

van Veldhuizen, Roel

**Article — Published Version**

## Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?

Journal of the European Economic Association

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* van Veldhuizen, Roel (2022) : Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?, Journal of the European Economic Association, ISSN 1542-4766, Oxford University Press, Oxford, Vol. 20, Iss. 4, pp. 1595-1618, <https://doi.org/10.1093/jeea/jvac031>

This Version is available at:

<https://hdl.handle.net/10419/261094>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# ONLINE APPENDIX FOR "GENDER DIFFERENCES IN TOURNAMENT CHOICES: RISK PREFERENCES, OVERCONFIDENCE OR COMPETITIVENESS?"

---

**Roel van Veldhuizen**

Lund University & WZB Berlin Social  
Science Center

---

---

E-mail: [roel.van\\_veldhuizen@nek.lu.se](mailto:roel.van_veldhuizen@nek.lu.se)

# CONTENTS

<b>Appendix A: Additional Analysis</b>	<b>A-1</b>
A1. Additional Results for Sections 3 and 4 . . . . .	A-1
A2. Measurement Error in Elicited Beliefs . . . . .	A-14
A3. Risk and Tournament Entry . . . . .	A-23
A4. Collecting Evidence . . . . .	A-31
A5. Decomposing the Gender Gap in Tournament Entry . . . . .	A-35
A6. Generating Figure 1 . . . . .	A-42
A7. Identifying Assumption . . . . .	A-48
A8. Power Calculation for Experiment 2 . . . . .	A-51
<b>Appendix B: Experimental Details and Procedures</b>	<b>B-1</b>
B1. Design of Experiment 2 . . . . .	B-1
B2. Instructions for Experiment 1 . . . . .	B-4
B3. Instructions for Experiment 2 . . . . .	B-15
<b>Appendix C: Original German Language Instructions</b>	<b>C-1</b>
C1. Instructions for Experiment 1 . . . . .	C-1
C2. Instructions for Experiment 2 . . . . .	C-11

## APPENDIX A: ADDITIONAL ANALYSIS

### *A1. Additional Results for Sections 3 and 4*

In this section, I present five sets of additional results related to the main analysis presented in Sections 3 and 4. First, I present a more detailed discussion of the tournament entry regressions in Table 2 and Table 4 along with several robustness checks. Second, I discuss the role of feedback aversion in experiment 1. Third, I provide a brief discussion of the similarities and differences in the main results of experiments 1 and 2. Fourth, I present formal tests for the role of order effects. Finally, I present additional figures summarizing gender differences in performance, probability of winning, and risk preferences.

#### **A1.1 Residualizing Competitiveness Using Regressions**

Table 2 and Table 4 estimate the role of competitiveness using the residual-based identification strategy of Niederle and Vesterlund (2007). As a first step, column (1) presents the gender gap in Stage 3 after controlling for gender differences in ability. Column (2) replicates the specification of Buser et al. (2014, Table 7) by adding the belief elicited in Stage 4 and two measures of risk preferences as controls. This specification is representative of the specifications used most commonly in the literature (see Table A11 in Appendix A6). Column (3) instead replicates the specification of Niederle and Vesterlund (2007, Table 8), who proxy for risk preferences and overconfidence using a choice that is similar to Stage 3 but lacks the competitive element. Finally, column (4) follows the approach of Gillen et al. (2019, Table 3) to control for multiple measures of risk attitudes and confidence.

To identify the importance of competitiveness, Niederle and Vesterlund (2007) and Gillen et al. (2019) compare the size of the gender gap in the respective column to column 1. In my experiments, the estimated importance of competitiveness varies from 49.8% to 90.0% depending on the experiment and specification used (Table 2 and Table 4). The main implication of these results is that regression-based estimates for competitiveness are larger than the corresponding treatment estimates. This is consistent with the theoretical

argument that the residualized estimates of competitiveness are biased upwards by measurement error (Gillen et al., 2019), whereas treatment-based estimates are not affected (see Section 5). In addition, the fact that my regression results closely track the results reported in previous work (summarized in Figure 1) suggests that it is the use of a different identification strategy, rather than some idiosyncratic characteristics of my data set, that explains the limited role of competitiveness in my data.

Table A1 to A5 present alternative estimates that demonstrate that these results are robust to excluding outliers (Table A1 and A2), using Stage 4 tournament entry as the dependent variable instead of Stage 3 (Table A3 and A4) and analyzing the pooled data from both experiments (Table A5). In particular, the estimated importance of competitiveness is greater than 50% in all specifications. Further robustness checks show that this remains true in robustness checks that vary the Stage (3 or 4), sample (restricted vs full) and experiment (pooled or individual) independently; the full results of this analysis are available upon request.

### **A1.2 Feedback Aversion**

In addition to being less competitive, more risk averse and less confident, women may also be more averse to receiving relative performance feedback. Niederle and Vesterlund (2007) refer to this as a gender difference in feedback aversion, and argue that it could be a fourth reason why women may want to avoid the tournament. To investigate its importance, I included a tournament entry decision in experiment 1 in which even participants who chose the piece rate were told whether they would have won the tournament (Stage 5). This treatment therefore removed the influence of feedback aversion, since choosing the piece rate was no longer a way for women to avoid relative performance feedback.

If feedback aversion is an important driver of tournament choices, the gender difference should be smaller in Stage 5 in experiment 1 than in Stage 4. However, this is not what I find. Instead, 63.9% of men and 33.3% of women choose the tournament in Stage 5. The gender difference is 31.7 percentage points, which is significantly different from zero

TABLE A1: TOURNAMENT ENTRY REGRESSIONS FOR EXPERIMENT 1 EXCLUDING OUTLIERS

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Dependent Variable: Tournament Entry (Stage 3)				
Female	-0.303*** (0.084)	-0.154* (0.084)	-0.218** (0.089)	-0.169* (0.086)
Elicited Belief in Stage 4		0.759*** (0.258)		
Eckel-Grossman		0.034 (0.024)		
SOEP		0.055*** (0.017)		
Treatment NOCOMP			0.273*** (0.097)	0.039 (0.140)
All Risk Measures				F=2.79** p=0.030
All Confidence Measures				F=2.34 p=0.101
Constant	0.341 (0.213)	-0.191 (0.211)	0.229 (0.201)	-0.091 (0.271)
Competitiveness Estimate		51.0%	72.0%	55.7%
Ability Controls	yes	yes	yes	yes
Exclude Outliers	yes	yes	yes	yes
Observations	123	123	123	123

*Notes.* OLS Estimates, robust standard errors in parentheses. Variable descriptions are presented in the notes to Table 2. All columns exclude the data from participants who violated expected utility, or displayed extreme levels of risk tolerance or overconfidence.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

TABLE A2: TOURNAMENT ENTRY REGRESSIONS FOR EXPERIMENT 2 EXCLUDING OUTLIERS

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Dependent Variable: Tournament Entry (Stage 3)				
Female	-0.247*** (0.051)	-0.210*** (0.050)	-0.213*** (0.050)	-0.190*** (0.052)
Elicited Belief in Stage 4		0.407*** (0.143)		
Eckel-Grossman		0.020 (0.015)		
SOEP		0.041*** (0.010)		
Treatment NOCOMP			0.219*** (0.052)	0.122 (0.080)
All Risk Measures				F=3.97*** p=0.004
All Confidence Measures				F=3.76** p=0.024
Constant	0.479*** (0.130)	0.093 (0.146)	0.458*** (0.126)	0.202 (0.157)
Competitiveness Estimate		85.1%	86.1%	76.9%
Ability Controls	yes	yes	yes	yes
Exclude Outliers	yes	yes	yes	yes
Observations	364	364	364	364

*Notes.* OLS Estimates, robust standard errors in parentheses. Variable descriptions are presented in the notes to Table 4. All columns exclude the data from participants who violated expected utility, or displayed extreme levels of risk tolerance or overconfidence.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

TABLE A3: TOURNAMENT ENTRY REGRESSIONS IN STAGE 4 FOR EXPERIMENT 1

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Dependent Variable: Tournament Entry (Stage 4)				
Female	-0.306*** (0.081)	-0.227** (0.089)	-0.208** (0.089)	-0.164* (0.093)
Elicited Belief in Stage 4		0.227 (0.231)		
Eckel-Grossman		0.036 (0.025)		
SOEP		0.018 (0.017)		
Treatment NoCOMP			0.270*** (0.089)	0.191 (0.134)
All Risk Measures				F=0.81 p=0.552
All Confidence Measures				F=0.01 p=0.992
Constant	0.041 (0.227)	-0.200 (0.231)	-0.070 (0.212)	-0.340 (0.267)
Competitiveness Estimate		74.4%	68.1%	53.8%
Ability Controls	yes	yes	yes	yes
Exclude Outliers	no	no	no	no
Observations	140	140	140	140

*Notes.* OLS Estimates, robust standard errors in parentheses. Dependent variable is the *Stage 4* tournament entry decision (1-tournament, 0-piece rate). Other variable descriptions are presented in the notes to Table 2.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



TABLE A4: TOURNAMENT ENTRY REGRESSIONS IN STAGE 4 FOR EXPERIMENT 2

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Dependent Variable: Tournament Entry (Stage 4)				
Female	-0.154*** (0.048)	-0.110** (0.046)	-0.114** (0.046)	-0.086* (0.047)
Elicited Belief in Stage 4		0.535*** (0.112)		
Eckel-Grossman		0.018 (0.014)		
SOEP		0.045*** (0.009)		
Treatment NoCOMP			0.341*** (0.047)	0.134* (0.076)
All Risk Measures				F=6.06 p<0.001
All Confidence Measures				F=3.96 p=0.020
Constant	0.413*** (0.100)	-0.099 (0.125)	0.301*** (0.107)	-0.117 (0.153)
Competitiveness Estimate		71.2%	74.1%	55.9%
Ability Controls	yes	yes	yes	yes
Exclude Outliers	no	no	no	no
Observations	424	424	424	424

*Notes.* OLS Estimates, robust standard errors in parentheses. Dependent variable is the *Stage 4* tournament entry decision (1-tournament, 0-piece rate). Other variable descriptions are presented in the notes to Table 4.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

TABLE A5: TOURNAMENT ENTRY REGRESSIONS FOR POOLED SAMPLE

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Dependent Variable: Tournament Entry (Stage 3)				
Female	-0.258*** (0.040)	-0.202*** (0.040)	-0.219*** (0.040)	-0.190*** (0.041)
Elicited Belief in Stage 4		0.439*** (0.102)		0.351*** (0.125)
Eckel-Grossman		0.015 (0.012)		
SOEP		0.039*** (0.008)		
Treatment NOCOMP			0.220*** (0.041)	0.072 (0.064)
All Risk Measures				F=5.61*** p<0.001
Constant	0.526*** (0.092)	0.113 (0.114)	0.448*** (0.097)	0.105 (0.127)
Competitiveness Estimate		78.4%	84.8%	73.6%
Ability Controls	yes	yes	yes	yes
Exclude Outliers	no	no	no	no
Observations	564	564	564	564

*Notes.* OLS Estimates, robust standard errors in parentheses. Variable descriptions are presented in the notes to Table 2. The fourth column only includes a single confidence variable because only one confidence variable (Elicited Belief in Stage 4) was elicited in both experiments.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

( $p < 0.001$ , Fisher’s exact test), and almost identical to the gender difference in Stage 4 (31.8 percentage points;  $p = 0.496$ , one-sided difference-in-difference test). Hence, I find no evidence that feedback aversion contributes to the gender difference in tournament entry. These results are consistent with Mobius et al. (2014), who find that women and men have a similar willingness to pay for (not) receiving performance feedback.

### **A1.3 Comparing the Results of Experiments 1 and 2**

The key results of experiments 1 and 2 are near-identical: the point estimate for competitiveness is close to zero in both cases. However, some differences are also worth discussing. Notably, the baseline gender gap in the preferred sample is smaller in experiment 2 (16.5pp) than in experiment 1 (31.8pp), though not significantly so ( $p = 0.128$ ). The fact that a similar decrease in the gender gap is also observed in the measures of beliefs and risk preferences and in treatment NOCOMP (which also captures beliefs and risk preferences) suggests that the smaller gender gap in tournament entry may have been driven by a smaller gender gap in risk attitudes and confidence in experiment 2.<sup>1</sup> Possible reasons for these differences include cohort effects, random sampling variation, and the timing of the experiment. Overall, it is reassuring that the treatment estimate of competitiveness remains similar in spite of these potential differences in the experimental population.

### **A1.4 Order Effects**

Table A6 contains the direct tests for order effects referenced in Section 4 for the order of Stage 3 and 4 (column 1) and the order of Stage 4 and 6 (column 2). The key result is that the treatment effect of competitiveness (captured by the “Female X NOCOMP” coefficient) does not differ significantly depending on which task came first. The  $p$ -values for the relevant three-way-interactions are  $p = 0.391$  and  $p = 0.305$  for column 1 and 2.

1. The decrease in the gender gap in experiment 2 compared to experiment 1 in treatment NOCOMP, the Eckel-Grossman measure and the SOEP measure of risk preferences are significant at the 10% level or greater. The decrease in the gender gap in the Holt-Laury measure, Stage 6 switch point and the elicited belief in task 4 are not significant.

TABLE A6: ORDER EFFECTS AND THE TREATMENT ESTIMATE OF COMPETITIVENESS

	Coefficient (Std. Errors)	
	Dependent Variable: Compensation Scheme Choice	
	(1)	(2)
Female	-0.148** (0.074)	-0.203*** (0.073)
NoCOMP	0.011 (0.057)	-0.027 (0.060)
Female X NoCOMP	-0.049 (0.082)	0.056 (0.084)
Task4first	0.020 (0.073)	
Task4first X Female	-0.035 (0.104)	
Task4first X NoCOMP	0.037 (0.081)	
Task4first X Female X NoCOMP	0.097 (0.113)	
Task6first		0.045 (0.073)
Task6first X Female		0.072 (0.104)
Task6first X NoCOMP		0.115 (0.080)
Task6first X Female X NoCOMP		-0.117 (0.113)
Constant	0.571*** (0.052)	0.559*** (0.052)
EU Violators	no	no
Extreme Risk Prefs	no	no
Extreme Confidence	no	no
Low Understanding	yes	yes
Observations	728	728
N(men)	184	184
N(women)	180	180

*Notes.* OLS Estimates, robust standard errors in parentheses. Dependent variable is the choice of compensation scheme in either Stage 4 (1-tournament, 0-piece rate) or treatment NoCOMP (1-lottery, 0-fixed payment). The first three independent variables are binary variables for gender (1-female, 0-male), treatment (1-NoCOMP, 0-baseline) and their interaction. “Task6first” is a dummy variable indicating whether Task 6 came before Task 4 in the experiment. “Task4first” is a dummy variable indicating whether Task 4 came before Task 3. The figure excludes participants who violated expected utility or displayed extreme levels of overconfidence or risk aversion, similar to the main analysis in section 4.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## A1.5 Additional Figures

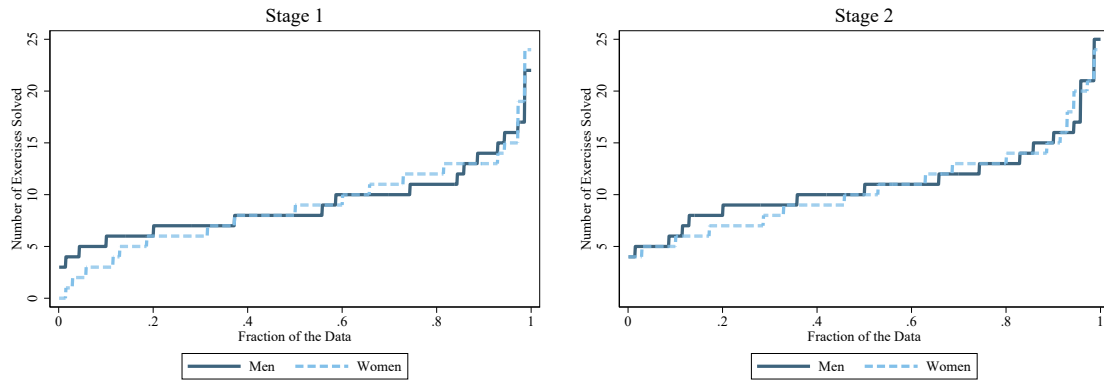


FIGURE A1: PERFORMANCE IN EXPERIMENT 1

*Notes.* The figure presents cumulative density plots for the number of addition problems solved in Stage 1 (left) and Stage 2 (right), by gender in experiment 1.

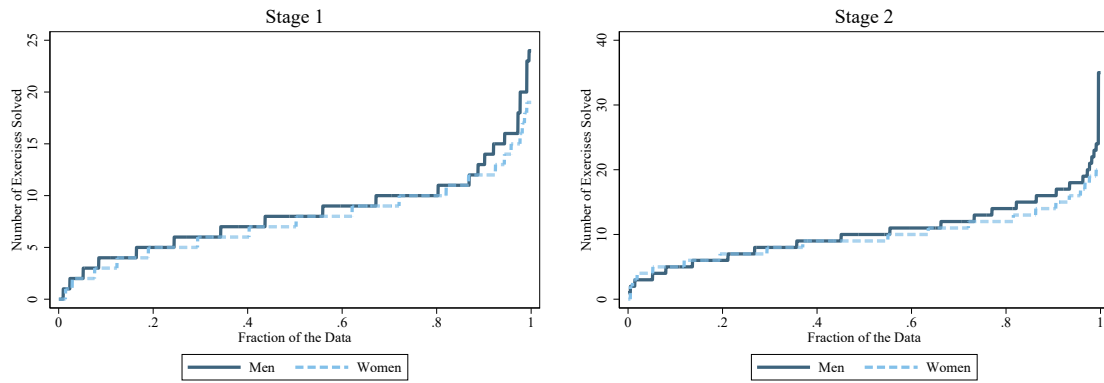


FIGURE A2: PERFORMANCE IN EXPERIMENT 2

*Notes.* The figure presents cumulative density plots for the number of addition problems solved in Stage 1 (left) and Stage 2 (right), by gender in experiment 2.

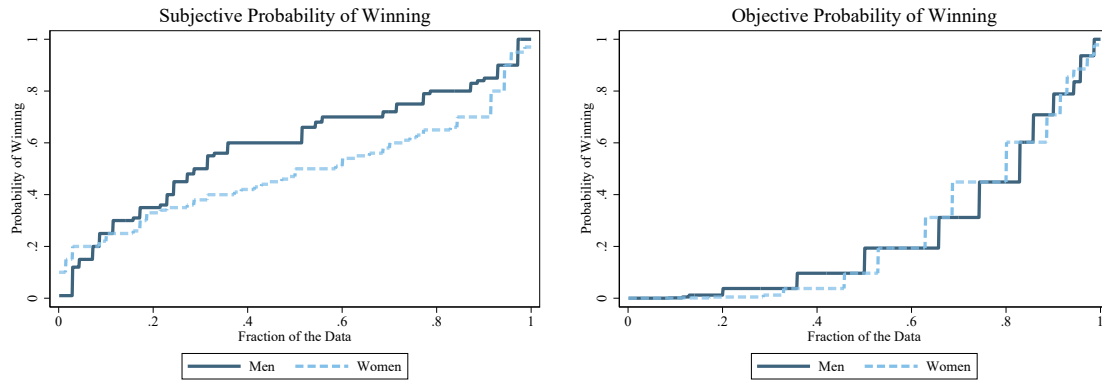


FIGURE A3: WIN PROBABILITY IN EXPERIMENT 1

*Notes.* The figure presents cumulative density plots for the subjective (left) and objective (right) probability of winning, by gender. The subjective probability of winning is the belief elicited in Stage 4. The objective probability of winning is computed by comparing each participant's Stage 2 performance with the Stage 2 performance of every other participant in experiment 1.

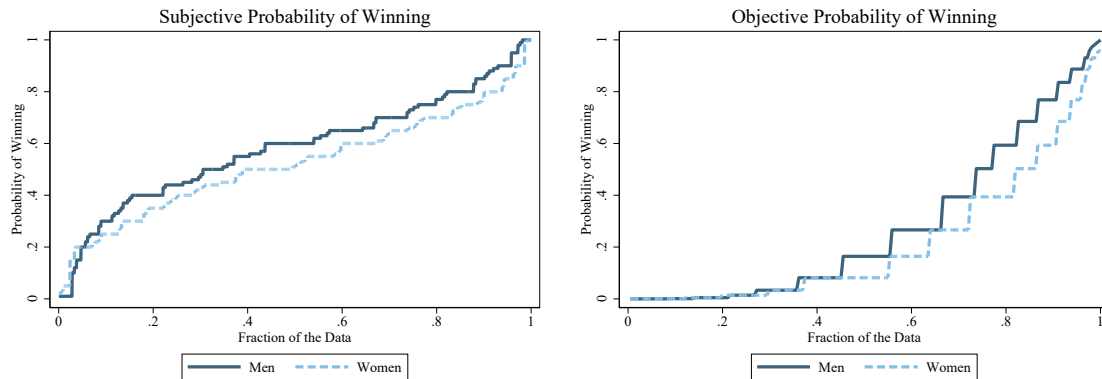


FIGURE A4: WIN PROBABILITY IN EXPERIMENT 2

*Notes.* The figure presents cumulative density plots for the subjective (left) and objective (right) probability of winning, by gender. The subjective probability of winning is the belief elicited in Stage 4. The objective probability of winning is computed by comparing each participant's Stage 2 performance with the Stage 2 performance of every other participant in experiment 2.

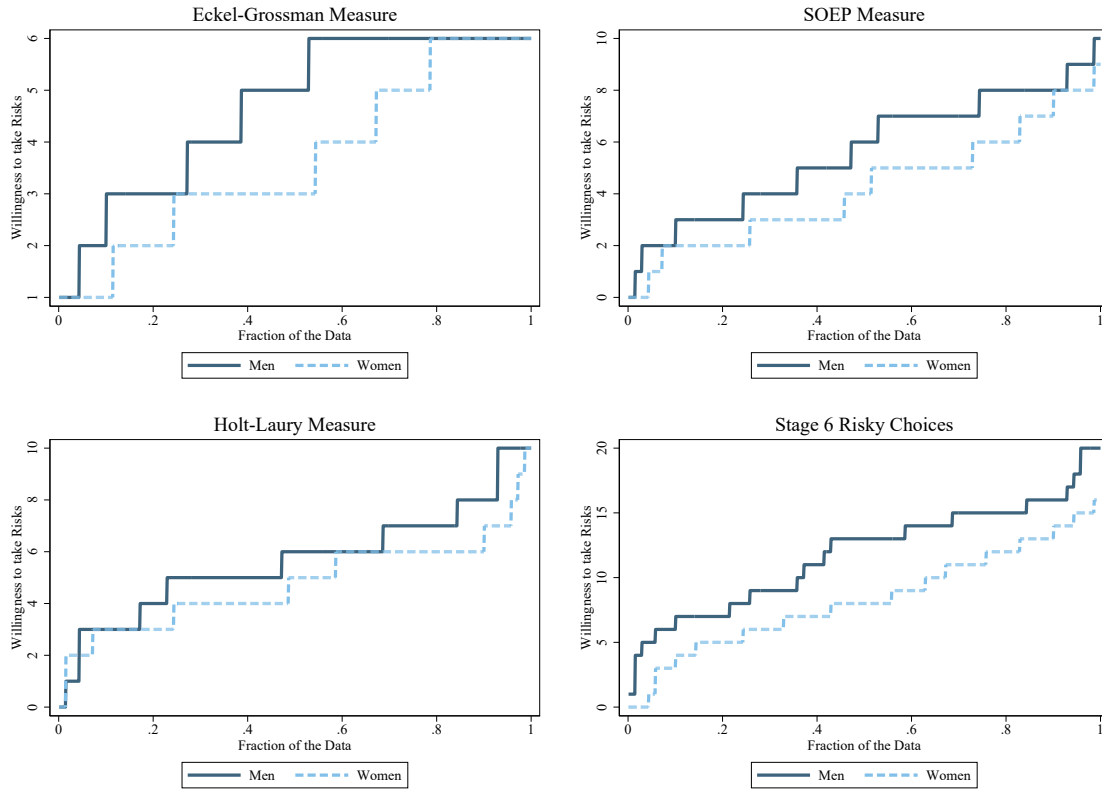


FIGURE A5: ELICITED RISK PREFERENCES IN EXPERIMENT 1

*Notes.* The figure presents cumulative density plots of the willingness to take risks for each of the four measures of risk preferences, by gender. For the Eckel-Grossman measure, the willingness to take risks is the risk rank of the chosen lottery (6-highest risk). For the SOEP measure, it is the answer to the general risk taking question. For the other two measures, it is the number of times the risky lottery was selected over the safer alternative.

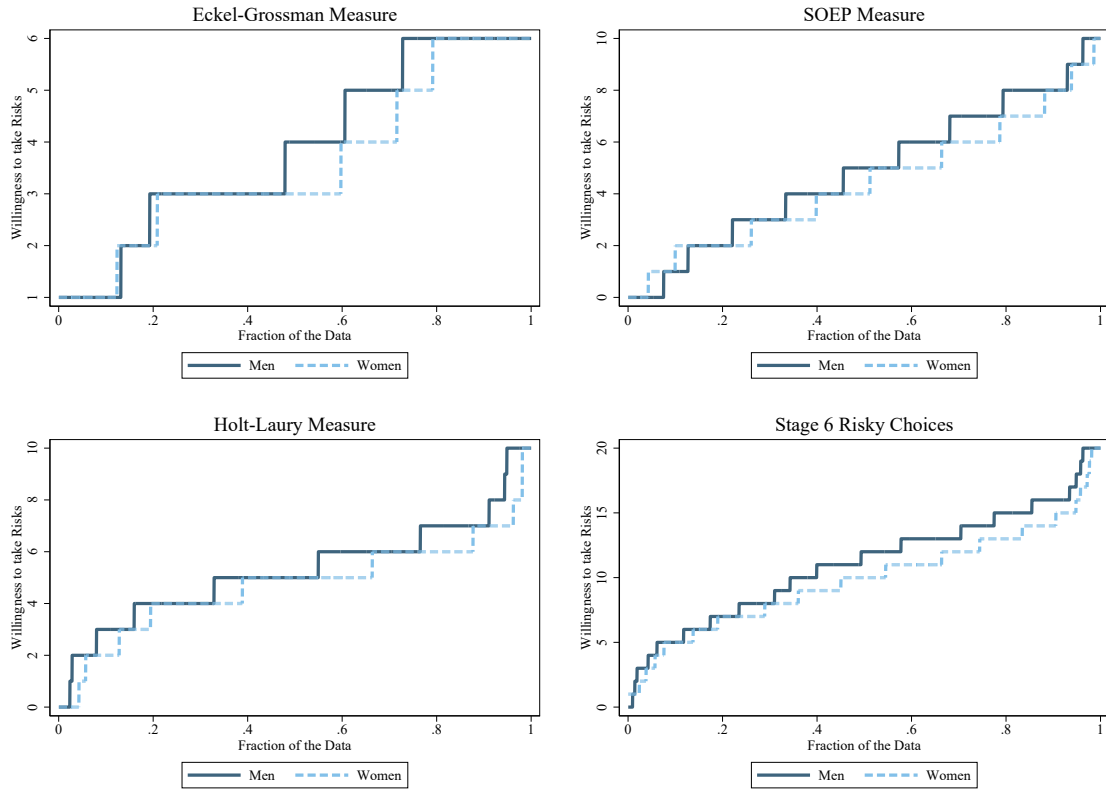


FIGURE A6: ELICITED RISK PREFERENCES IN EXPERIMENT 2

*Notes.* The figure presents cumulative density plots of the willingness to take risks for each of the four measures of risk preferences, by gender. For the Eckel-Grossman measure, the willingness to take risks is the risk rank of the chosen lottery (6-highest risk). For the SOEP measure, it is the answer to the general risk taking question. For the other two measures, it is the number of times the risky lottery was selected over the safer alternative.



## ***A2. Measurement Error in Elicited Beliefs***

In this section, I expand on the discussion in Section 5.1 and examine the role of measurement error in elicited beliefs (i.e., the subjective probability of winning) in greater detail. As a first step, I will expand upon the arguments presented in Section 5.1. I then present simulations that examine how measurement error impacts my treatment estimate for the importance of competitiveness for the case of classical and non-classical measurement error respectively. Finally, I will discuss the design and results of the robustness check based on Stage 7 and treatment NOCOMP7 in experiment 2 in greater detail.

### **A2.1 Measurement Error in Regression Analysis**

In Section 5.1, I discussed that my estimate of competitiveness is not susceptible to Gillen et al. (2019)'s critique. At first glance, this might seem surprising, given that my main estimate also relies on the use of elicited beliefs. In particular, the elicited beliefs determine which of the 20 binary choices in Stage 6 I use as treatment NOCOMP for a particular participant. Nevertheless, classical measurement error (i.e., adding a normally distributed mean zero noise term) in beliefs will not bias my results, because treatment NOCOMP (and hence elicited beliefs) serve as the y-variable in my analysis. Intuitively, my estimate of competitiveness is based on comparing the gender difference in choices across treatments. While classical measurement error in beliefs may affect the outcomes for individual participants, any resulting misclassifications will cancel out in expectation. Hence, the expected gender gap in treatment NOCOMP will not be affected, and neither will the expected treatment estimate for competitiveness.

An example might help illustrate this argument further. Consider two participants, and assume for simplicity that both are female and risk-neutral. Suppose that the first participant was fairly optimistic ( $p_s = 0.4$ ), but mistakenly entered a more pessimistic belief ( $\hat{p}_s = 0.2$ ). Due to the mismeasured elicited belief, this participant would be misclassified in treatment NOCOMP as someone choosing the safe payment. But now suppose that the second participant was more pessimistic ( $p_s = 0.16$ ) but mistakenly entered a too optimistic

belief ( $\hat{p}_s = 0.36$ ). In this case, both participants would be misclassified, but their errors would cancel out on average. Hence, the overall fraction of women choosing the lottery in treatment NOCOMP would not be affected.

It is important to note that measurement error does not cancel out when elicited beliefs are used as an x-variable in regression analysis. This is illustrated by comparing the estimated importance of competitiveness in Table 2 and 4 to the treatment estimate. Independent of the exact specification used, the treatment-based estimates for competitiveness are consistently much smaller (-2.9% for the pooled sample) than the corresponding estimates using Niederle and Vesterlund (2007)'s residual-based regression approach (between 50% and 90% depending on the experiment and specification). This is consistent with the theoretical argument that the latter estimates are biased upwards by measurement error (Gillen et al., 2019), whereas treatment-based estimates are not affected. In addition, the fact that my regression results closely track the results reported in previous work (summarized in Figure 1) suggests that it is the use of a different identification strategy, rather than some idiosyncratic characteristics of my data set, that explains the limited role of competitiveness in my data.

## A2.2 Classical Measurement Error

In Section 5.1, I also referred to previous work (summarized in Hausman, 2001) demonstrating that classical measurement error does not bias regression estimates when it only affects the dependent variable, as in my analysis. In this section, I will illustrate these results using simulations directly tailored to the present experiment. In particular, I assume a well-behaved linear data-generating process in which the variables of interest are subject to classical measurement error. Specifically, the data-generating process I use to generate tournament entry decisions  $Y$  is the following:

$$(A1) \quad Y = \alpha_0 + \beta_1 R^* + \beta_2 B^* + \beta_3 C^*$$

$$(A2) \quad R^* = \alpha_1 + \gamma_1 M + \nu_R^*$$

$$(A3) \quad B^* = \alpha_2 + \gamma_2 M + \nu_B^*$$

$$(A4) \quad C^* = \alpha_3 + \gamma_3 M + \nu_C^*$$

Here,  $M$  is equal to one if the participant is a man, and zero otherwise. I assume that beliefs ( $B^*$ ), risk preferences ( $R^*$ ), and competitiveness ( $C^*$ ) are partially determined by gender, and I normalize the effect of gender to be equal to one, i.e.,  $\gamma_1 = \gamma_2 = \gamma_3 = 1$ . The explanatory variables also reflect an idiosyncratic term  $\nu_R^*, \nu_B^*, \nu_C^* \sim N(0, 1)$  that captures all variation that is not due to gender. For illustrative purposes, I assume that  $\beta_1 = .60$ ,  $\beta_2 = .40$  and  $\beta_3 = 0$ . That is, the effect of competitiveness on tournament entry is zero and the effect of risk preferences and beliefs sum to 1. Without loss of generality, I further normalize all constants  $\alpha$  to zero.

To study the role of measurement error, I assume that behavior in treatment NOCOMP is governed by the true DGP and true risk preferences  $R^*$ , but by elicited and therefore possibly noisy beliefs ( $B$ ), where  $B = B^* + \eta_B$ , and  $\eta_B \sim N(0, \sigma_\eta^2)$ . Hence, choices in treatment NOCOMP ( $Y^1$ ) are constructed as follows.

$$(A5) \quad Y^1 = \alpha_0 + \beta_1 R^* + \beta_2 B$$

The effect of competitiveness can be obtained by subtracting the size of the gender gap in  $Y^1$  from the size of the gender gap in tournament choices  $Y$  within each simulated sample. In order to test for the effect of competitiveness, I generate 999 random samples for 200 different values of the measurement error variance and compare the size of the gender gap in  $Y^1$  and  $Y$  in each sample. In line with the size of the sample for my main analysis when

pooling across experiments, each sample consists of 245 men and 242 women.

For comparison purposes, I also simulate the effect of measurement error in the residual-based regression approach of Niederle and Vesterlund (2007) by estimating the following regression:

$$(A6) \quad Y = \hat{\alpha}_0 + \hat{\beta}_1 R + \hat{\beta}_2 B + \hat{\theta} M + e$$

Here,  $R = R^* + \eta_R$  is a noisy measure of risk preferences, and  $\eta_R \sim N(0, \sigma_\eta^2)$  as well. Since the baseline gender gap is equal to one, the effect of competitiveness is equal to the residual gender gap  $\hat{\theta}$ .

Figure A7 presents the results for both approaches. While measurement error increases the implied importance of competitiveness in the regression approach (Gillen et al., 2019), there is no such effect for treatment comparisons. Independent of the amount of measurement error, the treatment comparison on average correctly attributes none of the gender gap to competitiveness. In line with Hausman (2001), increased measurement error only affects the treatment-based estimator through an increase in its variance. All in all, these results illustrate that classical measurement error systematically biases the results of the regression approach, but does not affect the average treatment estimate.

### A2.3 Non-Classical Measurement Error

In contrast to the assumptions of the previous section, measurement error need not be classical (i.e., uncorrelated with the x-variable) in practice. This could both be due to measurement error not being classical per se (e.g., due to beliefs being bounded between 0 and 1), and due to the specific way that treatment NOCOMP is constructed. As an example of the former case, consider the case where the gender gap in latent beliefs is 20pp but 50% of participants fill out their beliefs in a totally random way. If the random beliefs are distributed uniformly between 0 and 1, the gender gap in elicited beliefs will only be 10 percentage points. As an example of the latter case, consider a situation where all

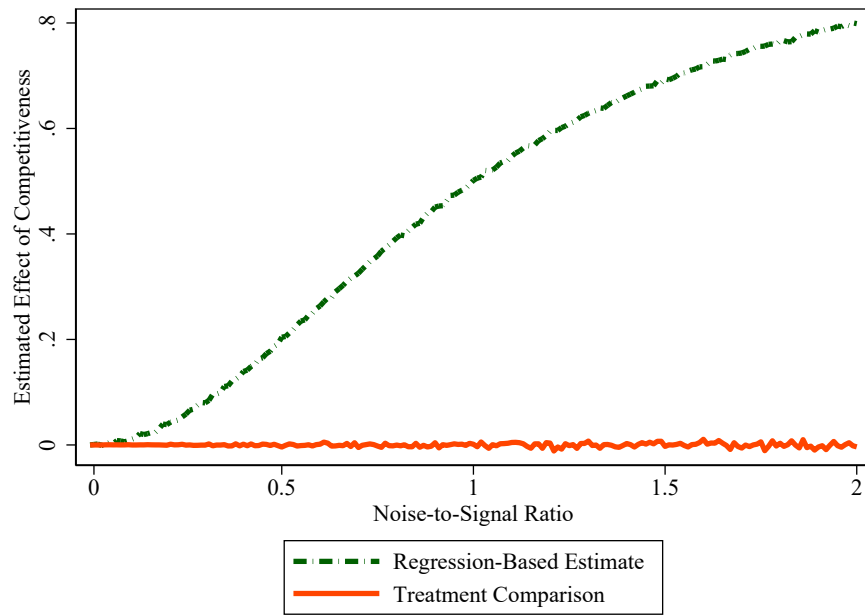


FIGURE A7: CLASSICAL MEASUREMENT ERROR SIMULATIONS

*Notes.* The dashed line displays the average estimated coefficient for gender in a regression of tournament entry choices on gender, risk preferences and beliefs (over 999 samples). The solid line is equal to the average difference between the simulated gender gap in tournament entry and the simulated gender gap in treatment NOCOMP (over 999 samples).

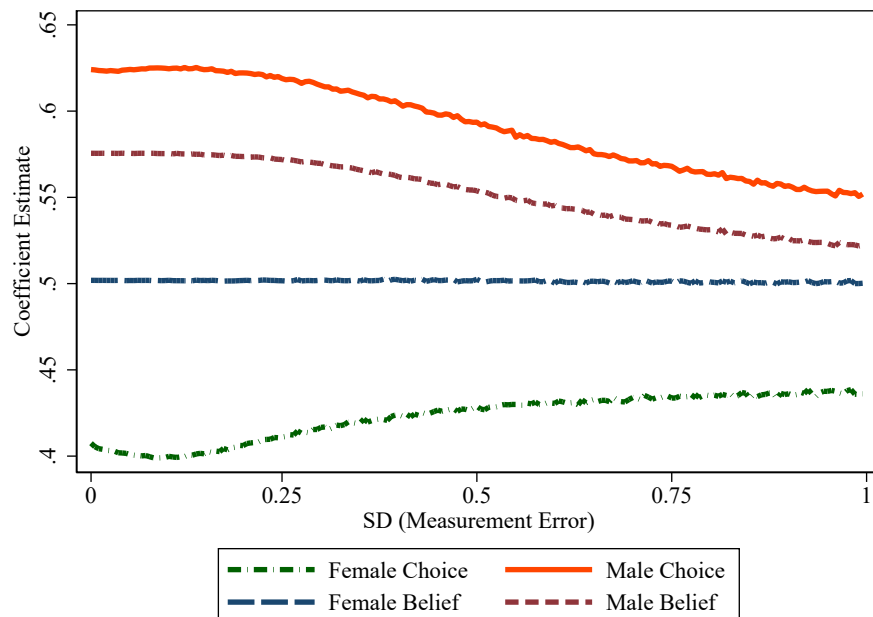


FIGURE A8: NON-CLASSICAL MEASUREMENT ERROR SIMULATIONS

*Notes.* The figure is based on simulations that perturb elicited beliefs from Stage 4 1000 times for every participant using a mean zero truncated normally distributed disturbance, for 200 different values for the standard deviation of the disturbance. The disturbance term is truncated such that only beliefs in the interval  $[0,1]$  are generated. I use the perturbed beliefs to determine which row in Stage 6 should be used as treatment NOCOMP, similar to the main analysis. The figure plots the perturbed beliefs and fraction of lottery choices made by each gender as a function of the standard deviation of the disturbance, averaged over all simulations and participants.

participants have their switch point in Stage 6 at 50% and all men (women) think they have a 55% (45%) win chance respectively. Without measurement error, the gender gap in treatment NOCOMP would be 100 percentage points. With measurement error, some men (women) may have elicited beliefs smaller (greater) than 50%, creating a downward bias in the gender gap in treatment NOCOMP. Hence, even if measurement error in beliefs is classical, measurement error in treatment NOCOMP choices need not be.

I examine the role of non-classical measurement error using simulations that perturb the elicited beliefs with a truncated normally distributed noise term and re-construct treatment NOCOMP based on the perturbed beliefs. Figure A8 plots the perturbed beliefs and resulting treatment NOCOMP choices as a function of the standard deviation of the noise term, separately for men and women. The results demonstrate that substantial amounts of

non-classical measurement error may in fact *reduce* the gender gap in both elicited beliefs and NOCOMP choices. This implies that, if anything, measurement error may lead me to underestimate the gender gap in treatment NOCOMP and hence overestimate the importance of competitiveness. However, it is important to note that a bias only emerges when measurement error is substantial. For example, even when adding measurement error that doubles the standard deviation of elicited beliefs (measurement error standard deviation of 0.20), the estimated gender gap is still similar to the gap reported in the main analysis.

In summary, these simulations suggest that only substantial amounts of non-classical measurement error have the potential to bias my estimates. Importantly, however, such a bias would actually *decrease* the gender difference in this treatment, which would imply that my estimates overstate the importance of competitiveness. The fact that my point estimate for competitiveness is already close to 0% suggests that such a bias did not play a major role in my data.

#### **A2.4 Estimating Competitiveness without Eliciting Beliefs**

In Section 5.1, I presented a robustness check that allowed me to estimate the importance of competitiveness without requiring the use of elicited beliefs. For this purpose, I included a third tournament entry decision in experiment 2 (Stage 7), where participants were informed of the Stage 2 performance of every participant in their session prior to making their entry decision. They were also told their “objective” probability of winning a tournament if everyone’s performance was as in Stage 2. Assuming that participants internalize this information, their entry decision would then be based on the objective win probability, a number that is known to the experimenter and therefore does not need to be elicited.

I compare tournament entry in Stage 7 to a version of treatment NOCOMP that I based on the same objective probability of winning (treatment NOCOMP7). To construct treatment NOCOMP7, I take the single decision row in Stage 6 for which the win probability of the lottery for a given participant most closely corresponds to that participant’s objective probability of winning a tournament. The binary choice made in this single decision row

is what I refer to as treatment NOCOMP7 in the main text. Note that this procedure is identical to the procedure I used to construct treatment NOCOMP, except that I use the *objective* rather than the *subjective* win probability to select the relevant decision row in Stage 6. By comparing the gender gap in Stage 7 to the gender gap in treatment NOCOMP7, I can then identify the importance of competitiveness without needing to elicit subjective beliefs. In particular, if competitiveness is important, the gender gap in treatment NOCOMP7 should be significantly smaller than in Stage 7.

To illustrate how this comparison worked in practice, consider the example of a participant  $i$  who solved 13 addition problems in Stage 2 ( $x_i = 13$ ). In Stage 7, this participant would be told that, based on her Stage 2 performance, her objective win probability was equal to 0.5.<sup>2</sup> When deciding to enter the tournament, this participant would then implicitly be choosing between a piece rate payment of 6.50 Euros and a 50% chance of obtaining a tournament payment of 26 Euros. Similarly, for this participant Stage 6 would consist of 20 decisions between a fixed payment of 6.50 Euros and a lottery paying either 26 Euros or nothing. I would then use the decision row in Stage 6 with a lottery win probability of 0.50 as this participant's treatment NOCOMP7 decision in my analysis.

The key assumption for this comparison is that participants internalize the information they receive in Stage 7, so that their subjective win probability in Stage 7 is indeed identical (at least on average) to the objective win probability that I use to construct treatment NOCOMP7. To test whether this is the case, I asked participants for their subjective win probability after receiving the information; this elicitation was not incentivized. Overall, the elicited subjective win probability was highly correlated with the objective win probability presented on screen ( $p=0.73$ ,  $p<0.001$ ). However, both men (37.7% versus 27.4%,  $p<0.001$ , t-test) and women (33.1% versus 21.2%,  $p<0.001$ , t-test) were significantly more optimistic than the objective probability. Rather than a systematic bias, however, these results likely reflect the fact that participants improve their score over the course of the experiment,

2. The actual number depended on the composition of a particular session. When averaging across all sessions, however, a participant solving 13 addition problems would have a 50% chance of winning a tournament.



something that they appear to anticipate. Indeed, the actual win chance in Stage 7 was 39.4% for men and 34.8% for women, neither of which differs significantly from the elicited subjective win probability (recall that participants compete against the Stage 2 performance of their opponents). To allow for these anticipated improvements, I construct treatment NOCOMP7 using the objective win probability based on participants' Stage 7 performance. However my results are robust to computing the objective win probability using Stage 2 performance instead.

If measurement error in beliefs was a key determinant of my main results, then I should observe a significantly larger estimate for competitiveness in the comparison that does not rely on eliciting subjective beliefs. However, the point estimate for competitiveness is in fact very similar. In particular, 39.7% of men and 26.1% of women chose to enter the tournament in Stage 7 ( $p=0.007$ , Fisher's exact test). In treatment NOCOMP7, the corresponding numbers are 39.9% for men and 26.9% for women ( $p=0.010$ , Fisher's exact test); the point estimates imply that competitiveness explains 4.1% ( $\frac{0.6}{13.6}$ ) of the gender gap in tournament entry in Stage 7 ( $p=0.456$ , one-sided difference-in-difference test). Note also that overall tournament entry is smaller than in Stage 4 (where it was 58.2% for men and 41.7% for women in the preferred sample). This is likely caused by the information reducing the average subjective probability of winning from 57.5% for men and 50.8% for women in Stage 4 to 37.7% for men and 33.1% for women in Stage 7.<sup>3</sup>

In summary, this robustness check removes concerns related to measurement error in beliefs by using a version of my identification strategy that does not rely on eliciting beliefs. The estimate of competitiveness is similar to my main identification strategy, suggesting that measurement error in elicited beliefs does not explain the limited importance attributed to competitiveness in my main analysis.

3. Instead of constructing treatment NOCOMP7 based on the true probability of winning in Stage 7, I can also use the elicited probability of winning in Stage 7. The resulting estimate for competitiveness is 18.5% ( $p=0.278$ , one-sided difference-in-difference test).

### ***A3. Risk and Tournament Entry***

In this section, I expand upon the discussion in Section 5.2 by presenting additional details regarding the robustness checks introduced in that Section. I will also discuss how a potential midpoint bias in the price list task could impact my treatment estimates. Finally, I will examine my identification strategy through the lens of non-expected utility theory.

#### **A3.1 Estimating Competitiveness Using Two Binary Choices**

The first robustness check that I discussed in the main text sought to harmonize the elicitation method by implementing treatment NOCOMP as a binary decision. At first glance, it might appear that this could be done by simply giving participants a binary choice between a fixed payment and a lottery with a win probability equal to the elicited subjective probability of winning. However, adding such a binary version of treatment NOCOMP to the experiment would pervert the incentives faced by participants in the belief elicitation task. In particular, it would incentivize participants to claim to have a 100% chance of winning the tournament, in order to maximize their earnings in the binary NOCOMP choice. While not making participants aware of future decisions during their belief elicitation task would prevent this issue, such a procedure would have violated the no-deception rules of the laboratory.

Instead, I therefore implemented a binary version of treatment NOCOMP7. Recall that treatment NOCOMP7 was based on the objective empirical probability of winning the tournament (which can be estimated in the data) as opposed to the subjective probability (which needs to be elicited). I included a binary version of treatment NOCOMP7 in experiment 2 as Stage 5, where I asked participants to choose between a safe amount of  $0.5x_i$  and a lottery that paid  $2x_i$  with probability  $p_i^o$ , where  $x_i$  was equal to Stage 2 performance and  $p_i^o$  (for objective) was the empirical probability that their Stage 2 performance would be good enough to beat three participants from experiment 1 or win a random tiebreaker. I can then identify the effect of competitiveness by comparing the gender gap in Stage 7 (where competitiveness plays a role) to the gender gap in Stage 5 (where it does not). The main

assumption for this comparison is that the lottery win probability in Stage 5 is identical to the tournament win probability in Stage 7.

In Stage 7, 39.7% of men and 26.1% of women chose to enter the tournament ( $p=0.007$ , Fisher's exact test). In treatment NOCOMP7, the corresponding numbers are 39.9% for men and 26.9% for women ( $p=0.010$ , Fisher's exact test). However, in Stage 5, 30.9% of men and 28.3% of women selected the lottery ( $p=0.646$ , Fisher's exact test). The comparison between NOCOMP7 and Stage 5 suggests that using a binary choice makes men less likely to choose the lottery ( $p=0.013$ , t-test), whereas for women there is essentially no effect ( $p=0.585$ , t-test). More importantly, the comparison between the gender gap in Stage 7 and Stage 5 implies that the point estimate for competitiveness is 80.5% ( $\frac{10.9}{13.6}$ ,  $p=0.010$ , one-sided difference-in-difference test).

Why is the estimate for competitiveness so much larger than the other estimates discussed in the main text? As a first step, note that the smaller gender gap in Stage 5 is driven by men who choose to compete in Stage 7 but do not choose the lottery in Stage 5 (women make similar decisions in both stages). If the identifying assumption holds, this would imply that men, on average, enjoy being in competitive environments, whereas women are, on average, indifferent to competitions. However, the data appear to tell a different story. In particular, the men who decide to switch from lottery to tournament pay also have a much greater actual likelihood of winning in Stage 7 (52.7% on average) than in Stage 5 (29.8% on average). This is because their performance has improved over the course of the experiment, something that they also anticipate (their elicited expected win chance in Stage 7 is 48.8%). Hence, it appears to be the increased win probability in Stage 7 that explains why these men switched from the safe payment in Stage 5 to tournament pay in Stage 7, and not their competitiveness.

Overall, these results imply that the main identifying assumption (that the win probability remain constant across the two stages) appears to be violated for the men who switch. In other words, men do not appear to 'compete' more in Stage 7 because they are competitive, but rather because they rationally expect to have a greater win chance in Stage 7 than

they had in Stage 5. This suggests that the reason I obtain different results in the Stage 5 vs Stage 7 comparison than I do in the NOCOMP7 vs Stage 7 comparison is not because binary choices generate different results than price lists, but rather because my specific implementation of the binary choice in Stage 5 was not able to fix the win probability at exactly the same level as in Stage 7.

### **A3.2 Estimating Competitiveness Using Two Price Lists**

The other way to harmonize the elicitation method for the tournament entry and treatment NOCOMP decisions is to use a price list for both. Bartos and Van Veldhuizen (2022) develop a novel method to decompose observed effects into mechanisms, and illustrate their approach in the context of gender differences in tournament entry. As part of their experiment, they implement price list versions of both decisions. For treatment NOCOMP, they use a price list that is similar to Stage 6 in this paper but includes a real effort task. In particular, they ask participants to choose between piece rate payment (equivalent to the safe payment in this paper) and a random payment scheme with varying probability (equivalent to the lottery in this paper). The random payment scheme is calibrated to have the same payment as the tournament, but lacks the competitive element.

To elicit tournament entry, Bartos and Van Veldhuizen (2022) ask participants to choose between piece rate pay and a tournament for which the win probability varies across the rows of the price list. They are able to do this by telling participants that they will be matched to either an inferior or a superior performer. In particular, they tell participants that they will be matched to one of two participants from a previous experiment, and are told the score of these participants. Importantly, Bartos and Van Veldhuizen (2022) selected these two participants to have scores that were either considerably smaller than the focal participant's (the "inferior performer") or were considerably larger (the "superior performer"). This meant that it was essentially impossible to win against the superior performer, or to lose against the inferior one. This in turn made it possible to vary the tournament win probability by varying the probability of being matched to the inferior

opponent.

To identify the relevant choice in each price list, Bartos and Van Veldhuizen (2022) follow a similar approach to the one used for treatment NOCOMP in this paper. In particular, they look at the choices made in the row of the respective price list that has the same win probability as the elicited subjective probability of winning the tournament. For example, for a participant who indicated that she thought she had a 40% chance to win the tournament, they used the choice with win probability 0.40 in each price list as their version of treatment NOCOMP and the baseline tournament entry decision. Note that, in addition to ensuring that both treatments were elicited using price lists, this procedure also implies that both treatments included a real effort task, and resolved uncertainty using a computerized random draw (a “nature risk”).

In the price list tournament entry decision, men (43.0%) were significantly more likely to choose tournament pay than women (21.4%,  $p < 0.0001$ , Fisher’s exact test). In treatment NOCOMP, men (38.8%) were more likely to choose random pay than women (11.0%,  $p < 0.001$ , Fischer’s exact test). The point estimate for competitiveness is -28.8% ( $\frac{-6.2}{21.6}$ ), which is not significantly greater than zero ( $p = 0.908$ , one-sided difference-in-difference test).

In summary, this robustness check harmonizes the decision environment of the two treatments by (a) using a price list, (b) including a real effort task and (c) having the same type of (“nature”) risk in both treatments. The estimate of competitiveness remains similar to my main identification strategy, suggesting that treatment differences in these three features cannot explain the main results of experiment 1 and 2.

### **A3.3 Midpoint Bias in Price List Tasks**

One specific reason why differences in the elicitation method could conceivably affect my results is that switch points on price lists may be biased towards the middle (see e.g., Harrison et al., 2005 or Andersson et al., 2016), perhaps because the midpoint serves as a cognitive default to cognitively uncertain individuals (Enke and Graeber, 2021). In particular, if participants display a midpoint bias, then the elicited switch point  $SP_i$  in the price

list may be biased in the following way:

$$(A7) \quad SP_i = (1 - \alpha_i)SP_i^* + \alpha_i D$$

Here,  $D$  is the cognitive default, which in price lists is assumed to be the midpoint. I normalize possible switch points to be between 0 and 1, so that the midpoint is given by  $D = 0.5$ . Equation A7 tells us that the midpoint bias induces the elicited switch point to lie somewhere between the true switch point  $SP_i^*$  and the midpoint  $D = 0.5$ . If this is the case, I can back out the true switch point by rewriting this equation as:

$$(A8) \quad SP_i^* = (SP_i - 0.5\alpha_i)/(1 - \alpha_i)$$

The problem with estimating equation A8 is that the value of  $\alpha_i$  is unknown. However, it is possible to systematically vary  $\alpha_i$  and then re-calculate the result of treatment NOCOMP for different values of  $\alpha_i$ . Figure A9 presents the results using data from the preferred sample in both experiments, where I assume that  $\alpha_i = \alpha$  for all participants. The main takeaway is that, for reasonable values of  $\alpha$ , the estimated gender gap in treatment NOCOMP is similar to the unadjusted gap (left side of the graph). In fact at its high point, the estimated gender gap has increased from 20.9 to 23.6 percentage points, which would put the estimated importance of competitiveness at -16.3% ( $\frac{-3.3}{20.3}$ ). Only when  $\alpha$  approaches 1 does the gender gap start falling below the gender gap in the data, up to a minimum of 15.4 percentage points (and an implied importance for competitiveness of  $\frac{4.5}{20.3} = 22.2\%$ ). However, such values imply that the latent switch point  $SP_i^*$  is either 0 or 1 for nearly every participant, which does not seem reasonable. Overall, the results therefore suggest that the midpoint bias cannot explain the low estimated importance of competitiveness in my data.<sup>4</sup>

4. In addition, the result of my main comparison between tournament entry and treatment NOCOMP is robust to removing participants who switched in the middle row or middle three rows of the price list; using data from both experiments the point estimates for competitiveness are 10.1 and 2.2 percentage points.

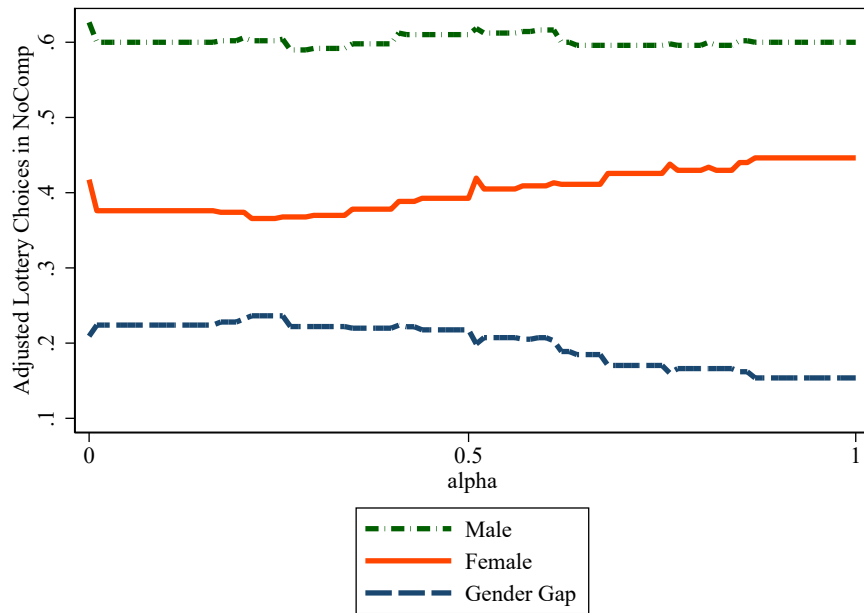


FIGURE A9: NOCOMP CHOICES AFTER ADJUSTING FOR A MIDPOINT BIAS

*Notes.* The figure plots the fraction of men and women who choose the risky option in treatment NOCOMP and the gender gap after adjusting for the midpoint bias. The size of the midpoint bias is represented by alpha ( $\alpha$ ). This analysis also assumes that  $SP_i^* = 1$  whenever  $(SP_i - 0.5\alpha_i)/(1 - \alpha_i) > 1$  and  $SP_i^* = 0$  whenever  $(SP_i - 0.5\alpha_i)/(1 - \alpha_i) < 0$ . The figure includes data from both experiments, removing participants who violated expected utility or were excessively risk tolerant or overconfident, in accordance with the main analysis.

### A3.4 Non-Expected Utility

In the main text, I discussed my identification strategy through the lens of expected utility theory. In this section, I discuss the more general case of non-expected utility maximization. Analogous to equation 1 in Section 2, under non-expected utility participant  $i$  should choose to enter the tournament if:

$$(A9) \quad \pi_i(p_i^s)v_i^T(2x_i) > v_i(0.5x_i)$$

Relative to expected utility, the two main differences are the value function  $v_i$ , which replaces the Bernoulli utility function  $u_i$ , and the probability weighting function  $\pi_i$ . Similarly, choosing the lottery payment in treatment NOCOMP requires that:

$$(A10) \quad \pi_i(p_i^s)v_i(2x_i) > v_i(0.5x_i)$$

Under expected utility theory, my identification strategy assumes that treatment NOCOMP removes the effect of competitiveness while keeping the riskiness of the environment and the *subjective win probability* the same as in the tournament entry decision. Under non-expected utility, it instead assumes that treatment NOCOMP removes the effect of competitiveness while keeping the riskiness of the environment and the *subjective probability weight* the same as in the tournament entry decision.

These two assumptions are equivalent as long as  $\pi_i$  is invertible. In particular, let  $p_i$  be the (lottery) win probability in treatment NOCOMP. Under expected utility, my identification strategy requires that  $p_i = p_i^s$ . Under non-expected utility, it instead requires that  $\pi_i(p_i) = \pi_i(p_i^s)$ . If  $\pi_i$  is invertible, however, we can write the second condition as  $p_i = \pi_i(\pi_i(p_i^s))^{-1} = p_i^s$ . Invertibility of the probability weighting function is not a strong assumption, since it is implied by standard probability weighting functions used in (cumulative) prospect theory.

Invertibility also implies that the belief elicitation task captures the subjective win



probability, so that indeed  $p_i = p_i^s$ , as required. In particular, this task elicits the point  $p_i$  at which a participant  $i$  is indifferent between obtaining 2 Euros with probability  $p_i$  and obtaining 2 Euros if their performance is good enough to win the tournament. Under non-expected utility, the indifference point is defined by  $\pi_i(p_i)v_i(2) = \pi_i(p_i^s)v_i(2)$ , so that participant  $i$  should report  $p_i = \pi_i(\pi_i(p_i^s))^{-1} = p_i^s$ . Hence, as long as the probability weighting function is invertible, the belief elicitation task will elicit the subjective win probability and my identification strategy remains valid.

When  $\pi_i$  is not invertible, the equivalence between the two sets of assumptions no longer holds. Nevertheless, my identification strategy remains valid under the assumption that the *subjective probability weights* are constant across treatments. An important example of a non-invertible probability weighting function arises in the case of ambiguity aversion in the sense of Gilboa and Schmeidler (1989), for which the probability weighting function  $\pi_i(P_i^s) = \min_{p \in P_i^s} p$ . Here,  $P_i^s$  is the set of possible subjective win probabilities, which captures situations where participants do not know their win probability exactly (the win probability is ambiguous). In such cases, the belief elicitation task will elicit the most pessimistic win probability, so that  $p_i = \min_{p \in P_i^s} p$ . Nevertheless, since  $\pi_i(p_i) = \min_{p \in P_i^s} p = p_i = \pi_i(P_i^s)$ , my identification strategy still ensures that *subjective probability weight* in treatment NOCOMP is identical to the tournament entry decision. Hence, my identification strategy remain valid even in cases where the belief elicitation task no longer elicits the (expected) subjective win probability, as long as  $\pi_i(p_i) = \pi_i(p_i^s)$ .

Apart from these differences, non-expected utility also changes the way that the identifying assumption is represented in the model. In particular, the riskiness of the environment is now captured by  $x_i$ ,  $v_i$  and  $\pi_i$ , where  $v_i$  also has more parameters (such as the reference point and loss aversion) than  $u_i$  under expected utility. Nevertheless, these changes do not affect my identifying assumption, which still requires the riskiness of the environment to be constant across treatments. These considerations also imply that the discussion of measurement error in subjective beliefs (Section 5.1) and risk attitudes (Section 5.2) remains valid under non-expected utility theory as well.

TABLE A7: POOLED ESTIMATE OF COMPETITIVENESS INCLUDING STAGE 5 AND 7

	Estimates	Sample Size
(1) Main comparison only	-2.9% ( $-\infty, 37.4\%$ )	487 (487)
(2) All comparisons from experiment 1 and 2	19.8% ( $-\infty, 54.7\%$ )	1215 (487)
(3) All data	4.9% ( $-\infty, 29.9\%$ )	1855 (1127)

*Notes.* See the notes to Table 6. The only difference is that this Table also includes the data from the comparison between Stage 5 and Stage 7 in rows (2) and (3).

#### ***A4. Collecting Evidence***

In Table 6 in Section 5.3, I estimate the effect of competitiveness using the pooled data from experiment 1, experiment 2 and two follow-up experiments. Table A7 repeats this analysis while adding the comparison between Stage 5 and Stage 7 discussed in Appendix A3.1 above. I did not include this comparison in Table 6 because the data suggest that the identifying assumption of an equal win probability in both treatments does not hold. Nevertheless, Table A7 shows that the pooled estimate for competitiveness when this comparison is included. In particular, when including all data the point estimate for competitiveness is 4.9% with a confidence interval of ( $-\infty, 29.9\%$ ). This is similar to the point estimate (-8.6%) and confidence interval ( $-\infty, 16.2\%$ ) in Table 6.

The limited importance of competitiveness is also consistent with several findings reported in previous work. First, previous studies that include a non-competitive control treatment along the lines of treatment NOCOMP in this study also do not consistently find a smaller gender gap in this treatment. Table A8 presents an overview of seven laboratory experiments that include such a treatment. The table includes studies that (1) report the raw gender difference in tournament entry in a laboratory experiment, (2) include a regression controlling for overconfidence and risk preferences, and (3) report the raw gender difference in a non-competitive control treatment.

TABLE A8: NON-COMPETITIVE CONTROL TREATMENTS IN OTHER WORK

Study	Gender Gap		Attributed to Competitiveness	
	Tournament	Control	Treatment Comps	Regressions
Dohmen and Falk (2011)	25	23.1	7.6	62.8
Reuben et al. (2017)	26.8	22.0	17.9	49.6
Niederle and Vesterlund (2007)	38	30	21.1	42.7
Lee et al. (2014)	7.6	5.3	29.7	122.4
Sutter and Glätzle-Rützler (2015)	21	14.5	31.0	79.5
Dargnies (2012)	33.3	10.6	68.2	89.8
Healy and Pate (2011)	53	15	71.7	97.0
Average			35.0	77.7

*Notes.* The first column presents the gender difference in the percentage of participants who chose the tournament. The second column presents the gender difference in the percentage of participants who chose the risky option in the non-competitive control treatment. The percentage attributed to competitiveness in treatment comparisons equals 1 minus the ratio of the second column to the first column. For more details on the regression approach, I refer the reader to Table A11.

Six of the studies included in Table A8 have a similar setup. First, participants decide between piece rate and tournament incentives and then solve addition problems, as in Stage 3 in this study. Second, they go through a control treatment in which they are asked to submit their performance from an earlier part of the experiment (typically Stage 1, the forced piece rate) to either tournament or piece rate incentives. Niederle and Vesterlund (2007) argue that this control treatment removes the competitive element, and use it in their regression analysis to control for gender differences in confidence and risk preferences. However, it is also possible to directly compare the gender gap in the control treatment to the gender gap in tournament entry to estimate the importance of competitiveness, along the lines of the main treatment estimate in this study. The main identifying assumption is that the control treatment removes the effect of competitiveness, but keeps the riskiness of the environment and the subjective probability of winning the same as in the baseline tournament entry decision. Neither Niederle and Vesterlund (2007) nor subsequent experiments with the same control treatment include this comparison.<sup>5</sup>

The seventh study (Dohmen and Falk, 2011) uses a different design. There, one-third of

5. Gillen et al. (2019) do include this comparison when re-analyzing the data from Niederle and Vesterlund (2007) (Table 4).

participants choose between tournament incentives and fixed pay. The other participants instead choose between fixed pay and either a piece rate or revenue sharing. Where the tournament is competitive, the piece rate and revenue sharing are not. However, all three options are riskier than fixed pay, and are also more likely to attract confident participants. If one is willing to assume that the riskiness and expected payment of the risky option is similar across treatments (a strong assumption), then the piece rate and revenue sharing treatments can serve as a non-competitive control for the tournament treatment.

The results (Table A8) are broadly in line with my treatment comparison results. On average, the gender gap is 35% smaller in the control treatment, which would imply that 35% of the gender gap in tournament entry in these studies is driven by competitiveness. This is considerably less than the corresponding average estimate obtained by residualizing competitiveness using regressions (78%). The main difference compared to my estimates is that the point estimate for competitiveness is somewhat larger. This appears to be driven by two studies, and could be due to order effects (the control treatment always came after the baseline entry decision) or gender differences in optimism about future performance (which are also eliminated in the control treatment).<sup>6</sup>

Flory et al. (2015) provide similar evidence from the field. In particular, they show that moving from competitive to non-competitive bonus payments reduces the gender gap in willingness to apply for administrative jobs from 12.9 to 12.6 percentage points. The difference is not significant, and the point estimate implies that competitiveness explains 2.6% of the gender gap in the willingness to apply for competitive jobs.<sup>7</sup>

Another set of studies include treatments that arguably leave the competitiveness of the environment constant, but reduce the gender difference in the subjective probability of winning by making men less or women more confident. If competitiveness is the main driver

6. The latter explanation is consistent with the data from Stage 5 in experiment 2. This stage also removed the role of gender differences in optimism, and we saw in the Appendix A3.1 that this may well explain the smaller gender gap in that stage compared to Stage 7.

7. Flory et al. (2015) present the results of these treatments in Table 8, but do not report the raw gender gap in application rates for the two types of jobs. These numbers are therefore based on own calculations using their data.

of the gender gap in tournament entry, then even treatments that eliminate all of the gender gap in confidence should only have a modest impact on the gender gap in tournament entry. If competitiveness is not important, however, then these treatments may well eliminate most of the gender gap in tournament entry.

Several studies decrease or eliminate gender differences in confidence by design, either by using a less stereotypically male task (Grosse et al., 2014; Shurchkov, 2012; and Dreber et al., 2014) or by providing performance feedback (Ertac and Szentes, 2011; and Wozniak et al., 2014). In all five cases, the treatment eliminates most of the gender gap in tournament entry; the results imply that respectively 62, 64, 70, 87, and 74% of the gender gap are due to gender differences in confidence.<sup>8</sup> This is not consistent with competitiveness being the main driver of the gender gap in tournament entry.

Also related are two studies by Niederle et al. (2013) and Balafoutas and Sutter (2012) that investigate whether various affirmative action policies such as quota can reduce the gender gap. Intuitively, these policies make it less likely for men and more likely for women to win, which should lower (and perhaps even reverse) the gender gap in confidence. However, the environment arguably remains competitive, and hence it seems reasonable to assume that the role of competitiveness remains similar. Nevertheless, both studies find that many policies can be very effective, with the quota in Niederle and Vesterlund (2007) even fully reversing the gender gap. Provided that the role of competitiveness indeed does not change due to these policies, these results are not consistent with competitiveness being the main driver of the gender gap in tournament entry, and suggest an important role for gender differences in confidence.<sup>9</sup>

Finally, Bartos and Van Veldhuizen (2022) measure competitiveness directly using a method not confounded by risk attitudes, performance and confidence. In particular, they

8. The numbers are calculated by taking one minus the ratio between the gender gap in the control treatment and the gender gap in the baseline. For Shurchkov (2012) I take the data from the high pressure treatments only.

9. As a potential alternative mechanism, it also seems possible that participants see the presence of a female quota as a signal that men should compete less and women should compete more (similar to an experimenter demand effect). This is in line with Kessel et al. (2021), who show that merely advising women to compete more and men to compete less already reduces the gender gap in willingness to compete.

ask participants to choose between a random payment scheme and a competition. Both alternatives are constructed so that they have a 50% chance of paying 2 Euro per addition problem solved, and a 50% chance of paying nothing. The difference is that in the former case uncertainty is resolved by a computer, whereas in the latter case it is resolved by a competition against either a superior or an inferior opponent (see Appendix A3.2). Bartos and Van Veldhuizen (2022) present participants with a price list that increases the incentives of one of the alternatives to up to 3 Euros per addition problem. If men are more competitive than women, they should be more likely to prefer the competition even when it pays a smaller amount than the random payment (e.g., 2 Euros versus 2.10 Euros per exercise). However, Bartos and Van Veldhuizen (2022) find no evidence that this is the case; similar results are reported by Lozano and Reuben (2022). In other words, neither study presents evidence that men are systematically more competitive than women.

#### ***A5. Decomposing the Gender Gap in Tournament Entry***

What explains the gender gap in tournament entry if not a gender difference in a competitiveness trait? One possibility is that women are less confident in their ability to win a tournament. Indeed, men (57.6%) are more confident across both experiments in the preferred sample than women (50.2%) in the belief elicitation task in Stage 4 ( $p < 0.0001$ , t-test). These results are in line with previous work that has almost universally found that men are more confident than women in these experiments; I also obtain similar results when looking at the Stage 5 elicitation task (experiment 1) or the elicitation task prior to Stage 7 (experiment 2). A second possibility is that women are more risk averse. In keeping with this, women are significantly more risk averse on the qualitative (SOEP) measure of risk attitudes ( $p = 0.016$ ), the Holt and Laury (2002) measure ( $p = 0.001$ ) and the Eckel and Grossman (2002) measure ( $p = 0.001$ ).

### A5.1 Risk Preferences and Confidence

To distinguish between these explanations, I examine the results of a second control treatment (treatment JUSTRISK) that builds on treatment NOCOMP but also removes the role of gender differences in confidence. To explain the logic of this treatment, let us go back to treatment NOCOMP. Recall that this treatment uses the single choice made by a particular participant in Stage 6 for which the win probability of the lottery most closely corresponds to their elicited belief  $p_i^s$ . By virtue of being more confident, men, on average, faced more attractive lotteries in treatment NOCOMP than women. Treatment JUSTRISK seeks to remove the role of gender differences in confidence. This treatment therefore requires that men and women should face equally attractive lotteries or, more precisely, the same distribution of lotteries.

How should this be implemented in practice? One way to do this would be to take the same lottery (i.e., the same row in Stage 6) for all participants. However, this approach is problematic because it would not just ensure that men and women face the same set of lotteries, but also changes the overall distribution of lotteries compared to treatment NOCOMP. This is important, because a key goal of treatment JUSTRISK is to keep the riskiness (and hence the role of risk attitudes) the same as in treatment NOCOMP (and the baseline). For example, suppose that for every participant I used the first row in Stage 6, where the lottery pays out with probability one. While this would indeed ensure that men and women face the same set of lotteries, it would be hard to argue that risk attitudes play the same role in this decision as they do in treatment NOCOMP.

Instead, my approach is to remove the gender gap in the distribution of lotteries while otherwise keeping the distribution as similar to treatment NOCOMP as possible. I do this by replacing men's distribution of lotteries with the distribution of lotteries faced by women. This requires several steps. First, I rank the subjective beliefs  $p_i^s$  elicited over all sessions in both experiments from largest to smallest, separately for each gender. For each man in the sample, I then replace his elicited belief with the belief of the woman with the same percentile rank. Because of ties and an unequal number of male (245) and female (242) participants, I

am unable to match participants exactly. Instead, I match them using simulations in which ties are broken randomly in every iteration.<sup>10</sup> Third, I then use the single choice made in Stage 6 that most closely corresponds to these adjusted beliefs as treatment JUSTRISK in my analysis. I will use the median gender gap in treatment JUSTRISK across all 9999 simulated samples to obtain the point estimates below.

It is useful to illustrate the logic of this method using a simple example that assumes there are an equal number of male and female participants and no ties. Consider the 35th most confident man and woman in the sample, and assume that they believe that they have an 80% chance and a 70% chance to win the tournament respectively. Treatment NOCOMP uses the rows corresponding to these beliefs, giving the man a more attractive lottery (80% vs 70% chance of winning). By contrast, treatment JUSTRISK uses the same row for both genders: it selects the row with a win chance of 70%, in line with the woman's elicited belief. When doing this for the whole sample, this procedure mechanically ensures that men and women face the same set of lotteries. This in turn eliminates the role of gender differences in confidence, which in turn implies that any residual gender gap in treatment JUSTRISK can be attributed to a gender difference in attitudes towards risk.

Figure A10 presents the results using the pooled data from both experiments and the same sample I use in the main analysis. If confidence is important, the gender gap should be significantly smaller in treatment JUSTRISK than in treatment NOCOMP. This is indeed what I find: 52.7% of men and 41.7% of women choose the lottery in this treatment (Figure A10). The resulting gender gap (10.9pp) is significantly smaller than in treatment NOCOMP (the median one-sided p-value in the simulations is  $p < 0.001$ ). The point estimate implies that the gender difference in confidence explains 49.2% ( $\frac{20.9-10.9}{20.3}$ ) of the gender difference in tournament choices. Risk attitudes play an important role as well: the residual gender gap in treatment JUSTRISK is still significant (the median p-value in the simulations is  $p = 0.007$ ). The point estimate implies that risk attitudes can explain 53.7% ( $\frac{10.9}{20.3}$ ) of the gender gap in

10. I match each man's belief with a weighted average of the two women with most closely matching percentile ranks to deal with the unequal number of male and female observations.



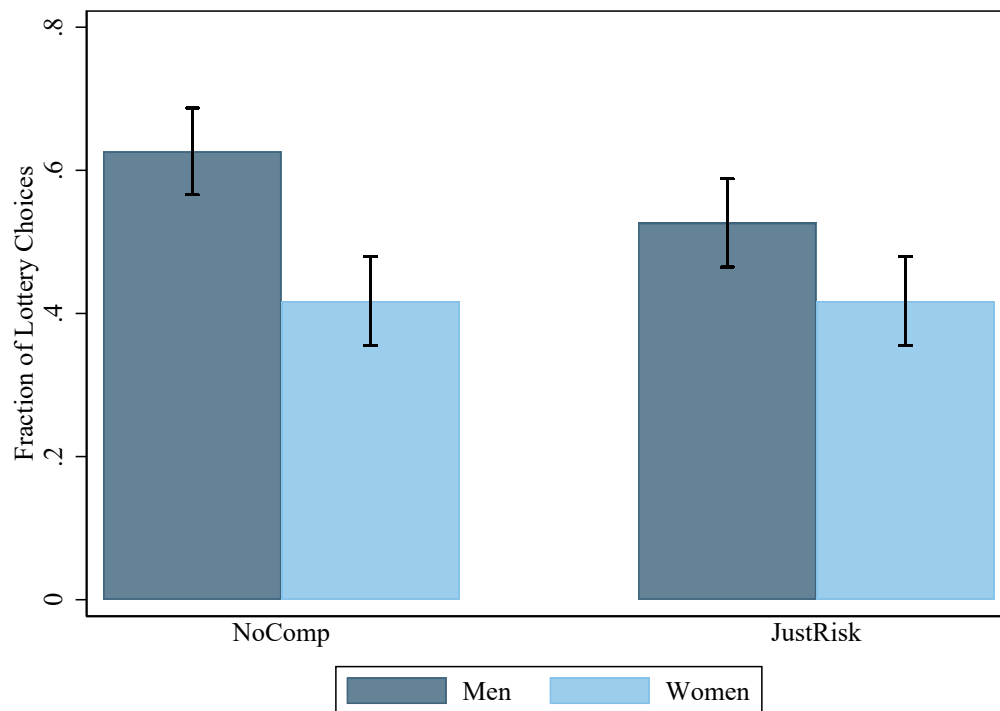


FIGURE A10: SUMMARY OF CHOICES IN TREATMENTS NOCOMP AND JUSTRISK.

*Notes.* This figure gives the fraction of participants choosing the lottery in treatments NOCOMP and JUST-RISK by gender. The error bars represent 95% confidence intervals. The figure includes participants from both experiments, and excludes participants who violated expected utility or displayed extreme levels of overconfidence or risk aversion.

tournament entry. Finally, the comparison between the baseline and treatment NOCOMP tells us that -2.9% ( $\frac{-0.6}{20.3}$ ) of the gender gap in tournament entry is due to competitiveness, as discussed extensively in the main text. These results are robust to using different exclusion criteria or using men as the reference sample, as I show in Section A5.3 below.

Two more remarks are in order. First, under non-expected utility the point estimate for ‘confidence’ may also capture gender differences in probability weighting or ambiguity aversion. Similarly, the point estimate for ‘risk attitudes’ will capture gender differences in loss aversion, the reference point, and the curvature of the value function. Previous research, however, suggests that ‘confidence’ and ‘risk attitudes’ are likely to be the most natural labels for these two effects.<sup>11</sup> Second, two other explanations discussed in the literature (gender differences in altruism and feedback aversion) are either excluded by the fact that participants’ tournament entry decisions do not affect the earnings of others (altruism) or are not supported by the data from experiment 1’s Stage 5 (feedback aversion, see Appendix A1.2).

## A5.2 Overconfidence and Performance

Why are men more confident than women? One potential reason is that men overestimate their relative performance more than women, i.e., that they are more overconfident. A second potential reason is that men actually perform better at the task than women. In the preferred sample, men (29.7%) are indeed slightly more likely to win the tournament than women (24.6%,  $p=0.053$ , two-sided t-test). By contrast, men (30.0pp) are not significantly more overconfident than women (27.7pp,  $p=0.17$ , one-sided t-test). Nevertheless, it may be useful to further decompose the effect of confidence ( $p_i^s$ ) into performance (captured by  $p_i^o$ , the true objective probability of winning) and overconfidence ( $p_i^s - p_i^o$ ).

11. There is robust evidence from a large number of studies linking the gender gap in tournament entry to gender differences in risk attitudes and confidence (see e.g., Niederle and Vesterlund, 2007 or Niederle, 2016). By contrast, I am only aware of two studies that have explored the role of any of the other variables. Both studies examine ambiguity version. Saccardo et al. (2018) find no correlation between ambiguity attitudes and tournament entry, whereas Balafoutas and Sutter (2019) find that ambiguity averse participants are less likely to choose the tournament. However, since Balafoutas and Sutter (2019) find no gender difference in attitudes towards ambiguity, ambiguity aversion cannot explain the gender gap in willingness to compete.

To distinguish between these two channels, I examine an extension of treatment NO-COMP (NOCOMP<sub>PERF</sub>) that adjusts beliefs for gender differences in performance. For this purpose, I first compute each participant’s objective probability of winning  $p_i^o$  in Stage 2 as the empirical probability that this participant would be randomly matched to three participants with an inferior score, or win a random tiebreaker against one or more participants with the same score. Second, I then rank participants by  $p_i^o$ , separately by gender. Third, for each man in the sample, I then replace his objective probability  $p_i^o$  by the objective probability of the woman with corresponding percentile rank ( $p_i^{o,f}$ ). Fourth, for each man in the sample I then construct the ‘performance-neutral’ subjective belief  $p_i^{s,n}$  by adding the man’s overconfidence ( $p_i^{s,n} = p_i^{o,f} + (p_i^s - p_i^o)$ ). I once again use simulations to deal with ties and non-equal numbers of male and female participants.

A brief example may be useful. Consider a man and woman with equal percentile performance rank. Suppose that the man thinks he has a 70% chance of winning the tournament ( $p_i^s=0.7$ ), but his objective probability of winning is only 36% ( $p_i^o=0.36$ ). Assume that his female counterpart is less confident ( $p_i^s=0.6$ ), but also had a slightly lower score and hence a smaller likelihood of winning the tournament ( $p_i^o=0.30$ ). In this example, the man is overconfident by 34 percentage points, the woman by 30. The man’s ‘performance-neutral’ subjective belief  $p_i^{s,n}$  will then be equal to 0.64. In other words, after adjusting for gender differences in the objective probability of winning, the man is still more confident, but the difference in confidence has decreased from 10 to 4 percentage points.

Table A9 presents the results of the full decomposition for different sample restrictions; the results for the sample used in the main analysis are presented in column 3. In this sample, the gender gap in treatment NOCOMP<sub>PERF</sub> is 15 percentage points. This implies that gender differences in performance can explain  $\frac{20.9-15.0}{20.3} = 29.3\%$  of the gender gap in tournament entry ( $p<0.0001$ ); gender differences in overconfidence explain  $\frac{15.0-10.9}{20.3} = 19.9\%$  ( $p=0.021$ ). Taken together, my results therefore suggest that about half of the gender gap in tournament entry can be explained by a gender difference in risk attitudes; overconfidence and performance differences explain about 20% and 30% respectively, and competitiveness

TABLE A9: DECOMPOSING THE GENDER GAP IN TOURNAMENT ENTRY

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Gender Gap in Tournament Entry	-0.213*** (0.041)	-0.205*** (0.042)	-0.203*** (0.044)	-0.218*** (0.046)
Gender Gap in NOCOMP	-0.191*** (0.041)	-0.191*** (0.042)	-0.209*** (0.044)	-0.233*** (.045)
Gender Gap in NOCOMP <sub>PERF</sub>	-0.147*** (0.040)	-0.138*** (0.041)	-0.150*** (0.043)	-0.161*** (.044)
Gender Gap in JUSTRISK	-0.106** (0.041)	-0.110*** (0.042)	-0.109** (0.044)	-0.105** (0.046)
Competitiveness	10.1% (22.1)	6.9% (23.3)	-2.9% (24.5)	-6.9% (23.8)
Performance	20.8%*** (7.0)	25.8%*** (7.4)	29.3%*** (7.9)	32.9%*** (7.8)
Overconfidence	19.1%** (8.3)	13.8%* (8.8)	19.9%** (9.7)	24.6%*** (9.9)
Risk Attitudes	50.0%** (19.4)	53.4%*** (20.5)	53.7%** (21.8)	48.3%** (20.9)
EU Violations	yes	no	no	no
Extreme Risk Prefs	yes	yes	no	no
Extreme Confidence	yes	yes	no	no
Low Understanding	yes	yes	yes	no
N(men)	283	276	245	231
N(women)	281	268	242	227

*Notes.* Robust standard errors in parentheses. The first four rows present the raw gender gap in the baseline (Stage 4) and three control treatments. For treatments NOCOMP<sub>PERF</sub> and JUSTRISK, these are the median estimates across 9999 simulated samples. “Competitiveness” is the point estimate for the importance of competitiveness, computed as the difference between the first two coefficients, normalized as a percentage of the baseline gender gap. “Performance” is the normalized difference between the third and second coefficient, “Overconfidence” is the normalized difference between the fourth and third coefficient, and “Risk Attitudes” is the normalized gender gap in treatment JUSTRISK. The first column includes all observations, the second removes participants who violated EU in Stage 6 (or Stage 5 in experiment 2), the third also removes participants with extreme risk preferences or extreme levels of overconfidence, and the fourth also removes participants flagged as having low understanding in experiment 2. See section 4.2.1 for more details regarding these exclusion restrictions.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

does not play a role.

### **A5.3 Robustness Checks**

Table A9 already demonstrates that the results are robust to using different exclusion restrictions. Table A10 reproduces the results of A9 while using men as the reference group to construct treatments JUSTRISK and NOCOMPPERF. In this case, I therefore replace each *woman's* subjective probability  $p_i^s$  (JUSTRISK) and  $p_i^o$  (NOCOMPPERF) by the probabilities of the man with the corresponding rank. The results are very similar; the main difference is that the point estimate for risk attitudes falls by about 5 percentage points and the estimate for overconfidence increases correspondingly.

Finally, note that treatment NOCOMPPERF computes participants' objective win probability  $p_i^o$  using their Stage 2 performance. As an alternative, I can also compute  $p_i^o$  using performance in a later part of the experiment. This allows me to capture learning effects, though it should be noted that performance in later stages may be endogenous to the choice of incentives. Using Stage 3 or 4 instead of Stage 2 performance slightly increases the gender gap in  $p_i^o$ , which increases the estimated importance of performance differences and lowers the importance of differences in overconfidence. The full results are available upon request.

## ***A6. Generating Figure 1***

Figure 1 in the main text presents the results of a number of previous studies that incorporate a variation of the regression specification used by Niederle and Vesterlund (2007) to identify the importance of competitiveness. In this section, I discuss how I selected the studies used in Figure 1 and present some additional details on their designs.

### **A6.1 Overview of Studies**

Table A11 summarizes the studies I incorporated in Figure 1. My inclusion criterion was for these studies to report both (1) the raw gender difference in tournament entry and (2) the residual gender coefficient after controlling for both risk preferences and confidence

TABLE A10: DECOMPOSITION WITH MEN AS THE REFERENCE GROUP

	Coefficient (Std. Errors)			
	(1)	(2)	(3)	(4)
Gender Gap in Baseline	-0.213*** (0.041)	-0.205*** (0.042)	-0.203*** (0.044)	-0.218*** (0.046)
Gender Gap in NoCOMP	-0.191*** (0.041)	-0.191*** (0.042)	-0.209*** (0.044)	-0.233*** (.045)
Gender Gap in NoCOMP <sub>PERF</sub>	-0.145*** (0.039)	-0.152*** (0.040)	-0.147*** (0.043)	-0.164*** (.044)
Gender Gap in JUSTRISK	-0.087** (0.041)	-0.093** (0.041)	-0.098** (0.044)	-0.103** (0.045)
Competitiveness	10.1% (22.1)	6.9% (23.3)	-2.9% (24.5)	-6.9% (23.8)
Performance	21.8%*** (9.0)	19.3%*** (8.6)	30.6%*** (9.3)	31.5%*** (9.5)
Overconfidence	27.5%** (9.1)	28.8%*** (8.8)	24.3%*** (9.0)	28.1%*** (9.1)
Risk Attitudes	40.5%** (19.0)	45.0%** (20.0)	48.0%** (21.7)	47.3%** (20.6)
EU Violations	yes	no	no	no
Extreme Risk Prefs	yes	yes	no	no
Extreme Confidence	yes	yes	no	no
Low Understanding	yes	yes	yes	no
N(men)	283	276	245	231
N(women)	281	268	242	227

*Notes.* Robust standard errors in parentheses. For variable definitions, see the notes to Table A9.  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

TABLE A11: OVERVIEW OF STUDIES IN FIGURE 1.

Study	Gender Gap		Attributed to Competitiveness	Controls	Source
	Raw	Controls			
Gillen, Snowberg, and Yariv (2019)	19***	4.8	25.3	GR,TP,PD,IG,CRT1	Table 3 col. 1 and 5
Zhang (2019)	14.6***	3.7	25.3	OC,EG,PW	Table 2 col. 1 and 5 (Han sample)
Dreber, von Essen, and Ranehill (2014)	19.1***	5.8	30.4	TP,GR,MPL	Table 4 col. 1 and 3
Kamas and Preston (2012)	17.6***	5.7	32.4	MPL,RQ,GR,EP,PC	Table 1 row 3 and Table 6 col. 7
Niederle and Vesterlund (2007)	37.9***	16.2**	42.7	TP,PD,GR,CT	Table 8 col. 1 and 3
Balafoutas, Kerschbamer, and Sutter (2012)	26.1***	11.6	44.4	TP,MPL,GR,OC	Table 5 col. 2 and 3
Buser, Geijtenbeek, and Plug (2018)	8.6*	4.0	46.5	RQ,BO	Table 3 col. 1 and 3
Buser, Niederle, and Oosterbeek (2014)	26***	12.2***	46.9	TP,PD,PW,GR,EG,RQ	Table 7 col. 4 and raw gap on p.1427
Niederle, Segal, and Vesterlund (2013)	36***	17**	47.2	TP,PD,GR,CT	Table 4 col. 1 and 3
Buser, Dreber, and Mollerstrom (2017)	23.9***	11.7	49.0	PP,TP,GR,GR1,RQ	Table 2 col. 1 and 4 (experiment 1) <sup>a</sup>
Reuben, Sapienza, and Zingales (2015)	26.8***	13.3**	49.7	TP,PD,GR,MPL	Table 3 col. 1 and 3
Gillen, Snowberg, and Yariv (2019)	19***	11***	57.9	GR,TP,PD,MPL,CRT2	Table 3 col. 1 and 4
Apicella, Demiral, and Mollerstrom (2017)	19.5*	11.4	58.5	RQ,PP,BO2	Table 2 col. 1 and 2 (lab)
Halko and Sääksvuori (2017)	24.1***	14.3*	63.1	PD,GR,RQ	Table 2 col. 4 and 5
Kessel, Mollerstrom, and Van Veldhuizen (2021)	26.0***	16.0***	61.5	TP,PD,IG,SPW	Own calculations
Reuben, Wiswall, and Zafar (2017)	29***	18**	62.1	PW,SPW,MPL	Table 2 col. 1 and 3
Dohmen and Falk (2011)	25***	15.7	62.8	EP,BO,RQ,SP	Table 2 col. 2 and Table 3 col. 2
Buser, Ranehill, and Van Veldhuizen (2021)	29.8***	18.8***	63.1	TP,GR,IG	Table 3 col. 1 and 3
Apicella, Demiral, and Mollerstrom (2017)	12.6**	9.0	71.4	RQ,PP,BO2	Table 2 col. 1 and 2 (online)
Almäs et al. (2016)	19.4***	13.9***	71.6	TP,BO,MPL,SP,PM	Table 1 col. 1 and 4
Sutter and Glätzle-Rützler (2015)	21.9***	16.5***	75.3	A,TP,PD,M,GR,MPL	Table 6 col. 1 and 2
Saccardo, Pietrasz, and Gneezy (2018)	36.4***	31.3***	86.0	EP,BO,MPL,RQ,AA	Table 2 col. 1 and 6
Flory, Gneezy, Leonard, and List (2018)	15.6	14.0	89.7	PP,PD,GR,CT	Table 1 col. 1 and 2 (American sample)
Dargnies (2012)	33.3***	29.9***	89.8	CT	Own calculation
Buser, Dreber, and Mollerstrom (2017)	28.9***	27***	93.4	PP,TP,GR,GR1,RQ	Table 7 col. 1 and 4 (experiment 2) <sup>a</sup>
Healy and Pate (2011)	52.7***	51.2***	97.2	TP,PD,GR,CT	Table 2 col. 2 and 4
Flory, Gneezy, Leonard, and List (2018)	7.9**	7.8**	98.7	PP,PD,GR,CT	Table 1 col. 1 and 3 (Malawi sample)
Zhang (2019)	22.9***	23.7***	103.5	OC,EG,PW	Table 2 col. 1 and 5 (Yi sample)
Zhang (2019)	27.5***	29.7***	108.0	OC,EG,PW	Table 2 col. 1 and 5 (Mosou sample)
Masclet, Peterle, and Larribeau (2015)	37.6**	41.4**	110.1	TP,EC,MPL,SP,GR	Table 1 col. 4 and 5
Lee, Niederle, and Kang (2014)	7.6***	9.3***	122.4	TP,PD,GR,CT	Table S2 col. 2 and raw gap on p. 476
Cassar, Wordofa, and Zhang (2016)	9.9**	12.2**	123.2	PW,MPL,BO3	Table S3 col. 1 and 4
Average			72.2		

*Notes.* The raw gender gap is the raw gender difference in tournament entry. The gender gap with controls is the gender coefficient in a regression of tournament entry on gender, risk preferences, ability, overconfidence, and potentially other factors. The fraction attributed to competitiveness is the ratio of the first two columns. The fourth column specifies the exact control variables used in the regression. More details on these variables are in the main text. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . (a) The entries for Buser et al. (2017) use regression output from the working paper version (Buser et al., 2015)

in a regression. Figure 1 plots the ratio of these two numbers, which is what is used as the point estimate for competitiveness when residualizing competitiveness using regressions (Niederle and Vesterlund, 2007).

The exact specification used differs considerably across studies. Different studies use different proxies for risk preferences and beliefs. Most studies also include at least one control for ability or past performance, and several studies also control for one or more additional factors. Whenever possible, I use the regression that controls for risk preferences and overconfidence, but otherwise controls for as few variables as possible. The relevant control variables used in each study are listed in Table A11.

Several types of risk measures are used. Most common are a non-incentivized question (RQ) concerning a general tendency to take risks, and one of several versions of a multiple price list (MPL). These are similar to the SOEP question and Holt-Laury measure used in this study respectively. A few studies use variations of the Eckel and Grossman (2002) measure (EG) or the investment game (IG, Gneezy and Potters, 1997). One study also controlled for ambiguity aversion (AA) using an MPL.

Many measures are considered for confidence. A popular measure is the guessed rank (GR), typically elicited for the forced tournament (Stage 2). One study includes two guessed rank measures, one for the forced tournament and one for the Stage 1 piece rate (GR1). Several studies subtract the actual rank from the guessed rank to get a measure of overconfidence (OC). Other studies ask participants to guess the probability with which they would beat another person in their session (BO). Two studies include a dummy for participants who thought their score exceeded the score of their opponent (BO2) or a continuous variable for the difference between the two scores (BO3) respectively. Another study elicits the subjective probability of winning (SPW). A few studies ask participants for their expected performance on the task (EP), and one study also asks participants how certain they are about their prediction (PC). One study also includes two measures of overconfidence in an unrelated cognitive reflection test (CRT1 and CRT2).

In addition, a few studies also include the choice in a control treatment (CT). This



variable is argued to pick up the effect of overconfidence, risk preferences, and possibly other variables including ability or feedback aversion. It should also be noted that even studies that are classified as using similar controls (e.g., MPL) typically use different versions, e.g., with differently sized incentives.

A large majority of studies also control for performance. Common measures include performance in a forced tournament (TP), performance in a forced piece rate (PP), and the difference between the two (PD). Several studies also include a measure for the objective probability of winning the tournament (PW). Finally, several studies also control for additional variables, including age (A), education (EC), math ability (M) various social preferences (SP) and patience (PM).

Several other remarks are in order. First, I include five studies that do not use the addition problem task, but use a similar design with similar outcomes. Buser et al. (2018) use a matrix addition task in which participants have to find two numbers in a 3x3 matrix that jointly add up to 10. Saccardo et al. (2018) use a ball-throwing task, where participants have to throw a ball into a basket from several meters away. Flory et al. (2018) let participants sort six blocks from smallest to largest. Masclet et al. (2015) let them decode numbers into letters. Apicella et al. (2017) use a captcha task for their online experiment. However, I do not incorporate studies that use different tasks with the stated aim of finding a smaller gender difference. Instead, I briefly discussed these studies in Appendix A4 above.

Second, several studies are included in the sample more than once. Buser et al. (2017), Zhang (2019), Apicella et al. (2017) and Flory et al. (2018) carried out multiple experiments within the same study that all fit the stated criteria. I treated each of these experiments as a single observation. Gillen et al. (2019) ran a single experiment, but present many estimates that fit my criteria. I selected two representative results for Table A11; each of these estimates only receive half the normal weight in computing the total average effect.

Third, in Dohmen and Falk (2011) and Masclet et al. (2015), participants choose between a tournament and a fixed amount of money (as opposed to a piece rate). Risk preferences, overconfidence, and competitiveness can still explain competitive choices in both cases, and

hence I include these studies in my survey as well. Fourth, Dargnies (2012) does not report a relevant regression, but sent me her raw data, allowing me to run the regression myself.

Fifth, a few studies do not report the raw gender gap in tournament entry, but always control for at least one variable (typically ability). Sixth, several studies report probit coefficients, which cannot directly be compared to the raw gender gap. In both cases, I compare coefficients between the regression controlling for risk preferences and overconfidence, and the regression controlling for the smallest number of other variables (typically only some measure of task ability).

## A6.2 Results

The key result in Table A11 is that the average fraction attributed to competitiveness is 72.2%. I discuss this result extensively in the main text. It is also interesting that the raw gender gap differs considerably across studies, ranging from 8 to 53 percentage points. This could be due to a number of factors, including changes in the design (e.g., shorter tasks, internet versus laboratory experiment, the number of opponents), as well as the population studied in the paper (students vs. non-students, Western vs. non-Western). Nevertheless, the raw gender gap is significant at the 10% level in all but one of the 31 experiments.

The fraction attributed to competitiveness also differs strongly across studies. Individual estimates fall anywhere between 25% and 122%. Part of this variation can likely be explained by differences in the sample and design of the experiments. Another explanation is that certain control variables may be better than others at filtering out the effect of risk preferences and overconfidence. Indeed, the two estimates I include from Gillen et al. (2019) use different control variables in the same data, and find very different results. A third potential reason is measurement error, which, in addition to generating an upward bias in the importance of competitiveness, also increases the variation in estimates that may be obtained, which may also explain the estimates that exceed 100%.

Finally, it is useful to note that Niederle and Vesterlund (2007) identify competitiveness as the “*significant*” residual in tournament entry after controlling for performances,

confidence and risk aversion” (Niederle, 2017, p.116, emphasis added). By this measure, 21 out of 32 estimates in Table A11 (65.6%) present evidence of a gender difference in a competitiveness trait. Since the purpose of Figure 1 and Table A11 is to estimate the importance of competitiveness across all studies, the p-values of individual studies are less relevant. Nevertheless, note that the fraction of studies that find a significant competitiveness effect (65.6%) is similar to the average point estimate for competitiveness across all 31 experiments (72.2%).

### ***A7. Identifying Assumption***

In this section, I formalize my identifying assumption using a potential outcomes framework. I do so by splitting the main assumption into distinct parts for the treatment variable (competitiveness) and the control variables (risk attitudes and confidence), and discussing several sets of sufficient (but not necessary) conditions under which the identifying assumption holds.

#### **A7.1 A Formal Framework**

Let  $r^i$ ,  $o^i$ ,  $c^i$  be the risk preferences, (over)confidence and competitiveness of individual  $i$ . Let  $x^i$  be a potential confounding variable that may affect the gender difference in tournament choices, such as ambiguity aversion. Further, let  $R$ ,  $O$ ,  $C$  and  $X$  be the empirical distribution of the respective variables over the whole population. For our purposes, it is convenient to separate the population into two parts – men and women – e.g.,  $R = \{R_M, R_F\}$ . Finally, let

$$(R, O, C, X) = (\{R_M, R_F\}, \{O_M, O_F\}, \{C_M, C_F\}, \{X_M, X_F\})$$

be the joint empirical distribution of the four variables across the population.

I am interested in explaining the gender difference  $G_T(R_T, O_T, C_T, X_T)$  observed in tournament choices  $T$ . In particular, the main goal in the experiment is to investigate

whether  $G_T$  changes when I remove the effect of competitiveness ( $C_T$ ).

## A7.2 Identification Strategy

This could, in theory, be done in two distinct ways. One approach is to prevent the variable of interest from affecting tournament choices. I use this approach to identify competitiveness. Specifically, the goal is to compare the tournament  $T$  to an otherwise identical environment  $T'$  where competitiveness can no longer explain gender differences in choices, i.e.,  $G_{T'}(R_T, O_T, X_T)$ . In the experiment, I approximate  $G_{T'}$  using the gender gap  $G_N$  in treatment NOCOMP. Hence, my assumption is the following:

$$\text{ASSUMPTION 1. } G_N(R_N, O_N, C_N, X_N) = G_{T'}(R_T, O_T, X_T)$$

In words, the gender difference in treatment NOCOMP ( $G_N$ ) is assumed to be identical to the difference that would have been observed in the tournament had competitiveness not been able to explain tournament choices.

A second approach is to remove the gender difference in the variable of interest. I could, in theory, also have removed a possible gender difference in competitiveness, and then examined  $G_T(R_T, O_T, C_0, X_T)$ , where  $C_0$  is such that  $C_M = C_F$ . In that case, the identifying assumption would have reduced to:

$$\text{ASSUMPTION 2. } G_N(R_N, O_N, C_N, X_N) = G_T(R_T, O_T, C_0, X_T)$$

In words, this assumes that the gender difference in treatment NOCOMP is identical to the difference that would have been observed in the tournament in the absence of a gender difference in competitiveness. I use this approach to decompose the gender gap in tournament entry into risk attitudes and confidence in Appendix A5.

## A7.3 Sufficient Conditions

Let us now examine several special cases of sufficient conditions for the control and treatment variables under which assumption 1 and 2 hold. Let us start with a sufficient condition for the treatment variable  $C$ .

CONDITION 3. For the treatment variable  $C$  at least one of the following holds

1.  $G_N() = G_T(., ., C_0, .)$  (No gender difference)
2.  $G_N() = G_{T'}()$  (No effect on choices)

Condition 3 states that treatment NOCOMP either removes the gender difference in competitiveness, or removes the effect of this variable on choices. In other words, condition 3 states that the treatment successfully removed the effect of the treatment variable of interest. I impose condition 3 in the remainder of this section to allow me to focus on several alternative sufficient conditions for the control variables, starting with condition 4:

CONDITION 4. For the control variables  $R$ ,  $O$  and  $X$ :

$$(r_N^i, o_N^i, x_N^i) = (r_T^i, o_T^i, x_T^i) \forall i. \text{ (1-to-1 correspondence)}$$

Condition 4 requires the risk attitudes, (over)confidence, and other control variables governing choices to be identical in the tournament entry decision and treatment NOCOMP for all individuals. Since treatment NOCOMP is constructed using elicited beliefs, in practice this would also require that beliefs are perfectly measured. However, condition 4 is unnecessarily strong and the following (weaker) condition is sufficient as well.

CONDITION 5. For the control variables  $R$ ,  $O$  and  $X$ :

$$(R_T, O_T, X_T) = (R_N, O_N, X_N) \text{ (Identical distribution)}$$

Condition 5 allows individual risk attitudes, beliefs, and other factors to vary across treatments, as long as the joint distribution is the same in both cases. This allows elicited beliefs to be noisy, provided that errors cancel out, on average, in a way that leaves the distribution unaffected. If I assume that the three variables are independent conditional on gender, I can also rewrite the condition to apply to the distribution of each variable individually.

Condition 5 may be violated if, for example, the distribution of elicited beliefs has a higher variance than the latent belief distribution. However, I can weaken the condition even further:

CONDITION 6.  $G_T(R_T, B_T, C, X_T) = G_N(R_N, B_N, C, X_N)$

(Changes in control variables do not affect the gender difference in outcomes)

Condition 6 says that assumptions 1 and 2 will also hold if changes in the control variables do not affect the gender difference  $G$ . For example, beliefs may change in a way that makes men and women adjust their behavior in a similar direction, leaving the gender gap unaffected. As discussed in the main text, the results from various robustness checks suggest that this condition is satisfied.

### ***A8. Power Calculation for Experiment 2***

The main purpose of the experiments in this paper is to use causal treatments to test whether competitiveness can explain part of the gender gap in willingness to compete. The purpose of these power calculations was to determine the appropriate sample size for experiment 2. For this purpose, I started by taking the fraction of men and women who entered the tournament in experiment 1 as the benchmark for willingness to compete in experiment 2 (men: 67.1%, women: 34.3%). I then specified several possible effect sizes for competitiveness, labeled in terms of the percentage of the original gender gap that competitiveness might eliminate. I used four values: 100% (competitiveness explains the entire gender gap in willingness to compete), 70% (the fraction approximately attributed to competitiveness in Figure 1), 50% (competitiveness explains half the gender gap) and 35% (half the fraction attributed to competitiveness in Figure 1).

Since my main test is a difference-in-difference test, I computed power  $P(p < 0.05|H_1)$  using simulations. For each simulation sample, I first determined tournament entry in the baseline by drawing tournament entry decisions from a binomial distribution with  $p = 0.671$  for men and  $p = 0.343$  for women. I then generated decisions in treatment NOCOMP using a distribution that ensures that the expected gender gap in NOCOMP was reduced by the hypothesized effect size (100%, 70%, 50% or 35%), while ensuring that the total fraction of participant who switched from the baseline to NOCOMP was identical to experiment 1 (31%). Allowing for switching helps capture both the fact that some participants may

TABLE A12: POWER CALCULATIONS FOR EXPERIMENT 2

$H_1$	N=50	N=100	N=150	N=200	N=250	N=300	N=350	N=400	N=450	N=500
100%	0.692	0.929	0.981	0.996	1	1	1	1	1	1
70%	0.461	0.671	0.822	0.912	0.963	0.978	0.991	0.993	0.999	0.998
50%	0.293	0.437	0.572	0.646	0.727	0.824	0.859	0.914	0.918	0.950
35%	0.191	0.269	0.339	0.436	0.447	0.554	0.584	0.625	0.699	0.748

*Notes.* The first column presents the target effect of competitiveness, expressed as a percentage of the baseline gender gap in tournament entry. The remaining numbers give the statistical power to detect an effect of the specified size, computed as the fraction of simulated samples in which the observed effect was significant at the 5% level (one-sided difference-in-difference test).

switch by mistake (or due to measurement error in beliefs) and the fact that there may be some competitive women and non-competitive men in the sample. I always assumed that removing competitiveness changes both genders' willingness to compete in a symmetric way, e.g., 42.5% of women and 58.9% of men competing in treatment NoCOMP if competitiveness explains 50% of the gender gap in the baseline.

Table A12 presents the results for various sample sizes. For each sample size and effect size combination, I simulated 1000 samples using the procedures described above; power is then equal to the fraction of simulated samples that have  $p < 0.05$  on a one-sided difference-in-difference test. For example, the table tells us that for a sample size of 150 participants (75 men and 75 women), I could have expected to have a power of 0.981 to detect a 100% decrease in the gender gap in NoCOMP. Based on these results, I determined that a target sample of 400 participants for experiment 2 would be appropriate. This would allow me to detect the average effect of competitiveness found in Figure 1 (70%) with very high power (0.993), while allowing me to still detect smaller effects with relatively high probability even when removing outliers from the analysis.

## APPENDIX B: EXPERIMENTAL DETAILS AND PROCEDURES

### *B1. Design of Experiment 2*

Experiment 2 was conducted at the experimental economics laboratory of the Technical University Berlin in October and November 2019. The experiment was programmed using PHP/MySQL, and participants were recruited using ORSEE (Greiner, 2015). Figure B1 summarizes the various stages of the experiment.

Stages 1, 2, 3, 4 and 6 were identical to experiment 1. The only change was that experiment 2 randomized the order of Stages 3, 4, 5 and 6 between sessions, allowing me to test for order effects. In order to facilitate comprehension, I divided the Stages into a lottery cluster (Stage 5 and 6) and a real-effort cluster (Stage 3 and 4), and made sure that the two stages within each cluster always followed each other (though the order within a cluster was randomly determined).

Stage 5 and 7 were new to experiment 2. Stage 5 presented participants with a decision that was similar to Stage 6, except that participants made only a single binary decision. For this purpose, I computed the probability that their score in Stage 2 would beat three randomly chosen opponents from experiment 1 (or win a random tiebreaker), and used this as the probability of payment in the lottery. For example, a participant with a score of 12 had a 31% chance of beating three random opponents from experiment 1. As a result, this participant would receive a choice between obtaining 6 Euro for sure and a 31% chance of obtaining 24 Euro in Stage 5. I calibrated the stakes in Stage 5 to match the expected payments of the tournament entry decision, similar to Stage 6.

Stage 7 was very similar to Stage 3 and Stage 4. The main difference was that, prior to making their decision, participants were informed of the Stage 2 performance of every participant in their session. They were also told their probability of winning a tournament if everyone's performance was as in Stage 2. The goal of this stage was to study tournament entry in a setting where participants knew their relative ability. Stage 7 also included two belief elicitation tasks, one before and one after receiving the feedback on relative ability.



FIGURE B1: EXPERIMENTAL TIMELINE FOR EXPERIMENT 2

Stage 1:	<u>Piece Rate</u> <ul style="list-style-type: none"><li>• Solve addition problems for 5 minutes with Piece Rate incentives (0.50€ per correct answer)</li></ul>
Stage 2:	<u>Tournament</u> <ul style="list-style-type: none"><li>• Solve addition problems for 5 minutes with Tournament incentives (2€ per correct answer if best performer in group of four)</li></ul>
Stage 3:	<u>Choice</u> <ul style="list-style-type: none"><li>• Choose between Piece Rate and Tournament incentives</li><li>• Solve addition problems for 5 minutes under chosen incentive</li></ul>
Stage 4:	<u>Choice + Belief Elicitation</u> <ul style="list-style-type: none"><li>• Belief elicitation task</li><li>• Choose between Piece Rate and Tournament incentives</li><li>• Solve addition problems for 5 minutes under chosen incentive</li></ul>
Stage 5:	<u>Binary Lottery Choice</u> <ul style="list-style-type: none"><li>• a single binary choice between <math>0.5x_i</math> and <math>2x_i</math> with probability <math>p_o</math> with <math>x_i</math> equal to performance in Stage 2</li></ul>
Stage 6:	<u>Non-Competitive Choice List</u> <ul style="list-style-type: none"><li>• 20 binary choices between <math>0.5x_i</math> and <math>2x_i</math> with probability <math>p</math> with <math>x_i</math> equal to performance in Stage 2 and <math>p \in \{0.05, 0.1, \dots, 0.95, 1\}</math></li></ul>
Stage 7:	<u>Choice After Feedback</u> <ul style="list-style-type: none"><li>• Belief elicitation task</li><li>• Feedback on relative ability in Stage 2</li><li>• Belief elicitation task</li><li>• Choose between Piece Rate and Tournament incentives</li><li>• Solve addition problems for 5 minutes under chosen incentive</li></ul>
	<u>Payment Screen</u> <ul style="list-style-type: none"><li>• One stage selected for payment, obtain feedback for this stage</li></ul>
	<u>Questionnaire</u> <ul style="list-style-type: none"><li>• Demographics and risk preference elicitation</li></ul>
	<i>Note.</i> The order of Stages 3–6 is randomized between sessions.

To reduce the length and complexity of this stage, only the former was incentivized. Note that Stage 7 was always the last stage in the experiment in order to prevent the performance feedback from influencing the decisions in Stage 3 and 4.

The questionnaire was also identical to experiment 1, with one difference: it now ended with a brief survey aimed at establishing whether each participant was from an East-German or West-German household. This survey was used by another team of researchers to study whether East-German and West-German participants differed in their willingness to compete. However, since they found no difference between the two groups, they decided to not write up their results as a paper.

Each session took between 75 and 90 minutes. There were eighteen sessions, two with 20 participants and sixteen with 24. Only participants who had taken part in ten or fewer previous experiments and had at a most a single instance of failing to show up for a previous experiment were allowed to register for the experiment. An equal number of men and women were invited to each session participants. There were a total of 424 participants (213 men and 211 women). All participants who showed up for a session agreed to sign a standard consent form required to take part in the experiment. Average earnings in the experiment were 17.06 € with a minimum of 7.20 € and a maximum of 60 €. 96.5% of participants indicated that they were students, most commonly majoring in engineering (33%), economics (11%) or dual majoring in economics and engineering or mathematics (19%). The mean and median ages of participants in experiment 2 were 23.5 and 22 respectively.

## ***B2. Instructions for Experiment 1***

This is the English version of the instructions used for experiment 1. The original German version is included in Appendix C. The welcome section is required for all experiments in the lab at the Technical University of Berlin. The instructions for Stage 1 to Stage 3 closely followed Niederle and Vesterlund (2007). The instructions for the belief elicitation task were based on Mobius et al. (2014). I used the word ‘task’ instead of ‘stage’ in the instructions.

### **Welcome**

Welcome to our experiment. During the experiment, it is not allowed to use electronic devices or to communicate with other participants. On your computer, please only use the experimental software. Please do not communicate with other participants. If you have a question, please raise your hand, and we will come and answer your question in private. Please do not ask the question in a way that everyone can hear it. If the question is relevant for all participants, we will repeat and answer it for everyone. If you break one of these rules, you will be excluded from the remainder of the experiment and will not receive any earnings.

### **Instructions**

In the experiment today you will be asked to complete six different tasks. None of these will take more than 5 minutes. At the end of the experiment you will receive €3 for having completed the six tasks, in addition we will randomly select one of the six tasks and pay you based on your performance in that particular task. Once you have completed the six tasks we will determine which task counts for payment by asking one of you to roll a six-sided die. The method we use to determine your earnings varies across tasks. Before each task we will describe in detail how your payment is determined.

Your total earnings from the experiment are the sum of your payment for the randomly selected task, your €3 payment for completing the tasks, and a €5 show up fee. At the end of the experiment you will be asked to come to the side room where you will be paid in private.

### **Task 1: Piece Rate**

For Task 1 you will be asked to calculate the sum of five randomly chosen two-digit numbers. You will be given 5 minutes to calculate the correct sum of a series of these problems. You cannot use a calculator to determine this sum, however, you are welcome to write the numbers down and make use of the provided scratch paper. You submit an answer by clicking the submit button with your mouse. When you enter an answer the computer will immediately tell you whether your answer is correct or not. Your answers to the problems are anonymous.

If Task 1 is the one randomly selected for payment, then you will get 50 cents per problem you solve correctly in the 5 minutes. Your payment does not decrease if you provide an incorrect answer to a problem. We refer to this payment as the piece rate payment.

Please do not talk to one another for the duration of the experiment. If you have any questions, please raise your hand.

*[Participants were brought to a wait screen, and waited until everyone had finished reading these instructions. On the wait screen, they could re-read the instructions printed above. After everyone had finished reading the instructions, they then had five minutes to solve addition problems. At the end of the task, they received feedback on the number of exercises they had solved and were notified that the next task would start in 20 seconds.]*

### **Task 2: Tournament**

As in Task 1 you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. However for this task your payment depends on your performance relative to that of a group of other participants. Each group consists of four people. If Task 2 is the one randomly selected for payment then your earnings depend on the number of problems you solve compared to the three other people in your group. The individual who correctly solves the largest number of problems will receive €2 per correct problem, while the other participants will not receive any payment. We refer to this as the tournament payment. You will not be informed of how you did in the tournament until all six tasks have been completed. If there are ties, the winner will be randomly determined.

Please do not talk to one another. If you have any questions, please raise your hand.

*[The wait screen, task, and feedback were identical to Task 1. Notably, participants only received feedback on their absolute performance, not on their relative.]*

### **Task 3: Choice**

As in the previous two tasks you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. However, you will now get to choose which of the two previous payment schemes you would prefer to apply to your performance in the third task.

If Task 3 is the one randomly selected for payment, then your earnings for this task are determined as follows. If you choose the piece rate you will receive 50 cents per problem you solve correctly. If you choose the tournament your performance will be evaluated relative to the performance of the other three participants of your group in the Task 2-tournament. The Task 2-tournament is the one you just completed. If you correctly solve more problems than they did in Task 2, then you will receive four times the payment from the piece rate, which is €2 per correct problem. You will receive no earnings for this task if you choose the tournament and do not solve more problems correctly now, than the others in your group did in the Task-2 tournament. You will not be informed of how you did in the tournament until all four tasks have been completed. If there are ties the winner will be randomly determined.

The next computer screen will ask you to choose whether you want the piece rate or the tournament applied to your performance. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Please do not talk to one another. If you have any questions, please raise your hand.

*[Participants were brought to a wait screen containing the Task 3 instructions, and waited until everyone had finished reading these instructions. They then decided between tournament and piece rate pay. After everyone had made their decision, they had five minutes to solve addition problems. At the end of the task they received the same feedback as in Task 1 and Task 2.]*

#### **Task 4: Choice 2**

As in the three previous tasks you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. As in the previous task, you can choose which payment scheme you would prefer to apply to your performance. There will also be an additional part of the task that will be explained on the next page.

If Task 4 is the one randomly selected for payment then your earnings for this task are determined the same way as in Task 3. In particular, you can choose between:

- Piece Rate: 50 cents per problem you solve correctly.
- Tournament: €2 per correct answer if you correctly solve more problems than your group members did in Task 2, €0 otherwise.

You will not be informed of how you did in the tournament until all six tasks have been completed. If there are ties, the winner will be randomly determined.

#### **Task 4: Robots**

Imagine that you live in a world full not only of TU students, but also full of robots. This is Bob the Robot. Bob is going to solve exercises too, along with all his clones – 100 robots in all. On average the robots are about as good at the exercises as TU students, but some are much better than others. In fact, they have been programmed so that

- Bob 1 has a 1% chance of scoring better than your three group members in Task 2.
- Bob 2 has a 2% chance of scoring better than your three group members in Task 2.
- ...etc...
- Bob 100 has a 100% chance of scoring better than your three group members in Task 2.

One of these robots will be assigned to be your robot. But we aren't going to tell you which robot it is until the end of the experiment – it could be any of the 100 models.<sup>12</sup>

For this part of Task 4 (the Robot part) you can earn €2 (in addition to the payment discussed on the previous page). For this, you can use either your performance on this task or the performance of your Bob. Whichever you use, you will earn €2 if that performance is larger than the number of exercises solved by your group members in Task 2.

12. These instructions, and the picture of the robot, were adapted or taken from Mobius et al. (2014)



FIGURE B2: BOB THE ROBOT (MOBIUS ET AL., 2014)

#### **Task 4: Robots (2)**

You will thus have to help us decide whether to use your score or the robot's score to determine your payment. We are going to ask you which robot you think you are most like. That means, which of the 100 Bob clones is as likely as you are to score better than your group members in Task 2. Based on the decision you have made, we will pick either your score or the robot's, depending on who is most likely to have a better score than your group members in Task 2.

For example, suppose that you say you are as good as Bob 60. If your actual robot is Bob 34, we would base your payoff on your score, since you are then more likely to solve more exercises. But if your actual robot is Bob 97 we will use the robot's score, since the robot is then more likely to solve more exercises.

Note also that since Bob X has an X% chance of having a higher score than your group members in Task 2, you are in effect estimating the probability that you will have a higher score than your group members in Task 2. The bottom line is that you are most likely to win €2 if you are as accurate as possible when you estimate your probability of solving more exercises than your group members did in Task 2.

#### **Task 4: Comprehension Question**

Suppose you think that you have a 44% chance of getting a higher score than your group members in Task 2. Given that you estimate your chance of winning at 44%, which Bob should you select to have the highest chance of winning the prize of €2?

#### **Task 4: Robots (3)**

The next computer screen will ask you to choose a robot. After that, you will be asked to choose between the piece rate and the tournament. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Please do not talk to one another. If you have any further questions, please raise your hand.

*[Participants went to a wait screen until all participants had finished reading these instructions. While on the wait screen, participants were able to reread the instructions for this task if they so wished. Once all participants finished the instructions, each participant then decided on his robot.]*

Please state which Bob you think you are most like. Remember, Bob X has an X% chance of having a higher score than your group members in Task 2, so you are in effect estimating the probability that you will have a higher score than your group members in Task 2. You are most likely to win the €2 if you are as accurate as possible.

I am as likely to have a higher score than my group members in Task 2 as Bob...

*[They then decided between piece rate and tournament pay. After everyone had made their decision, they worked on addition problems for five minutes, after which they received feedback on their absolute performance similar to the previous tasks. They did not receive any feedback on their relative performance, nor on whether the computer used a robot or their own performance to determine their earnings for the Robot task.]*

### Task 5: Choice 3 and Robot 2

As in the previous tasks you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. As in the previous task, you can choose which payment scheme you would prefer to apply to your performance. As in the previous exercise, you will be asked which Robot you are most similar to.

If Task 5 is the one randomly selected for payment, then your earnings for this task are determined the same way as in Task 4. You can choose between:

- Piece Rate: 50 cents per problem you solve correctly.
- Tournament: €2 per correct answer if you correctly solve more problems than your group members did in Task 2, €0 otherwise.

In addition, you may earn €2 for the Robot part. For this part, we will ask you which Robot is most similar to you in your opinion. In other words, which of the 100 Bob clones has the same probability as you to solve more exercises than your group members in Task 2.

The only new part about Task 5 is that you will find out whether your score was higher than the score of your team members in Task 2, even if you choose the piece rate.

You will not be informed of how you did in the tournament until all six tasks have been completed. If there are ties the winner will be randomly determined.

Please do not talk to one another. If you have any questions, please raise your hand.

[*Wait screen with Task 5 instructions; subsequent choices, feedback and task identical to Task 4.*]

### Task 6: Table

In Task 6 you will make 20 decisions. For each of these decisions, you will be choosing between a certain payment (option A) and a lottery (option B). Option A is identical for every decision problem: you will receive 0.5X Euro for certain. For option B you will receive either 2X Euro or 0 Euro. The probability with which you will receive 2X Euro will differ for every decision problem.<sup>13</sup>

If Task 6 is the one randomly selected for payment, then your earnings for this task are determined in the following way:

- First, one of your 20 decisions will be chosen at random. For this purpose, one participant in the experiment will be asked to roll a 20-sided die. If, for example, the number is 14, your 14th decision will be chosen for payment.
- In case you chose the certain payment (option A) on the selected decision problem, you will receive 0.5X Euro. In case you chose the lottery (option B), the computer will draw a random number from 0 to 100. If the randomly chosen number is smaller or equal to the probability of receiving 2X Euro in the selected decision problem, you

13. During the experiment, X was replaced by the participant's actual performance in Task 2. For example, a participant who solved 12 exercises, option A was equal to 6 Euro, and option B paid 24 Euro with some probability. This was true both for the instructions and for the Task itself.



will receive the 2X Euro. Otherwise, you will receive 0 Euro. Please have a look at the two examples below.

Example 1:

Decision problem 17 was selected for payment. You chose the lottery (option B).

Option A	Option B
0.5X Euro	20% chance of obtaining 2X Euro 80% chance of obtaining 0 Euro

In this example, a random number between 1 and 20 would yield a payment of 2X Euro, a higher number would give you a payment of 0 Euro.

Example 2:

Decision problem 12 was selected for payment. You chose the lottery (option B).

Option A	Option B
0.5X Euro	40% chance of obtaining 2X Euro 60% chance of obtaining 0 Euro

In this example, a random number between 1 and 40 would yield a payment of 2X Euro, a higher number would give you a payment of 0 Euro.

Please do not talk to one another. If you have any questions, please raise your hand.

*[Participants went to a wait screen until all participants had finished reading these instructions. While on the wait screen, participants were able to reread the instructions for this task if they so wished. Once all participants finished the instructions, each participant then made their 20 decisions. A literal translation of the screenshot is presented in Figure B3.]*

Please choose either option A or option B for each row below.

Decision	Option A	Option B
1	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 100% Chance 20,00 € to obtain; 0% Chance 0 € to obtain
2	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 95% Chance 20,00 € to obtain; 5% Chance 0 € to obtain
3	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 90% Chance 20,00 € to obtain; 10% Chance 0 € to obtain
4	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 85% Chance 20,00 € to obtain; 15% Chance 0 € to obtain
5	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 80% Chance 20,00 € to obtain; 20% Chance 0 € to obtain
6	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 75% Chance 20,00 € to obtain; 25% Chance 0 € to obtain
7	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 70% Chance 20,00 € to obtain; 30% Chance 0 € to obtain
8	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 65% Chance 20,00 € to obtain; 35% Chance 0 € to obtain
9	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 60% Chance 20,00 € to obtain; 40% Chance 0 € to obtain
10	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 55% Chance 20,00 € to obtain; 45% Chance 0 € to obtain
11	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 50% Chance 20,00 € to obtain; 50% Chance 0 € to obtain
12	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 45% Chance 20,00 € to obtain; 55% Chance 0 € to obtain
13	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 40% Chance 20,00 € to obtain; 60% Chance 0 € to obtain
14	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 35% Chance 20,00 € to obtain; 65% Chance 0 € to obtain
15	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 30% Chance 20,00 € to obtain; 70% Chance 0 € to obtain
16	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 25% Chance 20,00 € to obtain; 75% Chance 0 € to obtain
17	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 20% Chance 20,00 € to obtain; 80% Chance 0 € to obtain
18	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 15% Chance 20,00 € to obtain; 85% Chance 0 € to obtain
19	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 10% Chance 20,00 € to obtain; 90% Chance 0 € to obtain
20	<input type="radio"/> 5,00 € to obtain	<input type="radio"/> 5% Chance 20,00 € to obtain; 95% Chance 0 € to obtain
<input type="button" value="Submit"/>		

FIGURE B3: SCREEN SHOT OF TASK 6 (ENGLISH TRANSLATION)

### Overview of Payment

Task 4 was randomly chosen for payment.<sup>14</sup> During this Task, you solved 10 exercises and chose the tournament, and won it. You therefore receive 20€.

In addition, in Task 4 you chose Bob 52. Your randomly chosen Bob was 34. We therefore used your performance to determine your earnings.

You therefore earned 2€ for the robot part of Task 4.

In addition to the show-up fee of 5€ and the 3€ participation fee, you therefore earned a total of 30€.

### Questionnaire

Please answer the following questions:

- What is your gender?
- What is your age?

14. The choices, payment amounts and selected task presented here are an example.

- What is your major?
- Which finger on your right hand is longer, the index-finger or ring-finger?
- Which finger on your left hand is longer, the index-finger or ring-finger?

[Page 2]

Please answer the following questions:

- How do you see yourself: are you in general a person who is fully prepared to take risks or do you try to avoid taking risks?
- How do you see yourself: are you in driving a person who is fully prepared to take risks or do you try to avoid taking risks?
- How do you see yourself: are you in financial matters a person who is fully prepared to take risks or do you try to avoid taking risks?
- How do you see yourself: are you in your free time and in sports a person who is fully prepared to take risks or do you try to avoid taking risks?
- How do you see yourself: are you in your professional career a person who is fully prepared to take risks or do you try to avoid taking risks?
- How do you see yourself: are you in terms of your personal health a person who is fully prepared to take risks or do you try to avoid taking risks?

[Page 3]

For each of the following items, please choose either A or B. One of the items will be randomly selected for payment at the end of the experiment.

Item	Option A	Option B
1	1/10 Chance of obtaining 1.00€ 9/10 Chance of obtaining 0.80€	1/10 Chance of obtaining 1.90€ 9/10 Chance of obtaining 0.10€
2	2/10 Chance of obtaining 1.00€ 8/10 Chance of obtaining 0.80€	2/10 Chance of obtaining 1.90€ 8/10 Chance of obtaining 0.10€
3	3/10 Chance of obtaining 1.00€ 7/10 Chance of obtaining 0.80€	3/10 Chance of obtaining 1.90€ 7/10 Chance of obtaining 0.10€
4	4/10 Chance of obtaining 1.00€ 6/10 Chance of obtaining 0.80€	4/10 Chance of obtaining 1.90€ 6/10 Chance of obtaining 0.10€
5	5/10 Chance of obtaining 1.00€ 5/10 Chance of obtaining 0.80€	5/10 Chance of obtaining 1.90€ 5/10 Chance of obtaining 0.10€
6	6/10 Chance of obtaining 1.00€ 4/10 Chance of obtaining 0.80€	6/10 Chance of obtaining 1.90€ 4/10 Chance of obtaining 0.10€
7	7/10 Chance of obtaining 1.00€ 3/10 Chance of obtaining 0.80€	7/10 Chance of obtaining 1.90€ 3/10 Chance of obtaining 0.10€
8	8/10 Chance of obtaining 1.00€ 2/10 Chance of obtaining 0.80€	8/10 Chance of obtaining 1.90€ 2/10 Chance of obtaining 0.10€
9	9/10 Chance of obtaining 1.00€ 1/10 Chance of obtaining 0.80€	9/10 Chance of obtaining 1.90€ 1/10 Chance of obtaining 0.10€
10	10/10 Chance of obtaining 1.00€ 0/10 Chance of obtaining 0.80€	10/10 Chance of obtaining 1.90€ 1/10 Chance of obtaining 0.10€

[Page 4]

Please choose one of the following lotteries. Your chosen lottery will be paid out to you at the end of the experiment.

Item	Option
1	50% Chance of obtaining 1.40€ 50% Chance of obtaining 1.40€
2	50% Chance of obtaining 1.20€ 50% Chance of obtaining 1.80€
3	50% Chance of obtaining 1.00€ 50% Chance of obtaining 2.20€
4	50% Chance of obtaining 0.80€ 50% Chance of obtaining 2.60€
5	50% Chance of obtaining 0.60€ 50% Chance of obtaining 3.00€
6	50% Chance of obtaining 0.10€ 50% Chance of obtaining 3.50€

### Overview

In the experiment you already earned 22€. In the questionnaire, you earned an additional 1.80€.

Including the show-up fee of 5€ and the participation fee of 3€, you have therefore earned a total of 31.80€

Thanks for taking part in this experiment. Please enter your total payment on the receipt on your desk, and raise your hand when ready. The experimenter will then come to your desk to check the total amount, after which payment will be done in the office next door.

### ***B3. Instructions for Experiment 2***

Stages 1, 2, 3, 4 and 6 were identical in experiments 1 and 2, apart from some minor variations to take into account the now randomized order of tasks. I will present a brief summary of these changes below, followed by the full instructions for the two new stages (5 and 7). These instructions are English translations; the original German version is presented in Appendix C.

#### **Stages 1, 2, 3, 4 and 6**

I made several small changes to Stages 3, 4 and 6 to allow for different orders of experimental stages. For example, instead of saying “as in the previous two tasks” in Stage 3, the instructions would say “as in parts 1 and 2” if Stage 3 was randomly assigned to a later part of the experiment. I also changed the word ‘task’ to the word ‘part’ throughout the experiment. The main substantive change was in Stage 6, where I added a pop-up message in case participants submitted a response that violated expected utility (e.g., due to multiple switching), and encouraged them to raise their hand if they did not understand why they received the alert. The goal of this message was to reduce the number of violations, though participants were free to ignore the message if they so wished.

#### **Part 5: Choice 3**

For this part you will have to choose between two different payment schemes:<sup>15</sup>

- Certain payment of  $0.5x_i$  Euro.
- Lottery:  $p_i^o * 100\%$  Chance of obtaining  $2x_i$  Euro;  $(1 - p_i^o) * 100\%$  Chance of obtaining 0 Euro.

If part 5 is randomly chosen for payment, your payment will be determined as follows.

- If you chose the certain payment, you will receive  $0.5x_i$  Euro.
- If you chose the lottery, the computer will randomly draw a number between 1 and 100. If the randomly drawn number is smaller or equal to your chance of obtaining the  $2x_i$  Euro ( $p_i^o * 100\%$ ), you will receive  $2x_i$  Euro. Otherwise you will receive 0 Euro. You will therefore receive  $2x_i$  Euro with a  $p_i^o * 100\%$  Chance and no payment with a  $(1 - p_i^o) * 100\%$  Chance.

Please do not talk to one another. If you have any questions, please raise your hand.

15. The parameters in this part were participant-specific. In particular,  $x_i$  was equal to a participant  $i$ 's performance in Stage 2, and  $p_i^o$  was participant  $i$ 's empirical probability of beating three random opponents from experiment 2 based on his/her Stage 2 performance. Participants saw the actual money amounts and percentages and not the formulas on their screen.

### Part 7: Decision (1)

Like in Part 1, 2, 3 and 4 you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. As in Part 3 and 4 you will be able to choose between two payment schemes.

Before you choose between the two payment schemes, we would like to ask you how well you think you performed on the math task in Part 2 compared to other participants. These other participants solved the same exercises during another experiment in the same laboratory. (Recall that in Part 2 you and the other participants all received tournament pay.)

We ask you to estimate which rank you would have among 100 randomly chosen participants in Part 2. You have the chance to receive a bonus of 2 Euro. The closer your guess is to your true rank, the greater your chance of receiving the bonus payment. If you would like to know how exactly this probability will be calculated, please raise your hand and an experimenter will come to your desk. The more precisely you estimate your rank compared to the other participants, the higher is your chance to receive the bonus.

Please keep in mind that a lower number represents a good rank. For example, when you indicate your rank as “1”, this means that you believe that you were better than all 100 participants in the other experiment.

What do you think your rank in Part 2 is compared to the 100 other participants? Please choose a number between 1 (you believe you were the best) and 101 (you believe that you were the worst).

### Part 7: Decision (2)

As in Part 3 and 4 you will be able to choose between piece rate and tournament pay. The difference in this part is that you will be informed about the performance of your potential competitors before you make your decision.

When part 7 is randomly selected for payment, your payment will be determined in the same way as in Part 3 and 4. You can choose between:

- Piece Rate: 50 cents per correct answer.
- Tournament: 2 Euro per correct answer, if you solve more exercises than your team members in part 2, 0 Euro otherwise. In case of a tie the winner will be determined randomly.

Please do not talk to one another. If you have any questions, please raise your hand.

### Part 7: Decision (3)

The table below displays the scores of all participants in this experiment during Part 2. Your own performance is printed in bold. Three of the remaining numbers represent the performances of your team mates.

In addition, the bottom panel of the Table shows how likely you are to win a tournament, if your performance is the same as in Part 2 (12).<sup>16</sup> Your win chance therefore represents the probability that all of your team mates have a lower score than you or you are randomly selected as the winner in case of a tie. Please keep in mind that the win chance in the table is based on your performance in part 2. Your actual win chance will, however, be based on your performance in the current part. Your actual win chance will therefore be larger, if your performance in the current part is greater than in part 2 and smaller if your performance is smaller than in part 2.

All performances:	4	5	6	6	7	8	8	9	10	10	11	11	<b>12</b>	13	13	15	17	21
Your win chance:	32.35%																	

Please answer the following questions:

1. Assuming that your score in the current part is the same as in part 2, what is the chance that you will win the tournament? Your answer should be between 0% and 100%. [Correct answer: the win chance in the table.]
2. Recall that your score in this part may differ from your performance in part 2. How high would you estimate the chance to be that you actually win the tournament? Please keep in mind that your performance will still be compared to the performance of other participants in part 2. Your answer should be between 0% and 100%.

16. This score, the scores printed in the table, and the associated win chance are examples included for illustration.



## APPENDIX C: ORIGINAL GERMAN LANGUAGE INSTRUCTIONS

This section presents the original German language instructions for the two experiments. Detailed explanations of the procedures are presented in Section 3.1 and Appendix B.

### *C1. Instructions for Experiment 1*

This section presents the original German language instructions for experiment 1. In this experiment, all tasks were presented in a fixed order. The program used to run this experiment can be found in the supplementary materials.

#### **Willkommen**

Herzlich Willkommen bei unserem Experiment! Während des Experimentes ist es Ihnen nicht erlaubt elektronische Geräte zu benutzen oder mit anderen Teilnehmern zu kommunizieren. Bitte benutzen Sie nur die für das Experiment vorgesehenen Programme und Funktionen. Bitte sprechen Sie nicht mit den anderen Teilnehmern. Sollten Sie eine Frage haben, dann heben Sie bitte Ihre Hand. Wir werden dann zu Ihnen kommen und Ihre Frage im Stillen beantworten. Bitte stellen Sie Ihre Fragen auf keinen Fall laut. Wenn die Frage relevant für alle Teilnehmer ist, werden wir sie laut wiederholen und beantworten. Sollten Sie gegen diese Regeln verstoßen, müssen wir Sie vom Experiment und der Auszahlung ausschließen.

#### **Instruktionen**

Im heutigen Experiment werden Sie sechs verschiedene Aufgaben erledigen. Keine dieser Aufgaben wird mehr als 5 Minuten erfordern. Am Ende des Experiments erhalten Sie €3 für die Absolvierung der Aufgaben, zusätzlich werden wir eine der sechs Aufgaben zufällig auswählen und Sie abhängig von Ihrer Leistung in dieser Aufgabe auszahlen. Sobald Sie die sechs Aufgaben erledigt haben, werden wir die auszahlungsrelevante Aufgabe bestimmen, indem einer von Ihnen einen sechsseitigen Würfel wirft. Die Methode, nach der Ihre Auszahlung ermittelt wird, variiert von Aufgabe zu Aufgabe. Vor jeder Aufgabe werden wir im Detail beschreiben, wie Ihre Auszahlung jeweils bestimmt wird.

Ihr Gesamtverdienst aus dem Experiment ist die Summe Ihrer Auszahlung aus der zufällig ausgewählten Aufgabe, €3 für die Fertigstellung der Aufgaben und €5; für das Erscheinen. Am Ende des Experiments werden Sie im Nebenraum einzeln und nacheinander Ihre Auszahlung erhalten.

#### **Aufgabe 1 – Stücklohn**

In Aufgabe 1 werden Sie aufgefordert, die Summe aus fünf zufällig ausgewählten zweistelligen Zahlen zu berechnen. Sie haben 5 Minuten Zeit, um die korrekte Summe einer Reihe solcher Probleme zu berechnen. Sie dürfen keinen Taschenrechner zur Hilfe nehmen, können aber jederzeit die Zahlen niederschreiben und das zur Verfügung gestellte Schmierpapier verwenden. Sie schicken Ihre Antwort ab, indem Sie den “Abschicken”-Button mit der Maus betätigen. Wenn Sie eine Antwort abschicken, wird der Computer Ihnen sofort mitteilen, ob Ihre Antwort korrekt ist oder nicht. Ihre Antworten sind anonym.

Wenn Aufgabe 1 für die Auszahlung zufällig ausgewählt wird, erhalten Sie für jedes Problem, das innerhalb der 5 Minuten korrekt gelöst wurde, 50 Cent. Ihre Auszahlung verringert sich nicht, wenn Sie eine falsche Antwort geben. Wir beziehen uns auf diese Auszahlung als Stücklohnzahlung.

Bitte sprechen Sie während des Experiments nicht mit den anderen Teilnehmern. Falls Sie eine Frage haben, heben Sie bitte die Hand.

### **Aufgabe 2 – Wettbewerb**

Wie in Aufgabe 1 erhalten Sie 5 Minuten Zeit, um die korrekte Summe einer Reihe von jeweils fünf zweistelligen Zahlen zu berechnen. In dieser Aufgabe hängt Ihre Auszahlung jedoch von der Leistung einer Gruppe anderer Teilnehmer ab. Jede Gruppe besteht aus vier Personen. Wenn Aufgabe 2 für die Auszahlung zufällig ausgewählt wird, hängt Ihre Auszahlung von der Zahl der Probleme ab, die Sie im Vergleich zu den drei anderen Personen in Ihrer Gruppe gelöst haben. Die Person, welche die meisten Probleme korrekt löst, erhält 2 Euro pro korrekter Lösung, während die anderen Teilnehmer keine Bezahlung erhalten. Wir beziehen uns hierauf als die Wettbewerbszahlung. Sie erfahren nicht eher, wie Sie in dem Wettbewerb abgeschnitten haben, bevor alle sechs Aufgaben erledigt sind. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.

### **Aufgabe 3 – Entscheidung**

Wie in den beiden vorherigen Aufgaben erhalten Sie 5 Minuten, um die korrekte Summe aus einer Reihe von fünf zweistelligen Zahlen zu berechnen. Nun jedoch entscheiden Sie, welche der beiden vorherigen Auszahlungsmodi Sie für Ihre Auszahlung aus der dritten Aufgabe bevorzugen.

Wenn Aufgabe 3 für Ihre Auszahlung zufällig ausgewählt wird, werden Ihre Einnahmen wie folgt bestimmt. Wenn Sie die Stücklohnzahlung wählen, erhalten Sie 50 Cent pro korrekt gelöster Aufgabe. Wenn Sie die Wettbewerbszahlung wählen, wird Ihre Leistung relativ zur Leistung der anderen drei Teilnehmer aus Ihrer Gruppe in Aufgabe 2-Wettbewerb bewertet. Aufgabe 2-Wettbewerb ist diejenige, welche Sie gerade absolviert haben. Falls Sie mehr Aufgaben richtig lösen als dies die anderen Teilnehmer in Aufgabe 2 getan hatten, erhalten Sie das Vierfache der Stücklohnzahlung, also 2 Euro pro korrekter Lösung. Sie erhalten keine Auszahlung für diese Aufgabe, wenn Sie den Wettbewerb wählen und nicht mehr Probleme korrekt lösen als dies die anderen in Ihrer Gruppe in der vorangegangenen Aufgabe 2-Wettbewerb getan hatten. Sie erhalten keine Informationen darüber, wie Sie im Wettbewerb abgeschnitten haben, bevor alle sechs Aufgaben bearbeitet sind. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Der nächste Bildschirm wird Sie fragen, ob Sie die Stücklohnzahlung oder die Wettbewerbszahlung anwenden möchten. Sie erhalten dann 5 Minuten Zeit, um die korrekte Summe der fünf zufälligen zweistelligen Zahlen zu berechnen.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.

*[Once all participants finished the instructions, each participant then decided on their remuneration scheme:]*

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

#### **Aufgabe 4 – Entscheidung 2**

Wie in den vorherigen drei Aufgaben erhalten Sie 5 Minuten Zeit, um die korrekte Summe einer Reihe von fünf zweistelligen Zahlen zu bilden. Wie in der vorherigen Aufgabe können Sie zwischen den Auszahlungsmodi wählen. Es wird einen zusätzlichen Teil in dieser Aufgabe geben, der Ihnen auf der nächsten Seite erklärt wird.

Wenn Aufgabe 4 für die Auszahlung zufällig ausgewählt wird, wird Ihre Auszahlung auf die gleiche Weise bestimmt wie in Aufgabe 3. Sie können wählen zwischen:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

Sie erhalten keine Informationen darüber, wie Sie im Wettbewerb abgeschnitten haben, bevor alle sechs Aufgaben bearbeitet sind. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

*[Once all participants finished the instructions, each participant then decided on their remuneration scheme:]*

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

#### **Aufgabe 4 – Roboter**

Stellen Sie sich vor, Sie leben in einer Welt nicht nur voller TU-Studenten, sondern auch voller Roboter. Dies ist Bob, der Roboter. Bob wird ebenfalls Probleme lösen, zusammen mit all seinen Klonen – insgesamt 100 Robotern. Im Durchschnitt sind die Roboter bei der Lösung der Aufgaben ebenso gut wie die TU Studenten, aber einige sind auch viel besser als andere. Sie wurden so programmiert, dass

1. Bob 1 eine 1%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.
2. Bob 2 eine 2%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.
3. ...usw...
4. Bob 100 eine 100%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.



FIGURE C1: BOB THE ROBOT (MOBIUS ET AL., 2014)

Einer dieser Roboter wird Ihnen zugeteilt als „Ihr Roboter“. Wir werden Ihnen jedoch nicht vor dem Ende des Experiments mitteilen, welcher Ihr Roboter ist – es könnte eines der 100 Modelle sein.

In diesem Teil von Aufgabe 4 (dem Roboterteil) können Sie 2 Euro verdienen (zusätzlich zur Auszahlung, die auf der vorherigen Seite beschrieben wurde). Dafür können Sie entweder Ihre Leistung einsetzen oder die Ihres Bobs. Welche Sie auch nehmen, Sie erhalten 2 Euro, falls Ihre Leistung besser ist, d.h. falls Sie oder Ihr Bob mehr Aufgaben lösen als dies Ihre Gruppenmitglieder in Aufgabe 2 getan hatten.

#### **Aufgabe 4 – Roboter (2)**

Sie müssen uns also dabei helfen zu entscheiden, ob Ihre Punktzahl oder die Punktzahl des Roboters für Ihre Auszahlung herangezogen wird. Wir werden Sie fragen, welchem Roboter Sie glauben am ähnlichsten zu sein. Das heißt, welcher der 100 Bob-Klone hat die gleiche Wahrscheinlichkeit wie Sie, besser abzuschneiden als Ihre Gruppenmitglieder in Aufgabe 2.

Basierend auf Ihrer Entscheidung werden wir entweder Ihre Punktzahl oder die Ihres Roboters wählen, je nachdem wer die größere Wahrscheinlichkeit hat, eine höhere Punktzahl als Ihre Gruppenmitglieder in Aufgabe 2 zu haben.

Beispiel: Nehmen Sie an, Sie sind so gut wie Bob 60. Wenn Ihr tatsächlicher Roboter Bob 34 ist, legen wir für Ihre Auszahlung Ihre Punktzahl zugrunde, da Sie eine höhere Wahrscheinlichkeit haben, mehr Aufgaben zu lösen. Wenn aber Ihr tatsächlicher Roboter Bob 97 ist, benutzen wir die Punktzahl Ihres Roboters, da dieser mit höherer Wahrscheinlichkeit mehr Aufgaben löst.

Beachten Sie außerdem, dass - da Bob X eine  $x\%$ ige Wahrscheinlichkeit auf eine höhere Punktzahl als Ihre Gruppenmitglieder in Aufgabe 2 hat - Sie im Grunde die Wahrscheinlichkeit schätzen müssen, mit der Sie eine höhere Punktzahl als Ihre Gruppenmitglieder in Aufgabe 2 erreichen. Im Endeffekt haben Sie die höchste Chance 2 Euro zu gewinnen, wenn Sie so genau wie möglich Ihre Wahrscheinlichkeit schätzen, mehr richtige Lösungen abzugeben als dies Ihre Gruppenmitglieder in Aufgabe 2 getan haben.

#### **Aufgabe 4 – Frage**

Angenommen Sie denken, dass Sie eine 44%ige Chance auf eine höhere Punktezahl als Ihre Gruppenmitglieder in Aufgabe 2 haben. Unter der Bedingung, dass Sie Ihre Chance zu gewinnen auf 44% schätzen, welchen Bob sollten Sie auswählen, um die höchste Chance auf den Gewinn von 2 Euro zu haben?

#### **Aufgabe 4 – Entscheidung 2 und Roboter**

Der nächste Bildschirm fordert Sie auf, einen Roboter auszuwählen. Danach wählen Sie zwischen der Stücklohnzahlung und dem Wettbewerb. Anschließend erhalten Sie 5 Minuten Zeit, um die korrekte Summe aus einer Reihe von fünf zufällig gewählten zweistelligen Zahlen zu berechnen.

Bitte sprechen Sie nicht mit den anderen Teilnehmern. Falls Sie weitere Frage haben, heben Sie bitte die Hand.

[Once all participants finished the instructions, each participant then decided on their robot:]

Bitte geben Sie an, welchem Bob Sie glauben am ähnlichsten zu sein. Bedenken Sie, Bob X hat eine x%ige Chance eine höhere Punktzahl zu haben als Ihre Gruppenmitglieder in Aufgabe 2, so dass Sie im Endeffekt die Wahrscheinlichkeit schätzen, dass Sie eine höhere Punktzahl haben als Ihre Gruppenmitglieder in Aufgabe 2. Sie haben die höchste Wahrscheinlichkeit die 2 Euro zu gewinnen, wenn Sie so genau wie möglich schätzen.

Ich habe die gleiche Wahrscheinlichkeit, eine höhere Punktzahl als meine Gruppenmitglieder in Aufgabe 2 zu haben wie Bob...

[Once all participants finished choosing their robot, each participant then decided on their remuneration scheme:]

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

### **Aufgabe 5 – Entscheidung 3 und Roboter 2**

Wie in den vorherigen Aufgaben erhalten Sie 5 Minuten Zeit, um die korrekte Summe einer Reihe von fünf zweistelligen Zahlen zu bilden. Wie in den vorherigen zwei Aufgaben können Sie zwischen den Auszahlungsmodi wählen. Wie in der vorherige Aufgabe, werden Sie gefragt, welchem Roboter Sie am ähnlichsten sind.

Wenn Aufgabe 5 für die Auszahlung zufällig ausgewählt wird, wird Ihre Auszahlung auf die gleiche Weise bestimmt wie in Aufgabe 4. Sie können wählen zwischen:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

Darüber hinaus können Sie für den Roboterteil 2 Euro verdienen. Für diesen Teil werden wir Sie fragen, welcher Roboter Ihrer Meinung nach Ihnen am ähnlichsten ist. Das heißt, welcher der 100 Bob Klone hat die gleiche Wahrscheinlichkeit wie Sie, mehr Aufgaben als Ihre Gruppenmitglieder in Aufgabe 2 zu lösen.

Der einzig neue Teil in Aufgabe 5 ist, dass Sie erfahren werden, ob Ihre Punktzahl höher war als die Ihrer Gruppenmitglieder in Aufgabe 2, auch wenn Sie die Stücklohnzahlung gewählt haben.

Sie erhalten keine Informationen darüber, wie Sie im Wettbewerb abgeschnitten haben, bevor alle sechs Aufgaben bearbeitet sind. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Bitte sprechen Sie nicht mit den anderen Teilnehmern. Falls Sie weitere Frage haben, heben Sie bitte die Hand.

## Aufgabe 6 - Tabelle

Bei Aufgabe 6 werden Sie 20 Entscheidungen treffen. Bei jedem Entscheidungsproblem haben Sie die Wahl zwischen einer sicheren Auszahlung (Option A) und einer Lotterie (Option B). Option A ist bei jedem Entscheidungsproblem dieselbe: Sie erhalten 0.5X Euro mit Sicherheit. Bei Option B erhalten Sie entweder 2X Euro oder 0 Euro. Die Chance mit der Sie die 2X Euro (beziehungsweise 0 Euro) erhalten, ist bei jedem Entscheidungsproblem verschieden.<sup>17</sup>

Falls Aufgabe 6 zufällig für die Auszahlung ausgewählt wird, wird Ihr Verdienst folgendermaßen bestimmt:

1. Zunächst wird eine Ihrer 20 Entscheidungen zufällig ausgewählt. Hierzu wird ein Teilnehmer im Labor gebeten, einen 20-seitigen Würfel zu werfen. Zeigt der Würfel beispielsweise 14, so wird ist Ihre 14. Entscheidung auszahlungsrelevant.
2. Falls Sie sich bei Ihrem auszahlungsrelevanten Entscheidungsproblem für die sichere Auszahlung (Option A) entschieden haben, erhalten Sie 0.5X Euro. Falls Sie sich bei Ihrem auszahlungsrelevanten Entscheidungsproblem für die Lotterie (Option B) entschieden haben, wird der Computer zufällig eine Nummer zwischen 0 und 100 bestimmen. Falls die zufällig bestimmte Nummer kleiner oder gleich der Chance ist, 2 Euro bei dem auszahlungsrelevanten Entscheidungsproblem zu erhalten, dann erhalten Sie die 2 Euro. Andernfalls erhalten Sie 0 Euro. Betrachten Sie hierzu die folgenden Beispiele:

### Beispiel 1:

Entscheidungsproblem 17 ist auszahlungsrelevant. Sie haben sich für die Lotterie (Option B) entschieden.

Option A	Option B
0.5X Euro	20% Chance 2X Euro zu erhalten 80% Chance 0 Euro zu erhalten

In diesem Beispiel würde eine Nummer zwischen 1 und 20 zu einer Auszahlung von  $0.5 \cdot 2X$  Euro führen, eine höhere Zahl würde zu einer Auszahlung von 0 Euro führen.

### Beispiel 2:

Entscheidungsproblem 13 ist auszahlungsrelevant. Sie haben sich für die Lotterie (Option B) entschieden.

Option A	Option B
0.5X Euro	40% Chance 2X Euro zu erhalten 60% Chance 0 Euro zu erhalten

In diesem Beispiel würde eine Nummer zwischen 1 und 40 zu einer Auszahlung von 2X Euro führen, eine höhere Zahl würde zu einer Auszahlung von 0 Euro führen.

Bitte sprechen Sie nicht miteinander. Bei weiteren Fragen, heben Sie bitte Ihre Hand.

17. During the experiment, X was replaced by the participant's actual performance in Task 2. For example, a participant who solved 12 exercises, option A was equal to 6 Euro, and option B paid 24 Euro with some probability. This was true both for the instructions and for the Task itself.

## Überblick über Ihre Auszahlung:

Aufgabe 4 (Entscheidung 2) wurde für die Auszahlung ausgewählt. In dieser Aufgabe haben Sie 10 Aufgaben gelöst und die Stücklohnzahlung gewählt. Ihr Verdienst ist daher 5 Euro.<sup>18</sup>

Darüber hinaus haben Sie in Aufgabe 4 Bob 26. Ihr zufällig gewählter Bob war Bob 34. Daher haben wir die Leistung von Bob 34 genutzt, um Ihr Einkommen zu bestimmen.

Im Roboterteil aus Aufgabe 4 haben Sie 2 Euro verdient.

Mit den 5 Euro für das Erscheinen und den 3 Euro für die Teilnahme haben Sie insgesamt 15 Euro verdient.

## Fragebogen

Bitte beantworten Sie die folgenden Fragen:

- Was ist ihr Geschlecht?
- Was ist ihr Alter?
- Was ist ihr Studiengang?
- Welcher Finger Ihrer rechten Hand ist länger, der Zeigefinger oder der Ringfinger?
- Welcher Finger Ihrer linken Hand ist länger, der Zeigefinger oder der Ringfinger?

[Seite 2]

Bitte beantworten Sie die folgenden Fragen:

- Wie schätzen Sie sich persönlich ein: Sind Sie im allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?
- Wie schätzen Sie sich persönlich ein: Sind Sie bei ... ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?
  1. beim Autofahren
  2. bei Geldanlagen
  3. bei Freizeit und Sport
  4. bei Ihrer beruflichen Karriere
  5. bei Ihrer Gesundheit

[Seite 3]

Für jede der unteren Zeilen wählen Sie bitte zwischen Option A und B. Eine der (vom Computer zufällig bestimmten) Zeilen wird Ihnen am Ende des Experiments ausbezahlt.

18. The choices, payment amounts and selected task presented here are an example.



Zeile	Option A	Option B
1	1/10 Chance 1.00€ zu erhalten 9/10 Chance 0.80€ zu erhalten	1/10 Chance 1.90€ zu erhalten 9/10 Chance 0.10€ zu erhalten
2	2/10 Chance 1.00€ zu erhalten 8/10 Chance 0.80€ zu erhalten	2/10 Chance 1.90€ zu erhalten 8/10 Chance 0.10€ zu erhalten
3	3/10 Chance 1.00€ zu erhalten 7/10 Chance 0.80€ zu erhalten	3/10 Chance 1.90€ zu erhalten 7/10 Chance 0.10€ zu erhalten
4	4/10 Chance 1.00€ zu erhalten 6/10 Chance 0.80€ zu erhalten	4/10 Chance 1.90€ zu erhalten 6/10 Chance 0.10€ zu erhalten
5	5/10 Chance 1.00€ zu erhalten 5/10 Chance 0.80€ zu erhalten	5/10 Chance 1.90€ zu erhalten 5/10 Chance 0.10€ zu erhalten
6	6/10 Chance 1.00€ zu erhalten 4/10 Chance 0.80€ zu erhalten	6/10 Chance 1.90€ zu erhalten 4/10 Chance 0.10€ zu erhalten
7	7/10 Chance 1.00€ zu erhalten 3/10 Chance 0.80€ zu erhalten	7/10 Chance 1.90€ zu erhalten 3/10 Chance 0.10€ zu erhalten
8	8/10 Chance 1.00€ zu erhalten 2/10 Chance 0.80€ zu erhalten	8/10 Chance 1.90€ zu erhalten 2/10 Chance 0.10€ zu erhalten
9	9/10 Chance 1.00€ zu erhalten 1/10 Chance 0.80€ zu erhalten	9/10 Chance 1.90€ zu erhalten 1/10 Chance 0.10€ zu erhalten
10	10/10 Chance 1.00€ zu erhalten 0/10 Chance 0.80€ zu erhalten	10/10 Chance 1.90€ zu erhalten 1/10 Chance 0.10€ zu erhalten

[Seite 4]

Bitte wählen Sie eine der folgenden Lotterien aus. Die ausgewählte Lotterie wird Ihnen am Ende des Experiments ausbezahlt.

Zeile	Option
1	50% Chance 1.40€ zu erhalten 50% Chance 1.40€ zu erhalten
2	50% Chance 1.20€ zu erhalten 50% Chance 1.80€ zu erhalten
3	50% Chance 1.00€ zu erhalten 50% Chance 2.20€ zu erhalten
4	50% Chance 0.80€ zu erhalten 50% Chance 2.60€ zu erhalten
5	50% Chance 0.60€ zu erhalten 50% Chance 3.00€ zu erhalten
6	50% Chance 0.10€ zu erhalten 50% Chance 3.50€ zu erhalten

## Übersicht

Im Experiment haben Sie schon 7 Euro verdient. Im Fragebogen haben Sie dazu noch 2.30 Euro.

Mit den 5 Euro für das Erscheinen und den 3 Euro für die Teilnahme haben Sie insgesamt 17.30 Euro verdient

Vielen Dank für die Teilnahme an diesem Experiment. Bitte tragen Sie den Gesamtbetrag Ihrer Auszahlung 17.30 in die Quittung ein. Bitte heben Sie Ihre Hand, wenn Sie fertig sind. Der Experimentleiter wird dann zu Ihnen kommen und den Quittungsbetrag überprüfen. Die Auszahlung erfolgt danach im Raum rechts neben dem Experimentallabor.

## ***C2. Instructions for Experiment 2***

This section presents the original German language instructions for experiment 1. In this experiment, Stages 3 to 6 were presented in a randomized order; I use the same order as in Figure B1 for ease of exposition. The program used to run this experiment can be found in the supplementary materials.

### **Willkommen**

Herzlich Willkommen bei unserem Experiment! Während des Experimentes ist es Ihnen nicht erlaubt elektronische Geräte zu benutzen oder mit anderen Teilnehmern zu kommunizieren. Bitte benutzen Sie nur die für das Experiment vorgesehenen Programme und Funktionen. Bitte sprechen Sie nicht mit den anderen Teilnehmern. Sollten Sie eine Frage haben, dann heben Sie bitte Ihre Hand. Wir werden dann zu Ihnen kommen und Ihre Frage im Stillen beantworten. Bitte stellen Sie Ihre Fragen auf keinen Fall laut. Wenn die Frage relevant für alle Teilnehmer ist, werden wir sie laut wiederholen und beantworten. Sollten Sie gegen diese Regeln verstoßen, müssen wir Sie vom Experiment und der Auszahlung ausschließen.

### **Instruktionen**

Das heutige Experiment besteht aus sieben Teilen. Am Ende des Experiments erhalten Sie 3 Euro für das Bearbeiten der Teile. Zusätzlich werden wir einen der sieben Teile zufällig auswählen und Sie abhängig von Ihrer Leistung in diesem Teil auszahlen. Sobald Sie die sieben Teile erledigt haben, werden wir den auszahlungsrelevanten Teil bestimmen, indem einer von Ihnen einen Würfel wirft. Die Methode, nach der Ihre Auszahlung ermittelt wird, variiert von Teil zu Teil. Vor jedem Teil werden wir im Detail beschreiben, wie Ihre Auszahlung bestimmt wird.

Ihr Gesamtverdienst aus dem Experiment ist die Summe Ihrer Auszahlung aus dem zufällig ausgewählten Teil, 3 Euro für das Bearbeiten der Teile und 5 Euro für das Erscheinen. Am Ende des Experiments werden Sie im Nebenraum einzeln und nacheinander Ihre Auszahlung erhalten.

### **Teil 1 – Stücklohn**

In Teil 1 werden Sie aufgefordert, die Summe aus fünf zufällig ausgewählten zweistelligen Zahlen zu berechnen. Sie haben 5 Minuten Zeit, um die korrekte Summe einer Reihe solcher Aufgaben zu berechnen. Sie dürfen keinen Taschenrechner zur Hilfe nehmen, können aber jederzeit die Zahlen niederschreiben und das zur Verfügung gestellte Schmierpapier verwenden. Sie schicken Ihre Antwort ab, indem Sie den “Abschicken”-Button mit der Maus betätigen oder die Enter-Taste drücken. Wenn Sie eine Antwort abschicken, wird der Computer Ihnen sofort mitteilen, ob Ihre Antwort korrekt ist oder nicht. Ihre Antworten sind anonym.

Wenn Teil 1 für die Auszahlung zufällig ausgewählt wird, erhalten Sie für jede Aufgabe, die innerhalb der 5 Minuten korrekt gelöst wurde, 50 Cent. Ihre Auszahlung verringert

sich nicht, wenn Sie eine falsche Antwort geben. Wir bezeichnen diese Auszahlung als “Stücklohnzahlung”.

Bitte sprechen Sie während des Experiments nicht mit den anderen Teilnehmern. Falls Sie eine Frage haben, heben Sie bitte die Hand.

### **Teil 2 – Wettbewerb**

Wie in Teil 1 haben Sie 5 Minuten Zeit, um die korrekte Summe einer Reihe von jeweils fünf zweistelligen Zahlen zu berechnen. In diesem Teil hängt Ihre Auszahlung jedoch von der Leistung einer Gruppe anderer Teilnehmer ab. Jede Gruppe besteht aus vier Personen. Wenn Teil 2 für die Auszahlung zufällig ausgewählt wird, hängt Ihre Auszahlung von der Anzahl der Aufgaben ab, die Sie im Vergleich mit den drei anderen Personen in Ihrer Gruppe korrekt gelöst haben. Die Person, welche die meisten Aufgaben korrekt löst, erhält 2 Euro pro korrekter Lösung, während die anderen Teilnehmer keine Bezahlung erhalten. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt. Wir bezeichnen diese Auszahlung als “Wettbewerbszahlung”. Wir werden Sie erst darüber informieren, wie Sie im Wettbewerb abgeschnitten haben, wenn alle sieben Teile beendet sind.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.

### **Teil 3 – Entscheidung**

Wie in Teil 1 und 2 erhalten Sie 5 Minuten, um die korrekte Summe aus einer Reihe von fünf zweistelligen Zahlen zu berechnen. Nun entscheiden Sie jedoch selbst welche der beiden Auszahlungsarten (Stücklohn oder Wettbewerb) Sie für Ihre Auszahlung in diesem Teil bevorzugen.

Wenn Teil 3 für Ihre Auszahlung zufällig ausgewählt wird, werden Ihre Einnahmen wie folgt bestimmt:

1. Wenn Sie die Stücklohnzahlung wählen, erhalten Sie 50 Cent pro korrekt gelöster Aufgabe.
2. Wenn Sie die Wettbewerbszahlung wählen, wird Ihre Leistung relativ zur Leistung der anderen drei Teilnehmer aus Ihrer Gruppe in **Teil 2 (Wettbewerb)** bewertet. Falls Sie mehr Aufgaben korrekt lösen als dies die anderen Teilnehmer in Teil 2 getan haben, erhalten Sie das Vierfache der Stücklohnzahlung, also 2 Euro pro korrekter Lösung. Sie erhalten keine Auszahlung für diesen Teil, wenn Sie den Wettbewerb wählen und mindestens eine andere Person aus ihrer Gruppe in Teil 2 mehr Aufgaben gelöst hat als Sie in diesen Teil (Teil 3). Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Wir werden Sie erst darüber informieren, wie Sie im Wettbewerb abgeschnitten haben, wenn alle sieben Teile beendet sind.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.  
[*Once all participants finished the instructions, each participant then decided on their remuneration scheme.*]

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.

2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

#### **Teil 4 – Entscheidung 2**

Wie in Teil 1, 2 und 3 erhalten Sie 5 Minuten, um die korrekte Summe aus einer Reihe von fünf zweistelligen Zahlen zu berechnen. Wie bei Teil 3 entscheiden Sie selbst welche der beiden Auszahlungsarten (Stücklohn oder Wettbewerb) Sie für Ihre Auszahlung in diesem Teil bevorzugen.

Wenn Teil 4 für Ihre Auszahlung zufällig ausgewählt wird, werden Ihre Einnahmen wie folgt bestimmt:

1. Wenn Sie die Stücklohnzahlung wählen, erhalten Sie 50 Cent pro korrekt gelöster Aufgabe.
2. Wenn Sie die Wettbewerbszahlung wählen, wird Ihre Leistung relativ zur Leistung der anderen drei Teilnehmer aus Ihrer Gruppe in **Teil 2 (Wettbewerb)** bewertet. Falls Sie mehr Aufgaben korrekt lösen als dies die anderen Teilnehmer in Teil 2 getan haben, erhalten Sie das Vierfache der Stücklohnzahlung, also 2 Euro pro korrekter Lösung. Sie erhalten keine Auszahlung für diesen Teil, wenn Sie den Wettbewerb wählen und mindestens eine andere Person aus ihrer Gruppe in Teil 2 mehr Aufgaben gelöst hat als Sie im diesen Teil (Teil 4). Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Wir werden Sie erst darüber informieren, wie Sie im Wettbewerb abgeschnitten haben, wenn alle sieben Teile beendet sind.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.

#### **Teil 4 – Roboter**

Stellen Sie sich vor, Sie leben in einer Welt nicht nur voller TU-Studenten, sondern auch voller Roboter. Dies ist Bob, der Roboter. Bob wird ebenfalls Aufgaben lösen, zusammen mit all seinen Klonen - insgesamt 100 Robotern. Im Durchschnitt sind die Roboter bei der Lösung der Aufgaben ebenso gut wie die TU Studenten, aber einige sind auch viel besser als andere. Sie wurden so programmiert, dass

1. Bob 1 eine 1%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.
2. Bob 2 eine 2%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.
3. ...usw...
4. Bob 100 eine 100%ige Chance hat, besser abzuschneiden als Ihre drei Gruppenmitglieder in Aufgabe 2.

Einer dieser Roboter wird Ihnen zugeteilt als Ihr Roboter. Wir werden Ihnen jedoch nicht vor dem Ende des Experiments mitteilen, welcher Ihr Roboter ist- es könnte eines der 100 Modelle sein.

Im Roboterteil von Teil 4 können Sie 2 Euro verdienen (zusätzlich zur Auszahlung, die auf der vorherigen Seite beschrieben wurde). Dafür können Sie entweder Ihre Leistung einsetzen oder die Ihres Bobs. Welche Sie auch nehmen, Sie erhalten 2 Euro, falls Ihre Leistung besser ist, d.h. falls Sie oder Ihr Bob mehr Aufgaben lösen als dies Ihre Gruppenmitglieder in Teil 2 getan hatten.

#### **Teil 4 – Roboter (2)**

Sie müssen uns also dabei helfen zu entscheiden, ob Ihre Punktzahl oder die Punktzahl des Roboters für Ihre Auszahlung herangezogen wird. Wir werden Sie fragen, welchem Roboter Sie glauben am ähnlichsten zu sein. Das heißt, welcher der 100 Bob-Klone hat die gleiche Wahrscheinlichkeit wie Sie, besser abzuschneiden als Ihre Gruppenmitglieder in Teil 2.

Basierend auf Ihrer Entscheidung werden wir entweder Ihre Punktzahl oder die Ihres Roboters wählen, je nachdem wer die größere Wahrscheinlichkeit hat, eine höhere Punktzahl als Ihre Gruppenmitglieder in Teil 2 zu haben.

Beispiel: Nehmen Sie an, Sie sagen Sie sind so gut wie Bob 60. Wenn Ihr tatsächlicher Roboter Bob 34 ist, legen wir für Ihre Auszahlung Ihre Punktzahl zugrunde, da Sie eine höhere Wahrscheinlichkeit haben, mehr Aufgaben zu lösen. Wenn aber Ihr tatsächlicher Roboter Bob 97 ist, benutzen wir die Punktzahl Ihres Roboters, da dieser mit höherer Wahrscheinlichkeit mehr Aufgaben löst.

Beachten Sie außerdem, dass - da Bob X eine x%ige Wahrscheinlichkeit auf eine höhere Punktzahl als Ihre Gruppenmitglieder in Teil 2 hat - Sie im Grunde die Wahrscheinlichkeit schätzen müssen, mit der Sie eine höhere Punktzahl als Ihre Gruppenmitglieder in Teil 2 erreichen. Im Endeffekt haben Sie die höchste Chance 2 Euro zu gewinnen, wenn Sie so genau wie möglich Ihre Wahrscheinlichkeit schätzen, mehr korrekte Lösungen abzugeben als dies Ihre Gruppenmitglieder in Teil 2 getan haben.

#### **Teil 4 – Frage**

Angenommen Sie denken, dass Sie eine 44%ige Chance auf eine höhere Punktezahl als Ihre Gruppenmitglieder in Aufgabe 2 haben. Unter der Bedingung, dass Sie Ihre Chance zu gewinnen auf 44% schätzen, welchen Bob sollten Sie auswählen, um die höchste Chance auf den Gewinn von 2 Euro zu haben?

#### **Teil 4 – Entscheidung 2 und Roboter**

Der nächste Bildschirm fordert Sie auf, einen Roboter auszuwählen. Danach wählen Sie zwischen der Stücklohnzahlung und dem Wettbewerb. Anschließend erhalten Sie 5 Minuten Zeit, um die korrekte Summe aus einer Reihe von fünf zufällig gewählten zweistelligen Zahlen zu berechnen.

Bitte sprechen Sie nicht mit den anderen Teilnehmern. Falls Sie weitere Frage haben, heben Sie bitte die Hand.

[Once all participants finished the instructions, each participant then decided on their robot:]

Bitte geben Sie an, welchem Bob Sie glauben am ähnlichsten zu sein. Bedenken Sie, Bob X hat eine  $x\%$ ige Chance eine höhere Punktzahl zu haben als Ihre Gruppenmitglieder in Aufgabe 2, so dass Sie im Endeffekt die Wahrscheinlichkeit schätzen, dass Sie eine höhere Punktzahl haben als Ihre Gruppenmitglieder in Aufgabe 2. Sie haben die höchste Wahrscheinlichkeit die 2 Euro zu gewinnen, wenn Sie so genau wie möglich schätzen.

Ich habe die gleiche Wahrscheinlichkeit, eine höhere Punktzahl als meine Gruppenmitglieder in Aufgabe 2 zu haben wie Bob...

[Once all participants finished choosing their robot, each participant then decided on their remuneration scheme:]

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

## Teil 5 – Entscheidung ohne Matheaufgaben

Für diesen Teil müssen Sie sich zwischen zwei möglichen Auszahlungsarten entscheiden:

1. Sichere Auszahlung von  $0.5X$  Euro.
2. Lotterie:  $f(X)\%$  Chance  $2X$  Euro zu erhalten;  $(1 - f(X))\%$  Chance 0 Euro zu erhalten.<sup>19</sup>

Falls Teil 5 zufällig für die Auszahlung ausgewählt wird, wird Ihr Verdienst folgendermaßen bestimmt.

1. Falls Sie sich für die sichere Auszahlung entschieden haben, erhalten Sie  $0.5X$  Euro.
2. Falls Sie für die Lotterie entschieden haben, wird der Computer zufällig eine Nummer zwischen 1 und 100 bestimmen. Falls die zufällig bestimmte Nummer kleiner oder gleich der Chance ist,  $2X$  Euro zu erhalten ( $f(X)$ ), dann erhalten Sie die  $2X$  Euro. Andernfalls erhalten Sie 0 Euro. Sie erhalten also  $2X$  Euro mit der Wahrscheinlichkeit von  $f(X)\%$  und keine Auszahlung mit der Wahrscheinlichkeit von  $(1 - f(X))\%$ .

Bitte sprechen Sie nicht mit den anderen Teilnehmern. Falls Sie weitere Frage haben, heben Sie bitte die Hand.

## Teil 6 – Tabelle

Bei Teil 6 werden Sie 20 Entscheidungen treffen. Bei jedem Entscheidungsproblem haben Sie die Wahl zwischen einer sicheren Auszahlung (Option A) und einer Lotterie (Option B). Option A ist bei jedem Entscheidungsproblem dieselbe: Sie erhalten  $0.5X$  Euro mit Sicherheit.<sup>20</sup> Bei Option B erhalten Sie entweder  $2X$  Euro oder 0 Euro. Die Chance mit

19. Here,  $X$  is equal to the participant's score in part 2 of the experiment. The win chance  $f(X)$  is the win chance is computed as the empirical probability that the participant's score  $X$  is greater than 3 randomly chosen participants from experiment 1 (or win a random tiebreaker).

20. Here,  $X$  is equal to the participant's score in part 2 of the experiment.

der Sie die 2X Euro (beziehungsweise 0 Euro) erhalten, ist bei jedem Entscheidungsproblem verschieden.

Falls Teil 3 zufällig für die Auszahlung ausgewählt wird, wird Ihr Verdienst folgendermaßen bestimmt:

1. Zunächst wird eine Ihrer 20 Entscheidungen zufällig ausgewählt. Hierzu wird ein Teilnehmer im Labor gebeten, einen 20-seitigen Würfel zu werfen. Zeigt der Würfel beispielsweise 14, so wird Ihre 14. Entscheidung auszahlungsrelevant.
2. Falls Sie sich bei Ihrem auszahlungsrelevanten Entscheidungsproblem für die sichere Auszahlung (Option A) entschieden haben, erhalten Sie 0.5X Euro. Falls Sie sich bei Ihrem auszahlungsrelevanten Entscheidungsproblem für die Lotterie (Option B) entschieden haben, wird der Computer zufällig eine Nummer zwischen 1 und 100 bestimmen. Falls die zufällig bestimmte Nummer kleiner oder gleich der Chance ist, 2X Euro bei dem auszahlungsrelevanten Entscheidungsproblem zu erhalten, dann erhalten Sie die 2X Euro. Andernfalls erhalten Sie 0.5X Euro. Betrachten Sie hierzu die folgenden Beispiele:

Beispiel 1:

Entscheidungsproblem 17 ist auszahlungsrelevant. Sie haben sich für die Lotterie (Option B) entschieden.

Option A	Option B
0.5X Euro	20% Chance 2X Euro zu erhalten 80% Chance 0 Euro zu erhalten

In diesem Beispiel würde eine Nummer zwischen 1 und 20 zu einer Auszahlung von 2X Euro führen, eine höhere Zahl würde zu einer Auszahlung von 0 Euro führen.

Beispiel 2:

Entscheidungsproblem 13 ist auszahlungsrelevant. Sie haben sich für die Lotterie (Option B) entschieden.

Option A	Option B
0.5X Euro	40% Chance 2X Euro zu erhalten 60% Chance 0 Euro zu erhalten

In diesem Beispiel würde eine Nummer zwischen 1 und 40 zu einer Auszahlung von 2X Euro führen, eine höhere Zahl würde zu einer Auszahlung von 0 Euro führen.

Bitte sprechen Sie nicht miteinander. Bei weiteren Fragen, heben Sie bitte Ihre Hand.

## **Teil 7 – Entscheidung**

Wie bei Teil 1, 2, 3 und 4 erhalten Sie 5 Minuten, um die korrekte Summe aus einer Reihe von fünf zweistelligen Zahlen zu berechnen. Wie bei Teil 3 und 4 können Sie zwischen zwei Auszahlungsarten wählen.

Bevor Sie sich zwischen zwei Auszahlungsarten entscheiden, möchten wir wissen, wie gut Sie glauben, sich bei den Matheaufgaben in Teil 2 im Vergleich zu anderen Teilnehmern geschlagen haben. Diese anderen Teilnehmer haben bei einem anderen Experiment in



diesem Labor die gleichen Aufgaben gelöst. (Zur Erinnerung: Im zweiten Teil haben Sie und die anderen Teilnehmer die Wettbewerbszahlung erhalten.)

Wir bitten Sie zu schätzen, welchen Rang Sie bei 100 zufällig ausgewählten Teilnehmern in Teil 2 eingenommen hätten. Sie haben die Chance einen Bonus von 2 Euro zu bekommen. Die Wahrscheinlichkeit den Bonus zu bekommen steigt, je geringer der Abstand zwischen Ihrer Schätzung und Ihrem wahren Rang ist. Wenn Sie wissen möchten, wie genau die Wahrscheinlichkeit berechnet wird, melden Sie sich bitte und ein Experimentator wird zu Ihnen kommen. Je genauer Sie also ihren Rang im Vergleich zu den anderen Teilnehmern schätzen, desto höher ist für Sie die Chance, den Bonus zu erhalten.

Bitte beachten Sie: Eine niedrige Nummer steht für einen guten Rang. Beispiel: Wenn Sie einen Rang von "1" angeben, bedeutet dies, dass Sie glauben, Sie waren besser als alle 100 Teilnehmer des anderen Experiments.

Wie schätzen Sie Ihren Rang in Teil 2 verglichen zu den 100 anderen Teilnehmern ein?. Bitte wählen Sie einen Wert zwischen 1 (Sie glauben, Sie waren der/die Beste) und 101 (Sie glauben, Sie waren der/die Schlechteste):

### **Teil 7 – Entscheidung (2)**

Wie bei Teil 3 und 4 können Sie zwischen Stücklohn und Wettbewerbszahlung als Auszahlungsart wählen. Der Unterschied in diesem Teil ist, dass Sie über die vorherige Leistung ihrer potentiellen Gegenspieler informiert werden, bevor Sie Ihre Entscheidung über die Auszahlungsart treffen.

Wenn Teil 7 für die Auszahlung zufällig ausgewählt wird, wird Ihre Auszahlung auf die gleiche Weise bestimmt wie in Teil 3 und 4. Sie können wählen zwischen:

1. Stücklohnzahlung: 50 Cent pro korrekt gelöster Aufgabe.
2. Wettbewerbszahlung: 2 Euro pro korrekter Antwort, falls Sie mehr Aufgaben korrekt gelöst haben als Ihre Gruppenmitglieder in Teil 2, ansonsten 0 Euro. Im Falle eines Gleichstands wird der Gewinner zufällig bestimmt.

Bitte sprechen Sie nicht miteinander. Wenn Sie Fragen haben, heben Sie bitte die Hand.

### **Teil 7 – Entscheidung (3)**

Die Tabelle unten zeigt die erreichte Leistung aller Personen in diesem Experiment während Teil 2. Ihre eigene Leistung wird in fetten Buchstaben gezeigt. Drei der anderen Nummern beschreiben auch die Leistung ihrer Teammitglieder.

Darüber hinaus zeigt Ihnen der untere Teil der Tabelle mit welcher Wahrscheinlichkeit Sie den Wettbewerb gewinnen, falls Ihre erreichte Punktzahl die gleiche wie in Teil 2 (12) sein wird. Ihre Gewinnchance entspricht also der Wahrscheinlichkeit, dass alle Ihre drei Teammitglieder ein niedrigeres Ergebnis als Ihres erzielen werden, oder, dass es zu einem Gleichstand kommt und Sie zufällig als Gewinner ausgewählt werden. Beachten Sie, dass die angegebene Gewinnchance auf der Basis Ihrer Leistung in Teil 2 berechnet wird. Ihre tatsächliche Gewinnchance richtet sich aber nach Ihrer Leistung in dem aktuellen Teil. Ihre tatsächliche Gewinnchance ist also größer, falls Ihre erreichte Leistung im aktuellen Teil höher ist als in Teil 2 und kleiner, falls ihre erreichte Leistung niedriger ist.

Alle Ergebnisse:	4 5 6 6 7 8 8 9 10 10 11 11 <b>12</b> 13 13 15 17 21
Ihre Gewinnchance:	32.35%

Bitte beantworten Sie die folgenden Fragen:

1. Angenommen ihre erreichte Punktzahl in dem aktuellen Teil ist die gleiche wie in Teil 2, mit welcher Wahrscheinlichkeit werden Sie den Wettbewerb gewinnen? Bitte wählen Sie eine Wahrscheinlichkeit zwischen 0 und 100%. [Correct answer: the win chance in the table.]
2. Beachten Sie, dass ihr tatsächliches Ergebnis in dem aktuellen Teil von der erreichten Punktzahl in Teil 2 (12) abweichen kann. Wie hoch schätzen Sie die tatsächliche Wahrscheinlichkeit ein, dass Sie den Wettbewerb gewinnen? Beachten Sie, dass Sie in dem aktuellen Teil gegen die Leistung ihrer Teammitglieder aus Teil 2 antreten. Bitte wählen Sie eine Wahrscheinlichkeit zwischen 0 und 100%.

[Once all participants finished the instructions, each participant then decided on their remuneration scheme:]

Bitte entscheiden Sie sich zwischen einer der folgenden Auszahlungsmodi:

1. Stücklohnzahlung: 50 Cent pro richtig gelöstem Problem.
2. Wettbewerb: 2 Euro pro korrekter Antwort, falls Sie mehr Probleme korrekt gelöst haben als Ihre Gruppenmitglieder in Aufgabe 2, ansonsten 0 Euro.

Zur Erinnerung: Diese Tabelle zeigt die erreichte Leistung aller Personen in diesem Experiment in Teil 2, mit Ihrer eigenen Leistung in fetten Buchstaben. Die untere Zeile ist Ihre Gewinnwahrscheinlichkeit, falls Ihre Leistung die gleiche sein wird, wie in Teil 2.

Alle Ergebnisse:	4 5 6 6 7 8 8 9 10 10 11 11 <b>12</b> 13 13 15 17 21
Ihre Gewinnchance:	32.35%

### Überblick über Ihre Auszahlung:

Teil 7 wurde für die Auszahlung ausgewählt. In diesem Teil haben Sie 14 Aufgaben gelöst und den Wettbewerb gewählt, welchen Sie gewonnen haben. Ihr Verdienst ist daher 28 Euro.<sup>21</sup>

Leider erhalten Sie keine Bonus für das Schätzen Ihres Rangs.

Mit den 5 Euro für das Erscheinen und den 3 Euro für die Teilnahme haben Sie insgesamt 36 Euro verdient.

### Fragebogen

Bitte beantworten Sie die folgenden Fragen:

- Was ist ihr Geschlecht?

21. The choices, payment amounts and selected task presented here are an example.

- Was ist ihr Alter?
- Was ist ihr Studiengang?
- Welcher Finger Ihrer rechten Hand ist länger, der Zeigefinger oder der Ringfinger?
- Welcher Finger Ihrer linken Hand ist länger, der Zeigefinger oder der Ringfinger?
- Angenommen, Sie würden noch einmal die Additionsaufgaben bearbeiten. Was glauben Sie, wie würde sich Ihre neue erreichte Punktzahl von **Ihrer Punktzahl aus Teil 2 12** unterscheiden?

[Seite 2]

Bitte beantworten Sie die folgenden Fragen:

- Wie schätzen Sie sich persönlich ein: Sind Sie im allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?
- Wie schätzen Sie sich persönlich ein: Sind Sie bei ... ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?
  1. beim Autofahren
  2. bei Geldanlagen
  3. bei Freizeit und Sport
  4. bei Ihrer beruflichen Karriere
  5. bei Ihrer Gesundheit

[Seite 3]

Für jede der unteren Zeilen wählen Sie bitte zwischen Option A und B. Eine der (vom Computer zufällig bestimmten) Zeilen wird Ihnen am Ende des Experiments ausbezahlt.

Zeile	Option A	Option B
1	1/10 Chance 1.00€ zu erhalten 9/10 Chance 0.80€ zu erhalten	1/10 Chance 1.90€ zu erhalten 9/10 Chance 0.10€ zu erhalten
2	2/10 Chance 1.00€ zu erhalten 8/10 Chance 0.80€ zu erhalten	2/10 Chance 1.90€ zu erhalten 8/10 Chance 0.10€ zu erhalten
3	3/10 Chance 1.00€ zu erhalten 7/10 Chance 0.80€ zu erhalten	3/10 Chance 1.90€ zu erhalten 7/10 Chance 0.10€ zu erhalten
4	4/10 Chance 1.00€ zu erhalten 6/10 Chance 0.80€ zu erhalten	4/10 Chance 1.90€ zu erhalten 6/10 Chance 0.10€ zu erhalten
5	5/10 Chance 1.00€ zu erhalten 5/10 Chance 0.80€ zu erhalten	5/10 Chance 1.90€ zu erhalten 5/10 Chance 0.10€ zu erhalten
6	6/10 Chance 1.00€ zu erhalten 4/10 Chance 0.80€ zu erhalten	6/10 Chance 1.90€ zu erhalten 4/10 Chance 0.10€ zu erhalten
7	7/10 Chance 1.00€ zu erhalten 3/10 Chance 0.80€ zu erhalten	7/10 Chance 1.90€ zu erhalten 3/10 Chance 0.10€ zu erhalten
8	8/10 Chance 1.00€ zu erhalten 2/10 Chance 0.80€ zu erhalten	8/10 Chance 1.90€ zu erhalten 2/10 Chance 0.10€ zu erhalten
9	9/10 Chance 1.00€ zu erhalten 1/10 Chance 0.80€ zu erhalten	9/10 Chance 1.90€ zu erhalten 1/10 Chance 0.10€ zu erhalten
10	10/10 Chance 1.00€ zu erhalten 0/10 Chance 0.80€ zu erhalten	10/10 Chance 1.90€ zu erhalten 1/10 Chance 0.10€ zu erhalten

[Seite 4]

Bitte wählen Sie eine der folgenden Lotterien aus. Die ausgewählte Lotterie wird Ihnen am Ende des Experiments ausbezahlt.

Zeile	Option
1	50% Chance 1.40€ zu erhalten 50% Chance 1.40€ zu erhalten
2	50% Chance 1.20€ zu erhalten 50% Chance 1.80€ zu erhalten
3	50% Chance 1.00€ zu erhalten 50% Chance 2.20€ zu erhalten
4	50% Chance 0.80€ zu erhalten 50% Chance 2.60€ zu erhalten
5	50% Chance 0.60€ zu erhalten 50% Chance 3.00€ zu erhalten
6	50% Chance 0.10€ zu erhalten 50% Chance 3.50€ zu erhalten

[Seite 5]

- Wo wurden Sie geboren?
  - Bundesland (wenn in Deutschland):
  - Land (wenn nicht in Deutschland):
- Was zeichnet Ihren Geburtsort am besten aus? Ländlicher Raum oder Städtischer Raum?
- Wo haben Sie gewohnt, als Sie 15 - 18 Jahre alt waren?
  - Bundesland (wenn in Deutschland):
  - Land (wenn nicht in Deutschland):
- Was zeichnete Ihren Wohnort am besten aus? Ländlicher Raum oder Städtischer Raum?
- Wo wurde Ihre Mutter geboren?
  - Bundesland (wenn in Deutschland):
  - Land (wenn nicht in Deutschland):
- Was zeichnet den Geburtsort Ihrer Mutter am besten aus? Ländlicher Raum oder Städtischer Raum?
- Wo wurde Ihr Vater geboren?

- Bundesland (wenn in Deutschland):
- Land (wenn nicht in Deutschland):
- Was zeichnet den Geburtsort Ihres Vaters am besten aus? Ländlicher Raum oder Städtischer Raum?
- Sind Sie während ihrer Kindheit von ehem. Ost- nach Westdeutschland gezogen?
- Welchen Beschäftigungsstatus hatte Ihre Mutter während Ihrer Kindheit?
- Wie lautet der *höchste* Bildungsabschluss Ihrer Mutter?
- Welchen Beschäftigungsstatus hatte Ihr Vater während Ihrer Kindheit?
- Wie lautet der *höchste* Bildungsabschluss Ihres Vaters?
- In welchem Jahr wurde Ihre Mutter geboren?
- In welchem Jahr wurde Ihr Vater geboren?

## Übersicht

Im Experiment haben Sie schon 28 Euro verdient. Im Fragebogen haben Sie dazu noch 2.10 Euro.

Mit den 5 Euro für das Erscheinen und den 3 Euro für die Teilnahme haben Sie insgesamt 38.10 Euro verdient. Für die Auszahlung runden wir auf volle 50 Cent auf, d.h. Sie erhalten 38.50 Euro.

Vielen Dank für die Teilnahme an diesem Experiment. Bitte tragen Sie den Gesamtbetrag Ihrer Auszahlung (38.50 Euro) in die Quittung ein. Bitte heben Sie Ihre Hand, wenn Sie fertig sind. Der Experimentleiter wird dann zu Ihnen kommen und den Quittungsbetrag überprüfen. Die Auszahlung erfolgt danach im Raum rechts neben dem Experimentallabor.

## REFERENCES

- Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen, and Bertil Tungodden (2016). “Willingness to Compete: Family Matters.” *Management Science*, 62(8), 2149–2162.
- Andersson, Ola, Håkan J. Holm, Jean-Robert Tyran, and Erik Wengström (2016). “Risk Aversion Relates to Cognitive Ability: Preferences or Noise?” *Journal of the European Economic Association*, 14(5), 1129–1154.
- Apicella, Coren L., Elif E. Demiral, and Johanna Mollerstrom (2017). “No Gender Difference in Willingness to Compete When Competing against Self.” *American Economic Review*, 107(5), 136–140.
- Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter (2012). “Distributional Preferences and Competitive Behavior.” *Journal of Economic Behavior & Organization*, 83(1), 125–135.
- Balafoutas, Loukas and Matthias Sutter (2019). “How Uncertainty and Ambiguity in Tournaments Affect Gender Differences in Competitive Behavior.” *European Economic Review*, 118, 1–13.
- Balafoutas, Loukas and Mattias Sutter (2012). “Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory.” *Science*, 335(6068), 579–582.
- Bartos, Vojtech and Roel Van Veldhuizen (2022). “Non-Parametric Identification of Preferences through Experiments.” Mimeo (Lund University).
- Buser, Thomas, Anna Dreber, and Johanna Mollerstrom (2015). “Stress Reactions cannot explain the Gender Gap in Willingness to compete.” Tinbergen Institute Discussion Paper TI 2015-059/I.
- Buser, Thomas, Anna Dreber, and Johanna Mollerstrom (2017). “The Impact of Stress on Tournament Entry.” *Experimental Economics*, 20(2), 506–530.
- Buser, Thomas, Lydia Geijtenbeek, and Erik Plug (2018). “Sexual Orientation, Competitiveness and Income.” *Journal of Economic Behavior & Organization*, 151(9382), 191–198.
- Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2014). “Gender, Competitiveness, and Career Choices.” *The Quarterly Journal of Economics*, 129(3), 1409–1447.
- Buser, Thomas, Eva Ranehill, and Roel Van Veldhuizen (2021). “Gender Differences in Willingness to Compete: The Role of Public Observability.” *Journal of Economic Psychology*, 83.
- Cassar, Alessandra, Feven Wordofa, and Y. Jane Zhang (2016). “Competing for the Benefit of Offspring Eliminates the Gender Gap in Competitiveness.” *Proceedings of the National Academy of Sciences*, 113(19), 5201–5205.

- Dargnies, Marie-Pierre (2012). “Men Too Sometimes Shy Away from Competition: The Case of Team Competition.” *Management Science*, 58(11), 1982–2000.
- Dohmen, Thomas and Armin Falk (2011). “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender.” *American Economic Review*, 101(2), 556–590.
- Dreber, Anna, Emma von Essen, and Eva Ranehill (2014). “Gender and Competition in Adolescence: Task Matters.” *Experimental Economics*, 17(1), 154–172.
- Eckel, Catherine C. and Philip J. Grossman (2002). “Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk.” *Evolution and Human Behavior*, 23(4), 281–295.
- Enke, Benjamin and Thomas Graeber (2021). “Cognitive Uncertainty.” National Bureau of Economic Research Working Paper, 26518.
- Ertac, Seda and Balazs Szentes (2011). “The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence.” Working Paper.
- Flory, Jeff A., Andreas Leibbrandt, and John A. List (2015). “Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions.” *The Review of Economic Studies*, 82(1), 122–155.
- Flory, Jeffrey A., Uri Gneezy, Kenneth L. Leonard, and John A. List (2018). “Gender, Age, and Competition: A Disappearing Gap?” *Journal of Economic Behavior & Organization*, 150, 256–276.
- Gilboa, Itzhak and David Schmeidler (1989). “Maxmin Expected Utility with Non-Unique Prior.” *Journal of Mathematical Economics*, 18(2), 141–153.
- Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019). “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study.” *Journal of Political Economy*, 127(4), 1826–1863.
- Gneezy, Uri and Jan Potters (1997). “An Experiment on Risk Taking and Evaluation Periods.” *The Quarterly Journal of Economics*, 112(2), 631–645.
- Greiner, Ben (2015). “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE.” *Journal of the Economic Science Association*, 1(1), 114–125.
- Grosse, Niels D., Gerhard Riener, and Markus Dertwinkel-Kalt (2014). “Explaining Gender Differences in Competitiveness: Testing a Theory on Gender-Task Stereotypes.” Mimeo (University of Mannheim).
- Halko, Marja-Liisa and Lauri Sääksvuori (2017). “Competitive Behavior, Stress, and Gender.” *Journal of Economic Behavior & Organization*, 141, 96–109.
- Harrison, Glenn W., Morten I. Lau, Elisabet E. Rutström, and Melonie B. Sullivan (2005). “Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues.” In *Field Experiments in Economics*, pp. 17–50. Emerald, Greenwich, CT.



- Hausman, Jerry (2001). “Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left.” *Journal of Economic Perspectives*, 15(4), 57–67.
- Healy, Andrew and Jennifer Pate (2011). “Can Teams Help to Close the Gender Competition Gap?” *The Economic Journal*, 121(555), 1192–1204.
- Kamas, Linda and Anne Preston (2012). “The Importance of Being Confident; Gender, Career Choice, and Willingness to Compete.” *Journal of Economic Behavior & Organization*, 83(1), 82–97.
- Kessel, Dany, Johanna Mollerstrom, and Roel Van Veldhuizen (2021). “Can Simple Advice Eliminate the Gender Gap in Willingness to Compete?” *European Economic Review*, 138, 103777.
- Lee, Soohyung, Muriel Niederle, and Namwook Kang (2014). “Do Single-Sex Schools Make Girls More Competitive?” *Economics Letters*, 124(3), 474–477.
- Lozano, Lina and Ernesto Reuben (2022). “Measuring Preferences for Competition?” Mimeo (NYU Abu Dhabi).
- Masclet, David, Emmanuel Peterle, and Sophie Larribeau (2015). “Gender Differences in Tournament and Flat-Wage Schemes: An Experimental Study.” *Journal of Economic Psychology*, 47, 103–115.
- Mobius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat (2014). “Managing Self-Confidence: Theory and Experimental Evidence.” Mimeo (Stanford University).
- Niederle, Muriel (2016). “Gender.” In *Handbook of Experimental Economics*, edited by John Kagel and Alvin E. Roth, 2nd ed.
- Niederle, Muriel (2017). “A Gender Agenda: A Progress Report on Competitiveness.” *American Economic Review*, 107(5), 115–119.
- Niederle, Muriel, Carmit Segal, and Lise Vesterlund (2013). “How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness.” *Management Science*, 59(1), 1–16.
- Niederle, Muriel and Lise Vesterlund (2007). “Do Women Shy Away From Competition? Do Men Compete Too Much?” *Quarterly Journal of Economics*, 122(3), 1067–1101.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2015). “Taste for Competition and the Gender Gap Among Young Business Professionals.” NBER Working Paper 21695.
- Reuben, Ernesto, Matthew Wiswall, and Basit Zafar (2017). “Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender.” *The Economic Journal*, 127(604), 2153–2186.
- Saccardo, Silvia, Aniela Pietrasz, and Uri Gneezy (2018). “On the Size of the Gender Difference in Competitiveness.” *Management Science*, 64(4), 1541–1554.

- Shurchkov, Olga (2012). “Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints.” *Journal of the European Economic Association*, 10(5), 1189–1213.
- Sutter, Matthias and Daniela Glätzle-Rützler (2015). “Gender Differences in the Willingness to Compete Emerge Early in Life and Persist.” *Management Science*, 61(10), 2339–23354.
- Wozniak, David, William T. Harbaugh, and Ulrich Mayr (2014). “The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices.” *Journal of Labor Economics*, 32(1), 161–198.
- Zhang, Y. Jane (2019). “Culture, Institutions and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China.” *The Economic Journal*, 129(617), 509–552.