Stereotypes are only a Threat when Beliefs are Reinforced: on the Sensitivity of Gender Differences in Performance under Competition to Information Provision

Nagore Iriberri
Pedro Rey-Biel

This version: February 2016
September 2011

Barcelona GSE Working Paper Series

Working Paper nº 880

Stereotypes are Only a Threat when Beliefs are Reinforced: On the Sensitivity of Gender Differences in Performance under Competition to Information Provision

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

First version: September 2011

This version: January 2016

Abstract

We show that the existence of gender differences in performance is highly sensitive to the task used to measure performance, to existing stereotypes and to informational conditions. Out of sixteen purposely designed treatments we find that women underperform when competing only when two conditions are met: 1) the task used is perceived as favoring men and 2) the presence of a rival is strongly primed through information provided before competing. Such sensitivity sheds light on the contradictory evidence found on *stereotype-threat* causing gender differences in performance under competition.

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

JEL classification: C72; C91; D81.

[+] University of the Basque Country UPV-EHU, IKERBASQUE, Basque Foundation for Science. 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. Nagore Iriberri acknowledges financial support from Ministerio de Economía y Competitividad (ECO2012-31626), Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT869-13), Ministerio de Ciencia e Innovación (ECO2011-25295), and Instituto de la Mujer (2001-0004-INV-00037). Pedro Rey-Biel acknowledges financial support from Programa Ramón y Cajal, Ministerio de Economía y Competitividad (ECO2012-31962, ECO2015-63679-P and the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563)), Instituto de la Mujer (2001-0004-INV-00037) and Barcelona GSE.

1. **Introduction**

Gender differences in labor market outcomes persist, and are a continual subject for study among economists. In addition to the classical explanations based on gender differences in human capital and preferences, and on statistical discrimination, two seminal papers propose gender differences in *competitiveness* as a complementary explanation. Gneezy et al. (2003) show that women underperform compared to men in competitive environments, while Niederle and Vesterlund (2007) show that women are more likely to avoid competitive environments than men. 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 promotion. The importance of these results has led to many follow-up studies, which are reviewed in Croson and Gneezy (2009), Niederle and Vesterlund (2011) and in Niederle (forthcoming).

There are open questions in our understanding of underperformance by women in competitive environments, regarding both the sensitivity to the task used to measure competitiveness, and the gender composition of the competing group. Indeed, follow-up studies to the seminal paper by Gneezy et al. (2003) have shown contradictory findings: women's underperformance in competitive environments is not a robust finding but depends on the task (Günther et al., 2010, Dreber et al., 2011, Shurchkov, 2012, Cárdenas et al., 2012, and Wieland and Sarin, 2012), and on the gender composition of the competing group (Gneezy et al. 2003; Gneezy and Rustichini, 2004; Antonovics et al., 2009).

This paper revisits gender differences in performance under competition.[1] Our purpose is to study the importance of the main factors which could potentially influence the efforts of individuals when competing. In particular, we study whether *stereotype threat* (Steele, 1997), defined as the concern arising from a situation where a person confirms a negative stereotype about their social group, may be behind the conflicting results about the presence of gender differences in performance under competition. Steele and Aronson (1995) and Ryan and Ryan (2005) argue that very subtle manipulations can activate stereotype threat and affect performance. Therefore, our experiment varies the tasks used to measure performance, and thus the existing

[1] Most previous gender studies about competition have focused on policies and institutions that can change gender differences in tournament entry and not on performance once individuals are competing. See the papers by Cason et al. (2010), Wozniak et al. (2010), Ertac and Szentes (2010), Dargnies et al. (2012). Similarly, Brandts et al. (2015), show that information on relative performance differences transmitted through advice reduces the gender gap in tournament entry.

stereotypes about whether one particular gender outperforms the other in such tasks, and information regarding the competitive setting. That is, we seek to understand when (if) women underperform in competitive environments, and if so why.

We use incentivized elicitation of individuals' perceptions to confirm that perceptions about tasks indeed go in opposite directions. Although previous research hypothesizes that the male and female nature of tasks is behind the contradicting results found in the literature, none of them uses incentivized elicitation of perceptions to classify the different tasks. Dreber et al. (2011) and Wieland and Sarin (2012) elicit perceptions, but in contrast to our study they do so in a non-incentivized survey. In addition, we obtain demographic variables, measures of individuals' general attitudes toward competition, and incentivized measures of their confidence, which we use as additional controls in our analysis.

Subjects in our experiment perform each of two tasks twice, first under a piece-rate incentive scheme and then under a competitive scheme. The provision of information just before individuals participate in the competitive stage is the treatment variable in our paper. In order to run a comprehensive analysis of what type of information provision affects performance, in our control treatment individuals do not learn anything about their own performance under piece-rate incentives, or about their rivals. In the rest of the treatments, subjects are either primed about gender (both their own and the rival's), or receive information regarding their own performance, their relative performance or the presence of a rival, before they participate in the tournament. Additional treatments combine these pieces of information.

We find relative underperformance by women only in competitive environments in very specific contexts, and in just three out of 16 treatments. Specifically, this underperformance is found only in the task which is believed to favor men and only when the presence of a rival is primed, either by revealing the rival's gender or by informing that rivals are ready to start competing.

In order to address the possibility of experimental *demand* effects, which are a concern in any study where the treatment variable is the provision of information, we use the individual characteristics and beliefs of subjects as additional controls in our analysis. While there are no significant differences between men and women in standard demographics in our sample, we do observe gender differences in perceptions about tasks, attitudes towards competition, and overconfidence measures. When these variables are factored in as additional controls they partially explain gender differences in performance, so those differences become weaker. This is consistent with the presence of stereotype threat. In addition, the fact that these variables partly explain the strong gender differences in performance rules out the possibility that the gender differences observed may be caused by pure demand effects.

Freeman and Gelber (2010) and Gill and Prowse (2014) use relative performance information before a tournament to study whether women and men react to positive or negative information differently. The former finds no significant gender effects, but the latter finds that women and men react differently to positive and negative feedback. Geraldes et al. (2011) explicitly provide the gender of the opponent. The main difference between their paper and ours is that they always provide the gender of the opponent and vary the information regarding the existing stereotype threat, while one of our treatment variables is precisely information regarding gender. Our main contribution to the literature is that our treatment design allows us precisely to study how different degrees of manipulation regarding the information on both gender and ability, in combination with perceptions about which gender outperforms the other in different tasks, can explain gender differences in competitive environments. In particular, we show that the provision of information priming the presence of a competing rival explains women's underperformance in competitive environments and that this information provision interacts with the perception of the task, activating stereotype-threat.

The paper is organized as follows. Section 2 describes the experimental design and the procedures, and gives detailed information regarding the tasks and individuals' perceptions about them. Section 3 contains the results. Section 4 concludes. The Appendix contains translations of the instructions and the post-experiment questionnaire.

2. **Experimental Design and Procedures**

A total of 640 participants, 20 per one-hour session, were recruited using the ORSEE recruiting system (Greiner, 2004), ensuring that subjects had not participated in similar experiments in our laboratory in the past.[2] Experimental sessions were conducted in the Laboratori d'Economia Experimental (LEEX) at Universitat Pompeu Fabra using z-Tree experimental software (Fischbacher, 2007). Our recruiting method

[2] Some subjects suffered small computer glitches during the experiment that prevented them from having the full time to perform some of the tasks. We omit them from the analysis, which explains the small sample size variation across treatments.

ensured that in every session half of the subjects were men and half were women, without subjects noticing that the experiment involved a gender study.

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 randomly selected for payment and subjects were paid 15 euro cents for each correct solution they gave in that task. For the pair-wise tournaments, we followed a positive matching of participants based on their performance under the piece-rate task, which was public knowledge. Following Lazear and Rosen (1981), in order to study the pure effect of competition on all participants it is important for the competition to be similarly tight. This is ensured by our matching protocol.[3] For tournament payment, one of the two tasks was randomly chosen and the subject who performed best in each pair in that task earned 30 euro cents per correct answer, while the other subject earned nothing.[4] Additionally, once all tasks had concluded, subjects could earn 10 euro cents for each of 16 questions rewarding predictive accuracy. Finally, subjects also earned a 3 euro participation fee. Average total payments were 13.80 euro with a large standard deviation of 6.34 due to the competitive environment.

The two tasks were carefully chosen following an exhaustive reading of the Psychology literature on gender differences in performance in the absence of incentives. We aimed to find two distinct tasks in which each gender would perform better than the other and, at the same time, where there were 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 depend greatly on the specific details of the tasks.[5] Nevertheless, there is 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

[3] The average rank according to performance under piece-rate incentives was significantly different for men and women (9.17 and 11.80), but far from the extreme case (5 for men, 15 for women) for the mental rotation task. The average rank was not significantly different in the symbol digit substitution task (10.35 for men, 10.66 for women). This led to the following gender combinations: For the mental rotation task there were 87 male pairings, 87 female pairings and 146 mixed pairings, and for the symbol digit substitution task there were 80 male pairings, 81 female pairings and 160 mixed pairings.

[4] Since the subjects did not know until the end of the experiment whether they would be paid for each incentive-scheme according to their performance in the mental rotation task or symbol digit substitution task, regardless of their attitudes towards risk or their perceptions about how good they were in each of the two tasks it was in their best interest to always perform the best they could in both tasks.

[5] Hoffman et al. (2011) further show that nurture may be behind these gender differences.

description 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 and see Majeres (1983) for evidence on gender differences in this task. A non-incentivized pilot experiment conducted prior to our main experiment at a different university with 184 subjects of the same age, using paper and pencil, confirmed not only that men outperformed women in the mental rotation task and that women outperformed men in 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.[6]

We adapted the tasks to our computerized setting. The mental rotation task (MRT) consisted of showing pairs of three-dimensional figures to subjects, who had to answer whether such figures were "identical" or "mirror" figures. Figure 1 shows a pair of identical and a pair of mirror images from the experiment.

(a) Identical Figures (b) Mirror Figures

Figure 1. Mental Rotation Task (MRT)

The symbol digit substitution task (SDST) consisted of showing subjects codes which associated nine numbers with nine letters. 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.[7] 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.

[6] In our non-incentivized pilot, we gave subjects two minutes to perform each task. For the mental rotation task men solved 15.56 figures correctly on average and women 12.21. For the 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. As shown in Figure 6, a high proportion of subjects (43%) expected men to outperform women in the mental rotation task, while a high proportion (42%) expected women to outperform men in the codification task.

[7] Note that in adapting this task to the computer we modified two elements. First, our codes associate numbers with letters, while in the original task codes associate numbers with symbols. Thus, subjects in our experiment 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 the much longer strings commonly used. Shorter sequences enable us to make more precise performance measurements.

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

Given that our chosen tasks differ in several dimensions, such as their level of difficulty, processing and available strategies, we always perform the analysis separately for each task.

The experiment involved eight treatments with 40 male and 40 female subjects in each treatment.[8] We used a between-subject design. In all treatments subjects performed both tasks, MRT and SDST, under piece-rate incentives and then under a tournament scheme. In the "Control" treatment subjects received no information before they participated in the tournament. In the remaining treatments, subjects where either primed about gender (both own and rival's), or received information regarding their own performance, their relative performance or the presence of a rival, prior to participating in the tournament. Additional treatments combined these pieces of information. Figure 3 describes the treatment design.

Treatments / Information	Control	T1: Ability Differences	T2: Rival's Gender	T3: Ability Differences and Rival's Gender	T4: Own Ability Assessment	T5: Own Gender	T6: Own Ability Assessment and Own Gender	T7: Rival Ready
Info on performance differences	No	Yes	No	Yes	No	No	No	No
Info about rival's gender	No	No	Yes	Yes	No	No	No	No
Info on relative performance	No	No	No	No	Yes	No	Yes	No
Priming of Gender	No	No	No	No	No	Yes	Yes	No
Priming of Rival	No	No	No	No	No	No	No	Yes

Figure 3: Treatment Design

Once the four tasks were completed, subjects were given an incentivized questionnaire which included questions regarding their number of correct answers, their relative ranking and whether they believed women or men outperformed the other

[8] In the "Own Ability and Gender" treatment there were 30 male and 30 female subjects, as one session was cancelled due to software problems.

gender or not for each task. Additionally, in treatments where the information contained in these questions had not been provided earlier, subjects were asked questions regarding the gender of the rival and/or whether they believed they had outperformed their rival or not under piece-rate incentives. Finally, subjects filled in a questionnaire dealing with standard demographics (gender, age, nationality and studies), and questions about their attitude toward competition. All these variables are used as controls when analyzing the results. See Figure 4 for the timeline of the experiment and see the Appendix for experimental instructions and the questionnaire.

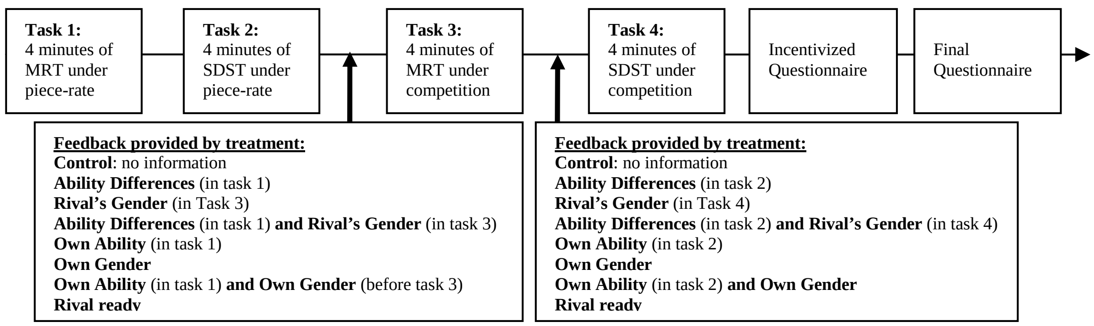


Figure 4. Timeline of the Experiment

We can now check whether our choice of tasks satisfies the inherent ex-ante gender differences we aimed for with our design. Notice that under piece-rate incentives, performance should not differ across all eight treatments for each task. We test for the validity of our randomization into treatments, confirming that we can aggregate data across treatments for both tasks (636 observations)[9]. Figure 5 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 0.00). However, for SDST, we cannot reject the notion that the two cumulative distributions are equal (two-sample Kolmogorov-Smirnov test for equality of distribution functions yields a *p*-value of 0.27). This 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. After adapting SDST to our computerized setting this is no longer the case (see footnote 7).[10]

[9] The average correct number of answers in MRT and SDST is 26.04 and 36.94, respectively. The *p*-value for the null hypothesis that the performance is not statistically different across the eight treatments is 0.11 for the MRT and 0.39 for the SDST.
[10] 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.

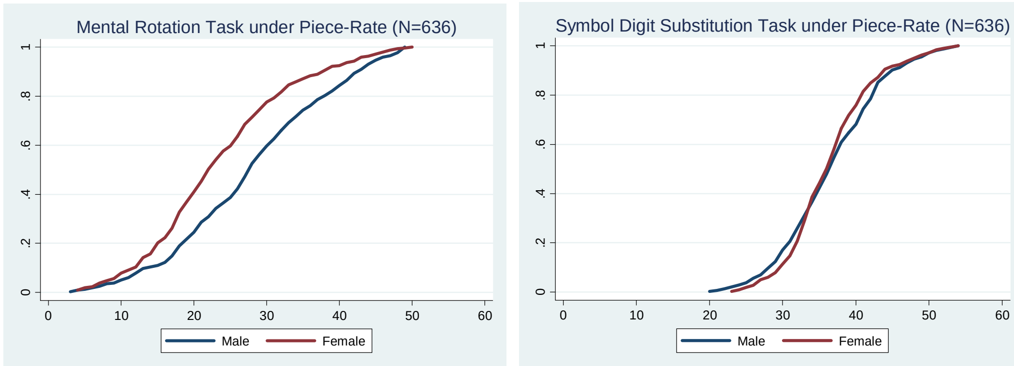


Figure 5. CDF of Number of Correct Answers in MRT and SDST by Gender under Piece-Rate

More importantly, perceptions regarding which gender is favored by each task do not change when the tasks are adapted to a computerized setting. Figure 6 uses answers from the questionnaire administered after participants had completed the experiment to graph the average frequency assigned by participants of each gender to each gender outperforming the other at each task under piece-rate incentives (see last question on 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.[11] Furthermore, while we find no gender differences in the perceptions for MRT, we do find that women perceive SDST to be a more female-favoring task than men do (*p*-values of 0.682 and 0.032 for the Kolgomorov-Smirnov test for the equality of perception distributions by gender, respectively).

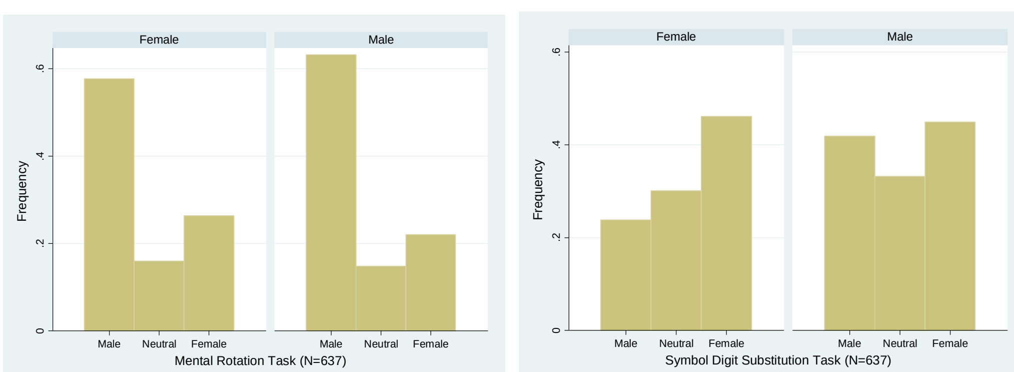


Figure 6. Histograms of Perceptions in MRT and SDST under Piece-Rate by Gender

The design of our experiment leaves us with two interesting cases. On the one hand MRT, in which men not only outperform women but there is a consensus that this

[11] Pearson's Chi-Squared tests, where the null hypothesis is that the both frequency distributions are consistent with a uniform distribution, conclude against the null with *p*-values smaller than 0.01.

is the case, and on the other hand SDST, where perceptions regarding a female advantage are not confirmed by performance data. This enables us to further explore the role of perceptions in explaining gender differences in performance under competition in section 3.2 below.[12]

3. Results

3.1. Do Women Underperform under Competition? If so, When?

Table 1 shows the mean of the main performance variable, i.e. the number of correct answers (*No. of Correct*), in each tasks under both piece-rate incentives (columns (2) to (4)) and competition (columns (6) to (8)), overall and separated by gender for each treatment.[13] The last block of columns, (10) to (12), show the mean values for the differences between performances under tournament and under piece-rate conditions, named *Improvement*. Finally, Table 1 also reports *p*-values for the tests comparing the average performances of men and women per treatment under piece-rate incentives (column (5)), under competition (column (9)), and for the improvement variable shown in column (13).

Overall, both male and female subjects improve their performances on average when they move from piece-rate incentives to competition. The average improvement in the number of correct answers in MRT and SDST are 5.23 and 3.03, respectively.[14] The improvement for male subjects is significantly different from that of female subjects in MRT, but the difference is not significant in SDST, with *p*-values of 0.02 and 0.91 for the overall improvement in MRT and SDST, respectively. Since subjects in the control group face exactly the same tasks in the same sequence, the treatment effects observed when providing information cannot be reconciled with pure learning. Per treatment, the improvement from piece-rate to tournament is different for men and women in three of

[12] Our measures of gender perceptions about the tasks are obtained once subjects have performed both tasks under piece-rate and tournament schemes, so they can be interpreted to some degree as an ex-post justification of their individual experiences. First, notice that subjects have monetary incentives to express their true perceptions. Second, perceptions are elicited before subjects are shown their performance results and thus only in the "Ability Differences", "Ability Differences and Rival's Gender", "Own Ability" and "Own Ability and Gender" treatments could subjects have a partial indication of whether they have an ex-ante advantage with respect to their rival in each tournament. We test whether the mean perception is different across the treatments: This is ruled out for MRT (*p*-value of 0.3558), but cannot be rejected for SDST (*p*-value of 0.0557). With respect to the latter, the main difference is in treatment 8, where almost 55% of the subjects believe it is a neutral task. Finally, the correlation between the number of responses submitted to each task and perceiving the task as favoring the opposite gender is negative, but is always below 0.15.

[13] We also analyze the number of answers submitted and their accuracy, which is calculated by the proportion of correct answers out of those submitted. See footnote 15.

[14] Low ability subjects, those performing below the median, improve more than high ability subjects in MRT (7.10 and 3.12 for low and high ability, respectively) but improve equally in SDST (3.07 and 2.99, respectively).

the treatments in MRT ("Rival's Gender", "Ability Differences and Rival's Gender" and "Rival Ready") and we find no evidence for any gender difference in improvement in SDST.

Table 1. Overview of Results on Performance per Gender and Treatment *(No. of Correct)*

	Task	Obs.	Piece-Rate Overall	Piece-Rate Male	Piece-Rate Female	Piece-Rate *p*-value	Competition Overall	Competition Male	Competition Female	Competition *p*-value	Improvement Overall	Improvement Male	Improvement Female	Improvement *p*-value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Overall	MRT	636	26.04	28.29	23.79	0.00	31.25	34.21	28.29	0.00	5.23	5.92	4.54	0.02
	SDST	636	36.94	37.02	36.87	0.77	40.14	40.24	40.04	0.68	3.03	3.02	3.05	0.91
Per treatment:														
Control	MRT	78	25.83	28.92	22.74	0.01	31.06	33.69	28.50	0.05	5.37	4.77	5.97	0.46
	SDST	79	37.20	37.10	37.30	0.89	39.22	38.89	39.53	0.63	2.10	1.97	2.23	0.76
Ability Differences	MRT	80	26.28	29.25	23.30	0.01	31.17	33.59	28.74	0.05	4.92	4.28	5.56	0.47
	SDST	79	35.66	35.05	36.25	0.38	38.84	38.62	39.05	0.74	3.18	3.56	2.80	0.37
Rival's Gender	MRT	80	24.23	25.00	23.45	0.53	29.99	33.40	26.58	0.02	5.76	8.40	3.13	0.00
	SDST	80	36.65	37.70	35.60	0.13	40.10	41.43	38.78	0.03	3.45	3.73	3.18	0.50
Ability Differences	MRT	79	24.75	25.72	23.80	0.39	31.13	33.79	28.53	0.02	6.38	8.08	4.73	0.04
and Rival's Gender	SDST	79	38.35	38.05	38.65	0.66	40.48	39.95	41.00	0.43	2.13	1.90	2.35	0.51
Own Ability Assessment	MRT	99	28.31	31.06	25.51	0.00	31.99	35.04	28.88	0.00	3.68	3.98	3.37	0.66
	SDST	99	37.00	37.86	36.12	0.19	41.84	43.49	40.23	0.01	3.47	3.59	3.35	0.80
Own Gender	MRT	80	25.70	28.43	22.98	0.02	30.25	33.28	27.23	0.01	4.55	4.85	4.25	0.70
	SDST	80	36.83	37.63	36.03	0.25	40.65	40.93	40.38	0.69	3.83	3.30	4.35	0.20
Own Ability Assessment	MRT	60	28.15	31.40	24.90	0.02	34.33	37.60	31.07	0.01	6.18	6.20	6.17	0.99
and Own Gender	SDST	60	38.20	36.70	39.70	0.06	40.55	39.10	42.00	0.08	2.35	2.40	2.30	0.93
Rival Ready	MRT	80	25.01	26.55	23.48	0.17	30.64	33.90	27.38	0.01	5.63	7.35	3.90	0.07
	SDST	80	35.96	35.73	36.20	0.73	39.54	39.20	39.88	0.57	3.58	3.48	3.68	0.82

We now turn to regression analysis using the improvement in the number of correct answers from piece-rate to tournament as the dependent variable. Columns (1) and (2) of Table 2 correspond to MRT while columns (3) and (4) correspond to SDST. In columns (1) and (3), the variable of interest is gender, while in columns (2) and (4) the variables of interest are gender and, more importantly, the interactions between *Female* and each treatment. The treatment omitted in all regressions is the control.[15]

[15] We have performed similar regressions with alternative dependent variables such as the improvement in the number of answers submitted and the improvement in the number of mistakes. We find qualitatively the same results, although the alternative dependent variables show higher variance and the results are slightly weaker. We have also run alternative specifications such as having the number of correct answers in the competitive stage as the dependent variable and including the number of correct answers under piece-rate as the independent variable. With this alternative specification we find qualitatively identical results, although some effects become weaker.

Table 2. Treatments versus Control: Differential Treatment Effect for Women

	MRT		SDST	
	Improvement	Improvement	Improvement	Improvement
	(1)	(2)	(3)	(4)
Female	**-1.387****	**1.205**	**0.0424**	**0.251**
	(0.596)	**(1.638)**	**(0.303)**	**(0.824)**
Ability Differences	-0.449	-0.487	1.075*	1.590*
	(1.211)	(1.661)	(0.587)	(0.849)
Rival's Gender	0.391	3.631**	1.348**	1.751**
	(1.181)	(1.734)	(0.579)	(0.835)
Ability Differences and Rival's Gender	1.017	3.308**	0.0243	-0.0762
	(1.151)	(1.673)	(0.534)	(0.804)
Own Gender	-0.822	0.0808	1.723***	1.326
	(1.141)	(1.663)	(0.578)	(0.831)
Own Ability	-1.702	-0.789	1.366**	1.616*
	(1.086)	(1.607)	(0.616)	(0.863)
Own Gender and Own Ability	0.812	1.431	0.248	0.426
	(1.340)	(1.775)	(0.708)	(1.005)
Rival Ready	0.253	2.581	1.473**	1.501
	(1.240)	(1.779)	(0.590)	(0.915)
Female*Ability Differences		**0.0769**		**-1.015**
		(2.398)		**(1.177)**
Female*Rival's Gender		**-6.480*****		**-0.801**
		(2.316)		**(1.164)**
Female*Ability Diff. and Rival's Gender		**-4.557****		**0.201**
		(2.286)		**(1.076)**
Female*Own Gender		**-1.805**		**0.799**
		(2.274)		**(1.159)**
Female*Own Ability		**-1.818**		**-0.491**
		(2.161)		**(1.238)**
Female*Own Gender and Own Ability		**-1.238**		**-0.351**
		(2.672)		**(1.425)**
Female*Rival Ready		**-4.655***		**-0.0513**
		(2.466)		**(1.189)**
Constant	6.065***	4.769***	2.081***	1.974***
	(0.883)	(1.181)	(0.441)	(0.610)
Observations	634	634	615	615
R-squared	0.022	0.044	0.030	0.035
H0: All (*Female*Treatments*) the same		0.05		0.83

Notes: the dependent variable is the *Improvement* from piece-rate to tournament, defined as the difference between the number of correct answers in the tournament and the number of correct answers in the piece-rate. *Female* is a dummy that takes value of 1 if the subject is female. The rest of the variables are dummy variables that take a value of 1 for each treatment. Robust standard errors are in parenthesis. * p<0.1; **; ** p<0.05; *** p<0.01.

For MRT, column (1) shows a negative, statistically significant coefficient for the gender variable *Female*, while column (3) shows no evidence of female underperformance for SDST. More interestingly, in column (2) *Female* becomes insignificant for MRT and the interactions between *Female* and the treatments prove to be significant for only three of the treatments in MRT ("Rival's Gender", "Ability Differences and Rival's Gender" and "Rival Ready"). Column (4) shows no gender differences for SDST. As shown by the hypothesis test at the end of Table 2, it cannot be rejected that female interactions with the treatments are the same for SDST but, more importantly, it can be rejected for MRT.

The strongest effect is observed for "Rival's Gender". When no information is provided, men on average improve in about 5 answers, shown by the constant, while women improve in about 6 answers, but the difference is not significant. However, when information about the rival's gender is provided, men on average improve in 8 answers while women improve only in 3, with this difference being highly significant. In other words, for performance effects in absolute terms, in the treatment including information about the rival's gender men's performance improves by 75% in the tournament, while women's performance decreases by almost 50% compared to the control when no information is provided. These effects are confirmed with separate regressions for male and female subjects (significance levels of 5% and 10%, respectively). Quantitatively similar effects are found for the "Gender Differences and Rival's Gender" and "Rival Ready" treatments, although they are qualitatively weaker.[16]

Two observations are noteworthy. First, it has already been seen that a crucial difference between MRT and SDST is the perception about which gender is favoured by the task: MRT is perceived as a task favouring men, while SDST is perceived as a task favouring women. Our results show that one necessary condition for women to underperform in competitive environments is for there to be a male-favouring task, as in SDST no underperformance by women is observed. However, this is not sufficient as even when performing the MRT there are several treatments in which women do not

[16] On separate regressions for male and female individuals, the coefficients for male subjects are always positive for these three treatments, while they are always negative for female subjects. For male subjects, the positive coefficients on the treatments are significant at 5% for "Gender Differences and Rival's Gender" and non-significant for "Rival ready" treatments. For female subjects, the negative coefficients on the treatments are not significantly for "Gender Differences and Rival's Gender" and non-significant for "Rival ready" treatments. Although the effects seem weaker, we cannot reject that the effects in all three treatments are the same.

underperform compared to men, including the control treatment, where no information is provided.

Second, the information primed before the competition is also crucial. The only cases in which women are observed to underperform compared to men is when the rival's presence is primed, either by priming the rival's gender or their existence, such as in the "Rival ready" treatment. Note that in the case of "Ability Differences and Rival's Gender" two pieces of information are provided, one that by itself has no effect (ability differences) and one that has the expected negative effect on improvement in women's performance (rival's gender). Separated regressions comparing each of these three treatments with the control show that female subjects underperform compared to men, significant at the 1%, 5% and 10% for "Rival's Gender", "Ability Differences and Rival's Gender", and "Rival ready" treatments, respectively.

Given that women have been observed to react very differently depending on the gender of their rival, i.e. on whether the competition is between women or is mixed, we have further analyzed these effects in the case of the two treatments in which the gender of the opponent is explicitly revealed ("Rival's Gender" and "Ability Differences and Rival's Gender"). Estimation results are shown in Table 3. The variables of interest are *Treatment*, which captures the treatment effect, *Female*Treatment*, which captures the differential treatment effect for females, *Treatment*Male Rival*, which captures the differential treatment effect when the rival is male, and, finally, *Female*Treatment*Male Rival*, which captures the differential treatment effect for women that depends on the rival's gender. For both informational treatments, we find that the female underperformance with respect to men does not depend on the rival's gender, as the triple interaction is insignificant in both cases.[17]

[17] We have also tested for differential gender and treatment effects that depend on the positive and negative nature of the information (advantageous/disadvantageous difference over the opponent) but found no significant effects.

Table 3.
Treatment versus Control: Differential Treatment Effect that Depends on the Gender of the Competitors

	Rival's Gender	Ability Differences and Rival's Gender
Female	5.706***	5.706***
	(1.493)	(1.493)
Male Rival	3**	3**
	(1.397)	(1.397)
Female*Male Rival	-2.524	-2.524
	(2.641)	(2.641)
Treatment	**9.222*** **	**10.26*** **
	(2.118)	**(1.558)**
Female*Treatment	**-12.20*** **	**-10.52*** **
	(2.925)	**(2.631)**
Treatment*Male Rival	**-4.495**	**-7.263*** **
	(2.970)	**(2.680)**
Female*Treatment*Male Rival	**4.903**	**5.337**
	(4.277)	**(4.099)**
Observations	158	157
R-squared	0.352	0.383

Notes: the dependent variable is the *Improvement* from piece-rate to tournament, defined as the difference between the number of correct answers in the tournament and the number of correct answers in the piece-rate, in MRT in the "Rival's Gender" and "Ability Differences and Rival's Gender" treatments. The control group is omitted. *Female* is a dummy that takes a value of 1 if the subject is female. *Treatment* takes a value of 1 if the treatment is the one given by the column. Robust standard errors are in parentheses. * p<0.1; **; ** p<0.05; *** p<0.01.

3.2. Why Do Women Underperform under Competition when Primed with a Certain Type of Information?

Using data on individual characteristics obtained during and after the experiment, we further explore whether differences in these variables explain the effect of information on women's underperformance.

We start by looking at whether there are differences between men and women in the control variables obtained during the experiment sessions. Table 4 summarizes the individual characteristics grouped into four categories for all subjects, and separated by gender. The last column includes the *p*-values for the Kruskal-Wallis equality-of-populations rank test of differences between the two genders.

Table 4. Control Variables for All, Male and Female Subjects

	Overall			Male			Female			
Variables	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	*P*-Value
Demographics:										
Age	637	20.89	2.90	318	20.98	3.21	319	20.81	2.55	0.45
Foreign	637	0.06	0.23	318	0.06	0.23	319	0.06	0.23	0.99
Fields of Study:										
Social Sciences	637	0.64	0.48	318	0.66	0.47	319	0.61	0.49	0.20
Humanities	637	0.23	0.42	318	0.20	0.40	319	0.25	0.43	0.19
Applied Sciences	637	0.04	0.21	318	0.06	0.24	319	0.03	0.16	0.02
Natural Sciences	637	0.05	0.22	318	0.03	0.16	319	0.08	0.26	0.00
Other Fields	637	0.04	0.20	318	0.04	0.21	319	0.04	0.20	0.84
Attitudes toward Competition:										
Experience Competing	637	0.32	0.47	318	0.49	0.50	319	0.15	0.36	0.00
Ability Competing	637	4.82	1.37	318	5.12	1.34	319	4.52	1.34	0.00
Enjoy Competing	637	4.80	1.67	318	5.29	1.60	319	4.30	1.60	0.00
Gender Perception about Tasks:										
MRT: Favors Opposite Gender	637	0.40	0.49	318	0.22	0.41	319	0.58	0.49	0.00
SDST: Favors Opposite Gender	637	0.31	0.46	318	0.37	0.48	319	0.24	0.43	0.00
Confidence:										
MRT: Guessed Rank	636	10.03	4.79	318	8.76	4.66	318	11.29	4.58	0.00
SDST: Guessed Rank	636	10.19	4.25	317	9.59	4.30	319	10.78	4.11	0.00
MRT: Confidence Rank	636	0.46	4.99	318	0.42	4.94	318	0.51	5.04	0.81
SDST: Confidence_Rank	636	0.32	5.50	317	0.76	5.47	319	-0.12	5.51	0.04

Notes: *Foreign* is a dummy variable that takes a value of 1 when the subject is non-Spanish. There are five fields of study. Each of them measures the proportion of subjects studying each of the fields. *Experience Competing* is a dummy variable that takes a value of 1 if the subject reveales he/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 (total agreement) of subjects in response to the following statement: "I am good at/enjoy competing". *Favors Opposite Gender* is a dummy variable that takes a value of 1 when the subject is male/female and reveals that he/she 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 out of 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.

The variables in the first category - Demographics – were elicited in the ex-post questionnaire. They include subjects' age, whether participants are foreigners or not, and their field of study, classified in five categories.[18] No significant differences are observed between female and male subjects, except in the proportions of subjects studying Applied and Natural Sciences, which have a low frequency in the sample. These differences go in the frequently observed direction, as more women are found to be studying Natural Sciences (i.e. Biology) and fewer studying Applied Sciences (i.e. Engineering). The second category – Attitudes Toward Competition – was also elicited in the ex-post questionnaire. It includes a dummy variable indicating whether subjects

[18] 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.

regularly participate in competitive activities (*Experience Competing*), and two variables ranging from 1 to 7 (where 1 indicates total disagreement and 7 total agreement) regarding whether subjects consider they are good at competing (*Ability Competing*) and whether they enjoy competing (*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 just after the subjects had concluded the tasks but before they could observe any result. We define *Favors Opposite Gender* as the proportion of subjects in each gender who think each task favors the opposite gender. Two observations are worth noting. First, on average both genders perceive MRT as a male-favoring task and SDST as a female-favoring task, as shown in the frequency distribution of beliefs in Figure 6. Second, a higher proportion of female subjects think MRT is a male-favoring task (58%) than the proportion of male subjects who think SDST is a female-favoring task (37%), so the negative stereotype concerning women performing in the male task is stronger than the negative stereotype concerning men performing in the female task.

The final category – Confidence – includes two types of variable. *Guessed Rank* is defined as subjects' incentivized beliefs about their ranking in each task. *Confidence Rank* 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 that they are ranked higher than they actually are as all 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.

In order to further explore gender differences under competition, we include all these variables as controls in our main regressions shown in Table 5 below, columns (1) and (2) for MRT and columns (3) and (4) for SDST. Note that among the controls we include those in which we find gender differences, such as fields of study, attitudes toward competition, perceptions about the tasks and confidence. We find that the variables *Ability Competing* and *Enjoy Competing* are highly correlated (correlation over 0.7), so we include only the former.

Table 5. Treatments versus Control with Controls: Differential Treatment Effect for Women

	MRT		SDST	
	Improvement	Improvement	Improvement	Improvement
	(1)	(2)	(3)	(4)
Female	**-0.999**	**1.094**	**0.0669**	**0.174**
	(0.667)	**(1.676)**	**(0.341)**	**(0.844)**
Ability Differences	-0.0992	0.00730	1.130*	1.525*
	(1.134)	(1.495)	(0.587)	(0.853)
Rival's Gender	0.485	2.970*	1.387**	1.726**
	(1.152)	(1.694)	(0.585)	(0.840)
Ability Differences and Rival's Gender	1.376	2.956*	0.208	0.0842
	(1.146)	(1.708)	(0.551)	(0.825)
Own Gender	-0.0513	0.502	1.858***	1.300
	(1.124)	(1.601)	(0.587)	(0.839)
Own Ability	-1.352	-0.731	1.560**	1.804**
	(1.095)	(1.619)	(0.625)	(0.882)
Own Gender and Own Ability	1.323	1.973	0.369	0.589
	(1.338)	(1.757)	(0.706)	(0.987)
Rival Ready	1.118	3.296*	1.703***	1.658*
	(1.205)	(1.711)	(0.605)	(0.914)
Female*Ability Differences		**-0.249**		**-0.776**
		(2.235)		**(1.192)**
Female*Rival's Gender		**-5.025****		**-0.659**
		(2.258)		**(1.175)**
Female*Ability Diff. and Rival's Gender		**-3.207**		**0.274**
		(2.332)		**(1.086)**
Female*Own Gender		**-1.208**		**1.133**
		(2.220)		**(1.159)**
Female*Own Ability		**-1.285**		**-0.464**
		(2.196)		**(1.264)**
Female*Own Gender and Own Ability		**-1.339**		**-0.424**
		(2.656)		**(1.405)**
Female*Rival Ready		**-4.423***		**0.114**
		(2.410)		**(1.193)**
Social Science	-1.614	-1.687	-0.879	-0.971
	(1.378)	(1.347)	(0.623)	(0.615)
Humanities	-0.873	-0.977	-0.0514	-0.117
	(1.439)	(1.397)	(0.667)	(0.656)
Applied Science	-1.468	-1.630	0.346	0.337
	(1.849)	(1.862)	(0.957)	(0.955)
Natural Science	-0.187	-0.517	-1.783*	-1.843**
	(1.827)	(1.775)	(0.933)	(0.932)
Experience Competing	0.132	0.157	0.0234	0.0333
	(0.704)	(0.702)	(0.363)	(0.363)
Ability Competing	0.386	0.377	0.0163	0.0155
	(0.236)	(0.239)	(0.119)	(0.120)
Favors Opposite Gender	-0.691	-0.616	0.116	0.0779
	(0.641)	(0.645)	(0.338)	(0.341)
Confidence Ranks	0.425***	0.410***	0.0181	0.0181
	(0.0575)	(0.0578)	(0.0307)	(0.0317)
Constant	4.933**	4.018**	2.448***	2.480**
	(1.931)	(1.997)	(0.943)	(1.005)
Observations	634	634	614	614
R-squared	0.110	0.123	0.048	0.054
H0: All (*Female*Treatments*) the same		0.21		0.78

Notes: the dependent variable is the *Improvement* from piece-rate to tournament, defined as the difference between the number of correct answers in the tournament and the number of correct answers in the piece-rate. *Female* is a dummy that takes a value of 1 if the subject is female. The rest of the variables are dummy variables that take a value of 1 for each treatment. For the controls see the notes to Table 4. Robust standard errors are in parentheses. * p<0.1; **; ** p<0.05; *** p<0.01.

The results for SDST remain the same, as expected. But for MRT the *Female* coefficient, shown in column (1), is no longer significant as it was in Table 2. The estimates in column (2) show that this is due to the differences in the three main treatments, “Rival’s Gender”, “Rival’s Gender and Ability” and “Rival Ready”: Both the treatment variables and the interactions between *Female* and the treatment variables are lower in magnitude and in significance. More importantly, the null hypothesis that all the interactions between the *Female* coefficient and the treatments are the same is no longer rejected, again in contrast with the results in Table 2. The control variables therefore partly explain the strong gender differences found in Table 2 for MRT. Among the controls, *Confidence Rank* is highly significant, meaning that those who believe they are higher ranked perform better.

We further investigate the interaction between being female, perceiving the task as male and being provided with the information. Column (1) in Table 6 shows the estimation results. In the “Rival’s Gender” and “Rival Ready” treatments, we see that the female underperformance is significantly worse and higher for those who believe they are performing a male task. The exception is the treatment “Ability Differences and Rival’s Gender”. This must be due to the interaction of different pieces of information as in this treatment not just the rival’s gender but also differences in ability between the competing rivals are provided. Similar exercises with other measures, such as perceived ability to compete, shown in column (2), and confidence, shown in column (3), do not show significant results, so we conclude that the main channel for the gender differences observed is related to the perceived male/female nature of the task.

We draw two conclusions from the addition of controls. First, it can be seen that women’s lower confidence, lower competitive attitudes and beliefs that the MRT task favors men partly explain the strong effect found when the rival’s presence is primed, either by providing rivals’ gender or by revealing that the rival is ready. This is consistent with stereotype threat, i.e. it is those women who believe that they are worse at competing and that the task favors male subjects who really underperform when the competition is primed. Second, the weakening of the strong gender underperformance found when the controls are added also rules out the idea that the strong gender effect is due purely to experimental demand effects. If this was the case, the inclusion of controls should leave this effect unchanged, as it should be uncorrelated with individual characteristics.

Table 6.
Treatments versus Control with Controls: Interactions Between Controls and Treatments

	Improvement (1)		Improvement (2)		Improvement (3)
Female	**1.066**	**Female**	**-0.00296**	**Female**	**1.286**
	(1.785)		**(3.082)**		**(1.686)**
Female*Favors Opp. Gender	**-0.733**	**Female*Ability Comp.**	**0.209**	**Female*Conf. Ranks**	**-0.192**
	(1.526)		**(0.524)**		**(0.132)**
Female*Favors Opp. Gender	**-4.416****	**Female*Ability Comp.**	**1.356***	**Female*Conf. Ranks**	**0.0196**
***Rival's Gender**	**(2.101)**	***Rival's Gender**	**(0.765)**	***Rival's Gender**	**(0.180)**
Female*Favors Opp. Gender	**1.902**	**Female*Ability Comp.**	**-0.0215**	**Female*Conf. Ranks*Ab. Diff**	**0.0742**
***Ab. Diff and Rival's Gender**	**(2.264)**	***Ab. Diff and Rival's Gender**	**(0.734)**	**and Rival's Gender**	**(0.196)**
Female*Favors Opp.	**-4.906****	**Female*Ability Comp.*Rival**	**1.214**	**Female*Conf. Ranks*Rival**	**0.0540**
Gender*Rival Ready	**(2.383)**	**Ready**	**(1.535)**	**Ready**	**(0.248)**
Ability Differences	-0.0627	Ability Differences	-0.0648	Ability Differences	0.115
	(1.526)		(1.519)		(1.488)
Rival's Gender	2.875*	Rival's Gender	2.988*	Rival's Gender	2.825*
	(1.720)		(1.724)		(1.712)
Ability Differences and Rival's Gender	2.872*	Ability Differences and Rival's Gender	3.017*	Ability Differences and Rival's Gender	2.864*
	(1.709)		(1.725)		(1.738)
Own Gender	0.488	Own Gender	0.499	Own Gender	0.604
	(1.607)		(1.609)		(1.604)
Own Ability	-0.897	Own Ability	-0.712	Own Ability	-0.749
	(1.639)		(1.627)		(1.635)
Own Gender and Own Ability	1.922	Own Gender and Own Ability	1.992	Own Gender and Own Ability	2.071
	(1.767)		(1.779)		(1.764)
Rival Ready	3.304*	Rival Ready	3.215*	Rival Ready	3.436**
	(1.738)		(1.718)		(1.725)
Female*Ability Differences	**-0.197**	**Female*Ability Differences**	**-0.182**	**Female*Ability Differences**	**-0.275**
	(2.257)		**(2.272)**		**(2.230)**
Female*Rival's Gender	**-2.401**	**Female*Rival's Gender**	**-11.34*****	**Female*Rival's Gender**	**-4.986****
	(2.579)		**(4.157)**		**(2.273)**
Female*Ability Diff. and Rival's Gender	**-4.259**	**Female*Ability Diff. and Rival's Gender**	**-3.191**	**Female*Ability Diff. and Rival's Gender**	**-3.225**
	(2.670)		**(3.825)**		**(2.348)**
Female*Own Gender	**-1.216**	**Female*Own Gender**	**-1.232**	**Female*Own Gender**	**-1.432**
	(2.220)		**(2.228)**		**(2.227)**
Female*Own Ability	**-1.127**	**Female*Own Ability**	**-1.312**	**Female*Own Ability**	**-1.345**
	(2.199)		**(2.201)**		**(2.203)**
Female*Own Gender and Own Ability	**-1.320**	**Female*Own Gender and Own Ability**	**-1.383**	**Female*Own Gender and Own Ability**	**-1.446**
	(2.646)		**(2.688)**		**(2.661)**
Female*Rival Ready	**-1.035**	**Female*Rival Ready**	**-9.924**	**Female*Rival Ready**	**-4.681***
	(2.655)		**(7.930)**		**(2.426)**
Social Science	-1.747	Social Science	-1.548	Social Science	-1.800
	(1.356)		(1.390)		(1.368)
Humanities	-1.018	Humanities	-0.752	Humanities	-1.138
	(1.413)		(1.433)		(1.424)
Applied Science	-1.737	Applied Science	-1.499	Applied Science	-1.707
	(1.874)		(1.891)		(1.882)
Natural Science	-0.568	Natural Science	-0.310	Natural Science	-0.526
	(1.785)		(1.796)		(1.815)
Experience Competing	0.00595	Experience Competing	0.177	Experience Competing	0.175
	(0.706)		(0.701)		(0.704)
Ability Competing	**0.346**	**Ability Competing**	**0.162**	**Ability Competing**	**0.405***
	(0.241)		**(0.355)**		**(0.238)**
Favors Opposite Gender	**0.330**	**Favors Opposite Gender**	**-0.523**	**Favors Opposite Gender**	**-0.607**
	(1.095)		**(0.646)**		**(0.646)**
Confidence Ranks	**0.407*****	**Confidence Ranks**	**0.401*****	**Confidence Ranks**	**0.497*****
	(0.0577)		**(0.0575)**		**(0.0910)**
Constant	4.164**	Constant	4.947**	Constant	3.918*
	(2.011)		(2.253)		(2.021)
Observations	634	Observations	634	Observations	634
R-squared	0.136	R-squared	0.128	R-squared	0.127

Notes: the dependent variable is the *Improvement* from piece-rate to tournament, defined as the difference between the number of correct answers in the tournament and the number of correct answers in the piece-rate. The rest of the variables are dummy variables that take a value of 1 for each treatment. For the controls see the notes to Table 4. Robust standard errors are in parentheses. * p<0.1; **; ** p<0.05; *** p<0.01.

4. Discussion

Our paper helps organize the sometimes conflicting evidence on the underperformance of women with respect to men in competitive situations. Our findings imply that female underperformance is more likely to appear when pre-existing differences in perceptions are exacerbated, for example with the provision of information. In particular, we observe gender differences only when the task used is already perceived as favoring men and when the presence of the rival is strongly primed. Indeed, it is precisely in environments in which the presence of rivals is most prominent (mixed gender head-to-head competitions) and in tasks that are perceived to be male (spatial ability and running) where underperformance by women has proven to be greatest, such as in Gneezy et al. (2003) and Gneezy and Rustichini (2004).

The rich design of our experiment enables us to confirm the extreme sensitivity of the results to informational manipulation. Since it is implausible to study all possible subtle manipulations which may create a self-fulfilling stereotype threat, policy recommendations should be cautious. Based on our evidence, it seems that omitting or emphasizing gender information can weaken or reinforce previous perceptions about gender differences in tasks as well as perceptions of competitive abilities, and therefore affect performance. It should also be considered how (possibly false) perceptions about each gender's skills at particular tasks change over time. In particular, the effect of correcting false preconceptions about women's relatively lower ability at jobs traditionally considered as male should be studied, when such perceptions are in fact not true.

5. References

Brandts, J. Gorenert, V., and Rott, C. (2015), "The Impact of Advice on Women's and Men's Selection into Competition," *Management Science* 61(5), 1019-1035.

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. (2012), "Men Too Sometimes Shy Away from Competition: The Case of Team Competition," *Management Science* 58(11), 1982-2000.

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.

Freeman, R.B., and Gelber, A.M. (2010), “Prize Structure and Information in Tournaments: Experimental Evidence,” *American Economic Journal: Applied Economics* 2(1): 149–164.

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

Geraldes, D., Riedl, A., and M. Strobel, (2011), “Sex and Performance under Competition: Is There a Stereotype Threat Shadow?,” mimeo.

Gill, D. and Prowse, V. “Gender Differences and Dynamics in Competition: The Role of Luck,” (2014), *Quantitative Economics* 5(2), 351-376.

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.

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.

Hoffman, M., Gneezy, U., and List, J. A. (2011), “Nurture Affects Gender Differences in Spatial Abilities,” *PNAS* 108(306), 14786-14788.

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.

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., forthcoming, “Gender” *Handbook of Experimental Economics*, second edition, Eds. John Kagel and Alvin E. Roth.

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.

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

Shurchkov, O. (2012), “Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints,” *Journal of the European Economic Association*, 10(5): pp. 1189–1213.

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.

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

Wieland, A., and Sarin, R. (2012), “Domain Specificity of Sex Differences in Competition,” Journal of Economic Behavior and Organization 83(1), pp. 151–157

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 is a translation of the instructions for the experiment (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 parentheses.

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 around is allowed. If you have any questions or need help please raise your hand and one of the researchers will assist you. If you do not follow the rules indicated, WE WILL ASK YOU TO LEAVE THE EXPERIMENT AND YOU WILL NOT RECEIVE ANY PAYMENT. Thank you.

Both the Pompeu Fabra and Autònoma de Barcelona universities have provided funds for use in this experiment. You will receive 3 euros for arriving 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 see when you arrived, your number has been assigned randomly. Participants will not be able to identify each other by their decisions or 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 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 whether you will be paid for task 1 or task 2 of the experiment, and for task 3 or task 4.

Press OK to continue with the 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 taking part + 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 taking part + your earnings in task 2 + your earnings in task 4.

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

Press OK to bring up the instructions for Task 1 of the experiment.

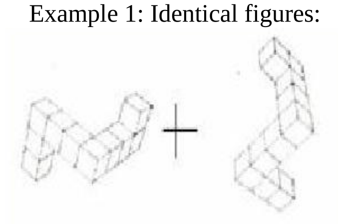
SCREEN 3

Task 1 Instructions

In task 1 of the experiment, you will see two geometric figures next to each other. These figures may either be "identical" or "mirror images". Your task is to indicate which is the case for each pair of figures.

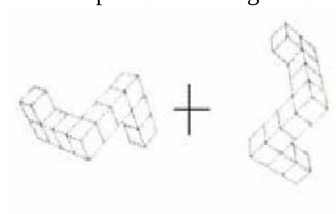
1. Identical: The two geometric figures are the same, although one of them may be rotated a certain number of degrees with respect to an axis, i.e., if rotated one figure is rotated we would get the other one.

Example 1: Identical figures:



2. Mirror images: The two geometric figures are different and, in fact, if one of them is rotated a reflection of the other would be obtained, i.e. if one of the figures is rotated two identical figures will never be obtained because one would be the reflection of the other.

Example 2: Mirror figures:



SCREEN 4

Over the next 4 minutes the computer will show you pairs of figures and your task will be to identify whether they are identical or mirror images. 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 be paid 15 euro cents for each correct answer.

Press OK to start Task 1 of the experiment.

SCREEN 5

Task 2 Instructions

In task 2 you will be given some codes. In each code letters of the alphabet correspond to the numbers from 1 to 9. Your task is to decode sequences of letters, i.e. to associate numbers with letters following the code provided.

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 that we give you is TWK
The correct answer is 469

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

Press OK to start Task 2.

SCREEN 6

Task 3 Instructions
In Task 3 you must do the same as in Task 1, i.e. for 4 minutes the computer will show you pairs of figures and you must identify whether they are identical or mirror images.

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

In this task you are matched with another participant in the experiment, based on the number of correct answers in Task 1 of the experiment. The computer will match participants from the highest to the lowest number of correct answers in Task 1.

Task 1
1 Participant with the highest number of figures correctly identified
2
3
4
5
6

…
18
19
20 Participant with the lowest number of figures correctly identified

Using this order, the computer will match participants as follows: The first with the second, the third with the fourth, the fifth with the sixth and similarly until the nineteenth participant is matched with the twentieth. You will not be told your position in the ranking, i.e., you will not know whether you are the 1st, 2nd … or 20th but, using this matching mechanism, you can be sure that the participant matched with you gave a similar number of correct answers to you in Task 1.

When Task 3 is completed the computer will compare your correct answers in Task 3 with the 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 that you will be paid for Task 3 you will earn double what you earned in Task 1, i.e. 30 euro cents per correct answer, provided that you give more correct answers than your matched participant.
- You will earn nothing if you give fewer correct answers than your matched participant.
- In the case of a tie, each participant will earn 15 euro cents per correct answer.

Press OK to start Task 3 of the experiment.

(In the control treatment the message shown was "Are you ready?"

In "Ability Differences" the following message appeared: "You provided XX more/fewer correct answers than your matched participant in Task 1. Are you ready?"

In "Rival's Gender" the following message appeared: "Your matched participant is female/male. Are you ready? "

In "Ability Differences and Rival's Gender" both these messages were shown.

In "Own Ability Assessment" the following message appeared: "Of the 20 subjects in this session, you were among the best/worst 10 in performing Task 1. Are you ready?"

In "Own Gender" the following message appeared before pairs of figures were shown: "Please indicate your gender for administrative purposes. Are you ready?"

In "Own Ability Assessment and Own Gender" both these messages were shown.

In "Rival Ready" the following message appeared: "Your matched participant is ready. Are you ready?")

SCREEN 7

Task 4 Instructions

In Task 4 you are required to do the same as in Task 2, i.e. for 4 minutes the computer will show you different codes and sequences of letters and you must decode sequences of letters, i.e. associate numbers with letters following the code provided.

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, based on the number of correct answers in Task 2 of the experiment. The computer will match participants from the highest to the lowest 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 as follows. The first with the second, the third with the fourth, the fifth with the sixth and similarly until the nineteenth participant is matched with the twentieth. You will not be told your position in the ranking, i.e. you will not know whether you are the 1st, 2nd … or 20th, but using this matching mechanism you can be sure that the participant matched with you gave a similar number of correct answers to you in Task 2.

When Task 4 is completed, the computer will compare your correct answers in Task 4 with the 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 that 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, provided that you give more correct answers than your matched participant.
- You will earn nothing if you give fewer correct answers than your matched participant.
- In the case of a tie, each participant will earn 15 euro cents per correct answer.

Press OK to start Task 4 of the experiment.

(In the control treatment the message shown was "Are you ready?"

In "Ability Differences" the following message appeared: "You provided XX more/fewer correct answers than your matched participant in Task 2. Are you ready?"

In "Rival's Gender", the following message appeared: "Your matched participant is female/male. Are you ready? "

In "Ability Differences and Rival's Gender" both messages were shown.

In "Own Ability Assessment" the following message appeared: "Of the 20 subjects in this session, you were among the best/worst 10 in performing Task 2. Are you ready?"

In "Own Gender", the following message appeared before pairs of figures were shown: "Please indicate your gender for administrative purposes. Are you ready?"

In "Own Ability Assessment and Own Gender" both messages were shown.

In "Rival Ready" the following message appeared: "Your matched participant is ready. Are you ready?")

SCREENS 8 TO 11

(The following three questions were asked of all participants in all treatments for each of the four tasks once the tasks had all been completed but before they were shown any results. Subjects were paid an extra 10 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 session, what do you think is your ranking in ordering results in Task 1 (2, 3, 4) of the experiment?
- Out of all the participants in the session, who do you think performed Task 1(2, 3, 4) best? Boys/Girls/Equally

SCREEN 12

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

- Who do you think correctly identified more figures in Task 1? Me/ My matched participant in Task 3.
- Do you think you were competing against a man or a women in Task 3?
- Who do you think correctly decoded more sequences of letters in Task 2? Me / My matched participant in Task 4.
- Do you think you were competing against a man or a woman 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. an activity in which you compete against others?
- If so, what is it?
- "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.