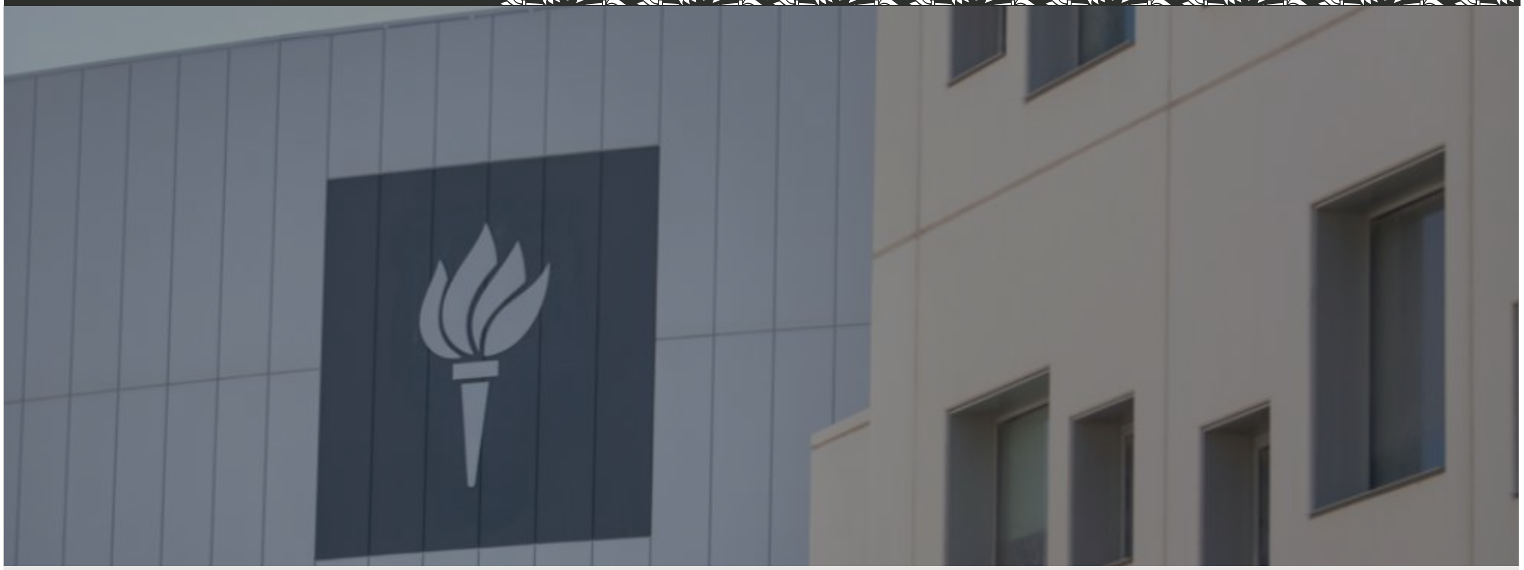
جامعة نيويورك أبوظبي
NYU | ABU DHABI



Does selection bias cause us to overestimate gender differences in competitiveness?

Aurélie Dariel, Nikos Nikiforakis & Jan Stoop

Working Paper # 0046

May 2020

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html

Does selection bias cause us to overestimate gender differences in competitiveness?*

Aurélie Dariel† Nikos Nikiforakis‡ Jan Stoop⁑

Abstract

Experimental evidence suggests there is a substantial difference in the willingness of men and women to compete that could help explain the gender gap in labor market outcomes. The use of volunteer samples, however, raises a question about whether self-selection into experiments biases the estimated difference in competitiveness. To address it, we first measure the willingness of 1,145 individuals to compete in a classroom experiment. We then identify among them the subset of 'lab volunteers' by observing who accepts an invitation to participate in lab experiments. To test for the existence of selection bias, we compare the gender gap among lab volunteers to that in the population from which they were recruited. We find that selection causes us to overestimate the gender gap in competitiveness by 16 percentage points in absolute terms and, in relative terms, by a factor of 2 to 3 depending on the econometric model. We also show that selection causes us to significantly overestimate the gender gap in risk attitudes and the tendency of low-performing men to select into competition. We present evidence men and women select differently into the lab, and discuss the implications of our findings for future research.

Key words: selection bias, laboratory experiment, external validity, competitive preferences,
JEL Classification: C90, D03

* We would like to thank Ernesto Reuben and Lise Vesterlund for useful comments, as well as participants at the 2019 Asia Pacific Meeting of the Economic Science Association, and the 2019 International Symposium for Experimental Economics. Special thanks go out to Bas Karremans, Omar Ricket and Tim Benning for their valuable support in conducting the experiment. We are also thankful for the financial support of NWO, the Netherlands (Veni grant 016.155.026 and Vidi grant VI.Vidi.195.061).

† Division of Social Science, New York University Abu Dhabi, PO Box 129188, United Arab Emirates. E-mail: apd5 @nyu.edu. Phone: +971 26285436

‡ Division of Social Science, New York University Abu Dhabi, PO Box 129188, United Arab Emirates. E-mail: nikos.nikiforakis@nyu.edu. Phone: +971 26285436

⁑ Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, the Netherlands. E-mail: stoop@ese.eur.nl. Phone: +31 104082708.

1. Introduction

One of the most significant insights in the past decade from experimental research in economics may be the existence of a pronounced difference in the willingness of men and women to engage in competition. Following the pioneering study of Niederle and Vesterlund (2007) (henceforth, NV2007), several experiments have documented the tendency of women to shy away from competition, and of men to compete 'too much.'[1] The importance of this finding can hardly be overstated as it suggests a possible explanation for women's lower wages and lower representation in high-level jobs (Blau and Khan 2017; Goldin, 2p014). The economic significance of the gender gap in competitiveness, however, is still a topic of discussion. As Blau and Kahn (2017) write in their authoritative review of the literature on the gender wage gap, "there are reasonable concerns about generalizing the results of such experiments outside the lab" (p. 837).

One of the main concerns when generalizing findings from lab experiments such as those on gender arises from the use of volunteer subjects (Falk and Heckman, 2009; Levitt and List, 2007). If people's attitudes toward competition affect selection into experiments, inferences about the gender gap in competitiveness drawn from volunteer samples could be biased, i.e., not reflect the gender gap in the population from which the volunteers were recruited. Specifically, inferences regarding the gender gap *will be* biased if attitudes toward competition affect differently the likelihood of men and women to participate in lab experiments. Selection bias could be one explanation (although not the only one) why attitudes toward competition, although substantial in the lab, appear to account for a relatively small portion of the gender pay gap in empirical studies (Blau and Kahn, 2017, Buser et al. 2014; Buser et al. 2017; Reuben et al. 2017; Reuben et al. 2019). Could selection bias cause us to *over*estimate the gender differences in the willingness to compete in controlled experiments?

The aim of the present study is to provide empirical evidence about whether self-selection into lab experiments biases the estimated gender differences in competitiveness and, if so, the direction of the bias. Selection bias need not cause us to overestimate the gender gap in competitiveness; it could also cause us to *under*estimate it. This could happen if women in the population are less willing to compete than men, but only men and women with a certain tolerance for competition volunteer for experiments. As the latter seems plausible, we start our investigation by looking for

[1] For reviews of the literature, see Dariel et al. (2017) and Niederle (2016).

evidence of selection bias in past studies in order to form a hypothesis about the potential impact of self-selection. Using the survey of Dariel et al. (2017), we classify previous studies into those using self-selected samples and those that do not.[2] Figure 1 presents the mean gender gap in competitiveness in studies relying on self-selected samples and otherwise.

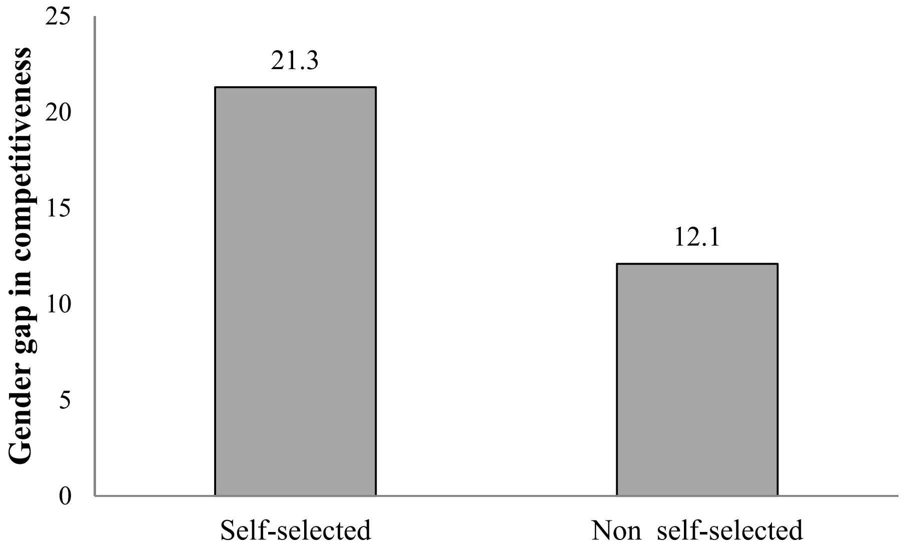


Figure 1 Gender gap in competitiveness (in percentage points) in previous studies surveyed in Dariel, Kephart, Nikiforakis and Zenker (2017) using designs similar to that in NV2007. *Self-selected* samples (N = 36) are those that could self-select into the experiment. *Non self-selected* samples (N = 17) are typically found in studies relying on classroom experiments in which participants did not receive prior information about the experiment. The gender gap is calculated as the unweighted average across different samples of the difference in the fraction of men and women choosing to compete in a design similar to that in NV2007.

As can be seen in Figure 1, the mean gender gap is 21.3 percentage points in studies using self-selected samples and 12.1 percentage points in studies using non self-selected samples.[3] That is, the gender gap in competitiveness is 76% *larger* in studies using self-selected samples. This suggests that selection bias causes us to substantially *over*estimate the gender gap in lab experiments. However, it is important to note that the data in Figure 1 do not permit us to rule out alternative explanations for the observed difference. In fact, there is a near perfect confound in the data which prevents us from drawing safe conclusions: all but one of the studies in which samples

[2] Dariel et al. (2017) provide an exhaustive list of all the studies that were either published or accepted for publication by the end of October 2017, i.e., ten years since the publication of NV2007, using designs similar to that of NV2007. Information on the classification of the different studies can be found in Appendix A.

[3] Using the data in Appendix A, we find that men are 68.8% more likely to compete than women in the self-selected samples, and 31.6% more likely in studies in which subjects do not self-select into the experiment.

do not self-select into the experiment, use samples of pre-university participants. Similarly, all but one of the studies using student samples rely on volunteer samples (see Appendix A for more details). For this reason, we designed an experiment to test the hypothesis that selection causes us to overestimate the difference in competitiveness between men and women.

In order to estimate the influence of self-selection, we use a method introduced by Cleave et al. (2013). At time t_0, we elicit individuals' willingness to compete in a large class with mandatory attendance such that we rule out self-selection into the experiment. For simplicity, we will refer to these individuals as the 'population'. At time t_1, we identify the subset of 'lab volunteers' in the population by observing who accepts an invitation to participate in lab experiments. To evaluate the effect of selection bias, we compare the gender gap among lab volunteers measured at time t_0 and compare it to that in the population. Specifically, let $\Delta_P^{t_0}$ denote the gender gap in competiveness in the population measured at t_0, i.e., in the classroom experiment. Let $\Delta_V^{t_0}$ denote the gender gap in competiveness among lab volunteers also measured at t_0. We will say self-selection biases our estimate of the gender gap if $\Delta_P^{t_0} \neq \Delta_V^{t_0}$. If $\Delta_P^{t_0} < \Delta_V^{t_0}$, we will say selection bias causes us to *over*estimate the gender gap. If $\Delta_P^{t_0} > \Delta_V^{t_0}$, we will say selection bias causes us to *under*estimate the gender gap. Note that, we compare $\Delta_P^{t_0}$ to $\Delta_V^{t_0}$ and *not* to $\Delta_V^{t_1}$ as learning, time and environmental effects would confound our estimates in that case.

Our data reveal that selection bias causes us to significantly *over*estimate the gender gap in competitiveness. The gap is 15 percentage points in the population (men are more willing to compete than women), and 31 percentage points in the subsample of lab volunteers. When we add controls for covariates that are known to affect the willingness of individuals to compete in our analysis, we find that selection bias causes us to overestimate the gender gap by a factor of 3 approximately. The bias is primarily driven by competitive men being more likely to select into the lab. In addition, we find that selection bias causes us to substantially overestimate the gender gap in risk attitudes in our population. This is due to the fact that risk tolerant women are less likely to select into the lab. Finally, we also find that selection bias causes us to greatly overestimate the tendency of low-performing men to select into competition.

The rest of this paper is organized as follows. Section 2 describes in detail the experiment. Section 3 presents the experimental results, and section 4 concludes.

2. The experiment

2.1 The population

Our population consists of students enrolled in the *Academic Skills* course at the Erasmus School of Economics (ESE), the Netherlands. Students in this course are taught presentation and discussion skills as well as how to study effectively. The main reason for choosing this course is that it is compulsory for all first-year students majoring in Economics, Econometrics, and Law and Economics. This implies that a large number of students are enrolled in the course (1,395). Another reason is that the course has a large number of (mandatory) lab tutorials, which simplifies the administration of a large-scale computerized experiment. Specifically, students in this course are divided into 96 tutorial groups with as many tutors. Each group consists of 15 students at most, implying that the tutor can easily monitor participants and ensure they do not communicate during the experiment.

The experiment took place during the first tutorial, in the first week of the academic year 2018-2019. Participants, therefore, had no previous exposure to laboratory experiments. To rule out self-selection, the experiment was not announced to students in advance. One week prior to the experiment, we gave tutors – who were unaware of our research question – precise instructions about how to run the experiment in class. Once in class, tutors guided students to a link on the course's webpage that directed them to the online experiment (see Section 1 in the *Online Supplementary Material, OSM*). Tutorials last 90 minutes and we purposefully limited the duration of the experiment (about 15 minutes) to ensure that students would not leave the class. If students did not wish to participate, tutors asked them to remain seated and be silent. Of the 1,216 students that showed up for the tutorials, 1,145 (94.2%) agreed to participate in the experiment. These 1,145 individuals – consisting of 811 males and 334 females – are our 'population.'

2.2 The experimental design

The experiment consists of four tasks and a short post-experiment questionnaire. The first three tasks are built on the paradigm introduced by Niederle and Vesterlund (2007) (NV2007). In each of them, participants are asked to add up a series of three two-digit numbers for 90 seconds. The tasks differ in how performance translates into earnings. In Task 1 (Individual Performance), participants are paid for each correct summation using a *piece rate* of 1 Euro. In Task 2 (Compared Performance), participants compete in a *tournament* against three other randomly selected anonymous students. The participant that correctly solves the highest number of additions receives 4 Euros per correct answer, while the other three people in the group receive nothing. Ties for the first place are randomly broken.

In the third task (Choice), participants must choose whether they wish their earnings to be determined by their individual performance (as in Task 1) or by their compared performance (as in Task 2). After making their choice, participants they have to add up a series of three two-digit numbers for a final time. The Choice task measures individuals' attitudes towards competition and is, therefore, our main variable of interest. Like in NV2007, if participants choose compared performance, their score is compared to that of participants in Task 2. In Task 4 (Lottery), students are given six lotteries and have to choose the one they prefer. The task is similar to that in Eckel and Grossman (2008), but lottery payoffs are chosen such that they reflect the incentives in the Choice task. The latter is important in light of a recent debate about the extent to which differences in the Choice task simply reflect gender differences in risk attitudes (Gillen et al. 2019; van Veldhuizen, 2018). The experimental instructions are available in the *OSM.*

The experiment concluded by asking participants to fill out a short questionnaire with questions about their gender, how much time they have for leisure activities, and the income group they belong to (from a scale from 1 to 10) (see Section 2 in the *OSM*). The survey also included a non-incentivized question aimed to capture participants' confidence in their abilities in Task 2. In particular, they were asked to state if they believed they solved the highest number of corrects summations in their group in Task 2.

Participants were informed that one in every four participants would be selected for payment and that, if selected, one of the four tasks would be selected for payment. While participants learned their individual score at the end of tasks 1-3, they were not informed about whether they won the tournament/lottery, their earnings or which task had been selected for payment. Participants were told that they will receive an email with information on who was selected for payment (see Section 3 in the *OSM*).[4]

[4] Students were informed that their earnings would be determined in a public event that would take place in the second week of the academic year. Subsequently, an announcement was posted on the course's webpage with the student ID numbers of those that were selected for payment. Students could pick up their earnings two weeks after the experiment took place, at which point, they learned their exact earnings but not of which task was selected for payment.

2.3 Identifying the lab volunteers in the population

We used a four-step process to identify lab volunteers in our population. We discuss each step in detail.

(1) Expressing an interest in experiments. We asked all participants in the classroom experiment if they would be interested to receive an e-mail with information about participating in future laboratory experiments at ESE.[5] To test whether the classroom experiment affected participants' willingness to volunteer in future lab experiments, half of the participants were randomly selected to be asked this question at the very first screen of the experiment, whereas the remaining half was asked the same question on the very last screen. As can be seen in Figure 2, 76.9% of the individuals in our population expressed an interest in receiving the email. The order of the question does not have a significant effect on who expresses an interest (Fisher Exact test, $N = 1{,}145$, $p = 0.14$) or volunteers for experiments (Fischer Exact test, $N = 1{,}145$, $p = 0.38$).

(2) Registering in the database of future volunteers. Two weeks after the last tutorial, all participants that expressed an interest in future experiments received an email by the lab manager (not us) to register to the ORSEE volunteer database (Greiner, 2015) (see Section 4 in the *OSM*). The email reminded them that they had expressed an interest in participating in lab experiments and contained a link for registering to the database. Reminder mails were sent twice and students who had provided us with multiple email addresses, received emails in all addresses. Note, that by the time they were invited to register in the database, students had been informed of their individual earnings from the classroom experiment. However, they did not receive any information about which task had been selected for payment or how their earnings were determined. Figure 2 shows that, similar to Cleave et al. (2013), only 19.7 percent of the entire population registered in the database of volunteers.

(3) Signing up for an experiment. Twelve weeks after the start of the classroom experiment, an email was sent to all those in our population that registered in the database of future volunteers (see Section 5 in the *OSM*). The email included links that enabled recipients to enroll for one of 20 laboratory sessions. The sessions were spread out over two weeks, the first being scheduled to run 10 days after the invitation email. None of the 20 sessions filled up to the maximum. We are

[5] As this was the first official recruitment drive of the academic year, it was extremely unlikely our first-year students would have signed up to the database prior to receiving this email.

therefore confident that lab constraints did not preclude individuals from participating in an experiment. Despite this, as can be seen in Figure 2, less than half of those who registered in the database signed up for an experiment (i.e., 8.6% of our population).

(4) Participating in an experiment. A large majority of those signing up for an experiment, participated in one. Specifically, 7.9% of the starting population participated in a lab experiment (90 out of 1,145 individuals).[6] This, again, is similar to the fraction found in Cleave et al. (2013) showing that the high level of attrition is not unique to our study. Following Cleave et al. (2013), we classify all those that participated in a lab experiment as 'lab volunteers'. This seems natural for our purposes as it is these subjects that we would normally use to estimate the gender gap in competitiveness in a lab experiment.[7] Importantly, we note that, while the fraction of lab volunteers is a small fraction of the population, in absolute numbers, lab volunteers are still more than in NV2007. Therefore, our tests using exclusively lab volunteers have slightly more power than those in NV2007.

We use the lab experiments only for the purpose of identifying who is a volunteer and who is not. We do *not* re-measure the willingness of volunteers to compete in these lab experiments as this would introduce several confounds in our comparisons (e.g., learning, different social environment). Instead, having identified who is a lab volunteer in our population, we estimate the gender gap among them using the decisions they made in the classroom experiment.

[6] Of the 90 lab volunteers, 57 are men (i.e., 7.0% of the men in our population) and 33 are women (i.e., 9.9% of the women in our population). In Appendix B, we discuss in detail attrition by gender, showing that attrition rates are similar for men and women.

[7] To ensure that we correctly identified lab volunteers in our population, we also checked a month after our lab sessions whether there were others in our population that participated in experiments arranged by other researchers, and classified those also as 'lab volunteers'.

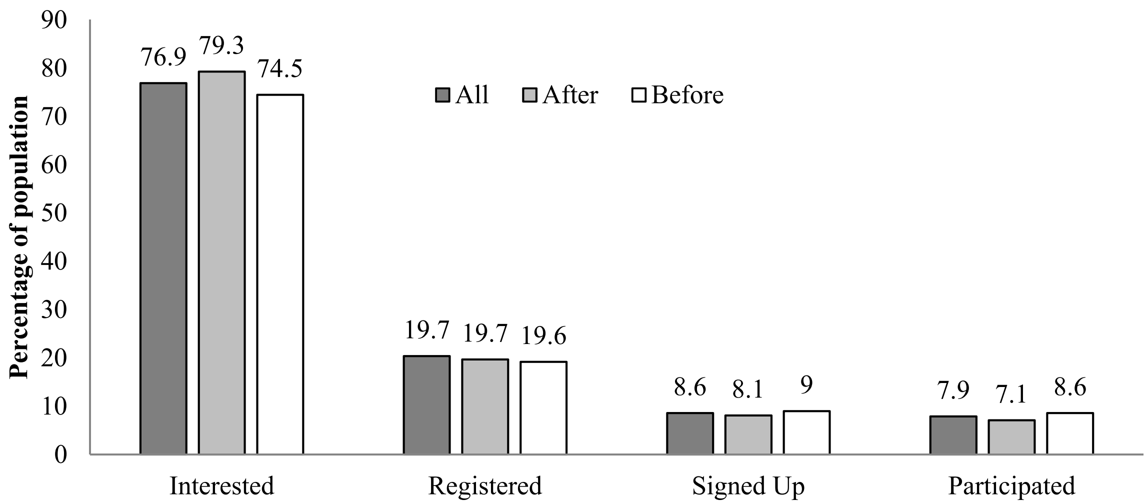


Figure 2 Attrition to the lab 'Interested' refers to the percentage of individuals in the population that indicated an interest in receiving an email with information about future experiments. 'Registered' refers to the percentage of individuals in the population that registered in the database of future volunteers. 'Signed Up' refers to the percentage of individuals in the population that signed up for one of the twenty lab sessions. 'Participated' refers to the percentage of individuals in the population that participated in an experiment – the lab volunteers. 'Before/After' refer to the order in which a question was asked during the classroom experiment about whether individuals wished to receive information about future experiments.

3 Results

We divide the analysis of our data into four subsections, each addressing a distinct question. The first addresses our main research question. The second subsection explores whether selection biases our estimates of gender differences in risk attitudes. The third investigates to what extent selection can account for the relatively higher tendency of low-performing men to select into competition. Finally, the last subsection explores factors that affect the selection of men and women into the lab.

3.1 Does selection bias cause us to overestimate the gender gap in competition?

We commence our analysis of the data by addressing our main research question. For clarity, the gender gap in competitiveness in the population is defined as $\Delta_P^{t_0} = m_P^{t_0} - w_P^{t_0}$, where $m_P^{t_0}$ and $w_P^{t_0}$ denote the fraction of men and women in the population, respectively, choosing to compete in Task 3 at time t_0, i.e., in the classroom experiment. Similarly, the gender gap in competitiveness among lab volunteers is defined as $\Delta_V^{t_0} = m_V^{t_0} - w_V^{t_0}$, where $m_V^{t_0}$ and $w_V^{t_0}$ denote the fraction of men and women lab volunteers, respectively, that choose to compete in Task 3 in the classroom experiment. We will say that selection bias causes us to overestimate the gender gap if $\Delta_V^{t_0} > \Delta_P^{t_0}$.

Figure 3 shows the fraction of men and women that chose to compete in Task 3. In line with past studies, we find that men clearly compete more often than women in the sample of lab volunteers (49% vs. 18%; Fisher exact test, two-tailed, N = 90, p < 0.01), i.e., $\Delta_V^{t_0} = 0.31$. A significant gender difference is also observed in our population (36% vs. 21%; Fisher exact test, two-tailed, N = 1,145, p < 0.01), i.e., $\Delta_P^{t_0} = 0.15$. Most importantly, the size of the gender gap is approximately twice as large among lab volunteers than among men and women in the population from which the volunteers were recruited, i.e., $\Delta_V^{t_0} > 2 * \Delta_P^{t_0}$. In other words, in line with our hypothesis, self-selection causes us to greatly overestimate the gender gap in competitiveness in our sample.

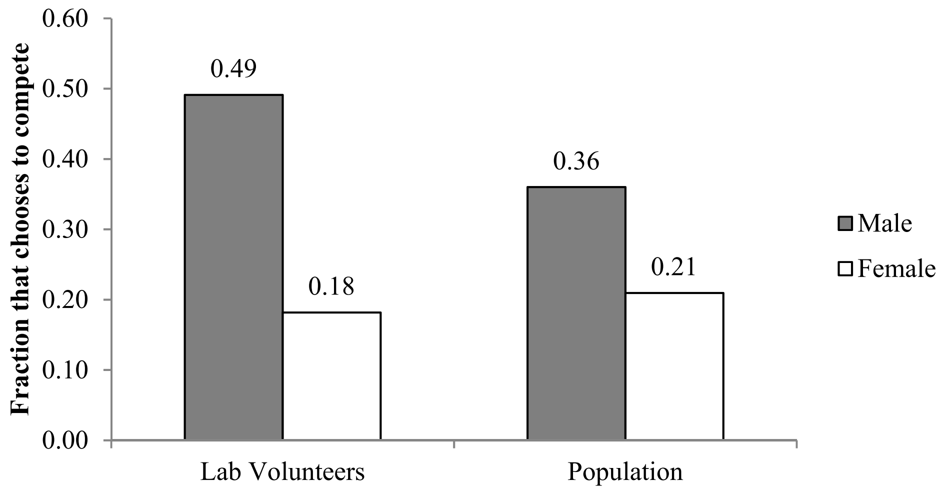


Figure 3 Gender gap in competitiveness Fraction of participants that chooses to compete in Task 3 of the experiment among volunteers and the population from which they were recruited.

Table 1 presents evidence from linear probability models to explore the significance of the bias in statistical terms. As a benchmark for comparison, Model (I) shows that women are 15 percentage points less likely than men to compete in the population (p <0.01). Model (II) shows that the *Female* variable is largely unchanged in the sample of non-volunteers. Most importantly, the interaction term *Female* × *Volunteer* in Model II shows that the bias in our estimate of the gender gap due to selection is substantial and statistically significant (p < 0.01). Model (II) also presents some first evidence suggesting that the bias is driven by competitive men who are more likely to volunteer for lab experiments ('Volunteer', p = 0.04). Evidence of this can also be seen in Figure 3: the fraction of competitive women is similar in the subset of volunteers and in the population, but not the fraction of competitive men. Men who choose to compete are more common in our sample of lab volunteers, i.e., substantially more competitive men select into our sample of volunteers, i.e., $m_V^{t_0} > m_P^{t_0}$.

	(I)	(II)
Female	-0.15***	-0.14***
	(0.03)	(0.03)
Volunteer		0.14**
		(0.07)
Female × Volunteer		-0.17**
		(0.08)
Constant	0.36***	0.35***
	(0.02)	(0.02)
Observations	1,145	1,145
R-squared	0.02	0.03

Table 1 Selection bias in competitiveness Linear probability regression. The dependent variable is a dummy variable indicating whether an individual competed in Task 3. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

The analysis in Table 1 does not control for covariates that could explain both the gender gap in competitiveness and self-selection into the lab. For instance, NV2007 show that individual performance in the summation tasks and beliefs about the probability of willing both affect positively the probability of choosing to compete in Task 3. To that end, Table 2 replicates the main statistical analysis in NV2007 (Table 6, p. 1089), separately for our sample of lab volunteers and for our population.[8] The estimated gender gap among lab volunteers is comparable to that reported in NV2007. Specifically, in Model (II), we find a point estimate for *Female* of -0.225 which is similar to the estimate of -0.278 reported in NV2007. Importantly, the additional controls further increase the estimated bias due to selection. Comparing the coefficient of *Female* in Model (II) and Model (IV) we observe that the absolute value of the coefficient for *Female* decreases from 0.225 (Model II) to 0.074 (Model IV). That is, when controlling for covariates, the gender

[8] In Appendix C, we present summary statistics for these and other control variables. Note, that in their Table 6, NV2007 present Probit estimates, whereas we present linear probability results to remain consistent with the remainder of our analysis. The results of a Probit analysis, however, are virtually identical.

gap in the sample of lab volunteers is found to be 3 times larger than that in the population, i.e., $\Delta_V^{t_0} = 3 * \Delta_P^{t_0}$.

	Volunteers (I)	Volunteers (II)	Population (III)	Population (IV)
Female	-0.21**	-0.23**	-0.09***	-0.07***
	(0.09)	(0.09)	(0.03)	(0.03)
Tournament	0.08***	0.07***	0.05***	0.04***
	(0.02)	(0.02)	(0.01)	(0.01)
Tournament – Piece rate	-0.04	-0.03	-0.01	-0.01
	(0.03)	(0.03)	(0.01)	(0.01)
Confidence		0.24*		0.32***
		(0.14)		(0.04)
Constant	-0.04	0.01	0.01	0.05
	(0.14)	(0.13)	(0.03)	(0.04)
Observations	90	90	1145	1145
R-squared	0.21	0.24	0.09	0.17

Table 2 The determinants of competition Linear probability regression replicating the analysis in NV2007 (Table 6, p.1089). The dependent variable is a dummy variable indicating whether an individual competed in Task 3. 'Tournament' measures the numbers of correct answers in Task 2, 'Tournament – piece rate' measures the difference between a subject's correct answers in Task 1 and 2, 'Confidence' is a dummy variable with value 1 if a subject believes he or she has a higher score in Task 2 than all competitors. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

3.2 The role of risk attitudes

Given the uncertainty inherent in tournaments, the decision to engage in competition is likely to be affected by individuals' attitudes toward risk. Indeed, two recent studies provided evidence that the gender gap in competitiveness disappears when differences in risk attitudes are taken into account in the analysis (Gillen et al. 2019; van Veldhuizen 2018).[9] For our purposes, this raises two questions: (*i*) To what extent is the observed *selection bias* in gender competitiveness driven by risk attitudes? (*ii*) Does self-selection bias our estimates of gender differences in risk attitudes?

[9] It is worth noting that the discussion about the role of risk attitudes is ongoing as other studies have found risk attitudes do not fully account for the gap in competitiveness (e.g., Datta Gupta et al. 2013, Buser et al. 2014, Wozniak et al 2014, Dariel et al. 2017 and Lowes 2018). Gillen et al. (2019) and van Veldhuizen (2018) attribute their findings to reducing the error in measuring risk attitudes.

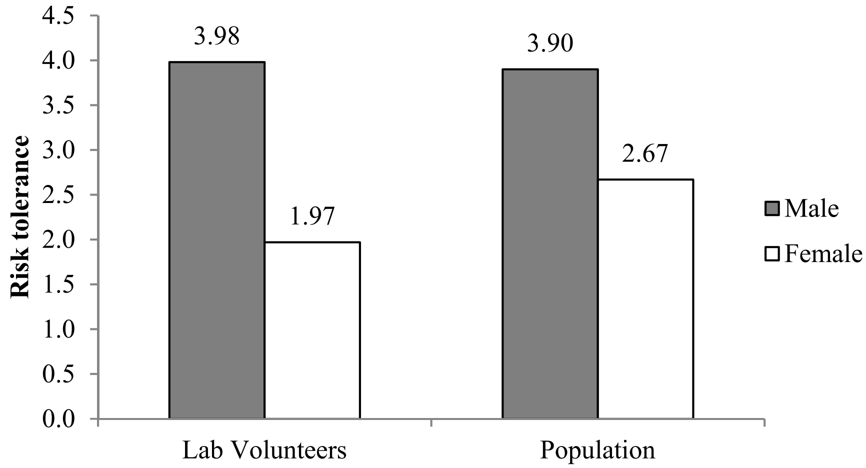


Figure 4 Mean risk tolerance

We first explore the latter question, i.e., whether self-selection biases the estimated gender gap in risk attitudes. This will happen if risk attitudes influence the decision of individuals to volunteer for experiments differently for men and women. We follow the same process as in the previous subsection. Figure 4 shows how risk tolerant men and women in our study are. As can be seen, men are more risk tolerant than women both in our sample of lab volunteers (Mann-Whitney U test, two-tailed, $N = 90$, $p < 0.01$), as well as in our population (Mann-Whitney U test, two-tailed, $N = 1{,}145$, $p < 0.01$). The data in Figure 4 suggest that selection bias causes us to *over*estimate the differences in risk attitudes between men and women by 63%. This appears to be driven by less risk tolerant women being more likely to volunteer for lab experiments.

	(I)	(II)
Female	-1.23***	-1.15***
	(0.12)	(0.12)
Volunteer		0.09
		(0.25)
Female × Volunteer		-0.87***
		(0.31)
Constant	3.90***	3.90***
	(0.08)	(0.08)
Observations	1,145	1,145
R-squared	0.0846	0.0890

Table 3 Selection bias in risk attitudes OLS regression. The dependent variable increases with the riskiness of the lottery chosen by participants. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

Table 3 presents evidence from linear regressions exploring the significance of the bias in statistical terms. As a benchmark for comparison, Model (I) shows that women are less risk tolerant than men in our population. Model (II) shows that the *Female* variable is similar in the sample of non-volunteers. Importantly, the interaction term in Model (II) illustrates that selection bias causes us to significantly overestimate gender differences in risk attitudes.

	Volunteers (I)	Population (II)	Population (III)
Female	-0.05	-0.02	-0.01
	(0.10)	(0.03)	(0.03)
Tournament	0.05**	0.03***	0.03***
	(0.02)	(0.01)	(0.01)
Tournament - Piece rate	-0.04	-0.00	-0.00
	(0.03)	(0.01)	(0.01)
Confident	0.24*	0.31***	0.31***
	(0.13)	(0.04)	(0.04)
Risk tolerance	0.09***	0.05***	0.05***
	(0.03)	(0.01)	(0.01)
Volunteer			0.15**
			(0.06)
Volunteer × Female			-0.17**
			(0.08)
Constant	0.25	0.51***	0.51***
	(0.31)	(0.09)	(0.09)
Observations	90	1,145	1,145
R-squared	0.32	0.20	0.21

Table 4 Risk attitudes and competition OLS regression. The dependent variable a subject's choice to compete in Task 3. Standard errors are clustered at the session level. 'Tournament' measures the amount of correct answers in Task 2, 'Tournament – Piece rate' measures the difference between a subject's score in Task 1 and 2, 'Confidence' is a dummy variable taking a value of 1 if a subject believes he or she has a higher score in Task 2 than all competitors, 'Risk tolerance' is a variable that increases in an individual's willingness to take on risk, on a scale 0-5. ***/**/*: significant at the 1%/5%/10%-level.

We now turn our attention to the first question: To what extent is the observed selection bias in gender competitiveness driven by risk attitudes? To answer this question, we extend the analysis presented in Table 2 to control for risk attitudes. The findings are presented in Table 4 which reveals three interesting findings. First, risk helps explain the willingness to compete both in the lab (Model I) and in the population (Model II). Second, in line with the results in Gillen et al. (2019) and van Veldhuizen (2018), we find that once we control for risk attitudes there does not appear to be a significant gender gap in competitiveness (captured by *Female*), either among lab volunteers or in the population. Finally, despite this, the coefficient for the interaction term *Volunteer* × *Female* in Model (III) is unchanged from that presented in Table 2 suggesting that risk attitudes cannot account for the selection bias in competitiveness seen in the previous subsection. In other words, risk attitudes and attitudes toward competition seem to affect separately the decision to volunteer for lab experiments. We explore this in section 3.4.

3.3 Performance and selection

One of the most intriguing findings in NV2007 was that high-performing women tended to compete 'too little', whereas low-performing men to compete 'too much.' In this subsection, we explore to what extent self-selection into experiments can account for this result.

Twenty-five percent is the probability that equalizes the expected earnings of an individual from the tournament and the piece rate. Therefore, we refer to people with a score in Task 3 that gives them a probability of at least 25% chance of winning the tournament as 'high performers'; if their score translates to a probability below 25%, we refer to them as 'low performers'. We follow NV2007 (Table III, p. 1085) in classifying individuals into two categories: (*i*) low performers that choose to compete ('over-competitors'), and (*ii*) high performers that choose not to compete ('under-competitors').

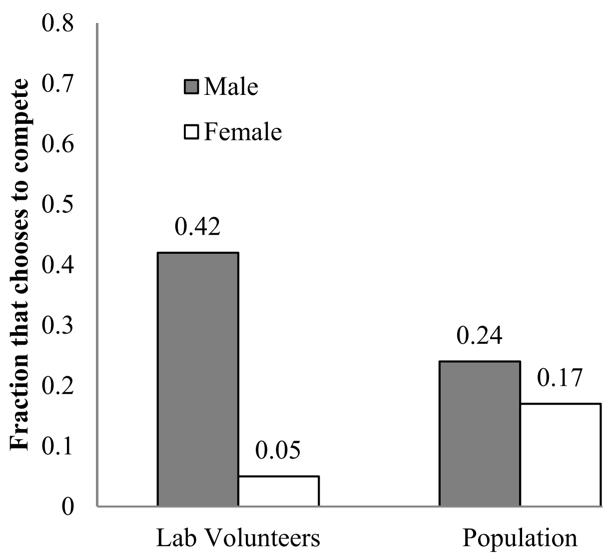


Figure 5A *Low performers*: subjects with a score that gives them a chance below 25% of winning the tournament.

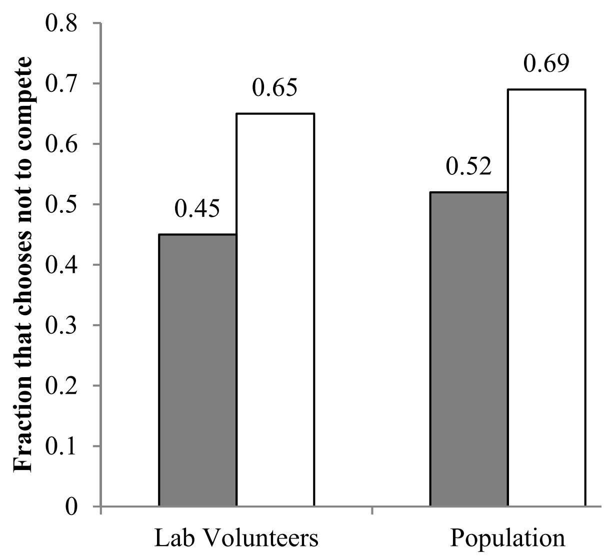


Figure 5B *High performers*: subjects with a score that gives them a greater than 25% chance of winning the tournament.

Figure 5 shows the fraction of low performers that compete (Panel A) and the fraction of high performers that abstain from competition (Panel B). Panel A reveals that, in line with NV2007, we find a substantial gender difference in over-competing among lab volunteers (Fischer Exact test, two-tailed, $N = 45$, $p < 0.01$). However, the gender difference in over-competing is about five times smaller in the population (Fischer Exact test, two-tailed, $N = 659$, $p = 0.05$). Panel B shows that selection does not seem to affect our conclusions regarding the tendency of high performers to avoid competition: high-performing women are more likely to abstain from competition than men to a similar extent in the population (Fischer Exact test, two-tailed, $N = 486$, $p < 0.01$) and among lab volunteers (Fischer Exact test, two-tailed, $N = 45$, $p = 0.34$), but the latter is not statistically significant due to the small number of high performers volunteering for experiments. Therefore, our data indicate that selection bias overestimates the willingness of low-performing men to compete relative to women, but not the tendency of high-performing women to avoid competition relative to men.

	Low performers			High performers		
	(I)	(II)	(III)	(IV)	(V)	(VI)
Female	-0.07**	-0.04	0.00	0.17***	0.18***	0.05
	(0.03)	(0.04)	(0.03)	(0.06)	(0.07)	(0.07)
Volunteer		0.20*	0.22**		-0.07	-0.08
		(0.10)	(0.09)		(0.09)	(0.08)
Volunteer × Female		-0.33***	-0.31***		0.01	-0.02
		(0.12)	(0.11)		(0.15)	(0.15)
Confident			0.42***			-0.26***
			(0.05)			(0.04)
Risk tolerance			0.04***			-0.06***
			(0.01)			(0.01)
Constant	0.24***	0.23***	0.03	0.52***	0.52***	0.89***
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.05)
Observations	659	659	659	486	486	486
R-squared	0.01	0.02	0.17	0.02	0.02	0.15

Table 5 The propensity to overcompete (low performers) and undercompete (high performers) Linear probability model. The dependent variable is a dummy capturing a subject's choice to compete in Task 3 for 'low performers', and a subject's choice *not* to compete for 'high' performers. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

To explore the statistical significance of the bias, Table 5 presents the results from a regression analysis, similar to that in the previous subsection. Model (I) indicates that low-performing men are 7 percentage points more likely to compete than low-performing women in the population.

Model (II) shows that this difference is significant among volunteers but not among non-volunteers. The interaction term (*Volunteer* × *Female*) in that model reveals that the selection bias is large and statistically significant. Model (III) shows that the bias cannot be fully explained by individual beliefs and attitudes toward risk. For high performers, we find women are more likely to *abstain* from competition (Model IV) and no evidence of selection bias affecting our estimates (Model V).[10] Finally, Model (VI) reveals that the relative unwillingness of high-performing women to compete can be explained by their lower tolerance to risk and confidence.

3.4 What determines selection into the lab?

The evidence of selection bias in the previous subsections raises the question of what factors influence the decision of individuals to volunteer for experiments and how these differ for men and women. Already in sections 3.1 and 3.2, we saw evidence suggesting that different factors may account for the willingness of men and women to select into the lab. In this subsection, we expand our analysis to consider the effect of competitiveness and risk attitudes on selection, while controlling for a number of other covariates.

Table 6 presents the results from a linear probability model where the dependent variable is a participant's choice to volunteer for lab experiments. The independent variables include our two main behavioral measures – that of an individual's competitiveness (*Compete*) and that of their willingness to take on risk (*Risk tolerance*) – a self-reported measure of one's leisure time (*Leisure*), a self-reported measure of their income (*Income*), a control for whether they were asked to express their interest for future lab experiments before the classroom experiment (*Before*), and, finally, a control for whether they were selected to be paid for their participation in the classroom experiment (*Paid*).

The results in Table 6 reveal that different factors drive the decision of men and women in our population to volunteer for experiments. For men, only the willingness to compete is a significant predictor of their choice to volunteer; risk attitudes appear to have no influence in their decision. For women, on the other hand, it is competitiveness that seems to have no influence on their willingness to volunteer; only their risk attitudes appear to determine selection as well as whether they were selected for payment. These findings are important for experimental economists, not

[10] As can be seen in Table 5, 486 individuals are classified as high performers, i.e., 42.4% of the population. The reason this number is higher than 25% is that there are learning effects across tasks which, although small in magnitude, exist for many individuals. We do not find evidence of different learning effects for men and women.

only because we find evidence that variables we are often interested in measuring such as competitiveness and risk attitudes affect selection, but also because men and women appear to select differently into the lab raising challenging questions for experimental studies on gender differences.

	Male	Female
Compete	0.04**	0.00
	(0.02)	(0.04)
Risk tolerance	-0.00	-0.03***
	(0.01)	(0.01)
Leisure	-0.00	0.03
	(0.01)	(0.02)
Income	-0.01	-0.00
	(0.00)	(0.01)
Before	0.01	0.02
	(0.02)	(0.03)
Paid	0.00	0.08**
	(0.02)	(0.04)
Constant	0.10**	0.07
	(0.05)	(0.10)
Observations	811	334
R-squared	0.09	0.04

Table 6 The determinants of selecting into the lab for men and women Linear probability models. The dependent variable is a dummy whether a subject volunteered for lab experiments. Standard errors are clustered at the session level. ***/**/*: significant at the 1%/5%/10%-level.

4. Conclusion

We have presented evidence showing that the widely-documented gender gaps in competitiveness and risk attitudes are substantially larger in a sample of volunteers that self-select into lab experiments than they are in the population from which the volunteers were recruited. Similarly, we found that the relative tendency of low-performing men to compete is much larger in the sample of lab volunteers than in the population. Finally, to our knowledge, our study provides the first evidence that men and women select differently into the lab. What are the implications of these findings?

On the specific issue of gender differences, our results suggest that selection bias in lab experiments may have caused us to overestimate the gap in competitiveness and also risk attitudes in previous studies (see also Figure 1). However, lest we be misunderstood, we emphasize that our study is *not* meant to be critical of controlled experiments as a method for empirical investigation; after all, we used a controlled experiment for our investigation. We also have no intention of being dismissive of controlled experiments exploring gender attitudes. Experiments such as that by Niederle and Vesterlund (2007) have a comparative advantage over other empirical methods which lies in their unrivaled ability to isolate factors hampering comparisons between men and women in daily life (e.g., differences in opportunities and skills), thus allowing researchers to focus on the variables of interest (e.g., underlying differences in competitiveness and risk attitudes).

The broader message from our study is the need to understand the limits of drawing *general* inferences from experiments using self-selected samples when the variable of interest has not been randomized by the experimenters (e.g., gender, race, ethnicity, academic background). Of course, someone might argue that, despite the evidence of substantial selection bias in our study, the problem posed by self-selection appears to be limited as we observe *qualitatively* similar results in our sample of volunteers and in the population from which we recruited them. While this is true, we believe that our 'population' is of no particular interest; in fact, we regard it as nothing more than an empirical benchmark for exploring the effect of self-selection. The existence of selection bias in our study, especially given its large magnitude, does not permit one to rule out the possibility that a gender gap found in a given lab experiment will not generalize to a broader population, either quantitatively or qualitatively. When the independent variable of interest cannot be randomized by the experimenters, researchers may wish to explore the limits of their findings either by using experiments such as the one presented in this paper or experiments with representative samples.

Like all empirical studies, it is important to remember that our paper offers but a data point to help answer a general research question. This implies that, while we find clear evidence of substantial bias in our estimates owing to selection, we *cannot* conclude that the same applies to estimates obtained from volunteer samples by other researchers. The characteristics of the population from which participants are recruited and the specifics of the recruitment method will ultimately determine whether selection bias plagues the estimates coming from a certain laboratory. To that end, established laboratories could usefully consider using the method employed in this paper to collect data on a broad range of individual attitudes (e.g., risk attitudes, social preferences, attitudes toward competition, time preferences) that will enable them to readily evaluate the extent of selection bias in different domains. This will also allow us to perform meta-analyses to better understand the process and impact of self-selection into lab experiments, calibrate our estimates, and reduce concerns about the generalizability of experimental findings.

Appendix A

		% selecting competition				
Study	Year	Male	Female	Gender Gap	Subjects	Self-selected
Niederle and Vesterlund	2007	73%	35%	38%	Students	Yes
Gneezy et al.	2009	50%	26%	24%	Other	Yes
Gneezy et al.	2009	39%	54%	-15%	Other	Yes
Healy and Pate	2011	81%	28%	53%	Students	Yes
Balafoutas and Sutter	2012	64%	30%	34%	Students	Yes
Balafoutas et al.	2012	59%	31%	28%	Students	Yes
Cardenas et al.	2012	44%	19%	25%	Children	No
Cardenas et al.	2012	39%	27%	12%	Children	No
Cardenas et al.	2012	35%	32%	3%	Children	No
Cardenas et al.	2012	26%	29%	-3%	Children	No
Dargnies	2012	85%	51%	34%	Students	Yes
Kamas and Preston	2012	41%	23%	18%	Students	Yes
Mayr et al.	2012	56%	36%	20%	Other	Yes
Mueller and Schwieren	2012	42%	26%	16%	Students	Yes
Price	2012	66%	49%	17%	Students	Yes
Shurchkov	2012	39%	30%	9%	Students	Yes
Shurchkov	2012	44%	19%	25%	Students	Yes
Andersen et al.	2013	52%	49%	3%	Children	Yes
Andersen et al.	2013	51%	39%	12%	Children	Yes
Cadsby et al.	2013	36%	9%	27%	Students	Yes
Datta Gupta et al.	2013	60%	34%	26%	Students	Yes
Niederle et al.	2013	74%	31%	43%	Students	Yes
Samak	2013	77%	83%	-6%	Children	No
Buser et al.	2014	49%	23%	26%	Children	No
Dreber et al.	2014	36%	17%	19%	Children	No
Dreber et al.	2014	33%	28%	5%	Children	No
Lee et al.	2014	30%	22%	8%	Children	Yes
Wozniak et al.	2014	54%	31%	23%	Students	Yes
Wozniak et al.	2014	50%	30%	20%	Students	Yes
Apicella et al.	2015	45%	30%	15%	Other	Yes
Apicella et al.	2015	52%	37%	15%	Other	Yes
Apicella et al.	2015	67%	29%	38%	Other	Yes
Brandts et al.	2015	59%	30%	29%	Students	Yes
Khachatryan et al.	2015	54%	52%	2%	Children	No
Khachatryan et al.	2015	57%	56%	1%	Children	No
Sutter and Rutzler	2015	40%	19%	21%	Children	No
Almås et al.	2016	52%	32%	20%	Children	Yes
Berlin and Dargnies	2016	63%	35%	28%	Students	Yes
Cassar et al.	2016	36%	26%	10%	Other	Yes
Sutter et al.	2016	44%	21%	23%	Children	No
Apicella et al.	2017	58%	38%	20%	Students	Yes
Buser et al.	2017	52%	28%	24%	Students	Yes

Buser et al.	2017	68%	51%	17%	Children	No
Dariel et al.	2017	50%	54%	-4%	Students	No
Halko and Saakvuori	2017	74%	54%	20%	Students	Yes
Reuben et al.	2017	54%	27%	27%	Students	Yes
Banerjee et al.	2018	22%	16%	6%	Other	Yes
Bönte et al.	2018	56%	45%	11%	Other	Yes
Buser et al.	2018	42%	26%	16%	Students	Yes
Zhong et al.	2018	49%	25%	24%	Students	Yes
Zhang et al.	2019	63%	48%	15%	Children	No
Zhang et al.	2019	60%	38%	22%	Children	No
Zhang et al.	2019	75%	48%	27%	Children	No

Table A.1 Summary and classification of studies surveyed by Dariel et al. (2017). Studies listed more than once use different samples. For details on the tasks used in these studies and sample sizes, see Dariel et al. (2017). Studies are presented in chronological order of publication. The publication year for studies that were 'forthcoming' at the time Dariel et al. (2017) was published has been updated.

Appendix B

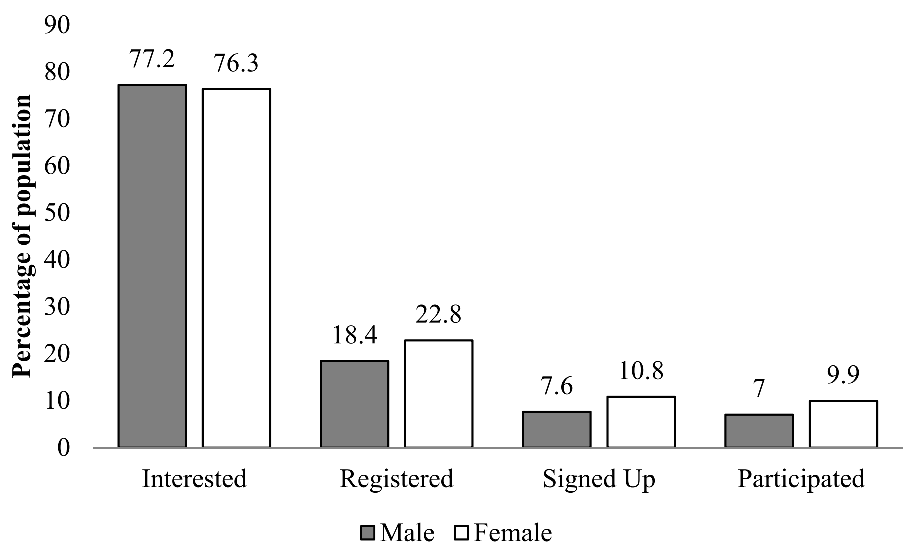


Figure B.1 Population attrition by gender 'Interested' refers to the percentage of the population who indicated interest to receive an email with more information about the ORSEE database. 'Registered' refers to the percentage of the population who registered in the ORSEE database. 'Signed Up' refers to the percentage of the population who signed up for one of the lab sessions. 'Participated' refers to the percentage of the population who showed up in one of the sessions.

Appendix C

Table C.1 presents summary statistics for the other variables collected during the experiment. In the population, males solve a higher number of math problems and are more willing to take risks (all are significant at $p < 0.01$ using Mann-Whitney U tests). The same holds for the sample of volunteers (Mann-Whitney U test at $p < 0.01$ for task 1 and task 4, and $p = 0.09$ for task 2). Men in the population also report to be more confident than women, and to have more leisure and higher income (Mann-Whitney U tests, all significant at $p < 0.01$). For the sample of volunteers, however, we do not find significant differences between men and women for confidence, leisure and income (Mann-Whitney U tests, $p > 0.29$).

	Males		**Females**	
	Population ($N = 811$)	Lab volunteers ($N = 57$)	Population ($N = 334$)	Lab volunteers ($N = 33$)
Task 1 (Piece rate)	5.84	6.18	4.68	4.66
	(2.55)	(2.34)	(2.30)	(2.13)
Task 2 (Tournament)	6.44	6.82	5.39	6.00
	(2.48)	(2.22)	(2.04)	(2.15)
Task 4 (Risk)	3.90	3.98	2.67	1.97
	(1.92)	(1.80)	(1.62)	(0.88)
Confident	0.28	0.21	0.15	0.18
	(0.44)	(0.41)	(0.36)	(0.39)
Leisure	3.32	3.32	3.01	3.15
	(0.88)	(0.93)	(0.81)	(0.76)
Income	6.15	5.80	5.75	5.61
	(2.02)	(1.92)	(1.87)	(1.68)

Table C.1 Average score in Tasks 1, 2 and 4 (standard deviation in parentheses). For Tasks 1 and 2 the score reflects the amount of correctly solved math problems in 90 seconds. Task 4 is a risk task that ranges from 1 (taking no risk) to 6 (taking the maximum amount of risk). *Confident* is a dummy variable indicating whether a subject believes s/he has outperformed the group member in Task 2. *Leisure* is a self-reported measure (on a scale of 1-5) of how much leisure a subject perceives to have. Finally, *Income* is a self-reported measure (on a scale of 1-10) regarding the income bracket a subject thinks s/he belongs.

References

Abeler. J. & Nosenzo. D. (2015). Self-selection into laboratory experiments: pro-social motives versus monetary incentives. *Experimental Economics* 18 (2), 195-214.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E., & Tungodden, B. (2016). Willingness to compete: Family matters. *Management Science, 62*(8), 2149–2162.

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics, 95*(4), 1438–1443.

Apicella, C. L., Demiral, E. E., & Mollerstrom, J. (2017). No gender difference in willingness to compete when competing against self. *American Economic Review*, *107*(5), 136-40.

Balafoutas, L., Kerschbamer, R., & Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior & Organization, 83*(1), 125–135.

Balafoutas, L., & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science,* 335(6068), 579–582.

Banerjee, R., Gupta, N. D., & Villeval, M. C. (2018). The spillover effects of affirmative action on competitiveness and unethical behavior. *European Economic Review*, *101*, 567-604.

Berlin, N., & Dargnies, M. P. (2016). Gender differences in reaction to feedback and willingness to compete. *Journal of Economic Behavior & Organization, 130,* 320–336.

Blau FD. Kahn LM (2017) The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55, 789-865.

Bönte, W., Procher, V., & Urbig, D. (2018). Gender differences in selection into self-competition. *Applied Economics Letters*, *25*(8), 539-543.

Buser, T., Gerhards, L., & Van Der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, *56*(2), 165-192.

Benndorf, V., Moellers, C. & Normann, H. (2017). Experienced vs. inexperienced participants in the lab: do they behave differently? *Journal of Economic Science Association* 3, 12–25.

Buser. T., Niederle. M. & Oosterbeek. H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics* 129 (3), 1409–1447.

Buser. T., Peter. N., & Wolter S.C. (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review* 107(5), 125-130.

Cadsby, C. B., Servátka, M., & Song, F. (2013). How competitive are female professionals? A tale of identity conflict. *Journal of Economic Behavior & Organization, 92,* 284–303.

Cárdenas, J. C., Dreber, A., Von Essen, E., & Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior & Organization, 83*(1), 11–23.

Cassar, A., Wordofa, F., & Zhang, Y. J. (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of Sciences, 113*(19), 5201–5205.

Cleave. B. L., Nikiforakis. N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics* 16 (3). 372-382.

Croson, R., & Gneezy. U. (2009). Gender differences in preferences. *Journal of Economic Literature*. *47*(2). 448-74.

Dargnies, M. P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science, 58*(11), 1982–2000.

Dariel, A., Kephart, C., Nikiforakis. N., & Zenker, C. (2017). Emirati women do not shy away from competition: Evidence from a patriarchal society in transition. *Journal of the Economic Science Association*. *3*(2). 121-136.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry, 51*(1), 816–835.

Dreber, A., von Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics, 17*(1), 154–172.

Eckel, C. C., & Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, *1*, 1061-1073.

Exadaktylos. F., Espín. A. M., & Branas-Garza, P. (2013). Experimental subjects are not different. *Scientific Reports*. *3*. 1213.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326* (5952), 535-538.

Falk. A., Meier. S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association 11*(4). 839-852.

Gillen. B., Snowberg. E., & Yariv. L. (2019). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy*. *127*(4). 1826-1863.

Gneezy, U., Leonard, K.L. & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica,* 77(5), 1637-1664.

Goldin. C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*. *104*(4). 1091-1119.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114-125.

Halko, M. L., & Sääksvuori, L. (2017). Competitive behavior, stress, and gender. *Journal of Economic Behavior & Organization, 141,* 96–109.

Healy, A. & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, 121(555), 1192–1204.

Harrison. G. W., Lau. M. I., & Rutström. E. E. (2009). Risk attitudes. randomization to treatment. and self-selection into experiments. *Journal of Economic Behavior & Organization*. *70*(3). 498-507.

Kamas, L., & Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization, 83*(1), 82–97.

Khachatryan, K., Dreber, A., Von Essen, E., & Ranehill, E. (2015). Gender and preferences at a young age: Evidence from Armenia. *Journal of Economic Behavior & Organization, 118,* 318–332.

Kocher. M. G., Schindler. D., Trautmann. S. T., & Xu. Y. (2019). Risk. time pressure. and selection effects. *Experimental Economics*. *22*(1). 216-246.

Lee, S., Niederle, M., & Kang, N. (2014). Do single-sex schools make girls more competitive? *Economics Letters, 124*(3), 474–477.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153-174.

Lowes. S. (2018). *Matrilineal kinship and gender differences in competition*. Working paper. Harvard University. Available online: https://scholar. harvard. edu/slowes/publications/ matrilineal-competition (accessed 9 March 2018).

Mayr, U., Wozniak, D., Davidson, C., Kuhns, D., & Harbaugh, W. T. (2012). Competitiveness across the life span: The feisty fifties. *Psychology and Aging, 27*(2), 278–285.

Müller, J., & Schwieren, C. (2012). Can personality explain what is underlying women's unwillingness to compete? *Journal of Economic Psychology, 33*(3), 448–460.

Niederle. M. (2016). Gender. In the *Handbook of Experimental Economics*. second edition. Eds. John Kagel and Alvin E. Roth. Princeton University Press.

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

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science, 59*(1), 1–16.

Price, C. R. (2012). Gender, competition, and managerial decisions. *Management Science, 58*(1), 114–122.

Reuben, E., Wiswall, W., & Zafar, B. (2017). Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. *Economic Journal* 127: 2153-2186.

Reuben, E., Sapienza, P., & Zingales, L. (2019). Taste for competition and the gender gap among young business professionals. NYU Abu Dhabi Working Paper.

Samak, Anya C. (2013). Is there a gender gap in preschoolers' competitiveness? An experiment in the US. *Journal of Economic Behavior & Organization*, 92, 22–31.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association, 10*(5), 1189–1213.

Slonim, R., Wang, C., Garbarino, E., & Merrett, D. (2013). Opting-in: Participation bias in economic experiments. *Journal of Economic Behavior & Organization*. *90*. 43-70.

Schulz, J., Sunde, U., Thiemann. P., & Thöni, C., (2019). Selection into Experiments: Evidence from a Population of Students. *GMU Working Paper in Economics* #19-39.

Sutter. M. & Glätzle-Rützler D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science.* 61(10). 2339–2354.

Sutter, M., Glätzle-Rützler, D., Balafoutas, L., & Czermak, S. (2016). Canceling out early age gender differences in competition—an analysis of policy interventions. *Experimental Economics, 19*(2), 412–432.

van Veldhuizen. R. (2018). Gender differences in tournament choices: Risk preferences. overconfidence or competitiveness? *LMU Discussion Paper* (No. 14).

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*. *32* (1). 161-198.

Zhang, Y. J. (2019). Culture, institutions and the gender gap in competitive inclination: Evidence from the communist experiment in China. *The Economic Journal*, *129* (617), 509-552.

Zhong, S., Shalev, I., Koh, D., Ebstein, R. P., & Chew, S. H. (2018). Competitiveness and stress. *International Economic Review*, *59*(3), 1263-1281.

Online Supplementary Material

Does selection bias cause us to overestimate gender differences in competitiveness?

List of Contents

1. Post with instructions for logging on to the experiment
2. Screenshots of the classroom experiment, including the experimental instructions
3. Post with information about earnings from the experiment
4. Email invitation to join volunteer database for lab experiments
5. Email invitation to participate in a lab experiment

1. Post with instructions for logging on to the experiment

Once in their tutorial room, students were instructed to go to the website for their course and read the following post.

Title of post: In-class experiment

Main text of post: Please select the link belonging to your work group. Select 'OTHER' in case your group is not listed below.

Work group	Link to Survey
IB01A	IB01A (Links to an external site.) Links to an external site.
…	…
IB12B	IB12B (Links to an external site.) Links to an external site.
OTHER	OTHER (Links to an external site.) Links to an external site.

2. Screenshots of the classroom experiment, including the experimental instructions

Below are shots of all the screens used in the classroom experiment. For clarity, each screen is presented on a separate page.

This is an experiment on individual decision-making conducted by academics at Erasmus University and New York University Abu Dhabi. I understand that the entire experiment is expected to last no more than 15 minutes, and that there is no risk from participation beyond that of everyday life.

Your payment for participating will depend on your actions in the experiment. The amount can be up to €60. After the experiment, we will randomly select one in every four participants and pay them their earnings. In addition, we will randomly select one of the four tasks in the experiment to determine your earnings.

No information will be provided to anyone other than the experimenter about the decisions, payment or identity of any participants. Participation is voluntary. You are free to leave the experiment at any time, but if you do you will not be eligible for payment.

This project has been approved by the Institutional Review Board at New York University Abu Dhabi, and the Erasmus Research Institute of Management. We attest that the information is accurate and truthful. If you have any questions about this research, feel free to contact the investigators at nn30@nyu.edu. If you have concerns about the experiment, you can contact the Institutional Review Board at IRBnyuad@nyu.edu.

Next

Task 1 Instructions

You will be presented with a series of three 2-digit numbers (e.g., 11, 22, and 33) which you will have to add (e.g. 11 + 22 + 33 = 66). You will have 90 seconds to add as many sets of numbers as you can.

Your earnings in this task will depend on your **individual performance**. In particular, if Task 1 is the task that is randomly selected to calculate your earnings, you will earn €1 for each correct answer you give. You will earn €0 for incorrect answers. For example, if you have 10 correct answers, your earnings from Task 1 will be €10.

You cannot use a calculator of any kind for answering questions, but you can use your own pen and paper. Please raise your hand if you have any questions. When you click "Next" your 90 seconds for Task 1 will begin.

Next

01 27

Task 1

Sum the numbers and enter the answer into the field below for as many questions as you can.

52 + 59 + 25 =

59 + 59 + 99 =

25 + 94 + 98 =

97 + 38 + 44 =

...

Your score in Task 1: 0 correct answers.

Next

Task 2 Instructions

You will again have 90 seconds to calculate the correct sum of as many sets of three 2-digit numbers as you can. Unlike in Part 1, your earnings in this task will depend on your **compared performance**.

For this task, your performance will be compared to that of three other participants that will be randomly and anonymously selected at the end of the experiment. These participants will either be in this room or in another room participating under similar circumstances.

If Task 2 is selected for payment, there are three possible outcomes:

- If you are the one with the highest number of correct answers in your group, you will earn €4 for each correct answer (if you are selected for payment). For example, if you have 10 correct answers and the other participants have 9, 8 and 7, you will earn 10 x €4 = €40. The others will earn €0.
- If another participant in your group has the highest number of correct answers, you will earn €0.
- In the case of a tie, one participant from those with the highest number of correct answers will be randomly selected to earn €4 per correct answer, while the others will earn €0.

-

Please raise your hand if you have any questions. When you click "Next" your 90 seconds for Task 2 will begin.

0127

Task 2

Sum the numbers and enter the answer into the field below for as many questions as you can.

83 + 79 + 63 =

90 + 47 + 89 =

28 + 75 + 33 =

72 + 14 + 83 =

…

Your score in Task 2: 0 correct answers.

Next

Task 3 Instructions

As in the previous tasks, you will have 90 seconds to calculate the correct sum of as many sets of three 2-digit numbers as you can.

Before this, however, you will have to choose how your earnings will be determined if Task 3 is selected for payment.

If you choose **individual performance** you will earn €1 for each correct answer (as in Task 1).

If you choose **compared performance** (as in Task 2) your performance will be compared to the performance of the other three participants in your group *in Task 2*. You will earn €4 for each correct answer if you are the individual with the highest number of correct answers. Otherwise, you will earn €0. In the case of a tie, one of the participants with the highest number of correct answer will be randomly chosen to earn €4 per correct answer.

Select your choice of earnings method:

- ○ **Individual performance**
- ○ **Compared performance**

When you click "Next", the 90 seconds for Task 3 will begin.

0126

Task 3

Sum the numbers and enter the answer into the field below for as many questions as you can.

90 + 80 + 55 =

40 + 26 + 61 =

20 + 19 + 93 =

36 + 77 + 71 =

...

Your score in Task 3: 0 correct answers.

Next

Task 4 Instructions

Consider the following situation: You are given six options, listed as options A, B, C, D, E and F. Your task is to choose *one* of the six options. If this task is selected for payment, you will be paid based on which option you choose. Each option involves a simple lottery with two possible outcomes that are equally likely to occur. The six options you must choose from are:

A) 50% to receive **€10** and 50% to receive **€10**
B) 50% to receive **€14** and 50% to receive **€8**
C) 50% to receive **€18** and 50% to receive **€6**
D) 50% to receive **€22** and 50% to receive **€4**
E) 50% to receive **€26** and 50% to receive **€2**
F) 50% to receive **€28** and 50% to receive **€0**

In each option, notice that there are two possible outcomes. To determine which of the two outcomes you receive, if this task is selected for payment, after the experiment, we will roll a six-sided die. If the die roll is 1, 2 or 3, then you will receive the first outcome in the option you choose. If the die roll is 4, 5 or 6 then you will receive the second outcome in the option you choose.

I choose:

- ○ Option A
- ○ Option B
- ○ Option C
- ○ Option D
- ○ Option E
- ○ Option F

Short Questionnaire
(You must answer all questions to be eligible for payment)

What is your gender?

- ○ Male
- ○ Female

I believe there is a high chance I will be the one who solved the highest number of correct sums in my group in Task 2 (compared performance).

- ○ Agree
- ○ Disagree

Please tell us how much time you have **for leisure activities** (e.g., sleeping, going out with friends, watching TV, social media, etc.) on an average day of the working week (i.e., not on weekends):

I never have enough time	More often than not, I don't have enough time	I usually have just enough time	More often than not, I have enough time	I always have enough time
○	○	○	○	○

Considering all your fellow students, on a scale from 1 to 10, which income group do you believe you are in? (1 indicates the lowest income group and 10 the highest income group)

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

Contact Details
In order to receive your payment, you have to fill out the information below. This information will be used to notify you about your payment only and will not be used in any way in the analysis of the data.

Note that on September 13th there will be a public drawing to determine who will receive payment for the experiment. This will be done in office H12-29 at 14.00 hours. You are welcome to attend. All participants will receive a private email about their earnings.

Please enter your Erasmus email account (must end with @student.eur.nl):

Please confirm your Erasmus email account:

In case you have another email address you use more frequently, please provide it in the field below:

Next

This screen was randomly set as either the final screen of the experiment, or the first screen of the experiment.

Would you like to be invited to take part in experiments in economics?
We take this opportunity to inform you about the possiblity to participate in future experiments at Erasmus University. Individuals are paid for their participation and **no experience in economics is necessary**. While payments depend on participants' decisions and the type of the experiment, *average* payments tend to be €15 per hour. Furthermore, by participating you will be helping promote research at the University.

Individuals who sign up for inclusion in the database are *randomly* invited on a monthly basis. **This means that each individual in the database will have the same chance to be invited to participate in experiments.** If at any point you wish to stop receiving invitations to experiments, you can do this easily by sending an email to admin@ese-econlab.nl.

- Yes, please send me an e-mail with information for signing up.
- No, please don't send me information.
- I am already registered.

Back

Next

We thank you for your time spent taking this survey.
Your response has been recorded.

3. Post with information about earnings from the experiment

As promised to students, the week following the classroom experiment, we uploaded a post on the course's website with information about earnings from the experiment.

Title of post: Earnings in the experiment

Main text of post: Dear student,

Here you can find an Excel document that contains the student numbers of the 25% of participants that was randomly selected for payment for last week's in class experiment.

If your student number does not appear in the list, this means that, unfortunately, you are not among the 25% of participants randomly selected for payment.

If your student number does appear on the list, then 1 of the 4 tasks of the experiments is randomly selected for you. It is still possible that your earnings are 0 Euro. If this is the case, then it means that you were not the person with the greatest score in Tasks 2 or 3. Or, it is possible that the coin flip caused your earnings to be 0 Euro in Task 4.

IMPORTANT

In case you have positive earnings, you can collect them on FRIDAY SEPTEMBER 21, between 10.00 – 14.00 hours in room H10-30.

Payment is done by bank account, so make sure that you have your bank account number and social security number (BSN) ready. We will ask you to fill out a short payment form, and show your student card and id card to verify your earnings.

4. Email invitation to join volunteer database for lab experiments

Two weeks after the final tutorial, the following email was sent to all those in our population that expressed an interest in participating in future lab experiments.

Dear student,

We are writing to you because in the first week of block 1 in Academic Skills/Academische Vaardigheden/Guidance/Mentoraat you expressed interest in participating in economics experiments.

If you would like to receive invitations to take part in upcoming experiments you need to register on our website

http://ese-econlab.nl

Make sure that you fully complete the registration form and that you sign up using your Erasmus address.

After you register, you will then start receiving invitations to participate in upcoming experiments.

The Economics Department runs experiments regularly for research purposes. Individuals are paid for their participation and no experience in economics is necessary. While payments depend on participants' decisions and the type of the experiment, *average* payments tend to be €15 per hour. Furthermore, by participating you will be helping promote research at the University.

If at any point you wish to stop receiving invitations to experiments, you can do this easily by sending an email to admin@ese-econlab.nl.

Sincerely,

Dr. Georg Granic and Dr. Chen Li
Behavioral Economics
Erasmus University Rotterdam – Erasmus School of Economics

5. Email invitation to participate in a lab experiment

Eleven weeks after the classroom experiment, the following email was sent to all those in our population that registered in the database of future volunteers.

Dear (first name, last name),

We would like to invite you to participate in an experiment.

Topic of the experiment: decision making

Expected length: 15-20 minutes

Payment: 7 euro show up fee, plus any earnings you make in the experiment

The sessions are scheduled for the following times (MONTH/DAY/YEAR and TIME):

- Monday 26th of November 10.00 hours
- Monday 26th of November 12.00 hours
- Monday 26th of November 14.00 hours
- Monday 26th of November 16.00 hours
- Tuesday 27th of November 10.00 hours
- Tuesday 27th of November 12.00 hours
- Tuesday 27th of November 14.00 hours
- Tuesday 27th of November 16.00 hours
- Wednesday 28th of November 10.00 hours
- Wednesday 28th of November 12.00 hours
- Wednesday 28th of November 14.00 hours
- Wednesday 28th of November 16.00 hours
- Thursday 29th of November 10.00 hours
- Thursday 29th of November 12.00 hours
- Thursday 29th of November 14.00 hours
- Thursday 29th of November 16.00 hours
- Friday 30th of November 10.00 hours
- Friday 30th of November 12.00 hours
- Friday 30th of November 14.00 hours
- Friday 30th of November 16.00 hours

If you want to participate, you can register by clicking on the following link:

#link#

(If you cannot click on the link, copy it to the clipboard by selecting it, rightclick and choosing "Copy", and then paste it into the address line in your browser by right clicking there and choosing "Paste".)

Kind regards,

The experimenters of the ESE-econlab