Gender and Selection into Self-Improvement Contracts

David Klinowski[†]
Santiago Centre for Experimental Social Sciences
Nuffield College, University of Oxford; and
Universidad de Santiago de Chile

February 2018

ABSTRACT

We examine whether men and women differ in their willingness to select into a contract that pays upon improving one's past performance. Experiment participants complete a task and choose between receiving a regular piece rate, and a larger piece rate provided they improve relative to a previous round. We find that once self-confidence and risk aversion are accounted for, men and women enter at equal rates into self-improvement. This is true both for a stereotypically male and a stereotypically female task, and for all levels of participants' ability. The results provide insight into the feasibility and potential of using self-improvement contracts as gender-neutral incentive mechanisms.

Keywords: Gender, self-improvement, competitiveness
JEL codes: C91, J16, J31, D02

[†] I am grateful for valuable comments from participants of the 2017 IMEBESS Conference and the 2017 Antigua Experimental Economics Conference, and from two anonymous referees.

1. INTRODUCTION

Across many countries there continues to be a gender gap in labor market outcomes. On average women earn less than men for comparable jobs, and are underrepresented in various high-paying occupations and leadership positions (Blau and Kahn, 2017; Bertrand, forthcoming). One explanation for this disparity is the possibility that women have a lower taste for competition than men, which may make them more reluctant to enter competitive 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 that depends on individual performance alone.[1] More recent work finds that a distaste for competition may affect actual labor market outcomes, as tournament entry 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 challenge oneself to do 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 taste for self-improvement, this difference may constitute an additional reason for the gender gap 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 in the labor market while potentially also promoting productivity growth.

This paper investigates whether men and women differ in their willingness to select into a self-improvement contract, 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

[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. (forthcoming). A comprehensive review is given by Niederle (2016).

rate at which men and women enter 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 their entry decision and whether these can explain any gender difference in entry. As detailed in Section 2, a feature of our design is that we create an individually-calibrated risk preference elicitation that approximately simulates the risk and stakes involved in the self-improvement contract, but that removes the need to self-improve. By comparing the participants' willingness to accept the contract to their willingness to accept the similar, context-free bet, we can classify participants as self-improvement-averse, self-improvement-seeking, or consistent, and can examine whether men and women differ in their distribution across types. We also employ two tasks, one that participants perceive as male-dominated and another that they perceive as female-dominated, which allows us to investigate whether beliefs about the "femaleness" of the task affect men's and women's decisions to select into the self-improvement contract.[2]

We find that women are less willing than men to select into self-improvement, regardless of the task. However, once we account for factors measured in the experiment—particularly risk aversion and to a lesser extent self-confidence—the gender gap disappears statistically. The comparison between preferences over contracts and preferences over similarly-valued monetary bets also suggests no gender differences in attitudes toward self-improvement: the distribution of self-improvement-averse, self-improvement-seeking, and consistent types is similar for men and women. An additional finding, relevant if one seeks to design incentives that leverage the willingness to self-improve, is that entry into the self-improvement contract does not depend on the participant's level of ability, thus potentially all participants are equally amenable to policy interventions. Finally, we conduct sessions in which we replace the self-improvement contract with one that pays conditional on beating another participant's performance, as is more typical in

[2] In the context of competition with other individuals, Shurchkov (2012), Dreber et al. (2014), Grosse et al. (2014), and Halladay (2017) find that women are equally, if not more, competitive than men when the task involved is stereotypically female. In the context of contributing test answers to the group, Coffman (2014) finds that an individual's confidence in her ability to answer correctly decreases in areas that are stereotypically outside of her gender's domain. We investigate whether these findings extend to the context of self-improvement—that is, whether men become less confident in their ability to improve and less willing to accept the self-improvement contract if they perceive the task as female-dominated (and whether the same is true for women for a task perceived as male-dominated).

the gender and competitiveness literature. We find that, after accounting for factors measured in the experiment, women are significantly less willing than men to compete on either task, and less likely to be classified as competition-seeking according to our type-categorization. This suggests that our finding that men and women select into self-improvement at equal rates is not driven by our sample or our analytical strategies.

Closest to our paper is Apicella et al. (2017), who in a laboratory and online experiment also find no gender differences in selection into self-improvement. We contribute to the line of work Apicella et al. started in several ways. Our individually-calibrated risk elicitation provides alternative evidence that a gender gap in entry into a self-improvement contract 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 perceptions about the femaleness of the task do not affect men's or women's self-confidence in their ability to improve or their willingness to accept the self-improvement challenge. We see these as valuable insights into the feasibility and potential of using self-improvement incentives as a way to build competitive environments (more broadly construed) that appeal to both men and women.

In what remains of the paper, Section 2 describes the experiment design, Section 3 presents the results, and Section 4 concludes with a discussion. An appendix to this paper contains supplementary analysis and the experiment instructions.

2. EXPERIMENT DESIGN

Our laboratory experiment consists of four incentivized parts plus an unincentivized questionnaire. Participants know from the start that there are these many parts to the experiment, but receive instructions for each part only at the beginning of the corresponding part. Four treatments are randomly assigned in a between-subjects design, with all participants in a given session receiving the same treatment. Treatments vary in the real-effort task that participants are asked to perform (a sums task or an anagrams task), and in the contract choice presented to them (a self-improvement contract or a competition contract). Here we describe an experiment session under

the sums task and the competition contract, and note how other treatments would vary when appropriate.

In Part 1, participants have 5 minutes to add as many sets of 5 randomly generated 2-digit numbers as they can, an earn 40¢ per correct sum (as in Niederle and Vesterlund, 2007). After the 5 minutes pass, each participant receives feedback on the number of correct sums she solved and her corresponding earnings. Before they begin Part 1, participants have 90 seconds of unincentivized practice with the task.

(For sessions that employ an anagrams task rather than a sums task, participants have five minutes to form as many 3- or more-letter anagrams as they can from a set of 6 given letters. Piece-rate incentives are unchanged.)

In Part 2, participants again have 5 minutes to add sets of 5 randomly generated 2-digit numbers (different sets from those in Part 1). Before solving the task, participants individually choose one of the following two options for generating earnings in this part.

- Option A: Receive 40¢ per correct sum, regardless of the number of sums solved correctly.
- Option B: Receive 80¢ per correct sum, provided the participant solves correctly a greater number of sums than the number solved in Part 1. Otherwise the participant receives no earnings in Part 2.

(For sessions that present a competition contract rather than a self-improvement contract, Option B offers a piece rate of 80¢ provided the participant solves correctly a greater number of sums than the number solved in Part 1 by another anonymous, randomly selected participant in the session.)

After selecting an earnings scheme, but before performing the task, each 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 sum in Part 2 than she did in Part 1. To avoid strategic interactions with performance in Part 2, we do not incentivize answers to these questions. These questions elicit measures of the participant's

confidence in her ability to improve. After responding to these questions, participants solve the task for 5 minutes, and receive feedback on their number of correct sums and their corresponding earnings in Part 2.

(For sessions that present a competition contract rather than a self-improvement contract, the participant is asked to state the probability with which she thinks she will solve correctly at least one more sum in Part 2 than what another anonymous, randomly selected participant solved in Part 1. We also ask where participants belief they rank based on their Part-1 performance in the session.)

The self-improvement contract (Option B) in Part 2 involves the risk of failing to improve, which would result in zero earnings for this part. It is possible then that risk preferences influence the choice of contract. If women are more risk averse than men, as the literature often finds (Croson and Gneezy, 2009; Charness and Gneezy, 2012), then women may be less willing to select into the self-improvement contract due to higher risk aversion, rather than due to lower taste for self-improvement per se. To explore this, in Part 3 we elicit risk preferences in a choice as close as possible to the one faced in Part 2, presenting participants with similar risks and stakes as Options A and B, but removing the need to perform the task. Thus, in Part 3 participants are faced with the following two options.

- Receive a sure payment of X.
- Make a bet that pays Y with probability p and pays nothing with probability $1 - p$.

X and Y are individually calibrated for each participant, so that X is her earnings from Part 1, and Y is the earnings she would receive in Part 2 under the self-improvement contract if she solved correctly exactly one more sum in Part 2 than she solved in Part 1. Under this calibration, the sure payment of X approximates Option A in Part 2, while the bet that pays either Y or 0 approximates the potential payoffs under the self-improvement contract.[3]

[3] The correspondence between Option A in Part 2 and the sure payment of X in Part 3 rests on the assumption that in Part 2 the participant is certain that she can replicate her Part-1 performance. In the Appendix we examine the validity of this assumption.

In Part 3, participants must indicate the minimum probability p for which they prefer to make the bet over receiving the sure payment X. The probability reported must be a percentage integer between 0 and 100. To incentivize this decision, after the participant reports her p, the experimenter randomly draws an integer probability value q between 0 and 100, which becomes the actual probability with which the bet pays Y. If q is smaller than the reported p, the participant receives X; otherwise, the participant enters the bet, and earns either Y or 0, depending on the outcome of the bet. To determine the outcome of the bet, the experimenter randomly draws a second number, between 1 and 100. If the number is in $[1, q]$ the bet pays Y; if the number is in $[q + 1,100]$ the bet pays 0. Just as in Part 2, where the participant always learns whether she improved her score relative to Part 1 regardless of her choice of contract, in Part 3 the participant always receives feedback on the outcome of the bet regardless of whether she enters the bet.

Since the belief about one's ability to improve may not be an objective, precise probability point estimate, the participant's choice of contract in Part 2 may be affected by her preferences toward ambiguity. In Part 4 we elicit ambiguity preferences by presenting participants 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% percent chance of receiving nothing. The ambiguous option is an undisclosed probability (known to be between 0% and 100%) of receiving Z and the complement probability 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 increments of 23¢. The items in the list are arranged in increasing order of Z.

We implement this elicitation using two physical jars, each containing 100 marbles. The risky jar contains 50 red marbles and 50 black marbles, with the contents 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 0 and 100, and they add to 100. The contents of the ambiguous jar are covered from the participants' view. Participants are asked to select the color on which they want to bet, and the item in the list (item 1 to 21) at which they prefer to switch from drawing a marble from the risky jar to drawing a marble from the ambiguous 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 series of unincentivized questions. We ask participants whether they would prefer to receive \$4 with certainty or to make a bet that pays \$8.80 with 50% probability and \$0 with 50% probability. This question serves as an alternative risk preferences measure that involves the same stakes for all participants. We also ask participants whether they think male or female participants on average perform better at the task (7-point Likert scale of the form "women(men) give a substantially/somewhat/slightly larger number of correct answers than men(women)" and a neutral option of "no gender difference"). This question captures beliefs about the femaleness of the task. Finally, we ask for the participant's student status, major, occupation, year of birth, and gender.

At the end of the session, one part from Parts 1, 2, and 3 is randomly selected for payment (participants knew this feature of the design at the beginning of the session). Earnings from this Part are added to those from Part 4, and to a show-up fee of \$3.15. Average total earnings were \$9.75. Sessions lasted approximately 45 minutes. The experiment was programmed in oTree (Chen et al., 2016).

3. RESULTS

We begin by presenting descriptive statistics of the variables used in the analysis. Table 1 documents means and proportions by gender and contract type (self-improvement or competition). We observe that, for the sums task in Part 1, on average men perform better than women in the self-improvement sessions, while women perform better than men in the competition sessions. If we aggregate all sessions—all sessions are identical in terms of action sets and information up to Part 1—we see that performance in the sums task is on average equal across gender (8.89 for men and 8.70 for women; p-value=0.732). On average, performance in Part 1 in the anagrams task is equal across gender. Men expect larger improvements in performance in Part 2 relative to Part 1 in all treatments, although the difference is statistically significant only for the anagrams task (for which women on average actually expect to decrease their performance in Part 2 relative to Part

1). For both tasks in the self-improvement sessions, men and women have equal average beliefs about the probability with which they will self-improve. In the competition sessions, on average men state slightly higher probability than women of beating a random opponent (statistically higher in the sums task).

Table 1: Descriptive statistics

	Self-Improvement			Competition		
	Male	Female	p-value	Male	Female	p-value
Part 1 (piece-rate) performance						
Sums	9.58	8.28	0.066	7.79	9.11	0.095
Anagrams	15.81	14.98	0.447	15.18	14.65	0.674
Expected improvement in Part 2 relative to Part 1						
Sums	1.21	0.88	0.323	1.65	1.07	0.321
Anagrams	0.47	-0.39	0.131	0.12	-1.8	0.030
Believed probability of self-improving / beating opponent						
Sums	0.68	0.66	0.771	0.60	0.48	0.009
Anagrams	0.58	0.59	0.903	0.57	0.51	0.168
Self-improve / Beat opponent (%)						
Sums	0.55	0.42	0.210	0.50	0.53	0.769
Anagrams	0.19	0.24	0.495	0.32	0.30	0.827
Perceived task femaleness (-3 to 3)						
Sums	-0.23	-0.14	0.530	-0.12	-0.33	0.211
Anagrams	0.86	0.80	0.746	0.74	0.53	0.344
Prefers risky choice (%)	0.58	0.40	0.013	0.54	0.29	0.002
Ambiguity tolerance score	10.59	11.07	0.390	11.03	10.81	0.743
Science major (%)	0.47	0.25	0.002	0.41	0.18	0.001
Sample size						
Sums	53	43	-	34	45	-
Anagrams	43	49	-	34	40	-

Notes: Values indicate subgroup means, or subgroup proportions for variables labeled with "%." For the calculation of the expected improvement in Part 1 relative to Part 2 in the *Competition* sessions, we remove one observation from a female participant who reports an expected performance in Part 2 of 1250 sums.

The actual fraction of participants who self-improve, or who beat the opponent, is on average equal across gender.[4] For both men and women, these fractions are smaller than their stated probabilities of self-improving or beating an opponent, which suggests that both men and women are overconfident in their ability to self-improve or to beat an opponent, but men tend to be more

[4] Participants are less likely to self-improve or to beat the opponent in the anagrams task than the sums task. This is likely because the set of six letters shown in Part 2 in the anagrams task turned out to be more challenging than the one shown in Part 1, while for the sums task difficult stays constant across Parts. The increased difficulty in Part 2 in the anagrams task was not intended in the design, and could not have affected the participants' choice of contract in Part 2, since participants choose the contract before seeing the set of six letters for Part 2.

overconfident than women. In terms of the perceived "femaleness" of the tasks, both men and women believe that on average men perform better than women in the sums task, and that women perform better than men in the anagrams task. Women are more risk averse than men, as seen by the fraction of participants who prefer the risky bet in the questionnaire section. There are no gender differences in ambiguity tolerance, as measured by the average switch point in the ambiguity-elicitation task. And finally, men are approximately twice as likely as women to be a science or math major, which roughly mirrors the gender proportion in STEM fields in Chilean universities (SIES, 2017). In the analysis, we do not control for being a STEM major, but check that doing so produces virtually identical results as those presented here.

3.1 Gender and selection into self-improvement

We now examine whether men and women differ in their willingness to choose the self-improvement contract, and what factors may explain this difference. We contrast the results to those from sessions in which participants instead choose whether to compete with another subject.

In total, 56 percent of participants choose the self-improvement option in the sums task, and 27 percent do so in the anagrams task.[5] Looking separately across gender, we see that 62 percent of the men and 49 percent of the women enter the self-improvement contract in the sums tasks, while 37 percent of the men and 18 percent of the women do so in the anagrams task. To formalize these results, Table 2 examines the gender gaps in a regression framework. We present marginal effects from probit regressions (separate regressions for each task) that estimate the probability with which a participant chooses the self-improvement contract. Column 1 is the simplest specification, which includes only a female indicator. The value on the female indicator is the estimated difference in percentage points between men and women in the probability of selecting into self-improvement. From this specification we see that on average women are 13 percentage points less likely than men to select into the self-improvement contract in the sums task, and 19 percentage points less likely to do so than men in the anagrams task.

[5] If forming anagrams was less familiar a task than adding numbers, this could explain why fewer participants select into self-improvement in the anagrams task than in the sums task. We have no data to test this conjecture, but we point out that participants are less optimistic about their probability to improve in the sums task than in the anagrams task (67% vs. 58% probability to improve, p-value=0.004).

Table 2: Marginal effects on the probability to select into self-improvement

	Baseline	Ability	Confidence	Risk	Ambiguity	Femaleness	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. SUMS TASK			
Female	-0.134**	-0.135***	-0.129	-0.087	-0.132**	-0.135**	-0.063
	(0.055)	(0.048)	(0.082)	(0.083)	(0.057)	(0.060)	(0.077)
Number of correct answers in Part 1		0.000					0.012
		(0.010)					(0.013)
Expected score improvement			0.298***				0.263***
			(0.049)				(0.044)
Chose risky option				0.316***			0.195***
				(0.088)			(0.074)
Ambiguity switch point					-0.003		0.005
					(0.011)		(0.005)
Agrees with perceived femaleness						0.009	0.100*
						(0.150)	(0.059)
Female* agrees femaleness						-0.117	-0.003
						(0.171)	(0.118)
Pseudo R^2	0.013	0.013	0.412	0.088	0.014	0.015	0.477
N	96	96	96	96	96	96	96
				B. ANAGRAMS TASK			
Female	-0.188**	-0.195***	-0.154**	-0.155*	-0.188***	-0.184**	-0.125
	(0.073)	(0.075)	(0.081)	(0.092)	(0.072)	(0.092)	(0.108)
Number of correct answers in Part 1		-0.006					-0.006
		(0.009)					(0.009)
Expected score improvement			0.044**				0.046**
			(0.019)				(0.016)
Chose risky option				0.155			0.166
				(0.108)			(0.110)
Ambiguity switch point					-0.001		0.000
					(0.010)		(0.008)
Agrees with perceived femaleness						0.032	-0.034
						(0.129)	(0.131)
Female* agrees femaleness						0.068	0.275
						(0.179)	(0.193)
Pseudo R^2	0.038	0.043	0.107	0.064	0.038	0.042	0.150
N	92	92	92	92	92	92	92

Notes: Marginal effects from probit regressions for each task separately. Standard errors clustered at the session level in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 3: Marginal effects on the probability to select into competition with opponent

	Baseline (1)	Ability (2)	Confidence (3)	Risk (4)	Ambiguity (5)	Femaleness (6)	Full (7)
			A. SUMS TASK				
Female	-0.293***	-0.313***	-0.234***	-0.262***	-0.291***	-0.295***	-0.243**
	(0.040)	(0.053)	(0.057)	(0.041)	(0.047)	(0.039)	(0.112)
Number of correct answers in Part 1		0.015					0.011
		(0.021)					(0.022)
Expected prob. beating opponent			0.004				0.003
			(0.003)				(0.003)
Chose risky option				0.126			0.107
				(0.139)			(0.113)
Ambiguity switch point					-0.016		-0.015
					(0.012)		(0.014)
Agrees with perceived femaleness						0.014	-0.005
						(0.082)	(0.090)
Female*agrees femaleness						-0.331**	-0.243*
						(0.135)	(0.131)
Pseudo R^2	0.081	0.093	0.109	0.097	0.101	0.103	0.156
N	79	79	79	79	79	79	79
			B. ANAGRAMS TASK				
Female	-0.275***	-0.280***	-0.255***	-0.247***	-0.272***	-0.277***	-0.215***
	(0.078)	(0.082)	(0.096)	(0.070)	(0.065)	(0.047)	(0.050)
Number of correct answers in Part 1		-0.007					-0.012
		(0.018)					(0.017)
Expected prob. beating opponent			0.003				0.005**
			(0.004)				(0.002)
Chose risky option				0.108			0.129
				(0.111)			(0.112)
Ambiguity switch point					0.013**		0.015**
					(0.007)		(0.007)
Agrees with perceived femaleness						0.042	0.010
						(0.117)	(0.090)
Female*agrees femaleness						-0.467*	-0.533**
						(0.241)	(0.252)
Pseudo R^2	0.064	0.069	0.076	0.074	0.075	0.113	0.175
N	74	74	74	74	74	74	74

Notes: Marginal effects from probit regressions for each task separately. Standard errors clustered at the session level in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 3 shows analogous results from regressions that estimate the probability that a participant selects into competition with a random other subject (in the competition sessions). From column 1 we see that women are 29 percentage points less likely than men to choose to compete in the sums task, and 28 percentage points less likely to do so in the anagrams task. Although the size of the gender gap is much larger for selection into competition than for selection into self-improvement, the difference between the two gaps is statistically significant for the sums task and both task jointly, but not for the anagrams task (as seen from the negative and significant estimate on the *Female*Competition* interaction term for the *Baseline* specification on Table A1 in the Appendix).

The results from this baseline specification suggest that women are less willing than men to select into the self-improvement contract, although the gender gap is smaller than what we observe for selection into a competition contract. We now examine whether the gap in entry into the self-improvement contract can be at least partially explained by gender differences in ability, self-confidence, risk and ambiguity tolerance, and perceived femaleness of the tasks.

3.1.1 Mechanisms: Ability

Willingness to select into the self-improvement contract may or may not depend on ability—both alternatives seem plausible ex-ante. On the one hand, regardless of their level of ability, participants may be equally motivated to challenge themselves to improve their performance in Part 2 relative to Part 1, in which case we expect no relationship between selection and ability. On the other hand, we may expect a relation, either because (i) low-performers in Part 1 may be more likely to believe it easy to improve in Part 2 and therefore be more willing to choose to self-improve, or because (ii) high-performers in Part 1 may be more driven and enthusiastic about the task and therefore be more willing to choose to self-improve. If there is a relation between ability and selection into self-improvement, and there are gender differences in ability, it is possible that ability may play a role in explaining the gender difference in contract selection.

To explore this, we conduct regression analysis. Column 2 of Table 2 presents the same regression as the *Baseline* specification discussed previously, but adds a control for the participant's performance in Part 1. Results suggest that ability does not affect the decision to self-improve and

does not explain any of the gender gap in contract selection for either task: The estimated effect on ability is indistinguishable from "0" on either task, and the estimated size of the gender gap—that is, the coefficient on the female indicator—remains largely unchanged in Column 2 relative to the *Baseline* specification. These results suggest that both low- and high-ability participants have similar drive to self-improve, and that ability does not explain entry into self-improvement. (Figure A1 in the Appendix gives more details about the distribution of ability by gender and task, and Figure A3 documents the proportion of men and women who select into self-improvement for different levels of ability.)

3.1.2 Mechanisms: Self-confidence

If participants select into the self-improvement contract only when they are confident enough that they can better their Part-1 performance, then a difference between men and women in self-confidence in their ability to improve may help to explain the gender gap in selection into self-improvement. We explore this possibility by looking at the two measures of self-confidence in ability to improve that we collected: the participant's expectation about the number of correct answers that she will obtain in Part 2, and her believed probability of improving her performance in Part 2 relative to Part 1. 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 expect to improve their Part-1 performance slightly more than women do, especially on the anagrams task: men and women expect to improve, respectively, by 1.21 and 0.88 correct answers on the sums task, and by 0.47 and -0.39 (i.e., a decline) correct answers on the anagrams task. These gender differences are not statistically significant. Men and women also report similar probabilities of improving their performance in Part 2 relative to Part 1: 68% and 66% in the sums task, and 58% and 59% in the anagrams task.[6]

[6] Men and women are also similarly accurate in their beliefs about their ability to improve. In the sums(anagrams) task, 51(79)% of men and 49(63)% of women expect to achieve a number of correct answers larger than they actually achieved. Figure A5 in the Appendix presents the cumulative distributions of the difference between the expected and actual performance in Part 2, which show no difference across gender (Kolgomorov-Smirnov test of gender equality of the empirical distribution functions p-value = 1.000 for the sums task and 0.219 for the anagrams task). In terms of the believed probability of improving, in the sums(anagrams) task 5(2)% of men and 4(2)% of women state a probability of improving smaller than 50% but actually improved their performance (i.e., are underconfident), while 25(47)% of men and 30(37)% of women state a probability of improving larger than 50% but actually failed to improve their performance

To examine the role of self-confidence in the ability to improve in explaining selection into the self-improvement option, we conduct probit regressions predicting the participant's probability of selecting into the self-improvement contract, on a female indicator and on the difference between the participant's expected number of correct answers in Part 2 and the actual number of correct answers in Part 2 (namely, on the participant's expected improvement in performance from Part 1 to Part 2).[7] This is a regression analogous to the *Baseline* specification in Table 2, with the additional control for the expected improvement in performance. The results, which appear in Column 3 of Table 2, indicate that participants with more self-confidence are significantly more likely to select into self-improvement: An expected increase by one additional correct answer in Part 2 relative to Part 1 is associated with an increase of 29.8 percentage points in the likelihood of choosing the self-improvement option in the sums task. For the anagrams task, this value is 4.4 percentage points. Both estimates are statistically significant. By comparing the female estimate in Column 3 to that in Column 1, we also see that, for the sums task, the gender gap in selection into self-improvement is only slightly changed by the introduction of the self-confidence control (-0.134 vs. -0.129). For the anagrams task, the gender gap is somewhat decreased after the introduction of the control (-0.188 vs. -0.154; an 18-percent decrease). Thus, the results suggest that while more self-confidence predicts a higher probability of entry into the self-improvement contract, self-confidence plays at most only a modest role in explaining the gender gap in entry.

3.1.3 Mechanisms: Risk tolerance

The self-improvement contract involves the risk of failing to improve, resulting in zero earnings. If women are more risk averse than men, they may be less willing to choose the self-improvement contract on account of differences in risk preferences rather than differences in the drive to self-improve per se. We use two measures of risk attitudes to examine possible gender differences in risk preferences and their role in explaining selection into self-improvement. The first measure is

(i.e., are overconfident). Figure A6 in the Appendix plots these distributions, which show no difference difference across gender for the sums task (chi-squared p-value = 0.787) or the anagrams task (chi-squared p-value = 0.622).

[7] The alternative self-confidence measure we could use in the analysis—the believed probability of improving from Part 1 to Part 2—does not significantly predict the choosing the self-improvement contract or the gender gap in this choice.

a hypothetical binary choice between a risky bet and a safe payoff, where payoffs are equal for all participants. The second is an incentive-compatible elicitation in which payoffs are individually calibrated to match what each participant faced in the contract choice. We examine each measure in turn.

For the hypothetical binary choice, 40% of the women and 58% of the men in the self-improvement sessions prefer the risky bet over the safe amount (p-value = 0.013). Women are thus more risk averse than men under this metric. We use a regression framework to examine whether this difference in risk aversion can help to explain differences in selection into self-improvement contract. We predict the probability that a participant chooses the self-improvement contract using probit regressions analogous to the *Baseline* specification in Table 2 Column 1, but now adding a control for whether the participant prefers the risky bet in the binary choice. Results appear in Table 2 Column 4. They indicate that more risk-seeking participants are more likely to select into the self-improvement contract: preferring the risky bet is associated with an increase of 31.6(15.5) percentage points in the probability of choosing the self-improvement option in the sums(anagrams) task. The effect is significant only for the sums task. We can examine how much of the baseline gender gap in selection into self-improvement is attributable to risk preferences by comparing the estimates for the female indicator across Columns 1 and 4. Accounting for the binary risk-elicitation choice shrinks the gender gap in entry from -0.134 to -0.087 (a 35-percent reduction) for the sums task, and from -0.188 to -0.155 (an 18-percent reduction) for the anagrams task. Therefore, the results suggest that gender differences in risk aversion play a moderate role in explaining the gender gap in selection into the self-improvement contract, but that nevertheless a significant residual gap remains unexplained after controlling for this measure of risk aversion alone.

We can also examine the role of risk preferences on the choice over the self-improvement contract using the alternative, individually-calibrated risk measure. For this we compare the participant's choice of contract given her stated probability of improving her performance, to her choice over a risky bet that involves similar stakes and the same stated probability, but that does not involve a requirement to improve one's performance for the bet to pay off. Comparing the two choices

allows us to classify participants into three types: self-improvement-averse, self-improvement-seeking, and consistent. The categorization is illustrated in Table 4, and is derived as follows.

Table 4: Participants' types based on comparing their willingness to select into the self-improvement contract in Part 2, and their willingness to make a similar bet in Part 3 that does not require a need to self-improve

	$\beta \geq \delta$	$\beta \leq \delta$
Selects into self-improvement	Consistent choices	Self-improvement seeking
Does not select into self-improvement	Self-improvement averse	Consistent choices

Notes: β denotes the probability with which the participant believes she will improve in Part 2 relative to Part 1, and δ denotes the minimum probability of the bet in Part 3 paying off for which the participant accepts to make the bet instead of receiving the sure payoff.

Suppose the participant completed Part 1 and is now deciding whether to select into the self-improvement contract in Part 2. Let X be the participant's earnings for her performance in Part 1; let Y be her earnings in Part 2 if she selects into the self-improvement contract and improves her Part-1 performance by "1" additional correct answer; and let β be the probability with which she believes she will improve in Part 2 relative to Part 1. Purely in terms of monetary risk, the choice of contract 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.[8] The risk elicitation in Part 3 presents the participant with a choice between these same risky and sure payments, and asks the participant to state the minimum probability of the bet paying off for which she would accept the bet over receiving the sure payoff. Denote the participant's reported probability as δ. In this risk elicitation, the participant reveals that she accepts the bet if it pays with probability of at least δ. We can contrast this revealed preference to her preference over contracts given β. If she accepts the bet when it pays with probability of at least δ but does not select into the self-improvement contract when her believed probability of improving is $\beta \geq \delta$, her choices suggest a distaste for the self-improvement contract net of the risk element. Similarly, if she accepts the bet when it pays with probability of at least δ but competes when her believed probability of improving is $\beta < \delta$, her choices suggest a taste for the self-improvement contract net of the risk element. In the remaining cases (selecting

[8] This assumes that the participant is certain that she can replicate her Part-1 performance. In the Appendix we test the validity of this assumption.

into the self-improvement contract when $\beta \geq \delta$, and not selecting into it when $\beta < \delta$) her choices are consistent with each other.

Comparing how men and women distribute across these three types provides a way to examine gender differences in preferences over self-improvement controlling for risk preferences. Among male participants, 56% made consistent choices, 16% are self-improvement-seeking, and 28% are self-improvement-averse. The distribution for female participants is 55% consistent, 12% self-improvement-seeking, and 30% self-improvement-averse. The distributions are statistically similar (χ^2 test p-value=0.679). Thus, this exercise suggests that men and women have similar preferences over the self-improvement contract once their risk preferences are accounted for.[9]

3.1.4 Mechanisms: Ambiguity tolerance

The belief in one's ability to improve in Part 2 relative to Part 1 is likely an imprecise probability estimate (rather than a point estimate). The choice of contract in Part 2 could therefore be affected by the participant's attitude toward ambiguity—if so, gender differences in ambiguity preferences could play a role in explaining the gender gap in selection into the self-improvement contract. To explore this, we predict the probability that the participant selects into the self-improvement contract with probit regressions analogous to the *Baseline* specification in Column 1 of Table 2, but now including as control the participant's switch point from the risky bet to the ambiguous bet in the ambiguity elicitation task. This switch point is a number between 1 and 21—a larger number indicates greater ambiguity aversion. The results are shown in Column 5 of Table 2. We see that the marginal effect of the ambiguity switch point on the probability of choosing the self-improvement option is -0.003, and statistically indistinguishable from zero. The estimated gender gap in selection into self-improvement is barely changed relative to the *Baseline* specification. This suggests that ambiguity preferences do not help to explain the decision to self-improve or the gender gap in selection into self-improvement in this experiment.

[9] Replicating the analysis for the case of selection into competition (comparing selection into the competition contract vs. the participant's choice in the risk elicitation) finds that, among men, 56% make consistent choices, 21% are competition-averse, and 24% are competition-seeking; while among women, 64% make consistent choices, 27% are competition-averse, and 9% are competition-seeking. These distributions are statistically significantly different (χ^2 test p-value=0.054). Thus, results from this alternative analytical strategy also suggest that women have a lower taste for competition than they have for self-improvement relative to men.

3.1.5 Mechanisms: Perceived femaleness of the task

If participants consider the task they perform in the experiment to be the domain of a particular gender (e.g., sums are the domain of males or anagrams the domain of females), one could imagine that a participant whose gender is incongruous with the gender type of the task may become less confident in her ability to improve and less willing to select into the self-improvement contract. To investigate this possibility, we examine whether participants systematically associate the sums and the anagrams tasks with a specific gender, and if so, whether agreement with this perception affects the participant's level of confidence in her ability to improve and her probability of selecting into the self-improvement contract.

In the final section of the experiment participants report in a 7-point Likert scale whether they believe men or women on average give a greater number of correct answers on the task. Here we code responses from -3 to 3 in ascending order of perceived femaleness, such that -3/-2/-1 = "men give a substantially/somewhat/slightly larger number of correct answers than women," 0 = "no gender difference," and 1/2/3 = "women give a slightly/somewhat/substantially larger number of correct answers than men." Among participants in the self-improvement sessions, average perceived femaleness is -0.188 for the sums task and 0.826 for the anagrams task. Both are significantly different from "0" (p-value<0.01 for both tests of equality of means and medians for both tasks). Table 1 shows that femaleness perceptions do not differ across gender. Thus, on average, both male and female participants agree that the sums task is male-dominated and the anagrams task is female-dominated.[10]

To examine whether agreement with these average perceptions of the femaleness of the tasks affects the participants' confidence in their ability to improve, we use OLS regressions to predict the participant's expected number of correct answers in Part 2, conditional on the number of correct answers she obtained in Part 1. We include a female indicator, an indicator of agreement with the average perceived femaleness of the task (this indicator equals "1" if the participant believes that men give a slightly/somewhat/substantially larger number of correct answers than women on the

[10] Participants from the competition sessions hold similar beliefs: perceived femaleness is -0.241 for the sums task and 0.622 for the anagrams task among these participants. Both values are statistically different from 0 (p-value<0.01 for both tests of equality of means and medians for both tasks) and statistically similar across gender, as Table 1 shows.

sums task, or that women give a slightly/somewhat/substantially larger number of correct answers than men on the anagrams task, and equals "0" otherwise), and the interaction of the two. We show results in Table 5. We find that, for both tasks, a woman on average expects to perform similarly in Part 2 regardless of whether she agrees with the average perceived femaleness of the task. The same is true for men in the sums task. However, for men in the anagrams task, we find that agreement with the average perception is associated with an *increase* in expected performance—that is, on average, a man who agrees that women perform better than men on the anagrams task expects to improve his performance by a *larger* amount in Part 2 relative to Part 1 than a man who disagrees with this average perception. Recall that, given the observed average perceptions, the motivating conjecture for this section is that a woman who agrees that the sums task is the domain of men might be less confident in her ability to improve on this task (and similarly for a man who agrees that the anagrams task is the domain of women). We do not find evidence of this. Women's confidence does not appear to be affected by her agreement with the average perceived femaleness of the task. Men's confidence is affected only for the anagrams task, and in the direction opposite to the conjecture. (Table A2 in the Appendix replicates these results using the believed probability of improving as the measure of self-confidence, and in Table A3 and Table A4 we show that neither the expected number of correct answers in Part 2 nor the believed probability of beating a random opponent are affected by agreement with the average perceived femaleness among participants in the competition sessions.)

Finally, we explore whether the participant's perceived femaleness of the task affects her willingness to select into the self-improvement contract. We use probit regressions to predict the probability that the participant selects into the self-improvement option, using a model analogous to the *Baseline* specification in Table 2, but adding an indicator that equals "1" if the participant agrees with the average perceived femaleness of the task, and the interaction between this indicator and the female indicator. Results appear in Column 6 of Table 2. We see that, for both tasks, neither agreement with the average perceived femaleness of the task nor its interaction with the participant's gender significantly predict the choice of contract. Moreover, the gender gap in entry rate remains virtually unchanged in this specification relative to *Baseline*. We therefore find that, though participants do agree on average that the sums task is dominated by males and the anagrams

task is dominated by females, agreement with this view does not help to explain selection into the self-improvement contract or the gender gap in this selection.

Table 5: Participant's expected number of correct answers in Part 2

	Sums task		Anagrams task	
	Men	Women	Men	Women
Agrees with average perceived femaleness	9.928 (0.436)	9.416 (0.283)	16.679 (0.335)	14.642 (0.366)
Disagrees with average perceived femaleness	10.348 (0.306)	9.974 (0.163)	14.544 (0.609)	15.200 (0.398)
P-value of the column difference	0.400	0.200	0.006	0.455
N	96		92	

Notes: Values are predictions from OLS regressions that control for the number of correct answers the participant obtained in Part 1, the gender of the participant, an indicator of agreement with the average perceived femaleness of the task (this indicator equals "1" if the participant believes that men give a slightly/somewhat/substantially larger number of correct answers than women on the sums task, or that women give a slightly/somewhat/substantially larger number of correct answers than men on the anagrams task, and equals "0" otherwise) and the interaction of the two. Sample restricted to participants from self-improvement sessions. Standard errors clustered at the session level in parentheses.

3.1.6 All mechanisms jointly

We conclude this section by considering a full model that jointly estimates the effect of ability, self-confidence, risk aversion, ambiguity aversion, and perceived femaleness of the task on the probability of selecting into the self-improvement contract. Column 7 of Table 2 presents the results from a probit regression analogous to the *Baseline* specification, but that includes all the mechanisms examined in Columns 2–6. Higher self-confidence in one's ability to improve continues to significantly predict a higher probability of entry into the self-improvement option, in both the sums and the anagrams task. Lower risk aversion continues to predict entry for the sums task. Once all factors measured in the experiment are accounted for, the gender gap in selection into the self-improvement contract shrinks by 53 percent in the sums task (from 13.4 to 6.3 percentage points) and by 34 percent in the anagrams task (from 18.8 to 12.5 percentage points). Neither residual gender gap is significantly different from zero. This suggests that selection into self-improvement is determined at least partly by the participant's self-confidence and risk tolerance, and that once these are accounted for, men and women have only slight

differences in their preferences over the self-improvement contract. In contrast, the gender gap in selection into competition with another participant remains large and significant after accounting for all measured factors (Table 3): these factors account for 17 percent of the gender gap in selection in the sums task (shrinking from 29.3 to 24.3 percentage points) and for 22 percent of the gap in the anagrams task (shrinking from 27.5 to 21.5 percentage points). The difference between the residual gender gap in selection into self-improvement contract and the residual gender gap in selection into the competition contract continues to be significant for the sums task, but is not significant for the anagrams task or both tasks jointly (see the *Female*Competition* interaction term for the *Full* specification in Table A1).

We reach a similar conclusion if we compare men and women in their probability of being classified as self-improvement averse, consistent, or self-improvement seeking, after controlling for the participant's ability, self-confidence, and ambiguity aversion. Recall that these types are constructed by comparing the participant's preference over the self-improvement contract to her preference over a similar monetary bet that involves no need to perform the task. Table A5 (left panel) in the Appendix presents results from a multinomial probit regression that estimates the probability that the participant falls into each category, given her gender, ability, self-confidence, and ambiguity tolerance. In Figure 1 Panel A we present the results graphically, plotting the predicted distribution over types for men and women. There is no gender difference in the distributions. In Figure 1 Panel B, we plot similar probabilities of being classified as competition-averse, consistent, or competition-seeking, for participants in the competition sessions (predictions from a multinomial probit regression presented in Table A5, right panel). Here we do observe gender differences in the distribution—in particular, women are significantly less likely than men to be classified as competition-seeking. Thus, the results suggest that, while women dislike the competition contract more than men relative to their individual risk preferences, such gender gap is not observe for the self-improvement contract.

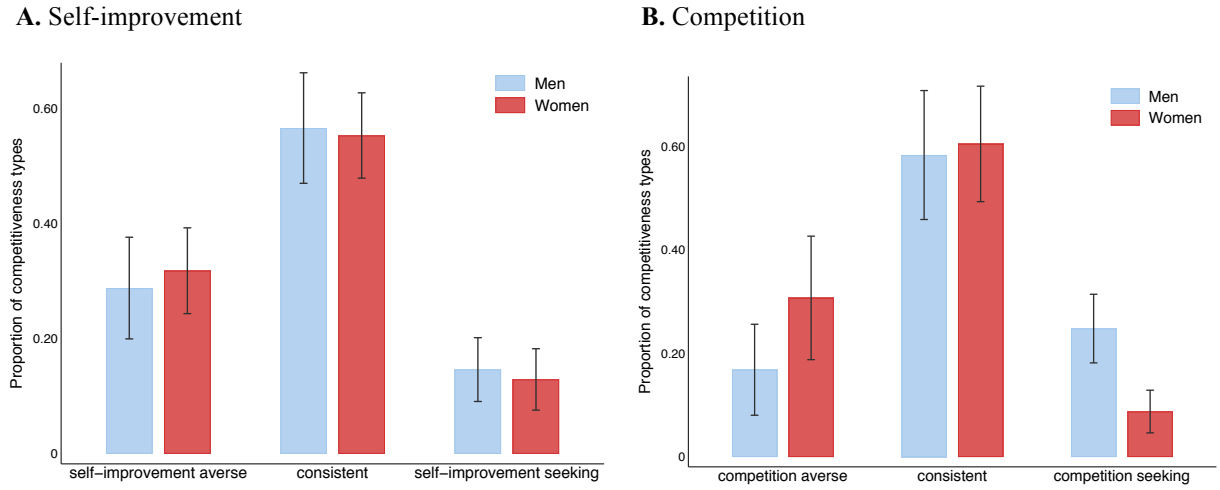

Figure 1: Distribution of participants over types, constructed by comparing their willingness to select into the self-improvement contract in Part 2, and their willingness to make a similar bet in Part 3 that does not require a need to self-improve. Estimates from probit regressions of the probability of falling into each category, controlling for the participant's gender, number of correct answers in Part 1, expected probability of improving performance in Part 1 relative to Part 2 (or expected probability of beating a random other participant in the competition sessions), and the switch point in the ambiguity tolerance elicitation, with clustered standard errors at the session level. Error bars show 90-percent confidence intervals of the predictions.

4 DISCUSSION

In this paper we examine whether men and women differ in their willingness to select into a contract that pays upon improving one's past performance in a real-effort task. We conduct a laboratory experiment in which participants choose between receiving a piece-rate independent of performance, and a larger piece-rate that pays only if the participant performs better than she did in 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 perceived femaleness of the task in the choice of contract and the gender difference in that choice. We explore whether results vary for a task perceived as male-dominated versus female-dominated. And we compare our results to a more standard paradigm in the literature in which participants choose whether to select into competition with another individual.

We find that greater self-confidence in one's ability to improve, and lower risk aversion, are significant predictors of a higher probability of entry into the self-improvement contract. We show with two different strategies that once self-confidence, risk aversion, and other factors are controlled for, men and women enter into the self-improvement contract at statistically similar rates, regardless of the task. One strategy finds no significant residual gender effect in a model that estimates the participant's choice of contract, conditional on the participant's ability, self-confidence, risk and ambiguity aversion, and agreement with the average perceived femaleness of the task. Another strategy finds that conditional on the participant's ability, self-confidence, and ambiguity aversion, men and women distribute equally across types (self-improvement averse, consistent, and self-improvement seeking), where types are constructed by comparing the participant's willingness to select into the self-improvement contract to her willingness to accept a similarly risky bet that requires no need to perform the task. We show that when applied to analyzing selection into competition with another individual, these two strategies find statistically significant gender differences (women are less competitive than men), as is often found in the literature. This suggests that neither our sample nor our analytical strategies are driving the finding of no gender differences in willingness to self-improve.

Our findings provide insight into the design of effective ways to attract men and women into challenging or competitive environments, more broadly construed. A mechanism that leverages people's willingness to self-improve may appeal to both men and women regardless of ability, the results suggest. Of course, the productivity gain for men and women from implementing such mechanism would depend on how output responds to self-improvement incentives, which is something that our design cannot examine. Doing so would require exogenous assignment (rather than self-selection) into the self-improvement contract, and the use of a task with demonstrated output elasticity.[11] Although some studies find that competition against others improves mens' but not women's performance (Gneezy et al. 2003; Günther et al. 2010; Shurchkov 2012), there is a basis to conjecture that self-improvement incentives 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

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

conditional on self-improvement, then, conditional on self-confidence and risk preferences, a self-improvement contract may represent a more gender-neutral way of introducing elements of competition (in this case against the self) into the incentive structures of organizations.

REFERENCES

Apicella, C., Demiral, E., and Mollerstrom, J., "No Gender Difference in Willingness to Compete when Cometing against Self." *American Economic Review: Papers and Proceedings*, 107.5 (2017): 136-140.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wand, S. W., and Wilson, A. J., "The Slider Task: An Example of Restricted Inference on Incentive Effects." *Journal of the Economic Science Association*, 2.1 (2016): 1-12.

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

Bertrand, M., "The Glass Ceiling: 2017 Coase Lecture" *Economica*, (forthcoming).

Blau, F. D., and Kahn, L M., "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature,* 55.3 (2017): 789-865.

Buser, T., Niederle, M., and Oosterbeek, H., "Gender, Competitiveness, and Career Choices." *Quarterly Journal of Economics*, 129.3 (2014): 1409-1447.

Buser, T., Peter, N., and Wolter, S., "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review: Papers and Proceedings*, 107.5 (2017): 125-130.

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

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

Chen, D. L., Schonger, M., and Wickens, C., "oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance*, 9 (2016): 88-97.

Coffman, K. B., "Evidence on Self-Stereotyping and the Contribution of Ideas." *Quarterly Journal of Economics*, 129.4 (2014): 1625-1660.

Croson, R., and Gneezy, U., "Gender Differences in Preferences." *Journal of Economic Literature*, 47.2 (2009): 448-474.

Dreber, A., von Essen, E., and Ranehill, E., "Gender and Competition in Adolescence: Tasks Matter." *Experimental Economics*, 17.1 (2014): 154-172.

Gneezy, U., Niederle, M., and Rustichini, A., "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, 118.3 (2003): 1049-1074.

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

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M., "Women Can't Jump? An Experiment on Competitive Attitudes and Stereotype Threat." *Journal of Economic Behavior and Organization*, 75.3 (2010): 395-401.

Halladay, B., "Perception Matters: The Role of Task Gender Stereotype on Confidence and Tournament Selection." *Working Paper*, (2017).

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

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

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

Moore, D. A. and Healy, P. J., "The Trouble with Overconfidence." *Psychological Review,* 115.2 (2008): 502-517.

Niederle, M., "Gender." In Kagel, J., and Roth, A. E., editors, *Handbook of Experimental Economics*, 2016: 481-553. Princeton University Press, 2nd Ed.

Niederle, M., Segal, C., and Vesterlund, L., "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science*, 59.1 (2013): 1-16.

Niederle, M., and Vesterlund, L., "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122.3 (2007): 1067-1101.

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

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

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

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

Servicio de Información de Educación Superior (SIES), *Brechas de Género en Educación en Chile 2016,* Ministerio de Educación de Chile, 2017. Data available at http://www.mifuturo.cl/index.php/estudios/estudios-recientes

Shurchkov, O., "Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints." *Journal of the European Economic Association*, 10.5 (2012): 1189-1213.

Sutter, M., and Rützler, D., "Gender Differences in Competition Emerge Early in Life and Persist." *Management Science*, 61.10 (2014): 2339-2354.

Wozniak, D., Harbaugh, W. T., and Mayr, U., "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices." *Journal of Labor Economics*, 32.1 (2014): 161-198.