Gender differences in willingness to compete: The role of public observability

Buser, Thomas; Ranehill, Eva; Veldhuizen, Roel van

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Journal of Economic Psychology 83 (2021) 102366

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Gender differences in willingness to compete: The role of public observability☆

Thomas Buser [a], Eva Ranehill [b], Roel van Veldhuizen [c,*]

[a] University of Amsterdam and Tinbergen Institute, School of Economics, Roetersstraat 11, 1018WB Amsterdam, the Netherlands
[b] University of Gothenburg, Vasagatan 1, 405 30 Gothenburg, Sweden
[c] Lund University, Scheelevägen 15B, 223 63 Lund, Sweden

ARTICLE INFO

JEL codes:
C91
D03
J16

Keywords:
Gender differences
Competitiveness
Social image
Experiment

ABSTRACT

A recent literature emphasizes that gender differences in the labor market may in part be driven by a gender gap in willingness to compete. However, whereas experiments in this literature typically investigate willingness to compete in private environments, real world competitions often have a more public nature, which introduces potential social image concerns. If such image concerns are important, and men and women differ in the degree to which they want to be seen as competitive, making tournament entry decisions publicly observable may further exacerbate the gender gap. We test this prediction using a laboratory experiment (N = 784) that varies the degree to which the decision to compete, and its outcome, is publicly observable. We find that public observability does not alter the magnitude of the gender gap in willingness to compete in an economically or statistically significant way.

1. Introduction

Gender differences in labor market outcomes remain a primary policy concern. Women have lower labor market participation, are underrepresented in positions of power, and earn lower wages even when occupying similar positions as men. Economists have proposed a number of reasons that may explain these differences, including discrimination, family constraints, and preferences for certain occupations (see Goldin, 2014, for an overview).

More recently, a large literature in experimental economics, starting with Gneezy, Niederle and Rustichini (2003), documents that men and women may differ in the way they respond to competitive environments (see e.g. Niederle 2016, for an overview of this literature). In particular, a large number of studies starting with Niederle and Vesterlund (2007) demonstrate that men tend to be more willing to seek out competitive environments. This gender difference in willingness to compete may at least partially explain why women are underrepresented in top positions, since obtaining highly remunerated and prestigious jobs often requires competing for

☆ We thank Kai Barron, Michel Maréchal, Johanna Mollerstrom, Renke Schmacker, Simeon Schudy, Justin Valasek and Roberto Weber for valuable comments. Thomas Buser gratefully acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a personal Veni grant. Eva Ranehill thanks the Swiss National Science Foundation (grant number 100010_149451) and the Jan Wallander and Tom Hedelius Foundation for generous financial support. Roel van Veldhuizen gratefully acknowledges financial support from the German Science Foundation (DFG) through collaborative research center CRC TRR 190. We are also grateful to Nina Bonge, Felix Bönisch, Friederike Heiny, Ornella von Matt, and Florian Wiek for their help in running the experiment.

* Corresponding author.
E-mail addresses: t.buser@uva.nl (T. Buser), eva.ranehill@economics.gu.se (E. Ranehill), roel.van_veldhuizen@nek.lu.se (R. van Veldhuizen).

https://doi.org/10.1016/j.joep.2021.102366
Received 24 February 2020; Received in revised form 30 November 2020; Accepted 13 January 2021
Available online 23 January 2021

them. Applying for promotions or new positions, and bargaining for higher wages, can similarly be thought of as competitive activities.

Most studies in this area rely on laboratory experiments that build on a paradigm introduced by Niederle and Vesterlund (2007). In these experiments, participants perform a real-effort task and choose between individual piece-rate compensation and a winner-take-all tournament. These studies typically find that, conditional on performance, men are substantially more likely to choose the tournament, particularly in male-stereotyped tasks such as solving math problems. More recently, several studies have shown that experimental measures of willingness to compete predict the selection into more challenging educational tracks (Buser, Niederle, and Oosterbeek 2014; Zhang 2013; Buser, Peter, and Wolter, 2017a, 2017b) and correlate with labor market outcomes (Buser, Geijtenbeek, & Plug, 2018; Reuben, Sapienza, & Zingales, 2015; Reuben, Wiswall, & Zafar, 2015; Buser, Niederle, & Oosterbeek, 2020). The laboratory results have also been replicated in field experiments using real labor market choices (Flory, Leibbrandt and List 2015; Samek 2019). Kleinjans (2009) further provides survey evidence suggesting that competitive preferences may increase educational outcomes and explain gender segregation in fields of study.

These studies share the feature that the decision of whether or not to compete is typically made in a private environment, where only the experimenter directly observes a participant's choices. In relevant career settings, however, the choice of whether or not to enter a competition—such as, for example, taking a competitive exam or participating in a promotion contest at work—is usually observable to others.

This is important, because public observability may increase the importance of social image concerns, and particularly concerns about gender-related norms regarding the appropriateness of competitive behavior (Eagly 1987). For example, women may shy away from *public* displays of competitive behavior if social norms prescribe that an aggressive or competitive attitude is inappropriate for women, or if such displays are at odds with gender stereotypes and self-perceived gender identity (Akerlof and Kranton 2000; Rudman and Glick 1999; Rudman and Glick 2001; Bénabou and Tirole 2006; Ridgeway 2009; Ridgeway and Correl, 2004). For example, it is possible that being overtly competitive is seen as a form of self-promotion at odds with norms about female modesty (e.g., Rudman 1998; Moss-Racusin and Rudman 2010). Men, on the other hand, may be encouraged to engage in competitive behaviors when the decision is publicly observable since gender stereotypes prescribe more agentic and competitive behavior on behalf of men (Phelan and Rudman 2010).[1]

This line of reasoning is supported by a growing body of literature that provides evidence that the gender gap in willingness to engage in competitive behaviors depends on social and contextual factors related to gender stereotypes. Notably, the large majority of studies documenting the gender gap in willingness to compete explore behavior in stereotypically male real-effort tasks. Studies using gender-neutral or stereotypically female tasks sometimes find substantially smaller gender gaps.[2]

Focusing individuals' attention on gender identity has also been shown to impact competitive and risky behavior. Cadsby et al. (2013) explore the gender gap in willingness to compete among MBA students who are either primed with their gender and family identity, or with their professional identity, and find that women primed with their professional identity are significantly more willing to compete than their female peers primed with a gender and family identity. Similarly, several studies document smaller or even reverse gender gaps in willingness to compete in matriarchal societies, where it is common for women to occupy prestigious positions (Gneezy, Leonard and List 2009, Andersen et al. 2013).

Taken together, these studies suggest that observability may increase concerns for social image and compliance with gender-role expectations and impact the gender gap in willingness to compete.[3] This is further supported by empirical evidence from related settings. For example, Bursztyn et al. (2017) find that single female MBAs express significantly lower professional ambition and leadership tendencies when they expect their answers to be available to their peers than when they answered the same questions privately. A number of studies also indicate that women who display gender incongruent behavior suffer backlash and adapt their behavior accordingly. For example, Bowles, Babcock and Lai (2007) show that women are more penalized than men for initiating negotiations. Women take this into account and are less likely than men to initiate negotiations on behalf of themselves, but not on behalf of others (Amanatullah and Morris 2010, Amanatullah and Tinsley 2013).[4]

If social image concerns are important, and public observability increases their effect, we would expect the gender gap in willingness to compete to further increase when decisions are publicly observable. Moreover, public observability of the competition *outcome* may also matter for social image. While engaging in competitions may correspond to existing male gender norms and therefore enhance male social image, a publicly observable loss may undermine it. Further, observability of the outcome might lead people to

[1] While we here emphasize the impact of observability on women's competitive choices, there is also evidence that men displaying counter-stereotypical behavior experience backlash (see Phelan and Rudman 2010 for an overview of this literature).

[2] Studies that find substantially reduced gaps include Kamas and Preston (2009), Grosse, Riener and Dertwinkel-Kalt (2014), Shurchkov (2012), Dreber, von Essen and Ranehill (2014), Boschini et al. (2019) and Apicella and Dreber (2015). Studies that find significant gender gaps also in stereotypically female or gender-neutral tasks include Cárdenas et al. (2012, in one of two countries explored), Wozniak, Harbaugh and Mayr (2014), Sutter and Glätzle-Rützler (2014) and Buser et al., 2017b.

[3] For the purpose of this study, we consider observability of choices to generate concerns for how the decision maker is perceived to others, generally. Observability may imply several, not necessarily mutually exclusive, mechanisms, such as signaling a value associated with the action itself, caring about adhering to gender stereotypes, or status concerns. For simplicity, we will refer to these different aspects as image concerns.

[4] Even the expressions of attitudes and beliefs that depart from gender-role expectations may come with a social image cost. Consistent with a penalty for gender incongruent behavior, Heatherington et al. (1993), and Ludwig et al. (2016) find that women are more modest than men when having to state their own abilities in public, but not in private. Research also demonstrates that a large gap in social confidence emerges during adolescence. Alan et al. (2016) find that female students are less willing to perform a more difficult, higher reward version of a numeric real effort task when they have to perform the task in public in front of the class compared to when their decision and performance are private.

think through their decision more carefully. Since in professional settings, the outcome of a competition is often observable to others, it is important to also investigate the effect of public observability of outcomes on the gender gap in willingness to compete.

Hence, we hypothesize that public observability will enhance the gender difference in willingness to compete by increasing the proportion of men and decreasing the proportion of women who choose the tournament. We further hypothesize that publicly announcing the outcome of a competition on top of the decision to compete will attenuate men's willingness to compete relative to a situation where only the decision to compete, but not the outcome of the competition, is publicly observable.

We test these hypotheses using an incentivized experiment that closely approximates the paradigm introduced by Niederle and Vesterlund (2007), except that we vary the degree to which an individual's competitive choices and outcomes are publicly observable. The approach of varying the degree of public observability in an experimental setting has previously been used successfully to generate image concerns (Ewers and Zimmermann 2015).[5] Based on the evidence cited above, we expect the gender gap in willingness to compete to be further exacerbated when choices are publicly observable in comparison to when they are private, but this effect to be attenuated when outcomes are also made public.

In our experiment, all participants (N = 784) perform an arithmetic task three times. As in the Niederle and Vesterlund (2007) design, participants first perform the task under an individual piece rate compensation scheme. This round serves as a baseline measure of participants' ability. In the second round, all participants perform under a competitive compensation scheme. They are randomly matched with another participant and gain twice their piece rate payment if they perform better than their opponent and nothing otherwise. This round serves as a measure of the participants' ability to solve exercises under competitive incentives. In the third and final round, participants choose whether to solve exercises according to the incentives in round 1 or round 2, and this binary choice serves as our measure of a participant's willingness to compete.[6]

We implemented four different conditions that varied the degree of public observability using a *between-subject* design. In the *Public Choice* condition, before the start of round 3, but *after* making their binding choice of whether to compete or not, participants had to stand up, introduce themselves and publicly announce their decision. In the *Control* condition, their decision remained private. To separate the effect of observability of the decision from potential gender differences in the aversion to speak in front of others (De Paola et al. 2020), participants in the *Control* condition still had to stand up and introduce themselves. Comparing these conditions allows us to investigate whether making the choice to compete publicly observable increases the gender gap in willingness to compete. In the *Public Outcome* condition, those participants who opted for competition had to publicly announce not only their choice but, at the end of the third round, also the outcome of the competition. This allows us to test whether making the outcome of a competition observable to others attenuates the gender gap in willingness to compete. Finally, to facilitate comparison with previous research, we implemented a *Private* condition, which omits any public interaction, and which is therefore close to the standard design of Niederle and Vesterlund (2007).

We do not find support for our hypotheses. The gender difference in willingness to compete does not differ in a statistically or economically significant way with the observability of the choice to compete in the context explored here. We also find no evidence that making the outcome of the competition observable as well has an impact on the gender difference. Our large sample size (N = 784 across two waves of data collection) and the power analysis in Appendix A, indicate that this null result is not driven by a lack of power.[7]

Why does making choices publicly observable not increase the gender gap in willingness to compete? To explore potential mechanisms of why observability may increase the gender gap, we also included a vignette task after the main part of our experiment. The vignette was based on Bowles, Babcock and Lai (2007) and elicited participants' attitudes towards competitive men and women. We incentivized the vignette using the technique developed by Krupka and Weber (2013). Since our hypothesis was that social image concerns influence men's and women's willingness to compete in opposite directions, it is informative to know more about how participants view competitive behavior in men and women and, in particular, how they think it is viewed by others. Overall, and consistent with our null result, we find no evidence of a gender-specific norm regarding the appropriateness of competitive behavior. This may at least partially explain why public observability does not affect the gender gap in willingness to compete in our sample. Instead, both competitive men and competitive women are rated less favorably than their non-competitive counterparts.

Two recent studies look into the effect of social status ranking on the gender difference in performance and investigate whether there is a gender difference in status-ranking aversion. Schram, Brandts and Gërxhani (2018) show that male participants perform better and female participants perform worse when their outcome on an experimental task is announced to another participant who can compare their outcome to several other performers, thereby introducing social image concerns. Brandts, Gërxhani and Schram (2020) show that in a context where incentives are non-competitive, there is no strong gender difference in the willingness to be ranked by others. The exception is that when the ranker is known to be male and the choice is imposed on all other participants, men are more likely than women to want to be ranked. While these studies are clearly related to our study, there are some important differences. In particular, in the studies by Schram, Brandts and Gërxhani (2018) and Brandts, Gërxhani and Schram (2020) the outcome of the competition is observed by only one other person (as opposed to everybody present in the lab) and payments do not depend on the

[5] Ewers and Zimmermann (2015) study whether people report more optimistic self-assessments when reporting to an audience than in the absence of one. They also briefly discuss gender differences, finding that women are less confident overall, though this effect does not depend significantly on the presence of an audience. Instead of looking at audience effects on self-assessment, we instead focus on the effect of public observability on the gender gap in tournament entry.

[6] Similar to Ewers and Zimmermann (2015), we implement only a single tournament entry decision in order to reduce repeated game effects.

[7] The power analysis was done before the second wave of data collection and presents power estimates for a number of plausible effect sizes.

ranking. That is, these studies are not interested in whether the gender difference in willingness to compete varies according to public observability but rather in whether there is a gender difference in the willingness (and reaction to) having one's performance compared to the performance of others.

An increasing number of scientific studies find that willingness to compete matters for real life outcomes related to educational choices and labor market outcomes. At the same time a large share of existing studies on the gender gap in competitiveness explore this gap in a specific setting, similar to the one introduced by Niederle and Vesterlund (2007). It is important for researchers, organizations, and policy makers to understand how different aspects of the decision-making context may reinforce, or mitigate, the gender gap in willingness to compete. Our results matter, for example, when deciding on the transparency of the institutions used to elect leaders, or to promote managers. From the perspective of the academic literature on willingness to compete, our results suggest that the standard design in the literature is robust to changing the level of public observability.

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

2. Experimental design

Our experiment was run at the experimental economics laboratory of the Technical University Berlin. Data for the Public Choice, Public Outcome and Control conditions were collected in two waves with 355 participants in June 2016, and 313 participants in April 2018; conditions were randomly allocated across sessions.[8] Data for the Private condition (116 participants) were collected in July 2016, after observing the data of the first wave of the other conditions. For each session, we invited 28 participants (14 men and 14 women), at most 24 of whom could participate; each session had between 16 and 24 actual participants. Participants' average age was 24. 31 percent majored in engineering, 17 percent double-majored in economics and engineering, 12 percent majored in economics, 12 percent in science or math, 11 percent in humanities or social sciences and 17 percent in something else. The experiment was programmed in Ztree (Fischbacher 2007) and participants were recruited using ORSEE (Greiner 2015). Average earnings were 19.49 Euros (including a 5 Euro show-up fee). Table 1 presents the number of male and female participants in our experiment by condition. There were 10 sessions in each of the first three conditions and 5 sessions in the Private condition.

These sample sizes give us sufficient power to detect effect sizes similar to previous work with high probability. For example, given a baseline rate of tournament entry similar to an earlier study run at the same laboratory (van Veldhuizen 2018), our power to detect an effect size of 25 percentage points (that is, a 25 percentage points increase in the gender gap in the Public Choice condition) equals 0.855. This effect size is comparable to the effect of introducing a gender quota observed in Balafoutas and Sutter (2012, 25 pp), and considerably smaller than the effect of culture in Gneezy, Leonard and List (2009, 39 pp) and the effect of a gender quota in Niederle, Segal and Vesterlund (2013, 81 pp). We present more details in Appendix A1, where we also show that our sample size is larger than nearly all previous studies examining treatment differences in the gender difference in willingness to compete. As a result, we are relatively well-equipped to detect even small to moderate treatment effects in the gender difference in willingness to compete.

Participants were randomly assigned to a computer upon entering the laboratory. Each participant received their show-up fee, and was told the experiment consisted of three separate parts, each of which would contribute to the final payment. Instructions for the respective parts were only provided after the previous part had finished. All payments in the experiment were displayed in experimental currency units (ECUs), which were converted to Euros at a rate of 10 ECUs per Euro. All instructions can be found in the online Appendix.

2.1. Part 1: Elicitation of risk preferences

In part one, we elicited participants' risk preferences using the investment game (Gneezy and Potters 1997; Charness and Gneezy 2012). Participants were given an endowment of 20 ECUs and were asked what fraction of their endowment they wished to allocate to a safe option and to a risky investment respectively. The safe option simply stored the endowment until the end of the experiment, whereas the investment returned 2.5 times the invested amount with 50% probability, and zero otherwise. This task has been used by a large number of studies to measure individual risk preferences. The more risk averse the participant, the less she should invest.

We elicit risk preferences because past studies typically find that attitudes towards risk influence individual decisions to compete and explain part of the gender gap. Past studies, starting with Niederle and Vesterlund (2007), have often interpreted the remaining gender gap in willingness to compete after controlling for performance, confidence and risk attitudes as a gender gap in taste for competition. Having a measure of risk preferences enables us to follow the analysis in this literature and study the impact of our public observability on this residual gender difference. We elicited risk preferences at the beginning of the experiment in order to prevent the outcomes of the other parts of the experiment from influencing participants' investment decisions. Participants were not told about the outcome of the investment until the end of the study.

2.2. Part 2: Elicitation of willingness to compete

Part two largely followed the design of Niederle and Vesterlund (2007) and consisted of three rounds. In each round, participants

[8] Fewer women than expected chose to compete in the first wave of our experiment in the Control condition, which resulted in a lower than expected statistical power. The second wave of data collection was implemented to increase power.

Table 1
Participants per condition.

Condition	Control	Public Choice	Public Outcome	Private	Total
Male	116	109	112	59	396
Female	107	115	109	57	388
Total	223	224	221	116	784

had four minutes to work on a task. The task involved solving as many addition problems consisting of five two-digit numbers as possible within the time limit of four minutes.

Performance was incentivized differently in the three rounds. In round one, participants were paid a piece rate, collecting 5 ECUs for each exercise they solved correctly. In round two, participants were paid according to a two-person winner-takes-all tournament. Each participant was compared with a random other participant from the same session. If her performance beat the score of her opponent, she received 10 ECUs per exercise she solved correctly. If her performance was worse, she did not receive any payment. In case of a tie, the computer randomly determined which of the two contestants won the tournament. In round three, participants could choose which of these two incentive schemes to apply to their performance. If a participant chose to compete, her performance was compared with the round three performance of a randomly chosen other participant, independently of whether this other participant had chosen to compete or not. At the end of the experiment, one out of the three rounds was randomly selected for payment. No feedback on relative performance was provided except at the end of round three, when all participants who chose to compete learned whether they won or lost.[9]

We ran four different conditions, which differed only in the degree of public observability in the third round. Participants took part only in one condition. In the *Public Choice* condition, participants were asked to first make their competition choice, and to then publicly announce their chosen incentive to all other participants in the session. Specifically, at the onset of the third round, participants were informed that after they had made their choice of payoff scheme privately on their computer, each of them would be asked to walk up to the front of the lab. There they would say their first name and announce their choice in front of all participants present. Participants were instructed to say "Hi, my name is ⟨name⟩ and I chose to compete" or "Hi, my name is ⟨name⟩ and I chose the piece rate", depending on their decision, and nothing else. The experimenter double-checked that all participants reported their true choice. Only three participants misreported their choice across all relevant conditions. The order in which participants came to the front was determined randomly. The first participant only made his announcement once everybody had registered their choice in the computer. After all participants had introduced themselves and announced their decisions, they were asked to sit down again and the third round proceeded based on their choices. Note that since all tournament entry decisions occurred before the public announcements, we treat each entry decision as one independent observation. This design feature also prevents participants from being influenced by the decisions taken by their peers.

In the *Control* condition, participants went through the same steps as in the Public Choice condition, but without announcing their competitive decision. Specifically, before entering their choice privately on the computer, participants were informed that they would subsequently be asked to come to the front of the lab and say "Hi, my name is ⟨name⟩", and nothing else. This allows us to separate the effect of publicly announcing the choice from the effect of having to go to the front of the lab to introduce oneself in a clean way.

The third condition, *Public Outcome*, was identical to Public Choice, except that we asked participants to stand up a second time after they finished working on the exercises in round three. This time, we asked them to come forward one by one and say "I chose the piece rate" or "I chose to compete and I won/lost the competition", depending on what they chose and the outcome of the competition. As in the previously described conditions, participants were informed about these steps before making the choice of incentive.[10]

To better compare our results to the existing literature, we also ran a fourth condition (the Private condition), in which choices and outcomes were private and participants did not announce their name in public. This allows us to determine whether any form of public statement changes people's willingness to compete. While the first three conditions were randomly allocated over the sessions, the Private condition was added after the first wave of data collection, and conducted three weeks after the initial sessions. We elected to not include this condition in the second wave of experiments because we wanted to maximize the number of participants in the other conditions, which allow us to test our hypotheses.[11]

Past studies have found that beliefs about relative performance (confidence) are an important determinant of individual and gender

[9] Participants were informed prior to their decision that they would receive immediate feedback at the end of round three in case they chose the competition.

[10] We did not ask participants to also report, e.g., their scores and score difference relative to a competitor. This is both because we had no a priori hypothesis that providing this type of information would attenuate the gender gap in willingness to compete, and because doing so would have changed not just the public observability of outcomes, but also the private knowledge of the score difference, which would have reduced comparability relative to the other conditions. We were not aware of the results of Schram et al. (2018) when we ran our first wave of sessions.

[11] All four of our conditions differ from Niederle and Vesterlund (2007) in that (1) participants competed against only one competitor instead of three, (2) they had four minutes for the task instead of five and (3) performance was compared with the round three performance of another participant, instead of round two. Similar changes implemented in previous work (e.g., Niederle, Segal and Vesterlund 2013; Buser, Niederle and Oosterbeek 2014 or Dreber, von Essen and Ranehill 2014) did not substantively affect the gender gap in willingness to compete. Further, competing against the current performance of the competitors is a feature of competition present in real world settings that we considered meaningful to include for the purpose of this study.

differences in willingness to compete (Gillen, Snowberg, & Yariv, 2018; Niederle & Vesterlund, 2011; Van Veldhuizen, 2018).[12] To measure confidence, we asked participants at the very end of Part 2, after the third round of the task, to guess their rank among all participants in their session in Round 2 (the forced tournament). Participants received a bonus of 10 ECUs (1 Euro) at the end of the experiment if their guessed rank was correct.[13] Payoff maximization therefore requires reporting the modal expected rank; the same approach has been used by Niederle and Vesterlund (2007) and others.

2.3. Part 3: Vignettes

For social image concerns to lead to a larger gender gap in the Public Choice condition, it needs to be true that men get positive and women get negative image utility from selecting into a tournament, as would be the case if there is a gender-specific norm regarding the appropriateness of competitive behavior. To study whether this is the case, we used another task to elicit participants' attitudes towards competitive and non-competitive women and men. Specifically, we followed Bowles, Babcock and Lai (2007) and gave participants the CV and some interview notes for a hypothetical candidate for an internship position at a Berlin-based bank. We experimentally varied the gender of the candidate and whether he/she was described as competitive, and then asked participants to rate the candidate on a number of general personality traits.[14] Participants then had to judge, on a seven-point scale, whether the applicant had the skills for the job, whether hiring the applicant would be beneficial for the company, whether working with the applicant would be enjoyable, and how likely they would be to hire the candidate. These questions were not incentivized. In addition, we also asked participants to guess how the candidate was assessed on the same four dimensions by other participants in the session. These guesses were incentivized using the method for eliciting social norms of Krupka and Weber (2013).

If gender stereotypes are important, we would expect participants to rate competitive men more favorably and competitive women less favorably than their non-competitive counterparts. Full instructions are presented in the online Appendix. Participants were randomly allocated to one of the four information conditions within each session.

2.4. Remaining procedures

After everyone had finished part three, we gave participants a brief questionnaire asking them about their gender, height, nationality, field of study, and age. They then received feedback on their income from the three parts of the experiment. Participants were then asked to collect their payment and leave the laboratory.

When studying gender differences, it is important to keep participants blind to the purpose of the study in order to prevent behaviors that arise due to, for example, experimenter demand effects or gender stereotypes. In line with this, participants were not aware that an equal number of participants were invited to each session and were only asked about their gender in the questionnaire at the very end of the experiment.[15] Nevertheless, it is possible that some participants anticipated our focus on gender by virtue of the observed gender-balance in a given session. What works against this explanation, however, is that many sessions were not exactly gender-balanced due to more women or men showing up for the session. In addition, gender-balanced sessions are fairly common in our laboratory also in experiments that do not study gender differences. As a result, we do not expect that our results are affected by participants being able to guess that our focus was on gender differences.

3. Results

We begin by presenting descriptive statistics of behavior in the risk and competition parts of the experiment and presenting treatment comparisons of the gender gap in willingness to compete. The vignette results will be discussed at the end of the section. Throughout this section we will use chi-square tests for binary outcome variables and *t*-tests for non-binary ones, except when otherwise indicated.

[12] Other potential drivers of the gender gap in willingness to compete include distributional preferences (Balafoutas et al. 2012; Dasgupta et al., 2019), ambiguity aversion (Balafoutas et al. 2017) and performance (Gneezy et al. 2003; Ivanova-Stenzel and Kübler 2011). We only elicit measures for risk preferences and confidence because these variables are typically thought to be the main determinants of willingness to compete (along with a preference for competition; see Niederle, 2016).

[13] In case of ties, tied participants were randomly assigned a rank from the set of appropriate ranks. For example, two participants tied for the 11th and 12th rank in the session would be assigned each of the two ranks with a probability of 0.5.

[14] We manipulated the candidate's competitiveness by adding the following sentence to her interview notes: "(S)he also said that (s)he found competitive environments stimulating, and asked if the bank provides a ranking of the interns hired for the year's summer internship program, after the program is completed."

[15] When signing up for our experiment in ORSEE, participants could only see whether they were allowed to sign up for an experiment, not what the inclusion restrictions were. In addition, we followed the standard procedure in our laboratory to only invite a subset of all participants to a particular session. This made it impossible for individual participants to know whether they were invited to a particular session due to random chance, or due to some other criterion like gender.

3.1. Main results

Table 2 presents average choices and outcomes across the three main conditions, separately by gender.[16] In line with previous research, men are more likely to choose the tournament scheme in round 3 (54 percent versus 22 percent). Men also score significantly higher than women in the two first rounds of the task, under piece rate and under competitive incentives. Further, also in line with previous research (including Ewers and Zimmermann, 2015 and most of the literature on gender and tournament entry), men are more confident than women. To make guessed rank comparable across sessions with different numbers of participants, we translate guessed rank into the implied guessed likelihood of winning against a randomly selected opponent.[17] On average, men see themselves as 14 percentage points more likely to win. This gender difference in confidence remains significant in a regression with guessed likelihood of winning as the dependent variable, controlling for gender and actual probability of winning, and equals 11 percentage points ($p < 0.001$). Men are also significantly more risk taking, investing an average of 13.5 ECUs (out of 20 ECUs) in the investment task while women invest 10.8 ECUs. Finally, note that both genders performed better in round 2 than in round 1; previous studies that alternated the order of these two rounds suggest that this is likely due to a learning effect (see e.g., Wozniak et al. 2016).

For obvious design reasons, conditions were randomized across sessions, increasing the chance of differences in the subject pool across conditions. Table B1 in the appendix shows descriptive statistics by experimental condition, showing that none of the background characteristics and pre-treatment experimental choices vary significantly across conditions. Moreover, data collection occurred in two waves with nearly two years in between. Table B2 in the appendix shows descriptive statistics by wave, showing that none of the background characteristics and pre-treatment experimental choices vary significantly across waves.

Turning to our main results, Fig. 1 shows the proportion of participants who choose the competition by gender and condition. For men, the results appear somewhat consistent with our hypothesis. Whereas only 49 percent of men choose to compete in the Control condition, this fraction increases to 60 percent when choices are publicly observable ($p = 0.09$, chi-square test). Similarly, making outcomes publicly observable decreases male willingness to compete from 60 to 52 percent ($p = 0.24$, chi-square test). However, while in the hypothesized direction, neither change is statistically significant.

The impact of treatment on women's behavior is even smaller. Specifically, the proportion of women who choose competition is nearly identical across conditions (20 percent in Control and 23 percent in Public Choice and Public Outcome; none of these differences are significant ($p = 0.59$ comparing behavior in the Control condition to the Public Choice condition, and $p = 0.95$ comparing Public Choice to Public Outcome; chi-square tests)).

Taken together, these results also imply that the gender gap varies across conditions in the predicted way but differences are modest and not significant. The gender gap in tournament entry is 30 percentage points in the Control condition ($p < 0.001$, chi-square test), 38 percentage points in the Public Choice condition ($p < 0.001$, chi-square test) and 30 percentage points in the Public Outcome condition ($p < 0.001$, chi-square test). Neither the difference between Control and Public Choice ($p = 0.33$, difference-in-difference test) nor the difference between Public Choice and Public Outcome ($p = 0.35$, difference-in-difference test) is statistically significant.[18]

Before moving on to the regression analysis, we briefly compare the results of the Control condition to the Private condition, in which choices were private and participants did not announce their name in public. In the Private condition, our point estimate suggests that both men and women are somewhat more likely to compete (12 percentage points for men and 8 percentage points for women), though the difference in tournament entry is not significantly different from zero ($p = 0.166$, chi-square test). More importantly, the gender gap is very similar to the gap in the Control condition. These results therefore suggest that having to stand up and announce one's name does not appear to have a meaningful effect on the gender gap in tournament entry.

Table 3 presents results from regressions that estimate the treatment effects for male and female participants and the change in the gender gap across conditions controlling for various sets of controls. In columns 1 to 3, we first estimate the overall gender gap across the three main conditions with various controls. Given that men have significantly higher scores on the task in our experiment, it is useful to adjust our estimates for gender differences in performance. We do this in all columns by controlling for performance in Round 2. Column 2 further introduces controls for risk preferences (measured as participants' chosen investment level in the investment task) and demographic controls for the participants' age, major, and study year; column 3 also controls for confidence.[19]

Conditional on performance, women are 30 percentage points less likely to choose competition compared to men (column 1) across the three main conditions. Further, in columns 2 and 3, we show that gender differences in risk attitudes, demographics, and confidence can at least partially explain this gender gap in compensation scheme choice. Conditional on these control variables, women are 19 percentage points less likely to choose competition (column 3). These findings are in line with the literature on gender and competitiveness (see, e.g., Van Veldhuizen, 2018).

In columns 4 to 6, we show the effects of our experimental conditions. Here, we regress a competition entry dummy on condition dummies, gender and the interaction of the two. Controlling for performance, the effect of publicly announcing the competitive choice

[16] Descriptive statistics for the Private condition are similar and available upon request.

[17] Guessed chance of winning is equal to (N-Guessed rank)/(N-1).

[18] The p-values of these difference-in-difference tests are from regressions of a competition dummy on treatment dummies, a gender dummy, and their interaction.

[19] The reason that we add confidence separately in a third specification is that we elicited confidence (measured as the guessed chance of winning in Round 2) after the condition had been revealed. This means that differences in confidence across conditions could reflect treatment effects (for example if having to publicly announce their decision makes some participants less confident) rather than underlying differences in confidence between participants.

Table 2
Descriptive statistics by gender (Pooled sample).

	(1) Men	(2) Women	(3) Gender difference:
Competing	0.540	0.218	0.323***
	(0.499)	(0.413)	(0.035)
Score round 1 (Piece Rate)	7.608	6.855	0.753**
	(3.159)	(2.803)	(0.231)
Score round 2 (Tournament)	8.691	7.988	0.703**
	(3.522)	(3.026)	(0.254)
Chance of winning (round 2)	0.527	0.473	0.054*
	(0.303)	(0.294)	(0.023)
Guessed chance of winning	0.646	0.511	0.135***
	(0.228)	(0.245)	(0.018)
Investment (risk)	13.454	10.758	2.696***
	(6.392)	(5.244)	(0.453)
Observations	337	331	668

Notes: Columns 1 and 2 show averages across the three main conditions (standard deviations in parentheses). Column 3 shows the gender difference (standard errors in parentheses). Investment is the amount invested in the part 1 investment task in ECU. Guessed chance of winning is equal to (N-Guessed rank)/(N-1). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; significance levels are from *t*-tests for the gender difference (chi-square test in case of competing).

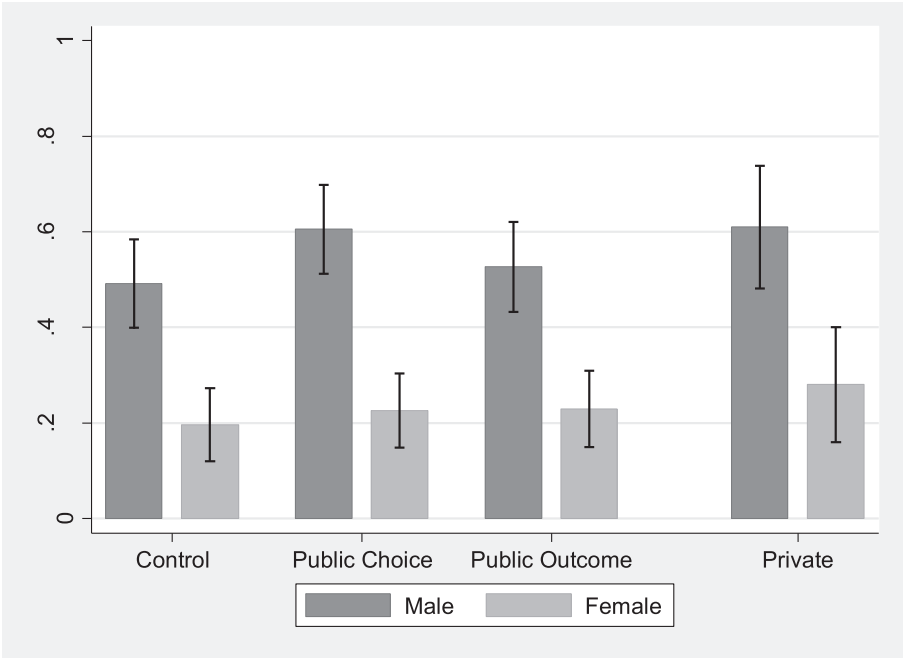

Fig. 1. Proportion choosing competition by condition and gender. Notes: The bars represent the proportion of participants who choose the competitive payment scheme over the piece-rate payment scheme. The error bars represent 95% confidence intervals.

(Public Choice) is 10 percentage points for men, which is significant at the 5-percent level (column 4). Men's competitive choices in the Public Outcome condition are virtually identical to those in the Control condition (but also not significantly different from the Public Choice condition, $p = 0.23$). After adding controls for risk preferences and demographic background variables (column 5) and confidence (column 6) the effect of the Public Choice condition for men is no longer statistically significant.

As with the non-parametric tests, the effects of the conditions on women's behavior are small and not significant and in fact now run opposite to the hypothesized directions. Across the different conditions and specifications, women's competitive choices change by at most 6 percentage points, and none of these changes are close to being statistically significant.

Our two hypotheses are based on interactions of the effects of the experimental conditions and gender. Our first hypothesis is that public observability of the choice increases the gender difference in willingness to compete. The observed gender difference increases by 5.3 percentage points (when controlling only for performance in column 4) but this difference is modest and not statistically significant. Our second hypothesis is that public observability of the outcome reduces the gender gap relative to when only the choice is observable. Relative to the control condition, the observed gender difference decreases by 3.2 percentage points (when controlling

Table 3
Regression results.

	(1)	(2)	(3)	(4)	(5)	(6)
Female	−0.298***	−0.251***	−0.188***	−0.293***	−0.242***	−0.181***
	(0.031)	(0.032)	(0.033)	(0.030)	(0.032)	(0.042)
Public Choice (T1)				0.099*	0.100	0.084
				(0.045)	(0.051)	(0.045)
Public Outcome (T2)				0.006	0.011	0.004
				(0.066)	(0.061)	(0.059)
Public Choice (T1) * Female				−0.053	−0.065	−0.064
				(0.059)	(0.066)	(0.067)
Public Outcome (T2)*Female				0.032	0.037	0.038
				(0.076)	(0.075)	(0.076)
Investment (risk)		0.013***	0.013***		0.013***	0.013***
		(0.003)	(0.003)		(0.003)	(0.003)
Guessed chance of winning			0.659***			0.654***
			(0.080)			(0.079)
Round 2 score	√	√	√	√	√	√
Age, major, year		√	√		√	√
T1-T2 (men)				0.093	0.088	0.080
P-val.(T1-T2) (men)				0.228	0.232	0.238
T1-T2 (women)				0.008	−0.013	−0.023
P-val.(T1-T2) (women)				0.873	0.777	0.642
R-squared	0.165	0.209	0.279	0.170	0.214	0.283
N	668	668	668	668	668	668

Notes: The table reports results from OLS regressions of a dummy for choosing the competitive payment scheme on gender and condition dummies. Round 2 score is the number of solved exercises in Round 2, Lottery is the chosen investment level in the investment task. Age is measured in years, and study major is a set of dummy variables for 8 different fields of study. Confidence is the guessed performance rank in Round 2. Standard errors in parentheses are clustered at the session level. T1-T2 is the difference between the coefficients for the Public Choice and Public Outcome conditions, the p-value tests whether this difference is significantly different from zero. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

only for performance in column 4), meaning that the gender gap shrinks by 8.5 percentage points relative to the public outcome condition. Again, this difference is not statistically significant. Adding controls in columns 5 and 6 does not alter these results. We also obtain very similar results if we analyze our data separately for each wave.

Why does public observability not have a greater effect on the gender gap in tournament entry? One possibility is that public observability makes participants consider their decisions and the associated outcomes more carefully. If true, one possibility is that more deliberation increases the quality of decision-making, increasing the fraction of participants choosing the payoff maximizing payment scheme. However, we find no indication that participants make better decisions in the Public Choice and Public Outcome conditions relative to control. In particular, neither the fraction of participants choosing the payoff maximizing payment scheme based on round 2 performance (62% in Control, 67% in Public Choice, 59% in Public Outcome) nor the correlation between performance and tournament entry (0.30 in Control, 0.33 in Public Choice, 0.16 in Public Outcome) differed significantly between Control and either of the two observability conditions.[20] Hence, we find no indication that our null result is driven by participants making improved decisions in the two conditions with public observability.

In Tables B3 and B4 in the appendix, we examine whether treatment differences vary depending on participants' relative ability by splitting the sample into the top and bottom 50 percent as defined by performance in Round 2. This analysis is exploratory in the sense that we did not plan to run it before seeing the results. We think this split is interesting for two reasons. First, it is important to know whether the conditions affect participants who, in expectation, financially profit from competing (top half) or those who financially lose from competing (bottom half). Second, because a large majority of high-performing men (72 percent) and a small minority of low-performing women (15 percent) compete in the Control condition, there is more room for our hypothesized treatment effects (positive for men and negative for women) in the subsamples of low-performing men and high-performing women.

Table B3 shows that the effect of the Public Choice condition on men indeed appears to be strongest for low performers. The point estimates for the Public Outcome condition (relative to Control) are very similar, suggesting that having to publicly announce the outcome of the competition does not diminish the effect. For this subsample, the Public Choice condition therefore increases the gender difference by 18–19 percentage points, depending on which controls are included; this difference is significant in all three specifications. However, given the exploratory nature of this analysis, these results should be interpreted with some caution. In Table B4, we look at high performers. Contrary to our expectations, we find no evidence that high performing women are less likely to compete in the Public Choice condition. If anything, our point estimate suggests that high-performing women are actually *more* likely to compete when they announce their choices in public.

In summary, we find no strong evidence to support our main hypothesis that public observability increases the gender difference in

[20] For the former measure, these results are based on chi-square tests. For the latter, they are based on the interaction terms in a linear regression of tournament entry on round 2 score and treatment dummies. The p-values for the differences between Public Choice and Public Outcome are p=0.060 and p=0.054 respectively.

willingness to compete. We also find no evidence for our second hypothesis that public observability of the competition outcomes decreases the gender gap.

3.2. Vignettes

In Part 3, we implemented a vignette study to determine whether people judge competitiveness in professional settings differently in women than in men. The results are reported in Table 4. Overall, participants rate both competitive men and competitive women as less enjoyable to work with, judge them less beneficial to have as employees, and state it less likely that they would hire them than less competitive candidates. These results are similar both when we ask participants to answer these questions themselves (without incentives) and when we incentivize them to guess the modal response made by other participants. In other words, participants dislike, and expect other participants to dislike, competitive people of both genders. Thus, consistent with our main results we find no evidence of a more pronounced backlash against competitive women than against competitive men.[21,22]

4. Conclusion

Our experiment examines whether the well-documented gender gap in willingness to compete increases when competitive decisions are observable to others, as compared to when they are private. If an aggressive and competitive attitude is incongruent with gender norms and expectations about female behavior, public observability may make women less willing to compete due to concerns about social image and a desire to conform to gender roles (Akerlof and Kranton 2000). Similarly, men may become even more likely to compete when their decision is public in order to conform to male gender roles.

If the gender gap in willingness to compete is larger when decisions are directly observable, the gender gap observed in previous studies may have underestimated the gender gap in willingness to compete in more realistic settings such as promotion contests, where the decision to enter the competition is observable. Knowing whether this is the case is important for managers, educators and policy makers, since previous experimental studies have found a positive correlation between willingness to compete in laboratory studies and educational and labor market related decisions. For example, application procedures for internal positions in a company can be made more or less confidential. Also, gossip could be detrimental to the advancement of women in organizations as they may be less willing to ask to be considered for promotions or bonuses if they think their colleagues will disapprove of their ambition. The decision to apply to a competitive school or university could be similarly affected.

Our experiment is carefully designed to isolate the public observability of the decision of whether to compete (and the public observability of the competition outcome) from all other factors. While we find a large gender gap in willingness to compete, this gap does not significantly increase when participants have to publicly announce their decisions. It is useful to note that our power calculation (presented in Appendix A1) demonstrates that we have sufficient (greater than 0.80) power to detect effects similar to those observed in previous work.

It is also useful to ask how much we can learn from this result. That is, given our data, how probable is it that public observability truly does not affect the gender gap in willingness to compete? In Appendix A2 we investigate this question using the post-study-probability (PSP) framework of Maniadis, Tufano and List (2014), which takes into account both the p-value and the sample size (and hence the power) of our study. Intuitively, the PSP tells us the posterior probability that our null hypothesis is true, given a prior probability and an alternative hypothesis. In analysis presented in Appendix A2 we show that our data in a meaningful way increase the probability that the null hypothesis is true, in particular for cases in which the alternative hypothesis is a moderate or large effect and the prior is small. For example, if our alternative hypothesis postulates a 20 percentage points larger gender gap in the Public Choice condition, and our prior is 0.5, the posterior probability that the null hypothesis holds increases from 0.5 to 0.93.[23] Hence, this analysis suggests that our data are informative in that they substantially decrease the posterior probability that public observability has an effect on the gender gap in willingness to compete in our setting.

In addition, our finding that the gender gap does not vary significantly across conditions is consistent with the lack of a gender difference in our vignette study, where participants report similar attitudes towards competitive women and competitive men. While

[21] We cannot exclude that participants to some extent understood the purpose of the study and adjusted their behavior in Part 3 in order not to display any bias. However, we judge it unlikely. In order to diminish these concerns each participant evaluated only one candidate in the vignette, and could therefore not compare across candidates and easily guess the purpose of the study. Moreover, the study did not mention gender in any explicit way. Addressing the concern that participants were made aware of the gender gap in competitiveness during the study by observing the other participants' choices, additional analysis does not indicate that participants in the Private and the Control conditions rated the job candidates differently than participants in the Public Choice and Public Outcome conditions.

[22] The fact that competitive men and competitive women are both perceived negatively suggests that increasing image concerns through public observability could have induced both genders to compete less. Several reasons may explain why we do not observe a similar effect, such as the relative importance of these norms compared to other factors guiding the decision to enter into competition, or other conflicting norms leading to different effects.

[23] We present several alternative examples in Appendix A2, exploring what our result tells us under various priors and hypothesized effect sizes. Since few studies explore settings that may increase the gender gap, we looked for examples of effect sizes in studies that to do the opposite, and explore experimental manipulations that aim to decrease the gender gap. While 20 percentage points is a fairly large effect, it is smaller than the effects found in, for example, studies exploring the effects of affirmative action (Balafoutas and Sutter 2012, Niederle, Segal and Vesterlund 2013) or cultural differences (Gneezy, Leonard and List 2009).

Table 4
Vignette results.

	(1) Skilled	(2) Beneficial	(3) Enjoyable	(4) Hireable
Self:				
Female, not competitive	0.077	0.057	0.294*	0.106
	(0.084)	(0.091)	(0.130)	(0.116)
Male, competitive	−0.140*	−0.216*	−0.517***	−0.375**
	(0.082)	(0.093)	(0.138)	(0.120)
Female, competitive	−0.122	−0.243*	−0.390**	−0.328**
	(0.090)	(0.104)	(0.139)	(0.120)
Others:				
Female, not competitive	−0.009	−0.015	0.142	0.040
	(0.078)	(0.081)	(0.101)	(0.092)
Male, competitive	−0.125	−0.211*	−0.513***	−0.315**
	(0.084)	(0.083)	(0.108)	(0.099)
Female, competitive	−0.086	−0.190*	−0.433***	−0.256**
	(0.087)	(0.088)	(0.110)	(0.098)
Observations	784	784	784	784

Notes: The table reports results from 8 OLS regressions using the vignette results as outcomes variables. All outcome variables are on a 7-point scale. "Skilled" is the answer to the question "How likely is it that the candidate has the skills for the job? (Scale: not at all likely-extremely likely)"; "Beneficial" is the answer to the question "How beneficial would it be to have this candidate working for you? (Scale: not at all beneficial-extremely beneficial)"; "Enjoyable" is the answer to the question "How much would you enjoy having this candidate working for you? (Scale: not at all enjoyable-extremely enjoyable)"; "Hireable" is the answer to the question "How likely is it that you would hire this candidate for the position? (Scale: not at all likely-extremely likely)". "Others" refers to participants' guess of the answer most commonly chosen by others. Independent variables are dummies for the "Female, not competitive", "Male competitive" and "Female competitive" treatments; the "Male, not competitive" treatment serves as the baseline. Robust standard errors are in parentheses. *p < 0.05, **p < 0.01, ***p < 0.001.

this suggests that there is no specific stigma associated with competitive women, it does not imply that other forms of stereotypes play no role for the participants in our sample. On the contrary, previous studies suggest that the large gender gap that is generally observed in this literature, and that we also observe across all of our conditions, is at least partly driven by women's lower confidence in the stereotypically male task used in most of the literature. Instead, our results indicate that willingness to compete is not rated more favorably in men, and hence emphasizing the importance of gender norms over competitive behaviors through public observability does not change the gender gap in willingness to compete.

In light of previous research, we find our results surprising. One possible explanation why we fail to find a larger gender gap in willingness to compete under public observability may be that public scrutiny causes individuals to consider their choices and the corresponding outcome more carefully and, for example, make more realistic estimations of their relative performance. However, we do not find evidence that the share of participants choosing the payoff maximizing option when deciding to compete or not increases in the conditions with public observability compared to the Control condition. Further, while the vignette study does not indicate that women are judged more harshly for competitive behavior in other contexts, it is of course also possibly that our results are driven by the specific way we implemented public observability, or by some specific characteristics of our sample. Another possibility is that the audience in our experiment (randomly selected other students) did not engender as strong social image concerns as relevant audiences in the real world (friends, colleague and family). Future work could replicate our experiment in different countries, with different audiences, or with different ways of manipulating public observability to explore whether our results are specific to our current design and setting, or apply more generally.

Many organizations and governments are currently considering different ways forward to increase gender equality. Our results help inform this process and indicate that increased transparency, which is sometimes proposed as a measure to decrease discrimination, does not imply a change in behavior on behalf of women (or men), nor does it necessarily imply backlash. It is also good news for countries like, for example, Norway, which are considering transparency legislation that makes public the identity of all applicants to government positions. The informed choice of institutions to promote and elect leaders can help increase the number of women at top positions.

Finally, our results are reassuring from the perspective of the existing experimental economics literature on gender differences in willingness to compete, since they suggest that the results of the standard design are largely robust to changes in the amount of public observability.

Declaration of Competing Interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joep.2021.102366.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics, 115*(3), 715–753.

Alan S., Ertac S., Kubilay E., & G. Loranth. (2016). "Understanding Gender Differences in Leadership." Working paper.

Amanatullah, E. T., & Morris, M. W. (2010). Negotiating Gender Roles: Gender Differences in Assertive Negotiating are Mediated by Women's Fear of Backlash and Attenuated when Negotiating on Behalf of Others. *Journal of Personality and Social Psychology, 98*(2), 256.

Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing Female Negotiators for Asserting too Much… Or not Enough: Exploring why Advocacy Moderates Backlash Against Assertive Female Negotiators. *Organizational Behavior and Human Decision Processes, 120*(1), 110–122.

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., & Dreber, A. (2015). Sex differences in competitiveness: Hunter-gatherer women and girls compete less in gender-neutral and male-centric tasks. *Adaptive Human Behavior and Physiology, 1*(3), 247–269.

Balafoutas, L., Davis, B. J., & M. Sutter (2017). "How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior." Working Paper.

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.

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review, 96*(5), 1652–1678.

Boschini, A., Dreber, A., von Essen, E., Muren, A., & Ranehill, E. (2019). Gender, Risk Preferences and Willingness to Compete in a Random Sample of the Swedish Population. *Journal of Behavioral and Experimental Economics, 83*(December), Article 101467.

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes it does Hurt to Ask. *Organizational Behavior and Human Decision Processes, 103*(1), 84–103.

Brandts, J., Gërxhani, K., & Schram, A. (2020). Are there gender differences in status-ranking aversion? *Journal of Behavioral and Experimental Economics, 84*. https://doi.org/10.1016/j.socec.2019.101485

Bursztyn, L., Fujiwara, T., & Pallais, A. (2017). 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review, 107*(11), 3288–3319.

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

Buser, T., Niederle, M., & Oosterbeek, H. (2020). "Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures." Working paper.

Buser, T., Geijtenbeek, L., & Plug, E. (2018). Sexual Orientation, Competitiveness and Income. *Journal of Economic Behavior & Organization, 151*(July), 191–198.

Buser, T., Peter, N., & Wolter, S. (2017a). Gender, Competitiveness and Study Choices in High School - Evidence from Switzerland. *American Economic Review, 107*(5), 125–130.

Buser, T., Peter N., & Wolter, S. (2017b). "Gender, willingness to compete and career choices along the whole ability distribution." Working paper.

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.

Charness, G., & Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization, 83*(1), 50–58.

Dasgupta, U., Mani, S., Sharma, S., & Singhal, S. (2019). "Can Gender Differences in Distributional Preferences Explain Gender Gaps in Competition?" Journal of Economic Psychology 70: 1-11.

De Paola, M., Lombardo, R., Pupo, V., & Scoppa, V. (2020). "Do Women Shy Away from Public Speaking? A Field Experiment." Working paper.

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

Eagly, A. H. (1987). *Sex Differences in Social Behavior: A Social-Role Interpretation*. Hillsdale, NJ: Erlbaum.

Ewers, M., & Zimmermann, F. (2015). Image and Misreporting. *Journal of the European Economic Association, 12*(2), 363–380.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made economic Experiments. *Experimental Economics, 10*, 171–178.

Flory, J. A., Leibbrandt, A., & List, J. A. (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.

Gillen, B., Snowberg, E., & Yariv, L. (2018). Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study. *Journal of Political Economy, 127*(4), 1826–1863. https://doi.org/10.1086/701681

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.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*(3), 1049–1074.

Gneezy, U., & Potters, J. (1997). An Experiment on Risk Taking and Evaluation Periods. *The Quarterly Journal of Economics, 112*(2), 631–645.

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.

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

Ivanova-Stenzel, R., & Kübler, D. (2011). Gender Differences in Team Work and Team Competition. *Journal of Economic Psychology, 32*(5), 797–808.

Heatherington, L., Daubman, K. A., Bates, C., Ahn, A., Brown, H., & Preston, C. (1993). Two Investigations of "Female Modesty" in Achievement Situations. *Sex Roles, 29*(11–12), 739–754.

Kamas, L., & Preston, A. (2009). "Social preferences, competitiveness and compensation: Are there gender differences?" Working paper.

Kleinjans, K. L. (2009). Do gender differences in preferences for competition matter for educational expectations? *Journal of Economic Psychology, 30*(5), 701–710.

Krupka, E. L., & Weber, R. A. (2013). Identifying Social Norms using Coordination Games: Why does Dictator Game Sharing Vary? *Journal of the European Economic Association, 11*(3), 495–524.

Ludwig, S., Fellner-Röhling, G., & Thoma, C. (2016). Do Women have more Shame than Men? An Experiment on Self-Assessment and the Shame of Overestimating Oneself. *European Economic Review, 92*(2), 31–46.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review, 104*(1), 277–290.

Moss-Racusin, C. A., & Rudman, L. A. (2010). Disruptions in Women's Self-Promotion: The Backlash Avoidance Model. *Psychology of Women Quarterly., 34*, 186–202.

Niederle, M. (2016). Gender. In J. H. Kagel, & A. E. Roth (Eds.), *Handbook of Experimental Economics* (2nd ed., pp. 481–553). Princeton: Princeton University Press.

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

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., & Vesterlund, L. (2011). Gender and Competition. *Annual Review of Economics, 3*(September), 601–630.

Phelan, J. E., & Rudman, L. A. (2010). Prejudice Toward Female Leaders: Backlash Effects and Women's Impression Management Dilemma. *Social and Personality Psychology Compass, 4*, 807–820.

Reuben, E., Sapienza, P., & Zingales, L. (2015). Taste for Competition and the Gender Gap Among Young Business Professionals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2677298

Reuben, E., Wiswall, M., & Zafar, B. (2015). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*. in press.

Ridgeway, C., & Correl, S. (2004). Unpacking the gender system - A theoretical perspective on gender beliefs and social relations. *Gender and Society, 18*(4), 510–531.
Ridgeway, C. (2009). Framed Before we Know it: How Gender Shapes Social Relations. *Gender and Society, 23*(2), 145–160.
Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personal Social Psychology, 74*, 629–645.
Rudman, L. A., & Glick, P. (1999). Feminized Management and Backlash toward Agentic Women: The Hidden Costs to Women of a Kinder, Gentler Image of Middle Managers. *Journal of Personality and Social Psychology, 77*(5), 1004–1010.
Rudman, L. A., & Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash Toward Agentic Women. *Journal of Social Issues, 57*(4), 743–762.
Samek, A. (2019). Gender Differences in Job Entry Decisions: A University-Wide Field Experiment. *Management Science, 65*(7), 3272–3281.
Schram, A., Brandts, J., & Gërxhani, K. (2018). Social-status ranking: A hidden channel to gender inequality under competition. *Experimental Economics.*
Shurchkov, O. (2012). Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints. *Journal of the European Economic Association, 10*(5), 1189–1213.
Sutter, M., & Glätzle-Rützler, D. (2014). Gender differences in the willingness to compete emerge early in life and persist. *Management Science, 61*(10), 2339–2354.
Van Veldhuizen, R. (2018). "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness?" Working paper.
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.
Wozniak, D., Harbaugh W.T., & Mayr, U. (2016). "The Effect of Feedback on Gender Differences in Competitive Choices." Working paper.
Zhang, Y. J. (2013). "Can Experimental Economics Explain Competitive Behavior Outside the Lab?" Working paper.