universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Effects of Frames on Performances in and Preferences for Competitions“

verfasst von / submitted by

Dominik Stelzeneder, Bakk.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 913
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Volkswirtschaftslehre
Betreut von / Supervisor:	Univ.-Prof. Dr. Wieland Müller

Acknowledgments

I would like to thank my supervisor Univ.-Prof. Dr. Wieland Müller for his support, his patience, his motivation and his trust in me. I could not have imagined having a better advisor for my thesis.

My sincere thanks also go to Kai Barron and Sebastian Schweighofer-Kodritsch, without them I would not have succeeded in finishing this thesis. They inspired me for experimental economics and gave me the opportunity to design and to perform a laboratory experiment with them and to analyze the collected data in this thesis. I am grateful for their helpful feedback and advice. It was and still is a great honor and pleasure for me to work with them.

I would also like to thank Sophie Redlinger for her invaluable support, she enriches my life like nobody else. I thank my fellow students and friends for their support and helpfulness, for the great time we have spent together and for all the fun we have had in the last years. And I thank Fabian for proofreading. Last but not least I would like to thank my family for supporting me throughout my life.

1 Introduction

In the last years a great number of articles has been published, in which a new potential explanatory factor for the widely observed differences in labor market outcomes between men and women (OECD 2015) was examined that had not been considered before: *Competitiveness*.

In their seminal article, Gneezy, Niederle, and Rustichini (2003) showed that when women and men have to perform certain tasks in a noncompetitive piece rate payment setting, their mean performances do not differ. However, if they have to perform these tasks in a competitive *tournament* setting, in which only the best performing participants receive a payment, women perform significantly poorer than men. Women also seem to "shy away from competition": Niederle and Vesterlund (2007) showed that women tend to self-select less into competitions than men, given a noncompetitive alternative.

One consequence of these findings is that institutions such as universities that perform competitive entrance exams to allocate places, for instance, put themselves in a delicate situation: On the one hand, university places should be alloted to the best performing applicants, but on the other hand, gender discrimination is not a desirable outcome either. In Austria in 2012 a system of gender quotas for medical studies was implemented in an attempt to square the circle. However, this quota was widely and controversially discussed and legally problematic (Hodasz 2015).

Our research tries to find and to open up new ways to reconcile the two potentially conflicting objectives of performance focus and gender equality. Many authors have shown that performances and tournament selection rates of men and women are task and stereotype dependent (Günther et al. (2010), Shurchkov (2012), Dreber, Essen, and Ranehill (2014), and Grosse, Riener,

and Dertwinkel-Kalt (2014)). But in order to promote the best performing examinees *and* to provide gender equality, it might not even be necessary to change the given tasks. We show that already the framing of tasks in competitive exams has a systematic impact on participants' performances and on tournament selection rates. To preview our results, we find that women who were asked to solve tasks that were framed as *intelligence* tasks performed significantly worse than men or than women who performed tasks that were framed as *creativity* tasks. When participants had the choice between a piece rate payment and a competitive incentive scheme, significantly more men selected the tournament in the *creativity* frame than in the *intelligence* frame or than women in general.

The thesis proceeds as follows: In Section 2 we review the literature about gender differences in tournament performances and different preferences to compete. We then present our experimental design in Section 3, formulate testable hypotheses in Section 4 and proceed with the main findings and results in Section 5. Section 6 concludes with a discussion of our findings.

2 Literature Review

Our research is related to two basic experimental observations, which we want to review in this chapter. First, several authors found differences in tournament performances between men and women, although their performances in a noncompetitive payment setting did not differ. The second observation is that men and women seem to have different preferences for performing in a competition. We begin by reviewing the literature about gender specific performance differences in tournaments and proceed with an overview of the literature about different preferences for competitions.

2.1 Gender Differences in Tournament Performances

In their influential experimental study, Gneezy, Niederle, and Rustichini (2003) asked participants to solve mazes under a piece rate payment and under a competitive (tournament) payment condition. In the piece rate payment task men and women did not perform differently but in the tournament payment task, in which only the best performing participant received payment, they did. Men's performance increased as competition was introduced and therefore they outperformed women significantly.

Many studies with the same experimental setting and just minor modifications followed. The findings of these subsequent studies are mixed, and some are even contradictory. Some authors were able to replicate the initial study results of Gneezy, Niederle, and Rustichini (2003) (see Günther et al. 2010), but when running tasks were used instead of mazes, several authors found men to increase their performances more than women, whereas others did not find such evidence (Gneezy and Rustichini (2004), Cárdenas et al. (2012), Dreber, Essen, and Ranehill (2011), Sutter and Rützler (2010)). When participants were asked to perform simple math tasks, performance changes from piece rate to tournament incentives were not different between man and women in most studies (Niederle and Vesterlund (2007), Dreber, Essen, and Ranehill (2014), Buser, Ranehill, and Veldhuizen (2017), Veldhuizen (2017), Wozniak, Harbaugh, and Mayr (2014), but also Grosse, Riener, and Dertwinkel-Kalt (2014), and Cárdenas et al. (2012) for different findings).

Günther et al. (2010) argue that this variety of results can be explained with *stereotype threat*. "Stereotype threat", as defined by Steele and Aronson, "is being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele and Aronson 1995, p. 797). This implies that an existing stereotype might negatively affect task performance of people who

are part of a negatively stereotyped group.

To test if the performances of men and women in tournaments are task and stereotype dependent, Günther et al. (2010) designed an experiment with what they considered a stereotypically male, a gender-neutral and two female tasks. In the "male" task, participants had to solve maze games, comparable to Gneezy, Niederle, and Rustichini (2003). In the "gender-neutral" task, participants were shown a slide with a letter on it and they were asked to write down as many words as possible that started with this letter. The first "female" task was a memory performance task. Participants had to read short texts and, after performing the second "female" task, they had to answer questions about the texts out of their memory. In the second "female" task, the participants were shown three pictures, and they had to identify the picture that was not identical to the others.

The experimental results were in line with the predictions of stereotype threat. In the "gender-neutral" task, both genders increased their performances from the piece rate to the competitive incentive scheme. In the "male" task, only men's and not women's performance increased, and in one "female" task women's, but not men's performance increased (in the memory performance task neither men's nor women's performance changed).

We show that it is not even necessary to change the tasks in order to evoke different gender specific performance patterns, but that a different framing of the given task is sufficient to change tournament performances. In our experiment, participants were asked to perform Raven's progressive matrices, a test in which participants have to identify visual patterns and to complete them with the missing elements. We use this task because Raven's matrices are not categorized as stereotypically male or female, as far as we know, in contrast to *math* or *fashion* tasks for instance. However, they can

easily be labeled with categories that might be *perceived* as favoring men or women to different extents.

We presume that Raven's matrices can be seen as *"creativity"* and/or as *"intelligence"* tasks, and we frame our experiment within these two categories and use these frames as our treatments. Many studies document that men are perceived as being more intelligent than women, especially by themselves but also by others (Hogan (1978), Bennett (1996), Furnham and Rawles (1995) and Furnham and Rawles (1999).[1] Creativity, in contrast, is not seen as stereotypically *male* or *female* (Goldsmith and Matherly (1988), Runco (1986), Chan (2005) and Henderson (2002)).[2]

2.2 Gender Differences in Tournament Entry Rates

In their seminal article, Niederle and Vesterlund (2007) extended the experiment of Gneezy, Niederle, and Rustichini (2003) and discovered that men and women have different preferences to opt for the competition, given a piece rate payment alternative. In their experiment participants were asked to add up sets of five two-digits numbers for five minutes in each task. The participants performed the first task under a noncompetitive piece rate payment condition and the second task under a competitive tournament incentive scheme. Thereafter, the participants were asked to choose their preferred remuneration scheme for the third task.

The results were remarkable, 73% of men and only 35% of women chose the *tournament*. This gender gap in tournament entry could not be explained by differences in ability, risk preferences or (over-) confidence. Hence, the authors inferred that men and women simply exhibit different preferences

[1]For an overview see Furnham, Hosoe, and Tang (2001).

[2]For an overview see Baer and Kaufman (2008).

for participating in competitions.

Many studies with a similar experimental design followed and basically confirmed the original finding of gender differences in tournament entry rates and consequently in preferences for competitions (see Niederle and Vesterlund 2011 for an overview and a discussion). However, some studies found evidence for cultural and/or task specific differences in the observed gender gaps in tournament entry rates: Gneezy, Leonard, and List (2009), for example, discovered that in a matrilineal society the gender gap was reversed and more women than men choose the competition. Other important factors for the tournament entry gap were the domain of the task (Wieland and Sarin 2012) and gender-task stereotypes (Shurchkov 2012, Dreber, Essen, and Ranehill 2014 and Grosse, Riener, and Dertwinkel-Kalt 2014, but also Wozniak, Harbaugh, and Mayr 2014 for a different result).

We contribute to this literature by showing that tournament entry rates can already be affected by different frames and hence, a different perception of a task by the participants. Like Niederle and Vesterlund (2007), we extended the experiment of Gneezy, Niederle, and Rustichini (2003) and asked participants to choose either a noncompetitive piece rate or a competitive tournament remuneration scheme. We tested whether men and women have different preferences to compete, and we examined whether these preferences depend on the *framing* and in turn on the *perception* of the task.

Contrary to the well documented effect of a negative stereotype on task performances, the effect on tournament entry decisions is less clear. However, in the experiments of Shurchkov (2012) and Dreber, Essen, and Ranehill (2014) the women's tournament decision rates varied under differently stereotyped tasks, whereas the men's tournament entry rates did not differ. We, therefore, expect that in our treatments women will prefer to enter the

competition less when they have to perform in the stereotypically male *intelligence* frame as compared to women in the *creativity* frame and men in either frame.

3 Experimental Design

We employed a 2 (gender) x 2 (frames) experimental design with five tasks and a questionnaire. In the first three tasks participants had to solve Raven's matrices with different remuneration conditions comparable to Niederle and Vesterlund (2007). In Task 1, participants were paid according to a piece rate incentive scheme, in Task 2 according to a competitive incentive scheme, and in Task 3 participants were asked to choose their preferred remuneration scheme. Subsequent tasks consisted of a self-confidence task and a risk preference elicitation lottery described by Eckel and Grossman (2002).[3]

At the end of the experiment the participants received a show-up fee of €5, their individual payment from one randomly selected task and a payment depending on their decision in the lottery.

3.1 Tasks with Varying Payment Schemes

In the first three tasks participants were asked to solve as many Raven's matrices as possible within 2 minutes. Before each task participants were informed about the number of available matrices: 16 matrices in the first two tasks and 15 matrices in the third task.[4] The matrices had four levels of difficulty and each of the matrices had either 6 or 8 answer options with

[3]In the lottery, the participants had the choice between 6 options with varying expected values and risk levels. Option 1 guaranteed a fixed payoff of €1,40; from Option 2 to Option 5 expected payoffs as well as risk levels increased, and Option 6 had the same expected value as Option 5 but it was riskier.

[4]The Raven's matrices were in the same order for all participants in all sessions.

exactly one correct answer. Participants were informed that wrong answers did not cause any losses and that only correct answers were counted. The number of correctly solved matrices in a task was counted and shown to the participants on their screens. At the beginning of the experiment, each participant was shown one example matrix.

In Task 1 (*piece rate*) participants received €0,50 for every correct answer. In Task 2 (*tournament*) the participants of each treatment group were randomly assigned to tournament groups of four and the best performing participant of each tournament group received €2 for each of his or her correct answers. The other group members got nothing and in case of a tie the computer randomly selected one of the best performing group members as the winner. Participants were informed that they wouldn't know whether they had won or lost the tournament until the end of the experiment.

Task 3 was the *decision* task and participants had to choose their preferred payment scheme. The available payment options corresponded to the above described *piece rate payment* and the *tournament payment*. To rule out the potential influence of social preferences, participants had to compete against their group members' past performances of Task 2 if they chose the tournament payment. This way, selecting the competition did not affect the outcome or winning chance of any other participant.

3.2 Treatments

At the beginning of the experiment participants were randomly assigned to the *creativity* or the *intelligence* treatment group. The tasks and their order were the same for all participants, the only differences that we introduced were:

1. The tasks were *labeled* as *"creativity"* tasks in one group and as *"intel-*

ligence" tasks in the other.

2. An italic font was used for the tasks' headlines of the *"creativity"* treatment group and a bold font for the *"intelligence"* group.

3. A small image was added to the headlines of Task 1 to Task 3, a colorful tree for the *"creativity"* frame and the shape of a head with gear wheels inside it for the *"intelligence"* frame.

3.3 Elicitation of self-confidence

Following Veldhuizen (2017) we used the self-confidence elicitation approach described by Mobius et al. (2011). The main advantage of this belief elicitation method is that it is incentive-compatible, meaning that it is in participants' best interest to state their true beliefs about their probability to win. Furthermore, it is independent of participants' risk preferences and it offers many distinct answer options for a more detailed statistical evaluation.

In Task 4 the participants were told that the remuneration for this task depended on their Task 2 performances: If the tournament in Task 2 had been won[5] the payment for Task 4 was €8 and €0 otherwise. But the participants were also informed that there were one hundred different "robots", for which the chances of winning in tournaments were programed to be 1%, 2%, 3%,...,99% or 100%. Instead of taking their own performances in Task 2, the participants could decide to let a randomly selected robot perform instead of them. To do so, they were asked to state a "winning chance threshold". If the randomly selected robot's winning chance was below this threshold, the participant's performance was taken for payment determination, but if

[5]At this stage of the experiment participants only knew about their absolute performances in Task 2 but not whether they had won the tournament or not.

the robot's winning chance was above this threshold, the robot's performance was used instead. The stated "winning chance thresholds" represent the subjective Task 2 (*tournament*) winning chances of each participant and were used as the measure of participants' levels of self-confidence.

4 Testable Hypotheses

4.1 The Effect of Framing on Tournament Performances

In accordance with the theoretical discussion about stereotype threat we expect women's tournament performances to be negatively affected by the *intelligence* frame. The experimental design of Gneezy, Niederle, and Rustichini (2003) and our treatments with a stereotypically male (*intelligence*) and a more neutral (*creativity*) frame enable us to test the following hypotheses about gender specific performance differences:

Hypothesis 1 *Women perform better in Task 2 (tournament) if tasks are framed as "creativity" instead of "intelligence" tasks.*

Hypothesis 2 *Men's mean performance in Task 2 (tournament) does not significantly differ if tasks are framed as "intelligence" or as "creativity" tasks.*

Hypothesis 3 *Men outperform women in Task 2 (tournament) in the intelligence frame.*

Hypothesis 4 *Men's and Women's mean performances in Task 2 (tournament) do not significantly differ in the creativity frame.*

4.2 The Effect of Framing on the Decision to Compete

Based on the finding of Shurchkov (2012) and Dreber, Essen, and Ranehill (2014) that mainly women's tournament entry rates are affected by treatments with differently stereotyped tasks, we test the following hypotheses about framing effects on tournament entry decisions:

Hypothesis 5 *The probability that a woman chooses the tournament payment is significantly larger in the creativity treatment than in the intelligence treatment.*

Hypothesis 6 *The probability that a man chooses the tournament payment does not differ between the creativity treatment and the intelligence treatment.*

Hypothesis 7 *The effect of the intelligence and the creativity frames on the probability to choose the tournament payment is significantly different for men and women.*

Or equivalent to Hypothesis 7:

Hypothesis 8 *The gender gap in tournament entry rates is significantly different between the creativity frame group and the intelligence frame group.*

5 Experimental Results

The experiment was programmed using oTree[6] and conducted at the experimental economics laboratory of the Technical University of Berlin in August 2017. There were four sessions with 24 participants each, in total 47 women and 49 men participated. 24 men and 24 women were in the *intelligence* treatment, and 25 men and 23 women in the *creativity* treatment. Gender

[6]Chen, Schonger, and Wickens (2016)

was never explicitly mentioned but participants could see each other in the laboratory and hence, determine the gender composition of the whole group within a session. The share of females was between 0.42 and 0.54 in the four sessions.

The participants were randomly assigned to a computer in the laboratory. They were informed about the approximate duration of the experiment and that they would receive a show-up fee of €5 and a variable payment depending on their performance in one randomly selected task.

5.1 Effect of Framing on Mean Performances

Prima facie, there were no performance differences between both genders, neither in Task 1 (p=0.90, Mann-Whitney U Test) nor in Task 2 (p=0.33, Mann-Whitney U Test). Looking at both treatment groups separately, however, revealed a significantly lower mean performance of women when tasks were framed as *intelligence* tasks in Task 2 (*tournament*) compared to men's or women's (in the *creativity* frame) mean performances. Other Task 2 mean performance comparisons did not show any significant differences (see Table 2) and we did not observe any performance differences in Task 1 (*piece rate*), neither between frames nor between genders (p>0.20 for all four comparisons, Mann-Whitney U Test, see Table 1).

We confirm Hypothesis 1 that women perform better in Task 2 (*tournament*) when tasks are framed as *creativity* rather than as *intelligence* tasks (6.00 compared to 7.52 solved mazes, p=0.02, Mann-Whitney U Test, see Table 2). Men's mean performances in Task 2 did not significantly differ between both frames and hence, Hypothesis 2 can also be confirmed (7.25 compared to 7.24 solved mazes, p=0.87, Mann-Whitney U Test, see Table 2).

Table 1: Task 1 (*piece rate*) mean performances by gender and frame

Frame	Mean number of correct answers in Task 1		
	Men	Women	MWU-Test
Intelligence	4.67	4.50	p=0.80
Creativity	4.88	5.04	p=0.64
MWU-Test	p=0.49	p=0.21	

Notes. MWU-Test reports p-values of the Mann-Whitney U Test

Table 2: Task 2 (*tournament*) mean performances by gender and frame

Frame	Mean number of correct answers in Task 2		
	Men	Women	MWU-Test
Intelligence	7.25	6.00	p=0.06
Creativity	7.24	7.52	p=0.57
MWU-Test	p=0.87	p=0.02	

Notes. MWU-Test reports p-values of the Mann-Whitney U Test

Consequently, differences in men's and women's mean performances were highly frame dependent: When tasks were framed as *intelligence* tasks, men's mean performance in Task 2 *tournament* was significantly higher than women's mean performance (7.25 compared to 6.00 correct answers, p=0.06, Mann-Whitney U Test, see Table 2). This better mean performance of men compared to the performance of women in the *intelligence* frame was predicted by Hypothesis 3, which, therefore, can be confirmed by our data. When tasks were framed as *creativity* tasks mean performances of men and women were not significantly different (7.52 compared to 7.24 correct answers, p=0.57, Mann-Whitney U Test, see Table 2), we thus also confirm Hypothesis 4.

Table 3: Decisions to compete by gender and frame

Frame	% choosing to compete				
	Men	N	Women	N	Fisher's exact test
Intelligence	37.5%	24	25%	24	p=0.534
Creativity	64%	25	22%	23	p=0.004
Fisher's exact test	p=0.089		p=1.000		

5.2 Gender Gaps in Tournament Entry Rates by Frame

5.2.1 Creativity Frame

When tasks were framed as *creativity* tasks 22% of the women chose the tournament, which was significantly less compared to 64% of all men (see Table 3). This gender gap of 42 percentage points between men's and women's tournament entry rates is highly significant (p=0.004, Fisher's exact test) and comparable in size to Niederle and Vesterlund (2007).

OLS and probit regressions[7] with tournament decisions as dependent variables show that *gender* is a major predictor for tournament entry and highly significant (see Table 4). Including only performance controls (Model 1, Table 4) or also risk preferences and self-confidence controls (Model 2, Table 4) into our regressions yields roughly the same estimate of -0.43 for the variable *"Female"*. This means that the probability that a woman chooses the tournament is 43 percentage points lower than a man's probability to choose the tournament - given the same risk preferences, performance improvements, Task 2 performances and participants' confidence levels.

Only performance improvement, i.e. the number of correctly solved mazes in Task 2 that exceeded the number of correct answers in Task 1, was another

[7]For the purpose of a more intuitive interpretation of the linear probability model compared to the probit model we describe our results using OLS estimates primarily. Note that the significance levels differ only once between both models.

factor with explanatory power for predicting tournament entry decisions. Solving one additional Raven's matrix in Task 2 compared to Task 1 increased the probability to choose the tournament by 10.1 percentage points. Eckel and Grossman (2002) measures of risk-preferences and self-confidence (the individual subjective belief about the probability that Task 2 *(tournament)* had been won, elicited in Task 4 (*belief elicitation*)) were no significant factors for participants' decisions to compete.

Table 4: Determinants for tournament entry in the *creativity* frame

	OLS		Probit	
	Model 1	Model 2	Model 1	Model 2
Female	-0.428***	-0.435***	-1.294***	-1.346***
	(0.126)	(0.126)	(0.420)	(0.433)
Perf. Improvement	0.101**	0.101**	0.342**	0.337**
	(0.047)	(0.049)	(0.166)	(0.168)
Performance Task 2	-0.025	-0.018	-0.096	-0.061
	(0.038)	(0.039)	(0.128)	(0.133)
Eckel-Grossman		-0.050		-0.172
		(0.047)		(0.181)
Confidence		0.001		0.002
		(0.003)		(0.010)
Constant	0.581**	0.680**	0.302	0.697
	(0.248)	(0.306)	(0.748)	(0.999)
Observations	48	48	48	48
Log-likelihood value	-	-	-25.449	-24.968
Pseudo R-squared	0.275	0.292	0.226	0.241

Notes. OLS Estimates: robust standard errors in parentheses, Pseudo R-squared reports R-squared. Probit Estimates: standard errors in parentheses. Dependent variable is the decision in Task 3 (1-tournament payment, 0-piece rate payment).
** p<0.05, *** p<0.01

5.2.2 Intelligence Frame

The experimental results for the participants whose tasks were framed as *intelligence* tasks were quite different to the *creativity* group's results. The gender gap in tournament entry rates, which was large and highly significant in the *creativity* group, was between -0.126 (with performance controls, Model 1, Table 5) and -0.143 (with performance, risk preferences and confidence controls, Model 2, Table 5) in the *intelligence* group, and not significant in any of the two regression models. Instead, confidence, and in the probit model performance in Task 2 *(tournament)*, were the only significant predictors for tournament entry.

5.3 Effect of Framing on Decisions to Compete

Hypothesis 5 predicted that relatively more women would choose the tournament in the *creativity* frame than in the *intelligence* frame. Women's decisions to compete, however, were not significantly different between both frames, 25% of women chose the tournament when tasks were framed as *intelligence* tasks compared to 22% when they were framed as *creativity* tasks. For the women, only performance improvements from Task 1 (*piece rate*) to Task 2 (*tournament*) significantly affected tournament entry decisions (Table 6). Therefore, we have to reject Hypothesis 5.

Hypothesis 6 predicted that men are as likely to choose the tournament in the *intelligence* frame as in the *creativity* frame. However, men's probabilities to compete were significantly different between both frames: 64% of men chose the tournament in the *creativity* frame and 37.5% in the *intelligence* frame (p=0.089, Fisher's exact test).

The framing effect accounted for a different tournament selection proba-

Table 5: Determinants for tournament entry in the *intelligence* frame

	OLS		Probit	
	Model 1	Model 2	Model 1	Model 2
Female	-0.126	-0.143	-0.364	-0.518
	(0.155)	(0.135)	(0.396)	(0.460)
Perf. Improvement	0.027	0.024	0.088	0.166
	(0.035)	(0.034)	(0.116)	(0.143)
Performance Task 2	-0.024	-0.054	-0.074	-0.274*
	(0.036)	(0.034)	(0.110)	(0.151)
Eckel-Grossman		0.039		0.253
		(0.048)		(0.169)
Confidence		0.010***		0.045***
		(0.003)		(0.014)
Constant	0.482*	0.038	-0.011	-2.266*
	(0.275)	(0.234)	(0.708)	(1.178)
Observations	48	48	48	48
Log-likelihood value	-	-	-29.032	-21.316
Pseudo R-squared	0.031	0.264	0.026	0.285

Notes. OLS Estimates: robust standard errors in parentheses, Pseudo R-squared reports R-squared. Probit Estimates: standard errors in parentheses. Dependent variable is the decision in Task 3 (1-tournament payment, 0-piece rate payment).
* p<0.10, *** p<0.01

Table 6: Women's decision-making factors for tournament entry

	OLS		Probit	
	Model 1	Model 2	Model 1	Model 2
Creativity Frame	-0.110	-0.123	-0.320	-0.313
	(0.142)	(0.153)	(0.468)	(0.481)
Perf. Improvement	0.085**	0.081**	0.357**	0.391**
	(0.033)	(0.034)	(0.170)	(0.187)
Performance Task 2	-0.003	-0.006	-0.054	-0.100
	(0.043)	(0.044)	(0.143)	(0.158)
Eckel-Grossman		0.032		0.165
		(0.056)		(0.181)
Confidence		0.001		0.004
		(0.003)		(0.011)
Constant	0.143	0.007	-1.041	-1.664*
	(0.252)	(0.332)	(0.795)	(0.997)
Observations	47	47	47	47
Log-likelihood value	-	-	-22.238	-21.674
Pseudo R-squared	0.126	0.140	0.130	0.153

Notes. OLS Estimates: robust standard errors in parentheses, Pseudo R-squared reports R-squared. Probit Estimates: standard errors in parentheses. Dependent variable is the decision in Task 3 (1-tournament payment, 0-piece rate payment).
* p<0.10, ** p<0.05

bility of 23.9 to 27.1 percentage points (OLS regressions with robust standard errors, dependent on the included control variables, see Model 1 and Model 2 of Table 7). Consequently, Hypothesis 6 has to be rejected.

Table 7: Men's decision-making factors for tournament entry

	OLS		Probit	
	Model 1	Model 2	Model 1	Model 2
Creativity Frame	0.271*	0.239*	0.698*	0.801*
	(0.143)	(0.136)	(0.369)	(0.409)
Perf. Improvement	0.027	0.016	0.074	0.034
	(0.045)	(0.044)	(0.124)	(0.130)
Performance Task 2	-0.016	-0.031	-0.043	-0.087
	(0.035)	(0.035)	(0.103)	(0.120)
Eckel-Grossman		0.008		0.062
		(0.043)		(0.167)
Confidence		0.011***		0.035***
		(0.003)		(0.012)
Constant	0.417*	-0.017	-0.197	-1.919*
	(0.235)	(0.208)	(0.620)	(1.013)
Observations	49	49	49	49
Log-likelihood value	-	-	-32.030	-26.538
Pseudo R-squared	0.077	0.261	0.057	0.218

Notes. OLS Estimates: robust standard errors in parentheses, Pseudo R-squared reports R-squared. Probit Estimates: standard errors in parentheses. Dependent variable is the decision in Task 3 (1-tournament payment, 0-piece rate payment).
* p<0.10, *** p<0.01

Hypotheses 7 and 8 predicted that men and women were differently affected by our treatments and that, therefore, the gender gap in tournament entry rates would significantly differ between the *creativity* and the *intelligence* frame. These hypotheses can be confirmed, the framing effects are significantly different for men and women as a difference in differences esti-

mation shows (p=0.083 and p=0.098 for Model 1 and Model 2 respectively, t-tests, see Appendix A.1 and Table 8 for details). The gender gap in tournament entry rates is thus significantly dependent on the frame and, therefore, on participants' perception of the tasks.

6 Discussion and Conclusion

We performed a laboratory experiment comparable to Niederle and Vesterlund (2007) in which participants had to solve tasks that were either framed as *creativity* or as *intelligence* tasks. Our research design enabled us to analyze the effects of different frames on performances and tournament entry rates of men and women, and to show that even minor changes in the framing of tasks can have significant effects on experimental outcomes.

In the piece rate payment task performances were not significantly different between men and women in any frame, but we observed significant performance differences in the competitive task: Women performed significantly worse when tasks were framed as *intelligence* tasks, but they performed as well as men when tasks were framed as *creativity* tasks.

In contrast, the decisions to compete were frame dependent for men but not for women. Men were significantly more likely to enter a competition when tasks were framed as *creativity* tasks, whereas women's probability to enter the competition was not different between both frames. The overall size of the gender gap in tournament entry rates was thus highly frame dependent.

These findings suggest that it is not even necessary to change tasks in order to evoke different tournament performance and tournament entry patterns of men and women. Changing the framing of a task and hence, the perception by the participants can already have an impact on these two varia-

bles. *Stereotype threat* seems to be an important influential factor with which a large share of tournament performance differences between both genders can be explained. However, contrary to our expectations, only men's and not women's tournament entry decisions were affected by the treatment. This also runs contrary to the results of Shurchkov (2012) and Dreber, Essen, and Ranehill (2014), and, from the authors' point of view, men were affected in a counterintuitive way. Based on the literature about subjective assessments of peoples' intelligence and creativity we concluded that *intelligence* is a male stereotype and that *creativity* is more gender neutral. We belief that this presumption was indirectly confirmed by the tournament performance data, since women - in line with predictions of stereotype threat - performed significantly worse in the *intelligence* frame, whereas men's performance was not affected. Our results suggest that the perception of a task is an important determinant for tournament entry, but they do not confirm the finding of other authors who argued that stereotypically male tasks in general increase the gender gap in favor of men (see Shurchkov (2012), Grosse, Riener, and Dertwinkel-Kalt (2014), or Dreber, Essen, and Ranehill (2014)).

In conclusion, the way in which competitive situations are framed can already affect peoples' willingness to enter a tournament and the performances that are realized. Many firms and institutions try to increase gender equality by asking job or study applicants to perform an assessment test or an entrance exam. Our study results may help them to design better tests since our findings suggest that shaping information and the presentation of these tests are important elements for achieving the objective of providing equal opportunities for men and women, and the promotion of the most qualified potential candidates.

A Appendix

A.1 Difference-in-differences estimation details

Our equation for analyzing the average difference in treatment effect on tournament entry rates between men and women is the following:

$$y = \beta_0 + \beta_1 \cdot female + \beta_2 \cdot creativity + \delta \cdot female \cdot creativity + \gamma \cdot C_{1,2} + \varepsilon$$

y denotes the decision in Task 3 (1 if tournament payment and 0 if piece rate payment was chosen by the participant), $female$ and *creativity* are dummy variables for the participant's gender and his or her randomly assigned frame (1 if tasks were framed as *creativity* tasks and 0 otherwise), $C_{1,2}$ represents several control variables for Model 1 and Model 2 and ε is the error term.

Let $\bar{y}_{M,I}$ be defined as the sample average of y for men in the *intelligence* frame, $\bar{y}_{M,C}$ as the sample average of y for men in the *creativity* frame and $\bar{y}_{F,I}$ and $\bar{y}_{F,C}$ as the sample averages of y for women in the *intelligence* and in the *creativity* frame, respectively.

Then, the OLS estimator $\hat{\delta}$ can be expressed as

$$\hat{\delta} = (\bar{y}_{F,C} - \bar{y}_{F,I}) - (\bar{y}_{M,C} - \bar{y}_{M,I})$$

and interpreted as the average difference in treatment effect on tournament entry decisions between men and women.

Several standard experimental arrangements were made to ensure correct causal inference: The participants were randomly assigned to one of the two treatments before the experiment began and the intervention was unrelated to any individual characteristic of the participants. Spillover effects, i.e. effects of the treatment status of one or more participants on other participants, were prevented by forbidding communication during the experiment

and ensuring that participants could only see their own screens. Information about the treatments of other participants were, of course, not given to any participant. The absence of spillover effects and identical treatments of every participant within each group ensure that the stable unit treatment assumption (SUTVA) holds. Since we do not compare outcomes at different points in time, the *Parallel Trend Assumption*[8] is not relevant for the interpretation of our difference-in-differences estimation results.

Table 8: Difference in differences estimation

	Model 1		Model 2	
Creativity Frame	0.277*	(0.143)	0.260*	(0.134)
Female	-0.087	(0.147)	-0.104	(0.135)
Female*Creativity	-0.337*	(0.192)	-0.313*	(0.187)
Perf. Improvement	0.056**	(0.027)	0.052*	(0.028)
Performance Task 2	-0.018	(0.026)	-0.031	(0.027)
Eckel-Grossman			0.006	(0.035)
Confidence			0.006***	(0.002)
Constant	0.362*	(0.203)	0.140	(0.205)
Observations	96		96	

Notes. OLS Estimates, robust standard errors in parentheses. Dependent variable is the decision in Task 3 (1-tournament payment, 0-piece rate payment).
* p<0.10, ** p<0.05, *** p<0.01

[8]The *Parallel Trend Assumption* ensures that two groups, i.e. the treatment and the control group, have parallel trends in outcome. Simply put, the treatment group *would have* the same outcome as the control group *if there were* no treatment or intervention.

A.2 Abstract

Beginning with Gneezy, Niederle, and Rustichini (2003) and Niederle and Vesterlund (2007) experimental economists have intensively discussed and explored gender differences in tournament performances and preferences for competitions. In this thesis we examine the hypothesis that these gender differences are highly dependent on the *perception* of the given tasks. We conducted a 2 (gender) x 2 (frame) experiment in which participants had to solve Raven's matrices that were either framed as *intelligence* or as *creativity* tasks. We find significant performance differences in the tournament setting (but not in the noncompetitive setting) as well as tournament entry rate differences between both treatments. In the *creativity* frame women performed as well as men with either frame but they performed more poorly in the *intelligence* frame. The tournament entry rates of women were not different between both frames, and the women chose the competition as often as the men when men's tasks were framed as *intelligence* tasks. Men who performed *creativity* framed tasks, however, had a significantly higher probability to opt for the tournament.

A.3 Zusammenfassung

Beginnend mit den Artikeln von Gneezy, Niederle und Rustichini (2003) und Niederle und Vesterlund (2007) wurden geschlechterspezifische Leistungsunterschiede in Turnieren sowie geschlechterspezifisch unterschiedliche Präferenzen für Wettbewerbe unter experimentellen Wirtschaftsforschern intensiv erforscht und diskutiert. In dieser Arbeit untersuchen wir die Hypothese, dass diese Unterschiede stark von der *Wahrnehmung* der zu bearbeitenden Aufgaben abhängig sind. Wir führten dafür ein 2 (Geschlechter) x 2 (Frames) Ex-

periment durch, in dem Proband_innen Raven Matrizen, die entweder als *Intelligenz-* oder als *Kreativitäts-* Aufgaben geframed waren, lösen mussten. Wir fanden sowohl Leistungsunterschiede im Turniersetting (aber nicht im nichtkompetitiven Setting) als auch unterschiedliche Präferenzen für Wettkämpfe zwischen beiden Treatments. Wenn die Aufgaben als *Kreativitäts-* Aufgaben geframed waren, dann lösten die Frauen im Durchschnitt gleich viele Aufgaben wie die Männer in beiden Frames. Waren die Aufgaben jedoch als *Intelligenz-* Aufgaben geframed, so war die durchschnittliche Leistung der Frauen im Turniersetting signifikant niedriger. Der Anteil der Frauen, die sich für das Turnier und nicht für das nichtkompetitive Setting entschieden, war nicht verschieden bei beiden Frames und entsprach dem Anteil der Männer, die unter dem *Intelligenz*-Frame das Turnier wählten. Im *Kreativitäts*-Frame hingegen war für Männer die Wahrscheinlichkeit, das Turnier zu wählen, signifikant höher.

References

Andersen, Steffen, Seda Ertac, Uri Gneezy, John A. List, and Sandra Maximiano. "Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society". In: *The Review of Economics and Statistics* 95.4 (2013), pp. 1438–1443.

Baer, John and James C. Kaufman. "Gender Differences in Creativity". In: *The Journal of Creative Behavior* 42.2 (2008), pp. 75–105.

Bennett, Mark. "Men's and Women's Self-Estimates of Intelligence". In: *The Journal of Social Psychology* 136.3 (1996), pp. 411–412.

Buser, Thomas, Eva Ranehill, and Roel van Veldhuizen. "Gender Differences in Willingness to Compete: The Role of Public Observability". In: *WZB Discussion Paper* SP II 2017.203 (2017).

Chan, David W. "Self-perceived creativity, family hardiness, and emotional intelligence of Chinese gifted students in Hong Kong". In: *Journal of Secondary Gifted Education* 16.2-3 (2005), pp. 47–56.

Chen, Daniel L., Martin Schonger, and Chris Wickens. "oTree—An open-source platform for laboratory, online, and field experiments". In: *Journal of Behavioral and Experimental Finance* 9 (2016), pp. 88–97.

Cárdenas, Juan-Camilo, Anna Dreber, Emma von Essen, and Eva Ranehill. "Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden". In: *Journal of Economic Behavior and Organization* 83 (2012), pp. 11–23.

Dreber, Anna, Emma von Essen, and Eva Ranehill. "Gender and competition in adolescence: task matters". In: *Experimental Economics* 17.1 (2014), pp. 154–172.

— "Outrunning the gender gap-boys and girls compete equally". In: *Experimental Economics* 14 (2011), pp. 567–582.

Eckel, Catherine C. and Philip J. Grossman. "Sex differences and statistical stereotyping in attitudes toward financial risk". In: *Evolution and Human Behavior* 23 (2002), pp. 281–295.

Furnham, Adrian, Tatsuro Hosoe, and Thomas Li-Ping Tang. "Male hubris and female humility? A crosscultural study of ratings of self, parental, and sibling multiple intelligence in America, Britain, and Japan". In: *Intelligence* 30 (2001), pp. 101–115.

Furnham, Adrian and Richard Rawles. "Correlations between self-estimated and psychometrically measured IQ". In: *The Journal of Social Psychology* 139.4 (1999), pp. 405–410.

— "Sex differences in the estimation of intelligence". In: *Journal of Social Behavior and Personality* 10.3 (1995), p. 741.

Gneezy, Uri, Kenneth L. Leonard, and John A. List. "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society". In: *Econometrica* 77.5 (2009), pp. 1637–1664.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. "Performance in Competitive Environments: Gender Differences". In: *The Quarterly Journal of Economics* 118.3 (2003), pp. 1049–1074.

Gneezy, Uri and Aldo Rustichini. "Gender and Competition at a Young Age". In: *The American Economic Review* 94.2 (2004), pp. 377–381.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel. "Women can't jump? - An experiment on competitive attitudes and stereotype threat". In: *Journal of Economic Behavior and Organization* 75 (2010), pp. 395–401.

Goldsmith, Ronald E. and Timothy A. Matherly. "Creativity and Self-Esteem: A Multiple Operationalization Validity Study". In: *The Journal of Psychology* 122.1 (1988), pp. 47–56.

Grosse, Niels Daniel, Gerhard Riener, and Markus Dertwinkel-Kalt. "Explaining gender differences in competitiveness: Testing a theory on gender-task stereotypes". In: *Unpublished Working Paper* (2014).

Henderson, Sheila J. "The Correlates of Inventor Motivation, Creativity and Achievement". PhD thesis. Stanford University, 2002.

Hodasz, Marta. "Gender-specific Scoring of Admission Tests for medical Studies is consistent with the Principle of Equality". In: *ICL Journal* 9.1 (2015).

Hogan, H Wayne. "IQ self-estimates of males and females". In: *The Journal of Social Psychology* 106.1 (1978), pp. 137–138.

Kaufman, James C. "Self-reported differences in creativity by ethnicity and gender". In: *Applied Cognitive Psychology* 20.8 (2006), pp. 1065–1082.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. "Managing self-confidence: Theory and experimental evidence". In: *NBER Working Paper Series* 17014 (2011).

Niederle, Muriel and Lise Vesterlund. "Do women shy away from competition? Do men compete too much?" In: *The Quarterly Journal of Economics* 122.3 (2007), pp. 1067–1101.

— "Gender and Competition". In: *The Annual Review of Economics* 3 (2011), pp. 601–630.

OECD. "Closing the gender gap: Act now". In: *OECD Publishing* (2015).

Raven, J. C., J. Raven, and J. H. Court. *A manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis, 1998.

Runco, Mark A. "Predicting Children's Creative Performance". In: *Psychological Reports* 59.3 (1986), pp. 1247–1254.

Shurchkov, Olga. "Under pressure: Gender differences in output quality and quantity under competition and time constraints". In: *Journal of the European Economic Association* 10.5 (2012), pp. 1189–1213.

Steele, C. M. and J. Aronson. "Stereotype threat and the intellectual test performance of African Americans". In: *Journal of personality and social psychology* 69.5 (1995), pp. 797–811.

Sutter, Matthias and Daniela Rützler. "Gender differences in competition emerge early in life". In: *Institute of Public Finance, University of Innsbruck. Working Papers in Economics and Statistics* 2010-14 (2010).

Veldhuizen, Roel van. "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?" In: *Working Paper* (2017).

Wieland, Alice and Rakesh Sarin. "Domain specificity of sex differences in competition". In: *Journal of Economic Behavior and Organization* 83 (2012), pp. 151–157.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. 2. ed. Cambridge, Massachusetts and London, England: MIT Press, 2010.

Wozniak, David, William T. Harbaugh, and Ulrich Mayr. "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices". In: *Journal of Labor Economics* 32.1 (2014), pp. 1–39.