Gender and the Willingness to Compete Against Own Past Performance

David Klinowski

Santiago Centre for Experimental Social Sciences
Nuffield College, University of Oxford
University of Santiago de Chile

February 2017

Abstract

We examine whether men and women differ in their willingness to select into competition against their past self. Laboratory participants complete a task and choose between receiving a regular piece rate, and a larger piece rate conditional on improving their past performance. We find with two different analytical strategies that once self-confidence and risk aversion are accounted for, men and women are equally self-competitive. This is true both for a stereotypically male and a stereotypically female task. We find that participants are equally willing to self-compete regardless of ability, and that confidence and self-competitiveness are not affected by gender self-stereotyping. The results provide insight into the feasibility and potential of using self-improvement contracts as gender-neutral incentive mechanisms.

JEL codes: C91, J16, J31, D02

1 Introduction

Across the world there is a persistent gender gap in labor market outcomes. On average women earn less than men for comparable jobs, and hold fewer leadership positions in business and other areas (Bertrand, 2011). One possible explanation for this disparity is that women dislike competition relative to men, which may make them more reluctant to enter high-paying careers and to seek promotions and pay raises (Niederle and Vesterlund, 2011). Indeed a literature starting with Niederle and Vesterlund (2007) finds that, when performing tasks for pay, women are less willing than men to enter tournaments against other individuals, preferring instead a piece rate dependent on individual performance alone.[1] More recent work finds that such unwillingness to compete bears on actual labor market outcomes, as it predicts career choices and earnings differentials (Buser et al., 2014; Reuben et al., 2015; Buser et al., 2017).

As much as the labor market rewards those who outperform the rest, it often also rewards individuals who better themselves, independently of others' performance. The drive to compete with oneself—to challenge oneself to do things better than before—is regarded as a common trait among business top performers, and a trait that managers ought to nurture in their employees in order to spur professional growth (Harvard Business Review, 2016). If men and women differ in their willingness to self-compete, this difference may constitute an additional reason for the gender gap we see in the labor market. But if instead we find that women are equally eager as men to select into challenges that pay upon self-improvement, we may be able to use this knowledge to design policies and incentives that help to close the gender gap while also promoting productivity growth.

This paper investigates whether men and women differ in their willingness to compete with the self, and if so, what accounts for the difference. We conduct a laboratory experiment where participants must choose, prior to performing a task, between a contract that pays a regular piece rate regardless of performance, and a contract that pays twice the piece rate provided performance improves relative to a previous round, and zero otherwise. We compare the rate at which men and women choose the self-improvement contract. We measure the participants' abilities at the task, self-confidence in their ability to improve, risk and ambiguity preferences, and perceptions about which gender dominates at the task, to explore whether these play a role in the decision to self-compete and whether these can explain any gender difference in competitiveness. As detailed in Section 2, one feature of our design is that

[1] See for instance Cason et al. (2010); Healy and Pate (2011); Kamas and Preston (2012); Niederle et al. (2013); Sutter and Rützler (2014); Wozniak et al. (2014); Petrie and Segal (2015); Saccardo et al. (2017). A comprehensive review is given by Niederle (2016).

we create an individually-calibrated risk preference elicitation that approximately simulates the risk and stakes involved in the decision to self-compete, but that removes the element of competition with one's past performance. By comparing the participants' willingness to accept the self-improvement contract to their willingness to accept the similar, context-free bet, we can classify participants as self-consistent, under-competitive, or over-competitive, and examine whether men and women differ in their distribution across types. Another feature of the design is that we employ two tasks that vary in the gender stereotypes associated to them (as confirmed by the participants' own perceptions about which gender performs better at the task), which allows us to investigate whether a gender gap in self-competitiveness varies with the task involved and whether gender self-stereotyping influences the decision to self-compete.[2]

We find that women are less willing than men to self-compete, primarily on the female-typed task. However, once we account for the factors measured—particularly risk aversion and to a lesser extent self-confidence—the gender difference disappears statistically. The competitiveness classification exercise also suggests no gender differences in competitiveness: the distribution of self-consistent, under-competitive, or over-competitive types is similar for men and women. An additional finding, relevant if one seeks to design incentives that leverage the willingness to self-compete, is that participants are equally competitive regardless of their level of ability, thus potentially they are all amenable to policy interventions. Finally, we find no evidence on self-stereotyping in the decision to self-compete.

Closest to our paper is Apicella et al. (2017), who in an experiment conducted concurrently to ours also find no gender differences in self-competition. We contribute to the line of work Apicella et al. started in several ways. Our individually-calibrated risk elicitation provides an alternative demonstration that the gender gap in self-competitiveness is largely accounted for by risk tolerance and self-confidence rather than by gender per se. We show that this is true both for a stereotypically male and a stereotypically female task. And we provide evidence that neither gender engages in self-stereotyping in a way that decreases their confidence

[2] In the context of competition with other individuals, Shurchkov (2012), Dreber et al. (2014), and Grosse et al. (2014) find that women are equally, if not more, competitive than men when the task involved is stereotypically female. As suggested by Baldiga Coffman (2014) and Bordalo et al. (2016), this task-dependency is consistent with a mechanism of self-stereotyping, whereby an individual's confidence in her ability decreases in areas that are stereotypically outside of her gender's domain. In the case of our experiment, self-stereotyping could imply that men are less willing to compete with the self in female-typed tasks, and that women are less willing to compete with the self in male-typed tasks. As explained in detail in Section 2, we examine the task dependency of the results and the possibility of self-stereotyping by using a male-typed task and a female-typed task in the experiment, and by measuring whether participants agree with the stereotypes associated to the tasks.

in their ability to improve or their willingness to self-compete. We see these as valuable insights into the feasibility and potential of using self-improvement incentives as a way to build competitive environments that appeal to both men and women.

In the remaining of the paper, Section 2 describes the experimental design, Section 3 presents the results, and Section 4 concludes with a discussion.

2 Design

The experiment consists of three parts, plus an additional elicitation procedure and a questionnaire. Two treatments are randomly assigned in a between-subjects design, with all subjects in a given session receiving the same treatment. Treatments vary only in the real-effort task the subjects are asked to perform. One is a math task, consisting in adding sets of 5 2-digit numbers as in Niederle and Vesterlund (2007); the other is a verbal task, consisting in forming anagrams by combining 6 given letters. Below we describe a session with the math task; all elements described below except the task itself apply to a session with the verbal task.

In Part 1 of the experiment, participants get 5 minutes to add sets of 5 randomly generated 2-digit numbers, and earn 40¢ per problem solved correctly. After the 5 minutes pass, each participant receives feedback on the number of problems she solved correctly and her corresponding earnings. Before beginning Part 1, participants get 90 seconds of unincentivized practice with the task.

In Part 2, participants again get 5 minutes to add sets of 5 2-digit numbers. The numbers in Part 2 are different from those presented in Part 1, and participants do not learn what Part 2 is about until they finish Part 1. Before the 5 minutes begin in Part 2, each participant must choose one of the following two options for generating earnings for this part.

o Option A: receive 40¢ point per problem solved correctly, regardless of the number of problems solved correctly.

o Option B: receive 80¢ per problem solved correctly, provided the participant solves at least one more problem than the number solved in Part 1. Otherwise the participant earns nothing in Part 2.

After making her choice of pay scheme but before performing the task, the participant is asked to state how many problems she thinks she will solve correctly in Part 2, and the probability (from 0% to 100% in integer values) with which she thinks she will solve correctly at least one more problem in Part 2 than she did in Part 1. To avoid strategic interactions with performance in Part 2, answers to these questions are unincentivized. Answers to these questions serve as measures of the participant's confidence in her ability to improve. After responding to these questions, the participant performs the task for 5 minutes, and receives feedback on the number of problems solved correctly and her corresponding earnings for Part 2.

The self-improvement contract (Option B) in Part 2 involves the risk of failing to improve and earning nothing; therefore risk preferences potentially influence the decision to self-compete. This in turn implies that if women are more risk averse than men, as the literature has found (Croson and Gneezy, 2009; Charness and Gneezy, 2012), then women may be less willing to choose the self-improvement contract because of higher risk aversion, rather than because of lower self-competitive drive per se. To explore this, Part 3 elicits risk preferences in an environment as close as possible to the one faced in the decision to self-compete. It presents the participant with similar risks and stakes to those faced in Part 2, but removes the element of competition against one's own past performance.

In Part 3 the participant does not perform the task again. Instead, she is faced with the following two options.

o Receive a sure payment of X.

o Make a bet that pays Y with probability p and pays nothing with probability $1 - p$.

X and Y are calibrated for each participant. X is the amount of money earned in Part 1, and Y is the amount of money that would be earned in Part 2 under the self-improvement contract if the participant solved correctly one more problem in Part 2 than in Part 1. With this calibration, the sure payment of X approximates Option A in Part 2, and the bet that pays either Y or 0 approximates the potential payoffs under the self-improvement contract.[3]

In Part 3 the participant must indicate the minimum probability p for which she prefers the bet over the sure payment of X. The probability reported must be a percentage integer between 0 and 100. To incentivize Part 3, after the participant reports p, the experimenter randomly draws a probability value between 0 and 100, which becomes the actual probability with which the bet pays Y. If this value is below the reported p, the participant receives X; otherwise the participant makes the bet and receives either Y or 0 depending on the outcome of the bet. Just as in Part 2, where the participant always learns whether she improved her score relative to Part 1 regardless of her choice of pay scheme, in Part 3 the participant always receives feedback on the outcome of the bet regardless of whether she makes the bet.

Since the probability of improving one's score is not objective, the participant's willingness to choose the self-improvement contract in Part 2 may be affected by her ambiguity preferences. We conduct an additional elicitation after Part 3 to measure ambiguity preferences. The

[3]The similarity between Option A in Part 2 and the sure payment of X in Part 3 assumes that in Part 2 the participant is certain that she can replicate her Part 1 score. In Section 3 we examine this assumption and provide a robustness check of the results.

participant is presented with a multiple price list between a risky option and an ambiguous option. The risky option is a 50% chance of receiving \$4 and a 50% chance of receiving nothing. The ambiguous option is an undisclosed chance (between 0 and 100 percent) of receiving Z and the remaining chance of receiving nothing. The price list has 21 items that keep the risky option fixed and vary Z from \$2.35 to \$7.00 in equal increments of approximately 23¢. The items in the list are arranged sequentially in increasing order of Z.

This elicitation is implemented with two physical jars, each containing 100 marbles. The risky jar contains 50 red marbles and 50 black marbles, and the content is visible to participants. The ambiguous jar contains an undisclosed number of red and black marbles; the number of marbles of each color can be anywhere between 1 and 100, and they add to 100. The content of the ambiguous jar is covered from the participants' view. Participants are asked to select the color for which they want to bet, and the item in the list (item 1 to 21) at which they prefer to start drawing a marble from the ambiguous jar rather than the risky jar. The experimenter then selects at random an item from the list, and draws a marble from the jar preferred by the participant for that item based on her reported switch point. If the color drawn matches the color selected by the participant, the participant receives either \$4 or Z, depending on whether the marble was drawn from the risky jar or the ambiguous jar; if the color does not match the one selected by the participant, the participant receives nothing.

The session concludes with a questionnaire that asks whether the participant prefers to receive \$4 with certainty or to make a bet that pays \$8.80 with 50% probability and \$0 with 50% probability. It also asks whether the participant thinks men or women on average score higher at the task (7 answer options in Likert format of the form "women score much/somewhat/slightly higher than men" and with the middle option being "no gender difference"). The final items of the questionnaire ask for the participant's student status, major, occupation, year of birth, and gender. Answers to the questionnaire are unincentivized. Responses to the first question serve as an alternative measure of risk preferences that presents the same stakes for all participants. Responses to the second question help to examine whether perceptions about the relative ability of men and women at the task interact with the participant's own gender to influence the decision to self-compete (namely, to examine whether participants engage in gender self-stereotyping in a way that affects their self-confidence and their willingness to self-compete).

One Part (1, 2, or 3) was randomly selected for payment at the end of the session. Average total earnings were \$9.75, which included a show-up fee of \$3.15 and the earnings from the ambiguity elicitation. Sixteen sessions were conducted at the Santiago Centre for Experimental Social Sciences from November 2016 to January 2017. Each session lasted approximately 45 minutes. The experiment was programmed in oTree (Chen et al., 2017).

3 Results

3.1 Descriptive statistics

Table 1 shows descriptive statistics of the participants. Men and women were similar in age and in the proportion who were students. Men were almost twice as likely as women to be a science or math major, which mirrors roughly the relative participation of men and women in STEM fields in Chilean universities.[4] Men scored higher than women in the math task, and equally in the verbal task. Women were more risk averse than men, as measured by the proportion who selected the safe payoff over the risky bet in the questionnaire. There were no gender differences in ambiguity preferences. Men expected larger score improvements for Part 2 than women in both tasks, although the differences are not statistically significant. Men and women reported similar believed probability that they would improve their score in Part 2. Both men and women believed that on average men scored higher than women on the math task, and that women scored higher than men on the verbal task. Finally, a similar proportion of men and women improved their score from Part 1 to Part 2, with improvement being more frequent in the math task.[5] In the analysis we control for being a STEM major, as it may play a role in the decision to self-compete, and significantly more men than women were STEM majors.

3.2 Baseline willingness to self-compete

In total, 56% of participants chose the self-improvement contract in the math task and 27% in the verbal task (difference p-value < 0.001).[6] Figure 1 shows men's and women's selection rates separately. In the math task, 62.3% of the men and 48.8% of the women chose to self-compete (difference p-value $= 0.187$), while in the verbal task 37.2% of the men and 18.4% of the women chose to do so (difference p-value $= 0.043$).

[4]25% of students enrolled in STEM fields in Chilean undergraduate programs in 2013 were women (Arias Rojas, 2016).

[5]The six-letter combination shown in Part 2 turned out to be more challenging than the one shown in Part 1. This made score improvement less likely for the verbal task. This feature, which was not intended in the design, could not have affected the choice of pay scheme, as participants chose the scheme before learning the combination of letters for Part 2.

[6]A possible explanation for why fewer participants selected into competition in the verbal task than in the math task is that forming anagrams was less familiar an exercise than adding numbers. This may have made it harder for participants to estimate their ability to improve in the verbal task, which may have led them to decrease their estimated chance of improvement. This is consistent with the fact that the mean estimated probability of improvement was 67% for the math task and 58% for the verbal task (p-value $= 0.004$).

Table 1: Descriptive statistics

	Men	Women	p-value
Year of birth	1993	1993	0.172
Proportion student	0.98	0.99	0.586
Proportion science/math major	0.47	0.25	0.002
Part 1 score			
math	9.6	8.3	0.066
verbal	15.8	15.0	0.447
Part 2 score			
math	10.2	8.5	0.009
verbal	13.5	13.0	0.561
Proportion risky choice	0.58	0.40	0.013
Ambiguity tolerance	10.6	11.1	0.390
Expected score improvement			
math	1.21	0.88	0.323
verbal	0.47	-0.39	0.131
Confidence in improving score			
math	0.68	0.66	0.771
verbal	0.58	0.59	0.903
Proportion who improve score			
math	0.55	0.42	0.210
verbal	0.19	0.24	0.495
Perceived task femaleness (-3 to 3)			
math	-0.23	-0.14	0.530
verbal	0.86	0.80	0.746
N			
math	53	43	-
verbal	43	49	-

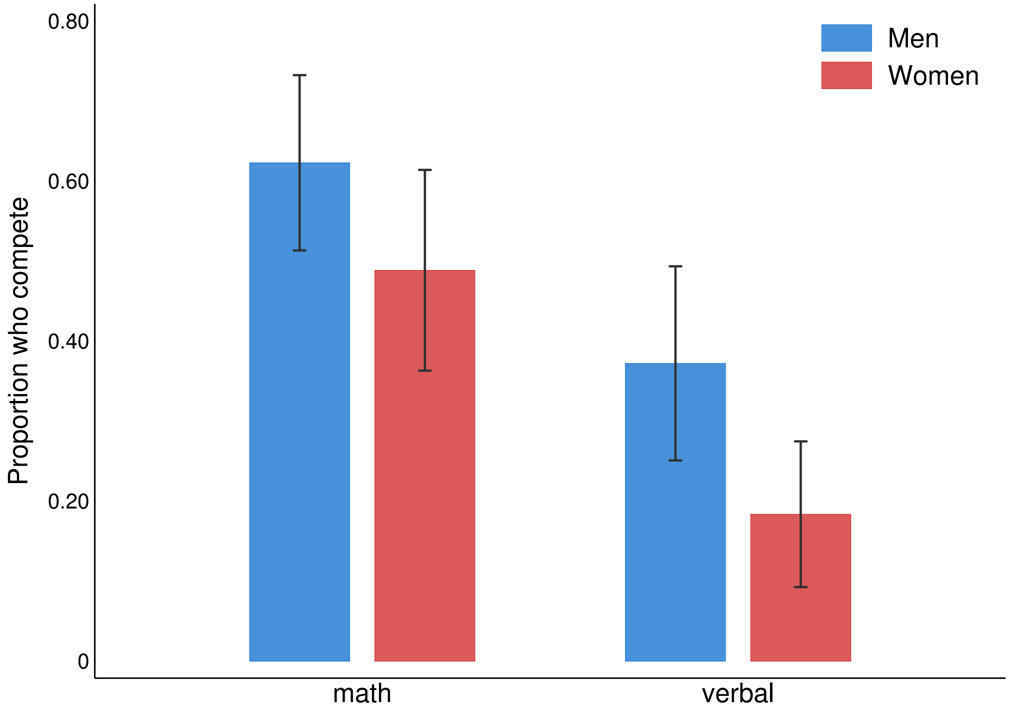


Figure 1: Baseline wilingness to self-compete

Table 2 examines this gender gap with regression analysis. It shows marginal effects from probit regressions estimating the probability of choosing the self-improvement contract under 7 different specifications. The upper panel shows results for the math task and the lower panel for the verbal task. The value shown on the female variable is the estimated difference in the probability to self-compete between women and men, in percentage points. All specifications control for whether the participant is a STEM major. Column (1) presents the baseline model, with a female indicator as the only regressor. This baseline estimate indicates that women were 12 percentage points less likely than men to self-compete in the math task, and 19 percentage points in the verbal task. The gender gap is statistically significantly different from zero only for the verbal task.

The results from this baseline specification indicate that women were less willing than men to choose the self-improvement contract, directionally so in the math task and statistically so in the verbal task. Next we examine whether this gap can be at least partially explained by differences in ability, self-confidence, risk and ambiguity preferences, and self-stereotyping.

3.3 Mechanisms: ability

It is reasonable to conjecture both that a person's choice of contract is independent of ability, or that it is systematically related to ability. On the one hand, participants may have exerted full effort in Part 1 regardless of their level of ability, and may have been equally willing to challenge themselves to do better in Part 2. On the other hand, those who obtained relatively low scores in Part 1 may have believed it easier to improve in Part 2 and therefore may have been more willing to self-compete; or, alternatively, those who obtained relatively high scores in Part 1 may have been more motivated by the prospect of improving and thus may have been more willing to self-compete. Here we examine whether ability predicts selection into the self-improvement pay scheme, and whether it can help to explain the gender gap in self-competitiveness.

Figure A1 in the Appendix shows the cumulative distribution of men's and women's Part 1 scores in each task. Men tended to score higher than women in the math task, and equally to women in the verbal task. This is also seen in Table 1, which shows that on average men scored 1.3 more points than women in the math task (p-value = 0.066) and 0.8 more points in the verbal task (p-value = 0.447). To explore a relationship between ability on the task and selection into competition, Figure A2 in the Appendix shows the proportion of men and women who chose to self-compete given their level of ability.[7] It shows that for both men and women, the proportion of individuals who chose to compete did not differ statistically by ability level, suggesting that both low and high ability individuals had similar drive to self-compete. For the math task, the highest-tercile performing women appear more willing to compete than the rest of the women, although neither this difference nor the difference between men and women at any level of ability are statistically significant.

Regression analysis confirms that ability does not drive the decision to self-compete and does not explain any of the gender gap in competitiveness. Column (2) of Table 2 includes Part 1 score as a predictor of the probability of choosing to self-compete. For both the math and verbal tasks, the effect of the Part 1 score is indistinguishable from 0, and the estimated female effect remains largely unchanged relative to the baseline specification.

3.4 Mechanisms: self-confidence

If participants decide to self-compete only when they are confident enough that they can improve, then a difference in self-confidence between men and women may help to explain the gender gap in competitiveness. Here we explore this possibility by looking at the two

[7] Part 1 scores ranged from 0 to 22 in the math task and from 0 to 29 in the verbal task. To aid exposition, Figure A2 aggregates Part 1 scores into three bins, each containing about a third of the observations.

Table 2: Probit regression estimates of the probability to self-compete (marginal effects)

	baseline	ability	confidence	risk	ambiguity	stereotype	all
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			___ *I. Math* ___				
female	-0.122	-0.123	-0.126	-0.087	-0.133	-0.135	-0.076
	(0.107)	(0.108)	(0.080)	(0.099)	(0.103)	(0.102)	(0.077)
Part 1 score	-	-0.001	-	-	-	-	0.013
		(0.015)					(0.013)
expected score improvement	-	-	0.298***	-	-	-	0.267***
			(0.027)				(0.029)
chose risky option	-	-	-	0.316***	-	-	0.188**
				(0.098)			(0.083)
ambiguity aversion	-	-	-	-	-0.003	-	0.005
					(0.014)		(0.010)
agrees with stereotype	-	-	-	-	-	-0.009	-0.102
						(0.116)	(0.085)
female x stereotype	-	-	-	-	-	0.117	0.011
						(0.236)	(0.171)
pseudo R^2	0.013	0.013	0.411	0.088	0.014	0.015	0.480
N	96	96	96	96	96	96	96
			___ *II. Verbal* ___				
female	-0.191**	-0.197**	-0.153*	0.153	-0.187**	-0.184**	-0.127
	(0.093)	(0.093)	(0.089)	(0.093)	(0.092)	(0.094)	(0.093)
Part 1 score	-	-0.006	-	-	-	-	-0.006
		(0.009)					(0.009)
expected score improvement	-	-	0.044***	-	-	-	0.046***
			(0.015)				(0.016)
chose risky option	-	-	-	0.159*	-	-	0.167*
				(0.093)			(0.095)
ambiguity aversion	-	-	-	-	-0.001	-	0.000
					(0.012)		(0.011)
agrees with stereotype	-	-	-	-	-	0.032	-0.036
						(0.093)	(0.091)
female x stereotype	-	-	-	-	-	0.068	0.272
						(0.189)	(0.187)
pseudo R^2	0.039	0.044	0.108	0.066	0.039	0.042	0.150
N	92	92	92	92	92	92	92

Table 3: OLS regression estimates of the self-confidence measures (marginal effects)

	expected Part 2 score		expected prob. of improving	
	math	verbal	math	verbal
female	-0.319	-0.660	1.533	1.769
	(0.382)	(0.596)	(5.049)	(4.551)
Part 1 score	0.940***	0.878***	0.355	-0.268
	(0.050)	(0.056)	(0.659)	(0.424)
R^2	0.838	0.778	0.107	0.120
N	96	92	96	92

measures of self-confidence we collected: the expected score in Part 2, and the believed probability of improving in Part 2. In the taxonomy of Moore and Healy (2008), these are measures of overestimation of one's ability, rather than overplacement or overprecision.

Table 1 shows that on average men expected to obtain slightly more points above their Part 1 score than women did, especially for the verbal task: men and women expected to improve, respectively, 1.21 and 0.88 points in the math task, and 0.47 and -0.39 (i.e., a decline) in the verbal task. These gender differences are not statistically significant. Men and women also reported similar probabilities of improving their scores: 68% and 66% in the math task, and 58% and 59% in the verbal task.

Table 3 provides more informative estimates of the gender differences in self-confidence than do averages alone. It reports OLS estimates of the two self-confidence measures when regressed on a female indicator and on the participant's Part 1 score (controlling for the participant's major). As with the averages, there is no statistical gender difference in the expected score improvement or in the believed probability of improving. Men and women were similarly confident in their ability to improve.[8]

[8] Men and women were also similarly accurate in their beliefs about their ability to improve. In the math(verbal) task, 51(79)% of men and 49(63)% of women expected to obtain a score higher than they actually achieved. Figure A3 shows the entire cumulative distributions of the difference between expected and actual score in Part 2, which are statistically similar for men and women (Kolgomorov-Smirnov test of equality of the empirical distribution functions p-value = 1.000 for the math task and 0.219 for the verbal task). In terms of the believed probability of improving, in the math(verbal) task 5(2)% of men and 4(2)% of women stated a probability of improving smaller than 50% but actually improved their score (i.e., were underconfident), while 25(47)% of men and 30(37)% of women stated a probability of improving larger than 50% but actually failed to improve their score (i.e., were overconfident). Figure A4 plots these distributions, which did not differ statistically between men and women for the math task (chi-squared p-value = 0.787) or the verbal task (chi-squared p-value = 0.622).

To examine the role of self-confidence in the decision to self-compete, Column (3) of Table 2 presents marginal effects from a probit model estimating the probability to compete on a female indicator and on the participant's expected score in Part 2 minus her score in Part 1 (namely, the expected score improvement). The results indicate that more confident participants were more likely to self-compete. An increase in the expected score improvement by one point is associated with an increase of 29.3 percentage points in the likelihood of competing in the math task, and with an increase of 4.4 percentage points in the likelihood of competing in the verbal task. Both effects are statistically significant. This positive effect also suggests that participants responded to this confidence elicitation seriously, despite it not being monetarily incentivized.[9]

To determine whether self-confidence helps to explain the gender gap in competitiveness, we compare the baseline gap estimate to the estimate in Column (3) of Table 2. The gap is practically unchanged for the math task, going from -0.122 to -0.126, but is somewhat decreased for the verbal task, going from -0.191 to -0.153 (a 19.67 percent decrease). Self-confidence therefore significantly predicts the decision to self-compete, and plays a small but existent role in explaining the gap in competitiveness.

3.5 Mechanisms: risk preferences

The decision to self-compete involves the risk of failing to improve and earning nothing. If women are more risk averse than men, they may be less willing to self-compete on account of their risk preferences rather than their competitiveness per se. We use two measures of risk attitudes to examine possible gender differences in risk preferences and their role in explaining the decision to self-compete: a binary choice between a risky payoff and a safe payoff that is the same for all participants, and an incentive-compatible elicitation that is individually calibrated to the payoffs each participant faced in the decision to self-compete. We examine each measure in turn.

Table 1 shows that 40% of women and 58% of men preferred the risky option over the safe amount in the binary choice elicitation (difference p-value = 0.013), indicating that women were more risk averse than men. To examine the role of risk aversion in the decision to self-compete, Column (4) of Table 2 shows marginal effects from a probit regression estimating the probability to self-compete on a female indicator and an indicator of choosing the risky option in the binary risk elicitation. Choosing the risky option is associated with an increase

[9]The alternative self-confidence measure—the believed probability of improving—does not significantly predict the decision to compete or the gender gap in competitiveness. Therefore, we only report results from using the expected score improvement as the measure of self-confidence.

Table 4: Competitive types formed by comparing willingness to compete to willingness to make similar non-competitive bet

	$\boldsymbol{\beta \geq \delta}$	$\boldsymbol{\beta < \delta}$
competed	consistent choices	over-competitive
did not compete	under-competitive	consistent choices

Notes: β denotes the participant's reported probability of improving her score in Part 2, and δ denotes the participant's elicited minimum probability of winning the bet at which the she prefers the bet over the sure payoff in Part 3.

of 31.6(15.6) percentage points in the probability to compete in the math(verbal) task. As expected given that self-competition is a risky proposition, these results indicate that more risk averse individuals were significantly less willing to self-compete.

Comparing the estimated baseline gender gap in competitiveness to the results in Column (4) of Table 2 gives an estimate of the fraction of the gap that is explained by risk preferences. Accounting for the binary risk elicitation choice reduces the gap from -0.122 to -0.087 (a 28.90 percent reduction) for the math task, and from -0.191 to -0.153 (a 19.76 percent reduction) for the verbal task. The latter residual gap is only marginally significant (p-value = 0.101).

The regression exercise above indicates that risk aversion affects the decision to self-compete, and that once this relation is accounted for, men and women display similar willingness to self-compete. We also examine the gender gap in competitiveness after accounting for risk preferences by looking at results from the individually-calibrated risk preference elicitation. Here we compare the participants' willingness to self-compete given their believed probability of improving, to their willingness to accept a risky bet that is similarly valued but that does not involve a competitive element. Comparing the two choices for different probabilities of winning the bet allows us to classify participants into three types: under-competitive, consistent, and over-competitive. The categorization is illustrated in Table 4.

The categorization is derived as follows. Let X be the payoff the participant earns for her performance in Part 1, let Y be the payoff the participant would earn in Part 2 if she decides to compete and improves her Part 1 performance by 1 point, and let β be the probability with which she believes she will surpass her Part 1 score. Purely in terms of risk, the choice of whether to compete is approximately a choice between a bet that pays Y with probability β and 0 with probability $1-\beta$, and a sure payment of X.[10] The risk elicitation in Part 3

[10]This assumes that the participant is certain that she can replicate her Part 1 score. In footnote 11 we test this assumption and provide a robustness check of the results.

presents the participant with a choice between these same risky and sure payoffs, absent the competition element, and asks the participant to state the minimum probability of the bet paying off for which she would accept the bet over the sure payoff. Denote the participant's reported probability by δ. In this risk elicitation, the participant reveals that she accepts the bet if it pays off with probability of at least δ. We can contrast this revealed preference to whether she chose to self-compete given β. If she accepted the bet when it pays off with probability of at least δ but did not compete when her believed chance of improving was $\beta \geq \delta$, her choices suggest a distaste for competition net of her risk tolerance. Similarly, if she accepted the bet when it pays off with probability of at least δ but competed when her believed chance of improving was $\beta < \delta$, her choices suggest a taste for competition net of her risk tolerance. In the remaining cases (competing when $\beta \geq \delta$, or not competing when $\beta < \delta$) her choices are consistent with each other.

Looking at how men and women distribute across these three types provides a way to examine the gender differences in competitiveness accounting for risk preferences. Among men, 56% made consistent choices, 16% were over-competitive, and 28% were under-competitive; among women, 55% made consistent choices, 12% were over-competitive, and 30% were under-competitive (chi-squared test of gender difference in distributions p-value = 0.679).[11] The pattern remains if we look at the math and verbal tasks separately. Thus, this exercise reaches a similar conclusion as the regression results: men and women are similarly self-competitive once their risk preferences are accounted for.

3.6 Mechanisms: ambiguity preferences

The belief in one's probability of improving in Part 2 is a subjective probability. The decision to self-compete may therefore be affected by the participant's attitude toward ambiguity. To estimate the gender gap in competitiveness controlling for ambiguity tolerance, Column (5) of Table 2 presents results from a probit regression estimating the probability to self-compete on a female indicator and the decision number, between 1 and 21, at which

[11] As mentioned previously, the equivalence between the choices in Parts 2 and 3 rests on the assumption that in Part 2 the participant believes with certainty that she can replicate her Part 1 score. This assumption is most likely to hold for participants who expected to obtain a score in Part 2 higher than in Part 1: 63(38)% of participants did so for the math(verbal) task. Restricting the analysis to these individuals finds that among men, 64% made consistent choices, 26% were over-competitive, and 10% were under-competitive, while among women 51% made consistent choices, 23% were over-competitive, and 26% were under-competitive. Although among this subset of participants women tended to be under-competitive in larger proportion and self-consistent in smaller proportion than men, the distributions are only marginally statistically different (chi-square test of gender difference in distributions p-value = 0.106).

the participant switched from preferring the risky bet to preferring the ambiguous bet in the ambiguity elicitation task. A larger switch point implies higher ambiguity aversion.

The results show that for both the math and verbal tasks, the marginal effect of ambiguity tolerance is indistinguishable from zero, and that the estimated gender gap in competitiveness does not change relative to baseline. This indicates that ambiguity preferences do not help to predict the decision to self-compete or the gender gap in competitiveness in this experiment.

3.7 Mechanisms: gender self-stereotyping

If participants considered that it is mostly men who excel at the math task or mostly women who excel at the verbal task, the threat of confirming this stereotype may have led a participant whose gender was incongruous with the task to become less confident in her ability to improve and less willing to self-compete. To explore this possibility, we examine first whether there are gender stereotypes associated with the math and verbal tasks in the experiment, and if so, whether the interaction between agreement with the stereotype and the participant's gender affected the level of confidence and the selection into competition.

At the end of the experiment participants reported on a 7-point Likert scale whether they believed men or women on average scored higher on the task. Here we label responses from -3 to 3 in ascending order of perceived "femaleness" (-3/-2/-1 = "men score much/-somewhat/slightly higher than men", 0 = "no gender difference" and 1/2/3 = "women score slightly/somewhat/much higher than men"). The average belief was -0.188 for the math task and 0.826 for the verbal task. Both are significantly different from zero (tests of means p-value = 0.007 and p-value $<$ 0.001 respectively, and tests of medians p-value = 0.001 and p-value $<$ 0.001 respectively). As Table 1 shows, the beliefs did not differ across gender. Therefore on average both male and female participants agreed that the math task was male-dominated and the verbal task was female-dominated.

We now examine whether holding this gender stereotype affected the participants' confidence in their ability to improve, especially when the participant's gender was not dominant at the task. Table 5 shows the estimated score the participant expects to get in Part 2, given her gender and whether she holds the stereotype for the corresponding task (controlling in addition for Part 1 score).[12] The self-stereotyping conjecture is that in the math task, which is perceived as male-dominated, a woman who agrees that "men are better at the math task"

[12]Estimated beliefs come from OLS regressions of the expected Part 2 score on a female indicator, Part 1 score, an indicator of whether the participant agrees with the stereotype, and the interaction of this indicator with the female indicator. The indicator of agreement with the stereotype equals 1 if the participant believes that men(women) score slightly/somewhat/much higher than the other gender in the math(verbal) task, and equals 0 otherwise.

Table 5: Expected Part 2 score conditional on agreement with stereotype

	math		verbal	
	men	women	men	women
agrees with stereotype	9.853 (0.414)	9.537 (0.539)	16.686 (0.494)	14.602 (0.565)
disagrees with stereotype	10.293 (0.264)	10.036 (0.279)	14.556 (0.643)	15.220 (0.511)
p-value	0.368	0.405	0.010	0.433

may be less confident in her ability to improve and less willing to self-compete than a woman who does not agree with the stereotype. Similarly, in the verbal task, which is perceived as female-dominated, a man who agrees that "women are better at the verbal task" may be less confident in his ability to improve and less willing to self-compete than a man who does not agree with the stereotype. Results in Table 5 suggest that neither case is borne out by the data. In the math task, a woman on average expects to score similarly regardless of whether she agrees with the stereotype (9.537 if she does, 10.036 if she does not; difference p-value = 0.405). For the verbal task, a man expects to score 16.686 in Part 2 if he agrees with the stereotype, and 14.556 if he does not. These values are statistically different from each other (p-value = 0.010), but in the direction opposite to the self-stereotyping conjecture. That is, we find that in the verbal task men are more self-confident when they agree that the verbal task is female-dominated than when they do not. The data do not allow us to explore this finding further, although we note that the finding persists when we use as the measure of self-confidence the probability of improving rather than the expected Part 2 score (see Table A1 in the Appendix).

Finally we explore whether agreeing with the gender stereotype affected the participant's willingness to compete. Column (6) of Table 2 shows results from probit regressions estimating the probability to self-compete on a female indicator, an indicator of agreement with the gender stereotype for the task, and their interaction. Neither agreement with the stereotype nor its interaction with the participant's gender significantly predict the decision to self-compete. Moreover, the gender gap in competitiveness in this specification is barely changed relative to baseline. We therefore find no evidence of self-stereotyping in the decision to self-compete: gender stereotypes about the tasks, which do exist in our participants, do not explain the decision to self-compete or the gender gap in competitiveness.

3.8 All mechanisms jointly

We compute more precisely the effects of the different mechanisms by considering a full model that jointly estimates the effect of ability, self-confidence, risk aversion, ambiguity aversion, and gender self-stereotyping. Column (7) of Table 2 presents the results from a probit regression estimating the probability to self-compete on all the mechanisms. Self-confidence and risk aversion continue to be the only measures that have significant explanatory power over the decision to compete. Higher self-confidence and lower risk aversion are associated with an increased likelihood of choosing to compete with oneself, in both the math and verbal tasks. Once these and the other measures are accounted for, the gender gap in competitiveness gets reduced by 37.90 percent (from -0.122 to -0.076) for the math task and by 33.50 percent (from 0.191 to 0.127) for the verbal task. Neither residual gap is significantly different from zero. This suggests that the decision to self-compete is determined at least partly by the participant's risk tolerance and self-confidence, and that once these are accounted for, men and women exhibit no differences in self-competitiveness.[13]

We reach a similar conclusion if we estimate the probability that the participant is classified as under-competitive, consistent, or over-competitive conditional on her gender and level of self-confidence. Recall from Section 3.5 that the classification is constructed by comparing the participant's willingness to self-compete to her willingness to accept a similar, non-competitive bet. Table A3 in the Appendix shows results from a multinomial probit model estimating the probability that the participant falls into each category, given her gender and level of self-confidence. The results show that higher self-confidence is associated with higher probability of being classified as over-competitive: a one-point increase in the expected score improvement is associated with an increase of 17.8 percentage points in the probability of being classified as over-competitive relative to being classified as consistent. This is in line with the full model estimation presented previously, which finds that self-confidence is positively related to self-competitiveness even after accounting for risk preferences.

The multinomial probit model in Table A3 in the Appendix finds no gender difference in the distribution of types, as the marginal effect of the female indicator is not significantly different from zero for either non-reference category. Figure 2 illustrates this by showing the estimated probabilities of being classified as under-competitive, consistent, and over-competitive for men and women separately. Men and women have similar distributions across types. Thus the results from this exercise indicate no gender differences in self-competitiveness

[13] An analysis on the pooled data, rather than for each task separately, obtains similar, if not stronger, results. The analysis of the pooled data interacts each explanatory variable (female indicator and the different mechanisms) with a task dummy in a single regression. See Table A2 in the Appendix for the results.

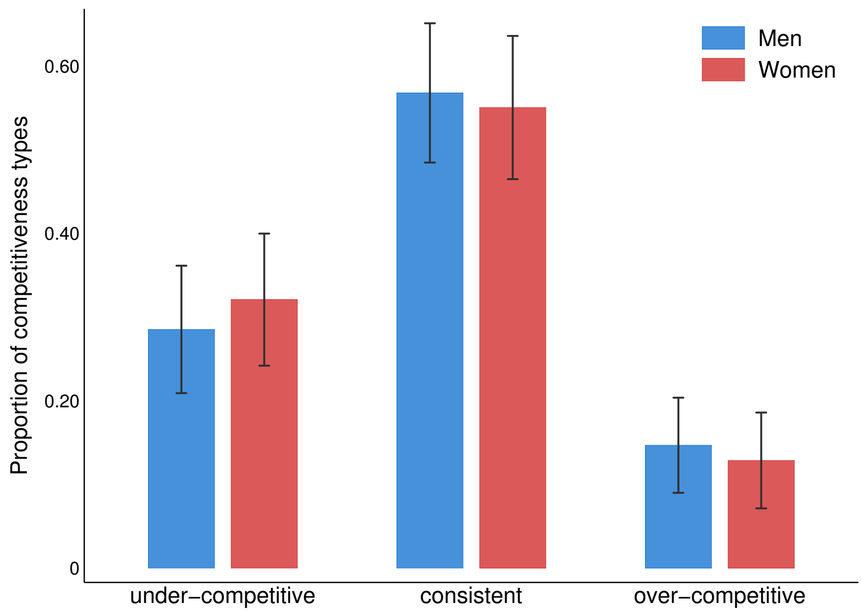


Figure 2: Distribution of competitiveness types

once we account for self-confidence and risk preferences.

4 Discussion

This paper examines whether men and women differ in their willingness to select into competition against their own past performance. We conduct an experiment where participants complete a task and choose between receiving a piece-rate payment regardless of performance, and a piece rate that is twice as large but that pays only if performance improves relative to a previous round. We document the rates at which men and women select into the self-improvement contract, and explore the role of ability, self-confidence, risk and ambiguity preferences, and gender self-stereotyping in explaining the choice of contract and the potential gender difference in that choice. We explore the dependency of the results to using a male-typed task versus a female-typed task.

We find that greater self-confidence and lower risk aversion are significant predictors of selection into self-competition, regardless of the task. We show in two different ways that once self-confidence and risk aversion are accounted for, men and women are statistically equally willing to self-compete. One way we show this is by finding no significant effect of gender in a model that estimates the participant's choice of contract, conditional on the factors measured in the experiment. The other way is by finding that conditional on self-confidence, men and women do not differ in their likelihoods of being classified as under-competitive, over-competitive, or self-consistent, defined by comparing their willingness to select into the self-improvement contract to their willingness to accept a similarly risky, but competition-free, bet.

These findings provide insights into the design of effective ways to attract men and women into competitive environments. A mechanism that leverages the willingness to compete against one's own past performance may appeal to both men and women regardless of ability, the results suggest. Of course the effectiveness of such mechanism would rest on men's and women's productivity responding equally to self-competition, which is something that we cannot test with our design. Doing so would require exogenous assignment (rather than self-selection) into self-competition, and using a task with demonstrated output elasticity.[14] Although some studies find that competition against other individuals improves men's but not women's performance (Gneezy et al., 2003; Günther et al., 2010; Shurchkov, 2012), there is a basis to expect that self-competition can push productivity for both genders: a look at sixteen lab and field experiments by Bandiera et al. (2016) finds that men and women respond equally, and positively, to performance pay. If a similar response is seen for performance pay conditional on self-improvement, then such pay scheme may represent a gender-neutral way of introducing competitiveness into the incentive structures of organizations.

References

Apicella, C. L., Demiral, E. E., and Möllerström, J. (2017). No Gender Differences in Willingness to Compete When Competing against Self. *American Economic Review Papers and Proceedings*, page Forthcoming.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., and Wilson, A. J. (2016). The Slider Task: An Example of Restricted Inference on Incentive Effects. *Journal of the Economic Science Association*, 2(1):1–12.

Arias Rojas, Ó. A. (2016). Brecha de género en matemáticas: el sesgo de las pruebas competitivas (evidencia para Chile). Working Paper.

Baldiga Coffman, K. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *Quarterly Journal of Economics*, 129(4):1625–1660.

[14]See Araujo et al. (2016) for a discussion on the importance of the choice of task when examining incentive effects.

Bandiera, O., Fischer, G., Prat, A., and Ytsma, E. (2016). Do Women Respond Less to Performance Pay? Building Evidence from Multiple Experiments. CEPR Discussion Paper 11724.

Bertrand, M. (2011). New Perspectives on Gender. In *Handbook of Labor Economics*, pages 1543–1590. 4 edition.

Bordalo, P., Baldiga Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics*, page Forthcoming.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, Competitiveness and Career Choices. *Quarterly Journal of Economics*, 129(3):1409–1447.

Buser, T., Peter, N., and Wolter, S. (2017). Gender, Competitiveness and Study Choices: Evidence from Switzerland. *American Economic Review Papers and Proceedings*, page forthcoming.

Cason, T. L., Masters, W. A., and Sheremeta, R. M. (2010). Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study. *Journal of Public Economics*, 94:604–611.

Charness, G. and Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior and Organization*, 83:50–58.

Chen, D. L., Schonger, M., and Wickens, C. (2017). oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments. Working Paper.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and Competition in Adolescence: Tasks Matter. *Experimental Economics*, 17(1):154–172.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, 118(3):1049–1074.

Grosse, N., Riener, G., and Dertwinkel-Kalt, M. (2014). Explaining Gender Differences in Competitiveness: Testing a Theory on Gender-Task Stereotypes. Working Paper.

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M. (2010). Women Can't Jump? An Experiment on Competitive Attitudes and Stereotype Threat. *Journal of Economic Behavior and Organization*, 75(3):395–401.

Harvard Business Review (2016). *HBR Guide to Delivering Effective Feedback*. Harvard Business Review Press.

Healy, A. and Pate, J. (2011). Can Teams Help to Close the Gender Competition Gap? *The Economic Journal*, 121:1192–1204.

Kamas, L. and Preston, A. (2012). Gender and Social Preferences in the US: An Experimental Study. *Feminist Economics*, 18(1):135–160.

Moore, D. A. and Healy, P. J. (2008). The Trouble with Overconfidence. *Psychological Review*, 115(2):502–517.

Niederle, M. (2016). Gender. In Kagel, J. and Roth, A. E., editors, *Handbook of Experimental Economics*, pages 481–553. Princeton University Press, 2 edition.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Vesterlund, L. (2011). Gender and Competition. *Annual Review of Economics*, 3:601–630.

Petrie, R. and Segal, C. (2015). Gender Differences in Competitiveness: The Role of Prizes. Working paper.

Reuben, E., Sapienza, P., and Zingales, L. (2015). Taste for Competition and the Gender Gap among Young Business Professionals. NBER Working Paper 21695.

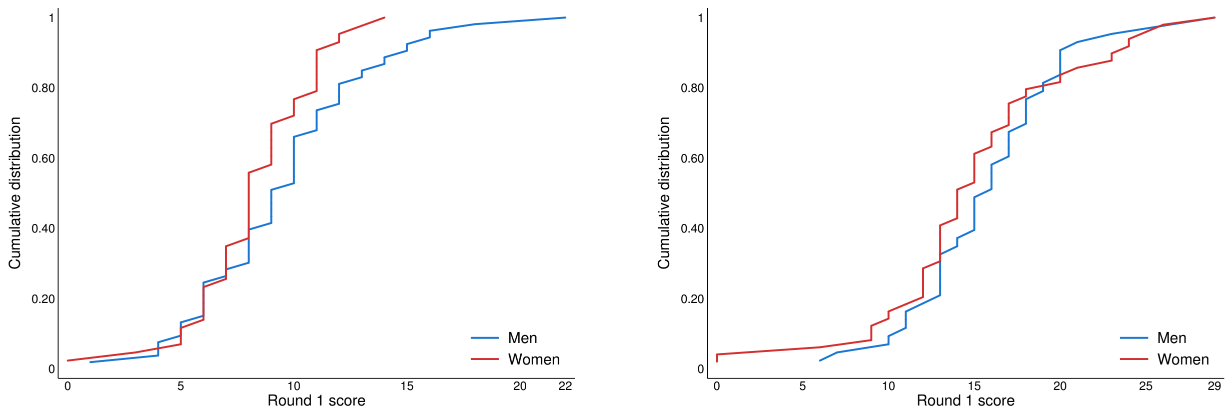
Saccardo, S., Pietraz, A., and Gneezy, U. (2017). On the Size of the Gender Difference in Competitiveness. *Management Science*, page Forthcoming.

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):1189–1213.

Sutter, M. and Rützler, D. (2014). Gender Differences in Competition Emerge Early in Life and Persist. *Management Science*, 61(10):2339–2354.

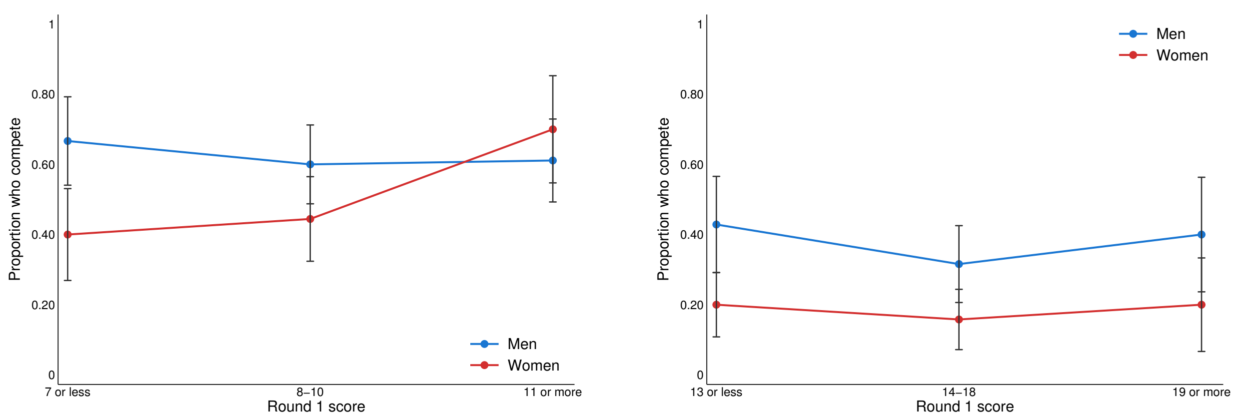
Wozniak, D., Harbaugh, W. T., and Mayr, U. (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. *Journal of Labor Economics*, 32(1):161–198.

A Supplementary Analysis



(a) Math (b) Verbal

Figure A1: Cumulative distribution of Part 1 scores

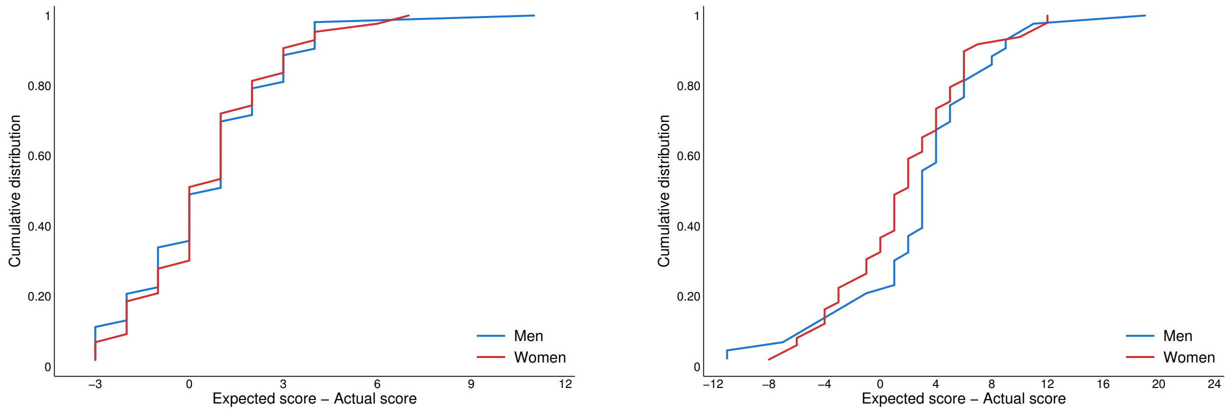


(a) Math (b) Verbal

Figure A2: Proportion of participants who compete given ability

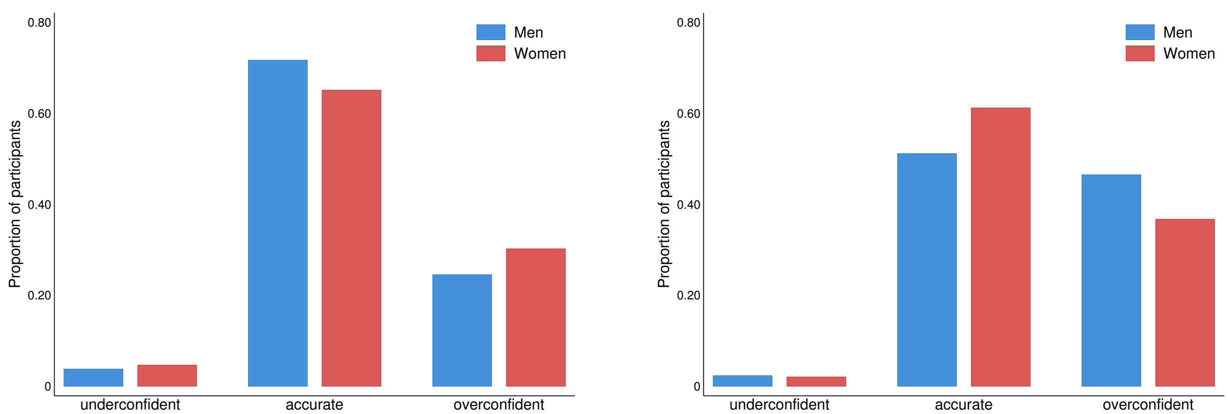
Table A1: Believed probability of improving conditional on agreement with stereotype

	math		verbal	
	men	women	men	women
agree with stereotype	62.431 (5.427)	62.386 (7.057)	62.778 (3.889)	56.560 (4.442)
disagree with stereotype	67.865 (3.463)	69.388 (3.650)	49.280 (5.054)	60.832 (4.019)
p-value	0.397	0.373	0.036	0.490



(a) Math (b) Verbal

Figure A3: Cumulative distribution of difference between expected and actual score in Part 2



(a) Math (b) Verbal

Figure A4: Accuracy of believed probability of improving given actual improvement in Part 2

Table A2: Probit regression estimates of the probability to self-compete, pooled data (marginal effects)

	baseline (1)	ability (2)	confidence (3)	risk (4)	ambiguity (5)	stereotype (6)	all (7)
			I. Math				
female	-0.122	-0.123	-0.096	-0.087	-0.133	-0.146	-0.062
	(0.107)	(0.108)	(0.061)	(0.099)	(0.103)	(0.105)	(0.063)
Part 1 score	-	-0.001	-	-	-	-	0.010
		(0.015)					(0.011)
expected score improvement	-	-	0.226***	-	-	-	0.215***
			(0.014)				(0.019)
chose risky option	-	-	-	0.316***	-	-	0.148**
				(0.098)			(0.066)
ambiguity aversion	-	-	-	-	-0.003	-	0.004
					(0.014)		(0.008)
agrees with stereotype	-	-	-	-	-	-0.004	-0.081
						(0.118)	(0.068)
female x stereotype	-	-	-	-	-	0.117	0.011
						(0.236)	(0.137)
pseudo R^2	0.088	0.090	0.322	0.138	0.088	0.090	0.375
N	188	188	188	188	188	188	188
			II. Verbal				
female	-0.191**	-0.203**	-0.162*	0.154	-0.187**	-0.177*	-0.106
	(0.093)	(0.097)	(0.094)	(0.094)	(0.092)	(0.096)	(0.106)
Part 1 score	-	-0.007	-	-	-	-	-0.006
		(0.009)					(0.010)
expected score improvement	-	-	0.048***	-	-	-	0.051***
			(0.017)				(0.018)
chose risky option	-	-	-	0.160*	-	-	0.185*
				(0.094)			(0.105)
ambiguity aversion	-	-	-	-	-0.001	-	0.000
					(0.012)		(0.012)
agrees with stereotype	-	-	-	-	-	0.029	-0.045
						(0.095)	(0.100)
female x stereotype	-	-	-	-	-	0.068	0.296
						(0.189)	(0.204)
pseudo R^2	0.088	0.090	0.322	0.138	0.088	0.090	0.375
N	188	188	188	188	188	188	188

Table A3: Multinomial probit regression estimates of the probability of being under- or over-competitive (marginal effects relative to being self-consistent)

	under-competitive	over-competitive
female	0.121	-0.067
	(0.274)	(0.076)
expected score improvement	-0.051	0.178**
	(0.063)	(0.076)

B Experiment Instructions [math task session]

Welcome to the experiment, and thank you for coming.

Informed consent

Please read carefully the Informed Consent form. It explains how your rights as a participant are protected. If you prefer not to participate, you are free to leave at any moment. If you accept to participate, please sign the form and leave it on the desk for the experimenter to take it. If you have any question, raise your hand.

Anonymity and risks

Your decisions in this experiment are anonymous. We will not record any personal information from you, and your identity cannot be linked to any decision you make in this experiment. There are no risks associated with participating in this experiment.

Rules of the laboratory

The other people in the room are also participating in the experiment. Please do not talk with them during the session. Please keep your cellphone turned off during the experiment. If at any moment you have any question, raise your hand and the experimenter will come to where you are to answer the question in private.

Payment

You have guaranteed a payment of $3.15 for your participation. In addition, you can significantly increase your payment depending on the decisions you make in the experiment. The experiment consists of 3 parts plus an extra part. At the end of the experiment we will randomly select only one of the three parts to add to your payment. That is, in addition to the $3.15 for your participation, you will received the earnings you generate from only one of the parts (Part 1, Part 2, or Part 3), selected at random, plus the extra part. You will receive your payment in private and in cash before leaving the laboratory.

The experiment

As was already mentioned, the experiment consists of 3 parts plus an extra part. Instructions for each part will be given at the beginning of the corresponding part.

B.1 Part 1

In Part 1 you will have 5 minutes to solve as many math problems as you can. Each problem consists in adding five numbers of two digits each.

For each problem solved correctly you will receive 40¢. At the end of the 5 minutes your answers will be automatically submitted, and you will see confirmation of the number of problems you solved correctly and your corresponding earnings for Part 1. You may use the scratch paper on your station if you want. The use of calculators or other electronic devices is not allowed.

Interface

Each problem shows five numbers of two digits each, one next to the other, as the example below. You must add them and enter your answer on the space provided to the right of the numbers.

58	55	47	91	45	_

Payment of randomly selected part

If at the end of the experiment Part 1 is randomly selected, you will receive as payment you earnings from Part 1 (plus your show-up fee and the earnings from the extra part).

Practice

Before beginning Part 1, you will have 90 seconds of practice to familiarize yourself with the interface. Answers on this practice will not affect your earnings.

B.2 Part 2

In Part 2 you will again have 5 minutes to solve as many math problems as you can. As before, each problem consists in adding five numbers of two digits each. You may use the scratch paper to help you. Calculators or other electronic devices are not allowed.

Choice to generate earnings

You can choose how to generate earnings in Part 2. You have two options:

- Option 1: Receive 40¢ per problem solved correctly, regardless of how many problems you solve correctly.

- Option 2: Receive 80¢ per problem solved correctly, provided you solve at least [Part 1 score plus 1] problems correctly. If you solve fewer than [Part 1 score plus 1] problems correctly, you earn $0.

Payment of randomly selected part

If at the end of the experiment Part 2 is randomly selected, you will receive as payment you earnings from Part 2 (plus your show-up fee and the earnings from the extra part).

B.3 Part 3

Part 3 does not involve adding numbers. Instead, for this part you have two options:

- Receive [Part 1 earnings] with certainty.
- Make a bet that pays [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score] with X% probability and $0 with (100-X)% probability.

X may be any whole number between 0 and 100. A larger X means that the bet pays [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score] with larger probability and $0 with smaller probability.

Your decision

For Part 3 you must indicate the minimum probability X for which you prefer to make the bet for [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score] instead of receiving [Part 1 earnings] with certainty.

There are no correct or incorrect answers. You must simply choose according to your preference.

How your earnings are determined

After you make your decision, the experimenter will choose at random a number between 0 and 100. This will be the actual probability with which the bet will pay [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score].

If the actual probability turns out to be smaller than the minimum probability you chose, you will not make the bet. Instead, you will receive [Part 1 earnings] as your earnings for Part 3.

If the actual probability turns out to be equal to or larger than the minimum probability you chose, you will make the bet. Your earnings for Part 3 will be either [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score] or $0, depending on the outcome of the bet.

How the outcome of the bet is determined

To determine the outcome of the bet, the experimenter will select at random another number, between 1 and 100. If this number turns out to be smaller than or equal to the actual probability, your earnings for Part 3 will be [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score]. If the number turns out to be larger than the actual probability, your earnings for Part 3 will be $0. This procedure guarantees that the bet pays [Part 2 earnings when Option 2 is chosen and score equals one point more than Part 1 score] with its actual probability.

Payment of randomly selected part

If at the end of the experiment Part 3 is randomly selected, you will receive as payment you earnings from Part 3 (plus your show-up fee and the earnings from the extra part).

B.4 Extra Part

Before we select the Part that counts for payment and conclude the experiment, you have an additional opportunity to generate earnings. These earnings will be added to the earnings from the Part that gets selected to count for payment.

Jar A and Jar B

The experimenter has two jars in the front of the room. Each jar contains 100 marbles. Each marble is either Red or Black.

Jar A contains 50 Red marbles and 50 Black marbles.

Jar B contains an unknown number of Red (and Black) marbles; it could be any number between 0 and 100.

Your decisions

You must first choose a color, Red or Black, for which you want to bet.

Then you must choose a jar, Jar A or Jar B, from which you want to draw a marble.

Once you make your decisions, the experimenter will draw a marble blindly (randomly) from the jar of your choice.

Earnings

If you choose to draw from Jar A and a marble of your color is drawn, you receive \$4. If a marble of the other color is drawn, you receive \$0.

If you choose to draw from Jar B and a marble of your color is drawn, you receive an amount that depends on the row you are in (1-21). If a marble of the other color is drawn, you receive \$0.

Below you can see your possible earnings for each row.

No.	Jar A: 50 Red Marbles 50 Black Marbles	Jar B: ? Red Marbles ? Black Marbles
1.	\$4 if your color is drawn \$0 if the other color is drawn	\$2.35 if your color is drawn \$0 if the other color is drawn
2.	\$4 if your color is drawn \$0 if the other color is drawn	\$2.58 if your color is drawn \$0 if the other color is drawn
3.	\$4 if your color is drawn \$0 if the other color is drawn	\$2.82 if your color is drawn \$0 if the other color is drawn
4.	\$4 if your color is drawn \$0 if the other color is drawn	\$3.05 if your color is drawn \$0 if the other color is drawn
5.	\$4 if your color is drawn \$0 if the other color is drawn	\$3.28 if your color is drawn \$0 if the other color is drawn
6.	\$4 if your color is drawn \$0 if the other color is drawn	\$3.51 if your color is drawn \$0 if the other color is drawn
7.	\$4 if your color is drawn \$0 if the other color is drawn	\$3.75 if your color is drawn \$0 if the other color is drawn
8.	\$4 if your color is drawn \$0 if the other color is drawn	\$3.98 if your color is drawn \$0 if the other color is drawn
9.	\$4 if your color is drawn \$0 if the other color is drawn	\$4.21 if your color is drawn \$0 if the other color is drawn
10.	\$4 if your color is drawn \$0 if the other color is drawn	\$4.44 if your color is drawn \$0 if the other color is drawn
11.	\$4 if your color is drawn \$0 if the other color is drawn	\$4.68 if your color is drawn \$0 if the other color is drawn
12.	\$4 if your color is drawn \$0 if the other color is drawn	\$4.91 if your color is drawn \$0 if the other color is drawn
13.	\$4 if your color is drawn \$0 if the other color is drawn	\$5.14 if your color is drawn \$0 if the other color is drawn
14.	\$4 if your color is drawn \$0 if the other color is drawn	\$5.37 if your color is drawn \$0 if the other color is drawn
15.	\$4 if your color is drawn \$0 if the other color is drawn	\$5.61 if your color is drawn \$0 if the other color is drawn
16.	\$4 if your color is drawn \$0 if the other color is drawn	\$5.84 if your color is drawn \$0 if the other color is drawn
17.	\$4 if your color is drawn \$0 if the other color is drawn	\$6.07 if your color is drawn \$0 if the other color is drawn
18.	\$4 if your color is drawn \$0 if the other color is drawn	\$6.30 if your color is drawn \$0 if the other color is drawn
19.	\$4 if your color is drawn \$0 if the other color is drawn	\$6.54 if your color is drawn \$0 if the other color is drawn
20.	\$4 if your color is drawn \$0 if the other color is drawn	\$6.77 if your color is drawn \$0 if the other color is drawn
21.	\$4 if your color is drawn \$0 if the other color is drawn	\$7 if your color is drawn \$0 if the other color is drawn

Rows 1-21

The amount of money you receive if a marble of your color is drawn from Jar B increases as you move down the rows. The amount goes from \$2.35 for row 1, to \$7 for row 21.

For each of the rows, you must choose if you want to draw the marble from Jar A or Jar B.

To choose between Jar A and Jar B for each row, you will make only one decision.

You must indicate at which row you would like to switch from Jar A to Jar B. For each row before your switch row, Jar A will be used to determine your earnings. For each row at and after your switch row, Jar B will be used to determine your earnings.

After you make your decisions, the experimenter will randomly select a row and will draw a marble from the jar of your choice on this row. Your earnings from this Extra Part will be given by the draw of this marble only.

B.5 Questionnaire

Before the experiment concludes, please take a moment to answer the following questions about you. Remember that your answers are anonymous and cannot be linked to your identity.

- If you are offered the following two options, which one do you prefer?
  - To receive \$4 with certainty
  - To make a bet that pays \$8.80 with 50% chance and \$0 with 50% chance
- Do you think men or women, on average, solve correctly a larger number of problems in this experiment?
  - Women solve many more problems
  - Women solve somewhat more problems
  - Women solve slightly more problems
  - There is no difference between men and women
  - Men solve slightly more problems
  - Men solve somewhat more problems
  - Men solve many more problems
- Are you a student?

- If a student, what is your major?
- If not a student, what is your occupation?
- What is your year of birth?
- What is your gender?