 4 of the experiment.

(In treatment "Rival's Gender", before showing the first code the following message appeared: "Your matched participant is a boy/girl". In treatment "Own Gender", before showing the first code the following message appeared: "Please, fill in your gender for administrative purposes").

SCREENS 8 TO 11

(The following three questions were asked to all participants in all treatments for each of the four tasks once the four tasks had concluded but before showing any result to them. Subjects were paid 10 extra euro cents per correct answer.)

- How many figures (sequences of letters) do you think you have correctly identified (decoded) in task 1 (2, 3, 4)?
- Out of the 20 participants in the experimental session, what do you think is your ranking when ordering results in task 1 (2, 3, 4) of the experiment?
- Out of all participants in the experimental session, who do you think performed task 1(2, 3, 4) best? Boys/Girls/Equally

SCREEN 12

(The following four questions were asked to subjects in treatments where the information contained in such questions had not been provided in the past. Subjects were paid 10 extra euro cents per correct answer.)

- Who do you think has correctly identified more figures in task 1? Me/ My matched participant in task 3.
- Do you think you have competed against a boy or a girl in task 3?
- Who do you think has correctly decoded more sequences of letters in task 2? Me / My matched participant in task 4.
- Do you think you have competed against a boy or a girl in task 4?

Final Questionnaire

Gender:
Language:
Studies:
Year of studies:
Age:
Nationality:

- Do you take part in any type of competitive activity (cultural, Sports, Entertainment), i.e., in which you compete?

- If so, in which one?

- “I am good at competing”, please indicate your degree of agreement with this sentence, using a 1 to 7 scale. 1 means you completely disagree, while 7 means you completely agree.

- “I enjoy competing”, please indicate your degree of agreement with this sentence, using a 1 to 7 scale. 1 means you completely disagree, while 7 means you completely agree.