# Overview

Compensation packages based on performance pay (i.e., bonuses, commissions, and piece-rate payments) have risen in popularity relative to hourly/salaried pay, especially among workers in the highest tiers of occupations (Cuñat & Guadalupe, 2005; Hall & Liebman, 1998; Lemieux, MacLeod, & Parent, 2009; Murphy, 1999). There is evidence that the increasing use of performance pay lends itself to wage inequality. Lemieux et al. (2009) showed that an increased dependence on performance pay during the late 1970’s and early 1990’s accounted for 21% of the observed growth in variance of male wages. Bonuses and commissions, arguably the most competitive compensation schemes, may be especially important in driving the large disparity between the highest and lowest percentile earners within organizations (Bell & Van Reenen, 2010, 2014; Bénabou & Tirole, 2016). Importantly, performance pay may contribute to the gender wage gap too. Using data from the National Longitudinal Surveys of Youth, McGee, McGee, & Pan (2015) show that women are less likely to be employed in occupations that receive bonuses, and simultaneously are more likely to receive piece-rate pay – the least competitive of all forms of performance pay, where workers are paid based on their absolute output.

Since competition is relevant to labor market outcomes, researchers began to focus on how a person’s gender affects their response to competition as a means of understanding persistent gender gaps in labor market outcomes (for review, see Niederle & Vesterlund, 2011). Seminal work on gender differences in competitiveness operationalized competitiveness as the choice of a tournament payment scheme that reaps potentially higher earnings over a piece-rate scheme but requires outperforming an opponent (Niederle & Vesterlund, 2007). This work found that women are less competitive than men, on average, even if they would have earned more by competing (Niederle & Vesterlund, 2007). Additionally, this laboratory measure of competitiveness predicts decisions that affect labor market outcomes, such as education choices (Buser, Niederle, & Oosterbeek, 2014; Zhang, 2012) and entrepreneurial decisions (e.g., investment, employment; Berge, Bjorvatn, Garcia Pires, & Tungodden, 2015), and has been shown to predict earnings (Reuben, Sapienza, & Zingales, 2015). Thus, these decisions can perpetuate gender differences in labor market outcomes (Blau & Kahn, 2017).

Follow-up research with nearly identical procedures has replicated the effect of gender on the choice to opt into tournaments (see Niederle & Vesterlund, 2011 for review). Notably, this effect has been replicated in diverse populations (Andersen, Ertac, Gneezy, List, & Maximiano, 2013; Apicella & Dreber, 2015; Buser et al., 2014; Buser, Peter, & Wolter, 2017; Dreber, Essen, & Ranehill, 2014; Mayr, Wozniak, Davidson, Kuhns, & Harbaugh, 2012; Sutter, Glätzle-Rützler, Balafoutas, & Czermak, 2016; Sutter & Rutzler, 2010) and with a diverse set of tasks (Apicella & Dreber, 2015; Bjorvatn, Falch, & Hernæs, 2016; Frick, 2011; Saccardo, Pietrasz, & Gneezy, 2018; Samek, 2019; Sutter & Glätzle-Rützler, 2015). However, there is evidence that the task used during competition affects the size of these gender differences. When the task is female-typed or gender-neutral, the gender gap in willingness to compete is either reduced or completely eliminated (Apicella & Dreber, 2015; Boschini, Dreber, Essen, Muren, & Ranehill, 2019, 2014; Dreber, Essen, & Ranehill, 2011; Dreber et al., 2014; Grosse & Riener, 2010; Günther, Ekinci, Schwieren, & Strobel, 2010; Iriberri & Rey-Biel, 2017; Shurchkov, 2012). Drawing from the psychology literature on stereotype threat (Spencer, Logel, & Davies, 2016; Spencer, Steele, & Quinn, 1999; Steele, 1997), some have suggested that women may be less willing to compete in male-typed tasks (e.g., math, mental rotation) because they do not want to confirm stereotypes that they will perform worse than men in male-typed tasks, so they decide not to engage in competitions with men, even if they would earn more by competing (Grosse & Riener, 2010; Günther et al., 2010; Iriberri & Rey-Biel, 2017; Shurchkov, 2012).

While competitions are generally motivating and are designed to improve performance through increased effort (Connelly, Tihanyi, Crook, & Gangloff, 2014; Miller, Petrie, & Segal, 2019; Murayama & Elliot, 2012), some research suggests that men perform better under competitive payment schemes relative to non-competitive payment schemes, while women’s performance does not respond to competitions (Gneezy, Niederle, & Rustichini, 2003; Gneezy & Rustichini, 2004; Günther et al., 2010; Samak, 2013). For instance, Gneezy et al. (2003) show that there is no gender difference in performance when participants are solving mazes following a piece-rate payment scheme, but a significant gender difference in performance arises under a tournament payment scheme. As suggested above, task type tends to interact with payment scheme to affect gender differences in performance during competition, as demonstrated by Günther et al. (2010), who replicate the effect of competition on gender differences in performance for the male-typed task, but find no gender differences in performance during competition for female-typed or gender-neutral tasks. Relatedly, during repeated competition, women tend to perform worse in subsequent performance rounds after losing, even if the monetary prize they lost was relatively meager, while men only perform worse in subsequent rounds if they lost the chance to win a large monetary prize (Gill & Prowse, 2014). Other research suggests women stop competing altogether after losing if given the choice, as shown by Buser & Yuan (2019), who examine the effects of losing while competing in the Dutch Math Olympiad on the choice to compete in subsequent years. They show that men are just as likely to compete even if they lost the previous year, while women are less likely to compete in the Dutch Math Olympiad again if they lost before.

Given the implications of gender differences in competitiveness for gender disparities in labor market outcomes (Berge et al., 2015; Blau & Kahn, 2017; Buser et al., 2014; Niederle & Vesterlund, 2011; Reuben et al., 2015; Zhang, 2012), most of the current research on gender differences in competitions has focused on either i) explaining the sources of the gender differences (e.g., Veldhuizen, 2017) or ii) designing interventions to encourage women to compete more (Alan & Ertac, 2018; Balafoutas & Sutter, 2012; Brandts, Groenert, & Rott, 2015; Brandts et al., 2015; Cassar, Wordofa, & Zhang, 2016; Healy & Pate, 2011; Niederle, Segal, & Vesterlund, 2013; Sutter et al., 2016). Less consideration has been paid to how competitions affect women. However, as the research on the gender differences in performance during competition suggests (Buser & Yuan, 2019; Gill & Prowse, 2014; Gneezy et al., 2003; Gneezy & Rustichini, 2004; Günther et al., 2010; Miller et al., 2019; Samak, 2013), it is also important to consider the potential downstream consequences of women’s entry into competitions.

The present proposal builds on prior research by examining how competitions affect gender differences in the amount of time spent preparing for competitions. We hypothesize that women will spend more time preparing than men, especially before competitions, in part because they are less risk-seeking (Bertrand, 2010; Croson & Gneezy, 2009; Dohmen et al., 2011; Eckel & Grossman, 2008) and confident (Barber & Odean, 2001; Bertrand et al., 2010; Croson & Gneezy, 2009; Lundeberg, Fox, & Puncochaf, 1994; Mobius, Niederle, Niehaus, & Rosenblat, 2011) than men on average. Confidence and risk attitude have been implicated in driving the gender gap in willingness to compete (Niederle & Vesterlund, 2011; Veldhuizen, 2017).

However, the extent to which confidence and risk attitude account for the gender gap in willingness to compete is still unclear. The seminal research in this literature suggested that confidence and risk attitude did not completely explain the gender gap in the choice to compete, since there was a residual gender gap in the choice to compete after controlling for these factors (Niederle & Vesterlund, 2007). The unexplained component of the original gender effect was taken as evidence of a distinct “competitiveness” trait, separate from risk attitude and confidence, that was allegedly driving the gender gap in the choice to compete (Niederle & Vesterlund, 2007, 2011). However, recent work correcting for measurement error (Gillen, Snowberg, & Yariv, 2019) and using experimental techniques to isolate the effects of the competitiveness trait (Veldhuizen, 2017) find that risk attitude and confidence fully explain the gender gap in the choice to compete. Regardless of whether competitiveness is a stand-alone trait, it is clear that confidence and risk attitude can generate differences in how men and women react to competitions, possibly including the decision to prepare before a competition.

Confidence is conceptualized as the accuracy of one’s perceived performance or ability on a task (Beyer & Bowden, 1997). Since competitions, by definition, compare the performance among two or more individuals, they naturally lead to self-evaluation and an assessment of one’s competitors - processes that are intimately linked to confidence. To the extent that confidence influences how much individuals think they need to prepare in order to win, we may expect to see women preparing more than men, particularly in competitive contexts. Independent of any ambitions to win, less confident individuals may prepare more in order to reduce the negative feelings caused by low confidence. Indeed, mastery is an important driver of confidence (for review, see Gist & Mitchell, 1992; Usher & Pajares, 2008) and there is no theoretical or empirical reason to suspect that women would be less concerned with mastery than men. In fact, research suggests that women are just as likely as men to compete when competing against their own past performance, suggesting an equal desire for improving performance (Apicella et al., 2017b).

There is ample research to suggest that women are less (over)confident on average than men across a number of domains (Bertrand, 2010; Beyer, 1990; Beyer & Bowden, 1997; Croson & Gneezy, 2009; Lundeberg et al., 1994; Mobius et al., 2011; Niederle & Vesterlund, 2007, 2011). Within the literature on the gender gap in competitiveness, confidence is operationalized as the belief about one’s relative performance during a competition, where individuals who have inaccurately high ratings of their performance are deemed overconfident. If an individual does not feel as though their performance is higher than individuals they are competing against, they are unlikely to make the decision to compete for fear of missing the opportunity to earn money, even if they would otherwise outperform their opponent. Although both men and women tend to be overconfident, men are far more likely to fall into the trap of overconfidence, which leads them to compete more often than they should, given their actual ability (Niederle & Vesterlund, 2007).

Another variable that has been identified as a possible explanation for gender differences in competitiveness is risk attitude, typically construed as the preference for a certain gain over a gamble, even if the gamble has an equal or greater monetary expectation (Kahneman & Tversky, 1982). For instance, a risk averse person would prefer a sure gain of $80 over a gamble where they have an 85 percent chance of winning $100 and a 15 percent chance of winning $0, even though the monetary expectation in the latter case is higher (e.g., average of $85 in earnings relative to an average of $80) (Kahneman & Tversky, 1982). Payment based on the outcomes of a competition are inherently riskier than non-competitive payment schemes (e.g., guaranteed payment per unit of output) because in most cases, there is uncertainty surrounding one’s relative performance (Niederle & Vesterlund, 2011). Several studies across diverse settings have documented a gender difference in risk attitudes, where women tend to be more risk-averse than men on average (Bertrand, 2010; Croson & Gneezy, 2009). Because competitions are inherently riskier and women tend to be more risk-averse, women may prepare more than men before competing to reduce some of the uncertainty of performing during competition.

The current proposal examines gender differences in preparation in competitive contexts. Because women tend to be more risk-averse (Bertrand, 2010; Croson & Gneezy, 2009; Dohmen et al., 2011; Eckel & Grossman, 2008), less confident (Barber & Odean, 2001; Bertrand et al., 2010; Croson & Gneezy, 2009; Lundeberg et al., 1994; Mobius et al., 2011) and prefer to opt out of competitions (Niederle & Vesterlund, 2011), they may engage in more coping strategies before entering competitions, like preparing more. Indeed, previous research suggests women are more likely to emphasize the importance of working hard, value dedication more, and spend more time preparing than men in general (Hirt & Mccrea, 2009; Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Kimble & Hirt, 2005; Leslie, Cimpian, Meyer, & Freeland, 2015; Lucas & Lovaglia, 2005; Mccrea et al., 2008a, 2008b). For instance, Mccrea et al. (2008b) show that a scale measuring one’s belief’s about effort predicts gender differences in the tendency to behaviorally self-handicap (e.g., fail to study before a test, leading a person to perform poorly), where women were more likely to value effort, and in turn, self-handicap. Additionally, Mccrea et al. (2008a) show that that gender differences in perceived value of effort are not driven by possible confounds, such as achievement orientation or implicit theories of intelligence.

In addition, my own research substantiates gender differences in effort, where women were more likely to choose to prepare than men before completing a multiplication task, even though they were less likely to compete (Richards et al., in prep).[[1]](#footnote-2) The goal of this work was to explore how preparation might influence women’s willingness to compete. The first preliminary study in this line of work manipulated participants’ (*N* = 1056) knowledge of whether they would have unlimited time to prepare before they made their decision to compete. We expected that participants who knew they had unlimited time to prepare would be more inclined to compete compared to participants who were not aware of the opportunity to prepare before they decided on a payment scheme. While we did not find that knowledge of preparation affected participants’ decision to compete, there was a sizable gender difference in the choice to prepare, while controlling for the decision to compete - women were 75% more likely to choose to prepare compared to men when offered the opportunity , 95% CI , , , (see Figure 1). Additionally, at the end of the experiment, participants were incentivized to correctly predict which gender they believed would be more likely to prepare in the study. Both men (74%) and women (84%) believed that women would spend more time preparing for the task (, ), with similar results when asked which gender prepares more in general, where 82% of men and 87% of women believed women generally prepare more (, ). These effects hold while controlling for participants’ own decision to prepare. Therefore, data from the first study provides evidence that people accurately believe women are more likely to choose to prepare.

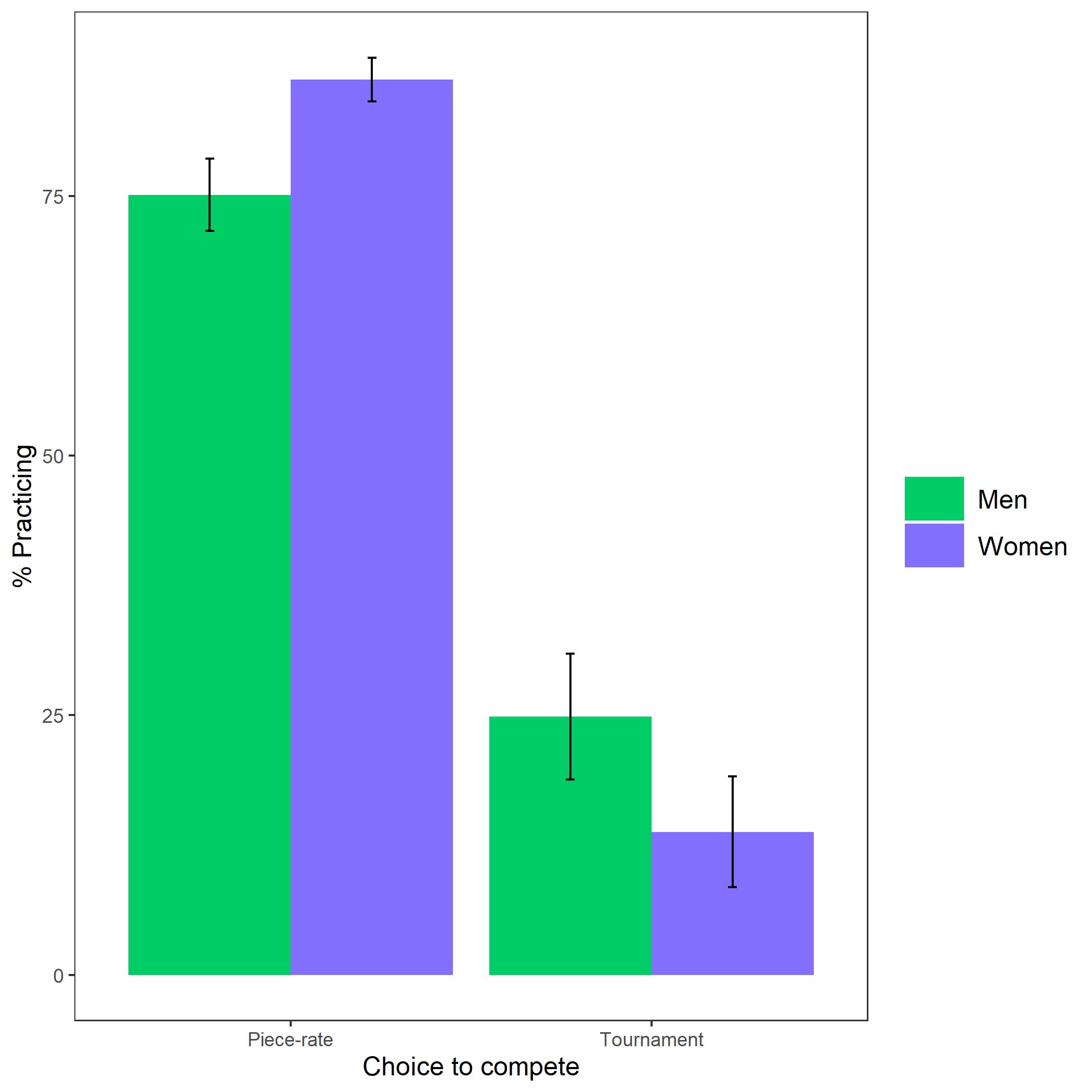


Figure 1. Proportion of participants who chose to compete based on participant gender from first study. Error bars represent standard error.

In a follow-up study designed to examine the role of forced preparation on the decision to compete, we recruited 1076 participants from Amazon Mechanical Turk (MTurk), who were assigned to either a condition where they were required to complete several rounds of practice for the upcoming paid multiplication task or several rounds of a filler task before choosing their payment scheme. Again, our manipulation had no effect on the choice to compete in men or women. However, after completing their respective rounds and choosing their payment scheme, participants in both conditions had the option to spend (extra) time preparing for the paid multiplication task. We replicated the effect of gender on preparation, where 40% of women and 35% of men chose to complete the optional preparation while controlling for the choice to compete (, 95% CI , , , ) (see Figure 2). In fact, even in the condition where participants were forced to prepare for 12 rounds lasting on average 2 minutes, there was a significant effect of gender on the choice to prepare (, 95% CI , , , ). Again, we find that these behaviors align with participants’ expectations that women will prepare more for the task (, ) and in general (, ). Overall, our previous work provides compelling evidence that women are more likely to choose to prepare than men, even after being forced to do so

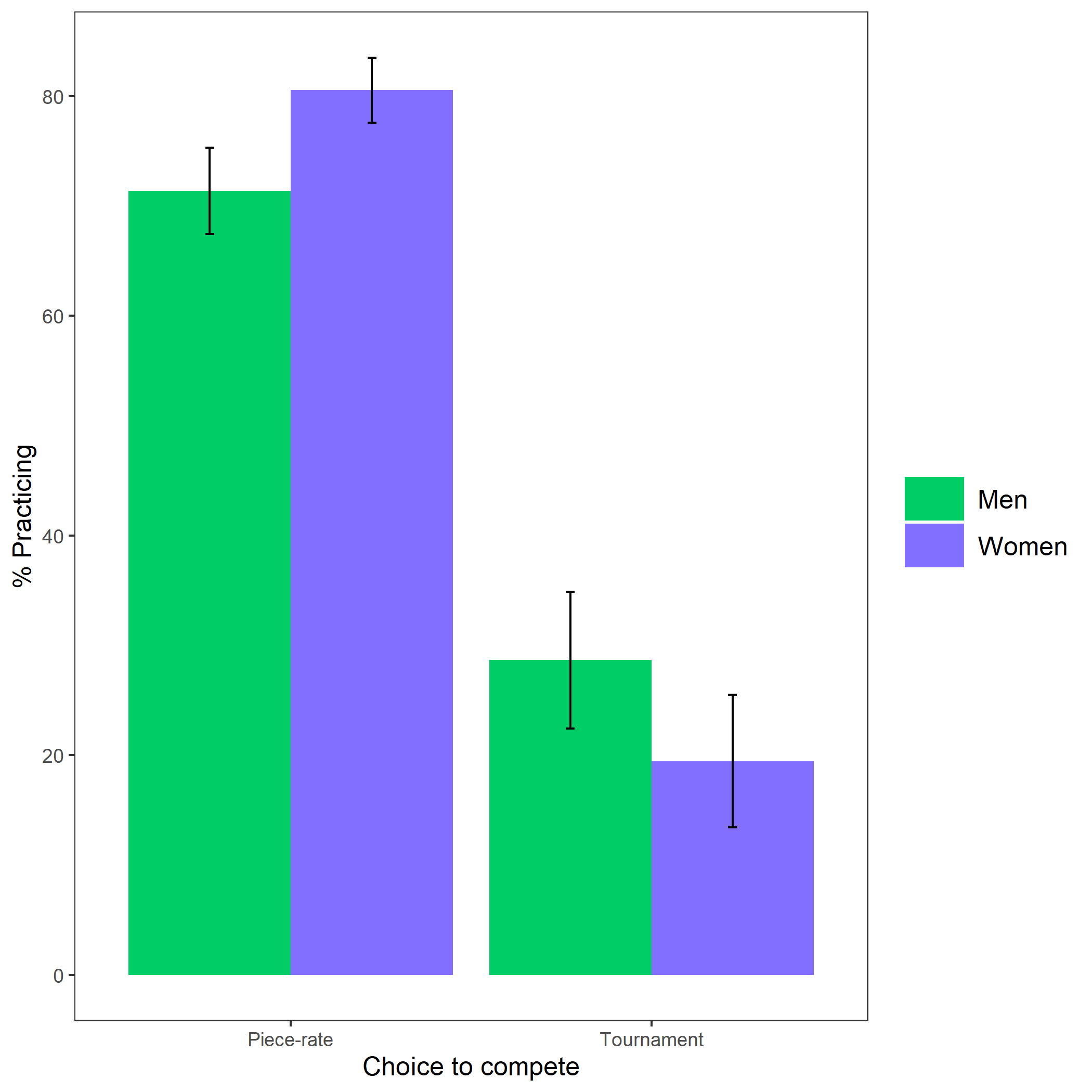


Figure 2. Proportion of participants who chose to compete based on participant gender from second study. Error bars represent standard error.

While this research did not find that preparation affects decisions to compete, we did uncover a sizable gender difference in decisions to prepare. In light of this discovery, we have flipped the research question **and instead** ask whether and how competitions affect men and women’s decisions to prepare. Notably, there was no interaction between gender and choice to compete on the choice to prepare **in either** of the previous studies. That is, women prepared more than men regardless of which payment scheme they had chosen. **However, it is impossible to draw conclusions** from this because i) we did not manipulate the payment scheme, so there could be selection effects on one’s choice to prepare across payment schemes and ii) there was little power to detect any possible interaction effects. For instance, in the first study, only 11% of women chose to compete, and in the second, 14% of women made this decision. Through the proposed experiments, we intend to address these limitations by directly manipulating participants’ payment scheme and recruiting a large sample to provide power to detect small effects.

We propose two studies. In Study 1, we will test whether competition exacerbates previously established gender differences in preparation by manipulating participants’ assigned payment scheme (i.e., competitive or non-competitive). We will also explore whether the effect of confidence and risk attitude on preparation differ based on participant gender and condition. Based on research suggesting that task-type affects women’s decision to compete (Apicella & Dreber, 2015; Boschini et al., 2019, 2014; Dreber et al., 2011, 2014; Grosse & Riener, 2010; Günther et al., 2010; Iriberri & Rey-Biel, 2017; Shurchkov, 2012), Study 2 will explore whether beliefs about how each gender performs on the task affects gender differences in preparation before competitions. Here, we expect women will spend significantly more time preparing on tasks when participants believe men outperform women, but the gender difference in preparation will be reduced when participants believe women outperform men on the task.

# Study 1: Does competition elicit gender differences in effort?

## Procedure

Given the difficulty of powering interaction effects (see [Simonsohn](http://datacolada.org/17) and [Giner-Sorolla](https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/)), we conducted a power analysis to determine an adequate sample size for the main hypothesized interaction effect in the primary analysis (simulations modeled after code [here](http://disjointedthinking.jeffhughes.ca/2017/09/power-simulations-r/)). We ran 5000 simulations while varying the sample size (*N* = 3000, 3250, 3500) and the effect size for the interaction effect (*b* = .2, .3, .4). Based on these simulated estimates, we will recruit 3250 participants to achieve at least 80% power for a relatively small effect (*b* = .2) (see Figure 3). Participants will be recruited to complete a study on “decision-making and performance” through MTurk, with a guaranteed payment and the opportunity to earn bonuses depending on their performance and the performance of others. Recruiting participants on this platform allows for efficient data collection while meeting acceptable psychometric standards, such as high test-retest and alpha reliability (Buhrmester, Kwang, & Gosling, 2011; Rand, 2012). Since we anticipate completing the required parts of the study will take no more than 10 minutes on average, we will pay participants $2.50 (i.e., double the federal and Pennsylvania minimum wage), with the opportunity for bonuses, outlined below. Participants will only be included if they indicate that they are 18 years or older, are American citizens, and identify as female or male while answering initial demographic questions.

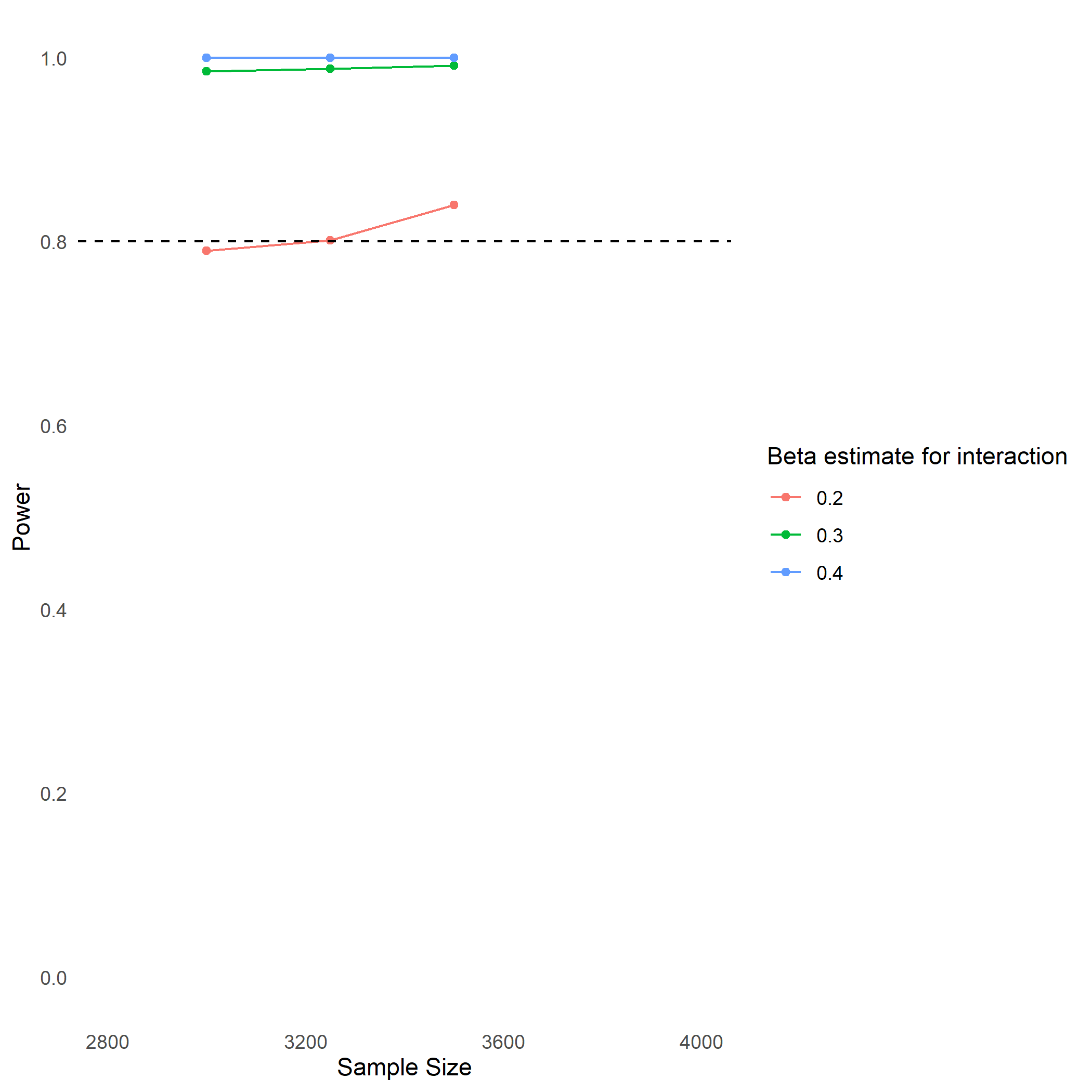


Figure 3. Plot of simulation output used to determine necessary sample size for at least 80% power in both proposed studies.

**Manipulation:** First, participants will be randomly assigned to follow either a competitive or noncompetitive payment scheme for one round (2 minutes) of multiplication problems. The payment scheme will be manipulated between subjects, where participants in the competition (tournament) condition will be paid 4 cents per problem on the task, but only if they beat another randomly assigned MTurker, while participants assigned to the noncompetitive (piece-rate) payment scheme will be paid 2 cents per problem. Although a within-subjects design would provide more power in detecting the hypothesized interaction effect, we opted to use a between-subjects design to avoid possible confounding. If we followed a within-subjects design, we would only be able to confidently interpret the results for whichever condition were presented first because there would be several confounds that could affect the decision to prepare (e.g., fatigue and/or learning effects reducing participants’ desire to prepare, demand effects for preparation if participants believe they are expected to prepare more in one condition compared to the other).

**Dependent variable:** After participants in each condition are told which payment scheme they will be following, they will have the option to complete unlimited “preparation” problems, which they will be told could improve their performance on the subsequent task. To measure their desire to prepare for the task, we will first ask participants whether they would like to spend any time practicing multiplication problems. For participants who agree to practice, we will ask them to enter how much time they would like to spend preparing using a slider scale (ranging from 0-5 minutes in 30 second increments), which will be granted. This measure is advantageous over using actual time preparing because participants may have unexpected interruptions while they are completing the preparation, which may lead us to erroneously conclude that they are expending more effort. Thus, we use chosen amount of time (in seconds) preparing as our primary dependent variable. Unbeknownst to the participants, the practice problems will be identical to the task itself, so preparing will likely improve performance on the task, as suggested to participants when they were presented the opportunity to practice. Once participants finish their optional preparation round, they will be asked to indicate whether they would like to continue preparing or move onto the task. If they choose the former option, they will be able to enter their desired time for preparing using the same slider scale. The dependent variable will be quantified as the total number of seconds of preparation across all optional preparation rounds.

**Task performance:** After practicing, participants in each condition will complete the paid multiplication task. Participants’ scores on the task will be quantified as the number of questions correct within the two-minute time frame allotted, without any penalties for incorrect responses. Afterwards, participants will be informed of the number of questions they answered correctly. We do not include any information about their relative performance since we ask them to guess their relative performance in the confidence measure. Thus, participants following the tournament pay scheme will not be told whether they won, since this serves as an indicator of relative performance.

There may also be the concern that participants will use a calculator to answer the multiplication questions, which may affect the interpretation of the results if there is a gender difference in calculator use and calculator use is related to the choice to practice. Our previous work suggests participants are unlikely to use calculators to complete the task and more importantly, there are no gender differences in the choices to use a calculator. In the first study we ran using the multiplication task, participants who completed the task were asked i) whether they thought using a calculator would help them answer the multiplication problems more quickly and ii) whether they used a calculator to complete the multiplication task (they were told their response would not affect their payment). Based on their responses, it is unlikely that participants will use calculators in the first place, since 86% of participants indicated that they thought using a calculator to answer the multiplication questions would slow them down and 93% of participants said they **did not** use a calculator. Importantly, there were no gender differences in perceptions of how calculators would affect performance (, ). Additionally, we did not find evidence of gender differences in actual calculator use (, ). Since we are recruiting participants through the same platform using the same task in the current study, we expect these findings will generalize to the current study, and thus, do not have evidence that gender differences in calculator use will be a confound when interpreting our results.

**Post-manipulation measures:** After completing the task, participants will complete a series of measures to be used for exploratory analyses. All questions will be counterbalanced. A confidence measure will incentivize participants to guess their relative performance compared to all other participants that completed the task by indicating the decile of their score relative to other participants. If correct, participants will earn $.25, which is 10% of their guaranteed earnings. We use a measure of relative performance, rather than a measure of absolute performance (e.g., asking participants to guess their score on the task) because perceptions of relative performance will likely be predictive of the choice to practice, especially when an individual is required to compete. The confidence measure draws from previous research (Niederle & Vesterlund, 2007), but instead of asking participants to indicate whether they won against a randomly selected opponent, we ask participant to guess their relative decile to provide us with more information about their relative confidence. Given the difficulty of guessing one’s exact percentile without any information about other participants, deciles are used rather than percentiles to make earning the bonus seem more achievable. Also, the item will be phrased so participants do not need to understand the word “decile,” but will be asked “If my performance is compared to that of all participants that completed the task, I think my score was…” with the options for responses ranging from “Better than all other participants” to “Better than none of the other participants” with 10% increments in between (e.g., “Better than 50% of participants”). Since task-specific confidence measures tend to be better predictors of behavior than general measures of confidence (see Oney & Oksuzoglu-Guven, 2015 for review), the confidence measure assesses participants’ beliefs within the context of the task used. We will also measure risk attitude by asking participants to indicate on a 0-10 scale “How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?” (Dohmen et al., 2011). There is evidence that risky behavior (i.e., lottery choices) is strongly associated with the risk measure included in the current proposal (Dohmen et al., 2011). Additionally, risk attitude tends to be explained by one underlying trait, with a relatively smaller amount of variation in risk attitude explained by context (e.g., risk attitude during career, health, or financial decisions). Thus, across contexts, risk attitude is likely to be stable and be predictive of behavior (Dohmen et al., 2011). These measures are included after completing the task largely because the confidence measure requires participants to state their perceived relative performance on the task.

## Attrition

Our previous work suggests that condition-dependent attrition is unlikely (Richards et al., in prep). In a prior study where participants completed one round of a similar multiplication task under each type of payment scheme, only a small proportion of participants (6%) dropped out during the study. Of the participants who dropped out, all did so at the end of the study, after completing both of the main tasks. Nevertheless, we will take several steps to counteract the possibility of condition-dependent attrition, which has the potential to lead to misleading conclusions (Zhou & Fishbach, 2016), especially if women and men drop out of the study at different rates based on condition. First, we will employ three costless strategies (i.e., personalization, forewarning of study content, and an appeal to participants’ conscience) suggested by Reips (2000) and shown in Zhou & Fishbach (2016) to be effective in reducing dropout rates by at least half. When participants enter the study, they will read a message that serves as both a forewarning and an appeal to their conscious (bolded) (modified from Zhou & Fishbach, 2016): “This is an anonymous survey consisting of multiple questions. **If a sizable number of people quit a survey partway, the data quality of that survey would be compromised. However, our research depends on good quality data,** so we ask that you are willing to participate in the survey for its entirety.” Then, participants will enter their MTurk ID as a means of establishing personalization. Notably, Zhou & Fishbach (2016) acknowledge that this is not a foolproof solution, since screening participants in advance in this way may reduce external validity. In this case, we want to have the capacity to establish the anticipated effect in the first place, so we are prioritizing internal validity.

In addition to these preventive measures, we will collect information about the rates of attrition during each study. Turkprime provides a metric for the overall rate of attrition, while Qualtrics offers the option to view partial responses from dropouts. For participants who drop out during or after learning about the manipulation, we will create an indicator variable for survey completion based on partial responses from Qualtrics, which will be coded as 1 if participants finish the study and 0 otherwise. This indicator will then be submitted as the dependent variable to a logistic regression with as predictors. If we find a significant interaction effect between gender and condition, this would suggest that we should interpret our results with caution because internal validity may be threatened, which will be explicitly stated in any reports on the studies, along with overall attrition rates and condition-dependent attrition rates (Zhou & Fishbach, 2016).

## Hypotheses and analyses

We will be using two-tailed tests during all hypothesis testing (*p* < .05) and all analyses will be conducted using *R*. To control the false-discovery rate during exploratory analyses, we will apply the Benjamini-Hochberg correction to all exploratory analyses. All analyses will be pre-registered on Open Science Framework.

**Primary analysis**: We expect that women will choose to prepare more than men, especially before a competition. We will test the interaction between gender and condition (competitive or noncompetitive pay) using a linear regression with amount of time a participant chose to prepare (summed across all rounds of preparation and then log-transformed) as the dependent variable. Thus, the following linear regression will be run: , where the piece-rate payment scheme and men will be coded as the reference groups for Condition and Gender, respectively. A positive beta coefficient for the interaction term () would support our hypothesis, indicating that the effect of gender on time spent preparing is greater in the tournament condition. Additionally, we expect positive beta coefficients for the main effects of gender and condition, suggesting the women and participants following the competitive pay scheme spent more time preparing. Finally, we will check that participants from different demographic groups were successfully randomized equally to each condition by running four separate logistic regressions with age, race/ethnicity, education, and income predicting condition (e.g., ). Since both proposed experiments include large sample sizes (*N* = 3250), it is unlikely demographic variables will become exceptionally imbalanced across conditions to the extent that they will explain our observed effects (Bowers, 2011). However, we will control for any variables that were not successfully randomized if we find significant differences in demographics across groups.

**Exploratory analysis 1:** We will explore whether the role of confidence on time spent preparing differs based on participant gender and condition. To this end, we will test whether confidence interacts with gender and condition by running a linear regression with time spent preparing (after log transformation) as a dependent variable. Therefore, the model will be structured as follows: . The reference groups will be the piece-rate payment scheme and men for condition and gender, respectively. Given the previous literature on gender gaps in confidence and competitiveness, we would expect a three-way interaction between gender, condition, and confidence on preparation, where women’s confidence plays a larger role in time spent preparing for women following the competitive payment scheme, relative to men following either the competitive or piece-rate payment scheme and women following the piece-rate payment scheme. A three-way interaction may be underpowered with the current sample size, so the analysis of the three-way interaction will serve as the foundation for future work using the effect sizes found.

**Exploratory analysis 2:** We will explore whether risk attitude interacts with participants’ gender and condition to affect their time spent preparing (log-transformed). The procedures for this test will be identical to those in Exploratory analysis 1, but risk attitude will replace confidence.

# Study 2: Do task stereotypes elicit gender differences in effort during competition?

## Procedure

Study 2 will explore the boundary conditions of gender differences in the choice to prepare during competition by manipulating stereotypes about gender differences in performance on the task. To attain power for the hypothesized interaction between condition and gender on time spent preparing (see Figure 3), we will recruit 3250 participants on MTurk to complete a study on “decision-making and performance.”

**Matching task**: Before seeing the manipulation, participants will learn about the paid task they will be completing. Participants will be paid to complete a one-minute matching task, where they are first presented a legend with numbers and corresponding letters. Using this legend, participants must enter letters that correspond to the sequences of two-digit numbers presented to them. This task was validated in a separate study online with MTurkers. Since my own research suggests participants complete the problems in the matching task twice as fast as problems in the multiplication task, we reduce the task time to one minute to reduce total study costs. This novel task is used instead of the multiplication task from Study 1 to increase the likelihood that participants will believe our manipulation (i.e., that men or women perform better) (Cvencek, Meltzoff, & Greenwald, 2011; Nosek, Banaji, & Greenwald, 2002; Swim, 1994). If we used the multiplication task or another task that participants were familiar with, it is possible participants may not believe there were gender differences in performance, or have pre-conceived ideas about which gender would perform better, based on any previous experience with the task they may have had. Also, our research showed that 80% of participants believed their score on the matching task would have improved with practice if they had been given the chance (, ) (Richards et al., in prep). Therefore, participants are motivated to practice before the matching task compared to other tasks where one’s score does not improve with practice.

In our study using the matching task with MTurkers, there was no significant gender difference in performance on the task, , 95% CI , , , , and we did not find a gender difference in competitiveness, , . These results provide tentative support for the notion that the task is perceived as gender-neutral both in terms of performance and the decision to compete, providing a more conservative estimate of the true effects of the manipulation, since a gender-typed task will likely amplify our anticipated effects. Notably, participants were significantly more likely to expect women performed better when asked about gender differences in performance (, ). However, we required participants to choose one gender in their response, instead of providing the option to guess that both genders performed equally well. Thus, it is entirely possible that participants would have accurately claimed that men and women performed equally well, if given that option. Even though it is unlikely that participants will not believe our statement about gender differences in performance since the task is novel and there were no gender differences in performance or competitiveness, we include a manipulation check and condition on participants’ responses to the manipulation check as a test of robustness.

**Manipulation**: After learning about the paid task, participants will read a statement about gender differences in performance supposedly found in a previous study, which will vary across conditions. Participants will be randomly assigned to learn that either men outperformed women or women outperformed men on the matching task (Fryer, Levitt, & List, 2008). We use deception in this study to manipulate condition rather than varying the task itself to avoid conflating the effects of different tasks with our hypothesized effects. The use of deception was one of the factors, in combination with the interdisciplinary nature of the proposal, that drove our decision to submit our DDIG proposal to the DRMS program instead of the Economics program.

Unlike Study 1, all participants will be entered into a competition

**Post-manipulation measures**: After completing the paid task, participants will complete the measures of risk attitudeParticipants will be incentivized to answer the confidence and manipulation check

## Demand effects

One may argue that the manipulation employed in Study 2 elicits demand effects, where participants may choose to prepare more when they are told their gender performs poorly on the task because they may be able to recognize our hypothesis and want to behave in ways that align with the hypothesis. We argue that demand effects are not problematic for interpreting the results for two reasons. First, it is unlikely participants will be sufficiently motivated by the unpaid preparation to succumb to demand effects, even if they know our hypothesis. Only 12% of U.S. MTurkers indicate that “MTurk money is irrelevant” and another 12% indicate that “MTurk is my primary source of income” (Mason & Suri, 2012), suggesting that many MTurkers try to maximize the amount of money they make in a given amount of time while on the platform, and are unlikely to be motivated by unpaid work. Even if participants are motivated to align their behavior with our hypothesis, there are many tasks in the real world where task stereotypes about gender differences in performance are either implicitly or explicitly stated (Grosse & Riener, 2010), with the assumption that one gender must exert more effort to “compensate” for a lack of ability. Thus, participants’ behavior in the study, even if driven by demand, will likely mirror effects we see in the real world, and as a result, will have real-world implications, especially if women are preparing more than necessary based on inaccurate stereotypes.

## Analyses and expected results

**Primary analysis**: We predict women will choose to prepare more than men before a competition, especially when the task is male-typed. The primary analysis will be parallel to Study 1, where the reference group for the condition variable will be the female-typed condition. A positive beta coefficient for the interaction term between the gender variable and the male-typed task condition would suggest that the manipulation elicited greater practice in women when they believed men performed better on the task. We do not expect the female-typed task will encourage men to practice significantly more than women, largely because men tend to be more confident on average than women (Bertrand, 2010; Beyer, 1990; Beyer & Bowden, 1997; Croson & Gneezy, 2009; Lundeberg et al., 1994; Mobius et al., 2011; Niederle & Vesterlund, 2007, 2011).

**Exploratory analyses 1 and 2:** Exploratory analyses 1 and 2 will be similar to the analyses in Study 1, with the female-typed task condition as the reference group for the condition variable.

**Exploratory analysis 3:** Additionally, we will compare participants’ time spent preparing (log-transformed) based on their responses to the manipulation check. Instead of excluding participants who fail the manipulation check, we will use this as a source of information about how participants’ beliefs affect their decision to prepare, even if they did not pay attention to the manipulation. To do this, we will run a linear regression with gender and participants’ response to the manipulation check as predictors and their time spent preparing as the outcome. Thus, the following linear regression will be run: , where responses that gender differences in performance were not found in the previous study and men will be coded as the reference groups for Manipulation check and Gender, respectively. A positive beta coefficient for the interaction term between the gender variable and the manipulation check responses where participants believed men performed better would suggest that the manipulation elicited greater practice in women when they held that belief. For the interaction between gender and the manipulation check responses where participants believed women performed better, we expect the coefficient will either be close to zero and nonsignificant or significant and positive (albeit smaller than the interaction between gender and beliefs that men performed better based on the manipulation check), like our predictions in primary analysis 1.

# Broader impacts

By identifying the gender difference in the choice to expend effort before a competition, we are tapping into a possible explanation for the robust gender difference in willingness to compete (Niederle & Vesterlund, 2011): women do not feel sufficiently prepared. Understanding how anticipated effort affects gender differences in competitiveness is important for reducing gender inequality that persists today. When women compete less than their male counterparts because they feel insufficiently prepared, they may be missing crucial economic opportunities, since competitiveness is relevant to one’s economic outcomes. For instance, the operationalization of competitiveness used in the current study is directly related to education choices, which is a crucial determinant of one’s career outcomes and may contribute to persistent horizontal job segregation, which in turn, explains a large proportion of the gender wage gap (Blau & Kahn, 2017). Competitiveness predicted Dutch students’ choice of track for the last three years of secondary school (Buser et al., 2014). In further support of the importance of competitiveness in explaining education choices, the gender differences in track choices were reduced by 20% after controlling for competitiveness for individuals matched on objective and subjective measures of ability. Similarly, Zhang (2012) showed that when admission to high school was dependent on performance in an entrance exam in China (Ninglang county), competitive students were more likely to take the entrance exam than their non-competitive counterparts. Additionally, a longitudinal study showed that the standard measure of competitiveness was a positive predictor MBA students’ earnings two years later, where students who chose the more competitive payment scheme earned 9% more (Reuben et al., 2015). The researchers also showed that the gender gap in earnings they observed was partially explained by competitiveness. Overall, these studies demonstrate the relevance of competitiveness to labor market outcomes, so factors that may affect competitiveness, like beliefs about effort, are important for improving gender equality in labor market outcomes. In fact, our research uses relatively unimportant tasks that are unlikely to affect participants’ labor market outcomes outside of the MTurk studies. Yet, our previous work shows a gender difference in preparation, suggesting that our study likely *underestimates* gender differences in choices to prepare for tasks that are more important for one’s career and/or labor market outcomes. In this way, our study is providing a conservative test of the gender differences in effort in the real world.

Additionally, Keana Richards is dedicated to promoting diversity in STEM both in her service and research activities. Outside of research, Richards serves as a mentor with the University of Pennsylvania College Achievement Program Graduate School Mentoring Initiative, which helps undergraduate students from disadvantaged backgrounds (e.g., first-generation, low-income) apply to graduate school. As research coordinator for the Upward Bound Math and Science Summer Scholars Academy at the University of Pennsylvania, Richards led a group of first-generation and/or low-income high school students in preparing a competitive application for the George Washington Carver Science Fair by providing guidance and feedback, improving their chance of earning an academic scholarship from Temple University. Over the course of the 6-week program, Richards also cultivated the students’ passion for research by teaching them how to review background literature, generate novel hypotheses and appropriate methodology to test them, analyze results, and craft a scientific poster based upon their research. Finally, the NSF DDIG will improve Richards’ ability to produce high-impact work that will enable her to pursue a tenure-track faculty position as a woman of color. As a tenure-track professor, she will be able to serve as an important role model for young women and people of color in the academy.

# Future directions

This work is critical because women who do not feel as though they are sufficiently prepared before entering competitive industries, jobs, or even majors in college may engage in costly and unnecessary preparation time, which may reduce their interest in pursuing these endeavors. This would suggest women’s greater willingness to prepare before they enter a competition may actually be harming them and their economic outcomes. To determine which gender is preparing more (or less) than needed, future research should test whether gender and time chosen to prepare interact to affect a participants’ probability of winning a competition (see Niederle & Vesterlund, 2007). Another follow-up study based on these experiments can manipulate whether there is a monetary cost for preparing to explore whether gender differences in the choice to prepare persist, and whether this leads to gender differences in earnings within the study. Finally, it is important to understand whether gender differences in preparation are universal (Apicella & Barrett, 2016; Henrich, 2015; Henrich, Heine, & Norenzayan, 2010), especially with recent work suggesting established psychological phenomenon in certain cultures may not be generalizable to other populations (Apicella et al., 2014; Falk, Heine, & Takemura, 2014; Henrich et al., 2005; Marcinkowska et al., 2014; Muthukrishna et al., 2020; Paulus, 2015; Pisanski & Feinberg, 2013), including gender differences in competitiveness (Gneezy, Leonard, & List, 2009). Given Dr. Apicella’s expertise working with populations across cultures (Apicella, 2014, 2018; Apicella et al., 2014, 2017a, 2007a, 2007b, 2018; Apicella & Barrett, 2016; Apicella & Crittenden, 2015; Apicella & Dreber, 2015; Apicella & Feinberg, 2009; Apicella, Marlowe, Fowler, & Christakis, 2012), she is well-equipped to explore whether this phenomenon exists across cultures in future work.

Alan, S., & Ertac, S. (2018). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association*, *17*(4), 1147–1185.

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. (2014). Upper-body strength predicts hunting reputation and reproductive success in Hadza hunter-gatherers. *Evolution and Human Behavior*, *35*(6), 508–518. <https://doi.org/10.1016/j.evolhumbehav.2014.07.001>

Apicella, C.2018). High levels of rule-bending in a minimally religious and largely egalitarian forager population. *Religion, Brain and Behavior*, *8*(2), 133–148. <https://doi.org/10.1080/2153599X.2016.1267034>

Apicella, C. L., Azevedo, E. M., Christakis, N& Fowler, J. H. (2014). Evolutionary origins of the endowment effect: Evidence from hunter-gatherers. *American Economic Review*, *104*(6), 1793–1805. <https://doi.org/10.1257/aer.104.6.1793>

Apicella, C. L., & Barrett, H. C. (2016). Cross-cultural evolutionary psychology. *Current Opinion in Psychology*, *7*, 92–97. <https://doi.org/10.1016/j.copsyc.2015.08.015>

Apicella, C. L., & Crittenden, A. N. (2015). Hunter-gatherer families and parenting. In *The handbook of evolutionary psychology* (pp. 578–597).

Apicella, C. L., Crittenden, A. N., & Tobolsky, V. A. (2017a). Hunter-gatherer males are more risk-seeking than females, even in late childhood. *Evolution and Human Behavior*, *38*, 592–603.

Apicella, C. L., Demiral, E. E., & Mollerstrom, J. (2017b). No gender difference in willingness to compete when competing against self. *DIW Discussion Papers*.

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. <https://doi.org/10.1007/s40750-014-0015-z>

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1077–1082. <https://doi.org/10.1098/rspb.2008.1542>

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007a). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, *3*(6), 682–684. <https://doi.org/10.1098/rsbl.2007.0410>

Apicella, C. L., Little, A. C., & Marlowe, F. W. (2007b). Facial averageness and attractiveness in an isolated population of hunter-gatherers. *Perception*, *36*(12), 1813–1820. <https://doi.org/10.1068/p5601>

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, *481*(7382), 497–501. <https://doi.org/10.1038/nature10736>

Apicella, C. L., Rozin, P., Busch, J. T. A., Watson-Jones, R. E., & Legare, C. H. (2018). Evidence from hunter-gatherer and subsistence agricultural populations for the universality of contagion sensitivity. *Evolution and Human Behavior*, *39*(3), 355–363. <https://doi.org/10.1016/j.evolhumbehav.2018.03.003>

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

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, *116*(1), 261–292.

Bell, B., & Van Reenen, J. (2010). Bankers’ pay and extreme wage inequality in the UK. *Centre for Economic Performance (CEPR)*, (April), 1–34. <https://doi.org/DOI:>

Bell, B., & Van Reenen, J. (2014). Bankers and their bonuses. *Economic Journal*, *124*(574). <https://doi.org/10.1111/ecoj.12101>

Berge, L. I. O., Bjorvatn, K., Garcia Pires, A. J., & Tungodden, B. (2015). Competitive in the lab, successful in the field? *Journal of Economic Behavior and Organization*, *118*, 303–317. <https://doi.org/10.1016/j.jebo.2014.11.014>

Bertrand, M. (2010). New Perspectives on Gender. In *Handbook of labor economics* (Vol. 4b, pp. 1545–1592).

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Review*, *2*(3), 228–255.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, *59*(5), 960–970. <https://doi.org/10.1037/0022-3514.59.5.960>

Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, *23*(2), 157–172.

Bénabou, R., & Tirole, J. (2016). Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, *124*(2), 305–370.

Bjorvatn, K., Falch, R., & Hernæs, U. (2016). Gender, context and competition: Experimental evidence from rural Uganda. *Journal of Behavioral and Experimental Economics*, *61*, 31–37.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865. <https://doi.org/10.1257/jel.20160995>

Boschini, A., Dreber, A., Essen, E. V., 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*(September), 101467. <https://doi.org/10.1016/j.socec.2019.101467>

Boschini, A., Dreber, A., Essen, E. von, Muren, A., & Ranehill, E. (2014). Gender and economic preferences in a large random and representative sample. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2443315>

Bowers, J. (2011). Making effects manifest in randomized experiments. In *Cambridge handbook of experimental political science* (pp. 459–480).

Brandts, J., Groenert, V., & Rott, C. (2015). The impact of advice on women’s and men’s selection into competition. *Management Science*, *61*(5), 1018–1035. <https://doi.org/10.1287/mnsc.2013.1877>

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. <https://doi.org/10.1177/1745691610393980>

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competition and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447. <https://doi.org/10.1093/qje/qju009.Advance>

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

Buser, T., & Yuan, H. (2019). Do women give up competing more easily? Evidence from the lab and the Dutch Math Olympiad. *American Economic Journal: Applied Economics*, *11*(3), 225–252.

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. <https://doi.org/10.1073/pnas.1520235113>

Connelly, B. L., Tihanyi, L., Crook, T. R., & Gangloff, K. A. (2014). *Tournament theory: Thirty years of contests and competitions* (Vol. 40, pp. 16–47). <https://doi.org/10.1177/0149206313498902>

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474. <https://doi.org/10.1257/jel.47.2.448>

Cuñat, V., & Guadalupe, M. (2005). How does product market competition shape incentive contracts? *Journal of the European Economic Association*, *3*(5), 1058–1082.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550. <https://doi.org/10.1111/j.1542-4774.2011.01015.x>

Dreber, A., Essen, E. V., & Ranehill, E. (2011). Outrunning the gender gap — boys and girls compete equally. *Experimental Economics*, *14*(4), 567–582. <https://doi.org/10.1007/s10683-011-9282-8>

Dreber, A., Essen, E. V., & Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Experimental Economics*, *17*, 154–172. <https://doi.org/10.1007/s10683-013-9361-0>

Eckel, C. C., & Grossman, P. J. (2008). Men, women, and risk aversion: Experimental evidence. In *Handbook of experimental economics* (Vol. 1, pp. 1061–1073). <https://doi.org/10.1016/S1574-0722(07)00113-8>

Falk, C. F., Heine, S. J., & Takemura, K. (2014). Cultural variation in the minimal group effect. *Journal of Cross-Cultural Psychology*, *45*(2), 265–281. <https://doi.org/10.1177/0022022113492892>

Frick, B. (2011). Gender differences in competitiveness: Empirical evidence from professional distance running. *Labour Economics*, *18*(3), 389–398. <https://doi.org/10.1016/j.labeco.2010.11.004>

Fryer, B. R. G., Levitt, S. D., & List, J. A. (2008). Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study. *American Economic Review*, *98*(2), 370–375.

Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, *5*, 351–376. <https://doi.org/10.3982/QE309>

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.

Gist, M. E., & Mitchell, T. R. (1992). Self-Efficacy: A Theoretical Analysis of Its Determinants and Malleability. *The Academy of Management Review*, *17*(2), 183–211.

Gneezy, B. 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. <https://doi.org/10.3982/ECTA6690>

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

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *The American Economic Review*, *94*(2), 377–381.

Grosse, N., & Riener, G. (2010). Explaining gender differences in competitiveness: Gender-task stereotypes. *Jena Economic Research Papers*.

Günther, C., Ekinci, N. A., Schwieren, C., & Strobel, M. (2010). Women can’t jump?-An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior and Organization*, *75*(3), 395–401. <https://doi.org/10.1016/j.jebo.2010.05.003>

Hall, B., & Liebman, J. (1998). Are CEOs really paid like bureaucrats? *The Quarterly Journal of Economics*, *113*(3), 653–691.

Healy, A., & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, *121*(555), 1192–1204. <https://doi.org/10.1111/J.1468-0297.2010.02409.X.>

Henrich, J. (2015). Culture and social behavior. *Current Opinion in Behavioral Sciences*, *3*, 84–89. <https://doi.org/10.1016/j.cobeha.2015.02.001>

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. <https://doi.org/10.1017/S0140525X05000142>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83.

Hirt, E. R., & Mccrea, S. M. (2009). Man smart, woman smarter ? Getting to the root of gender differences in self-handicapping. *Social and Personality Psychology Compass*, *3*, 260–274.

Iriberri, N., & Rey-Biel, P. Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. *Journal of Economic Behavior and Organization*, *135*, 99–111.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160–173.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children’s approach to schoolwork. *Developmental Psychology*, *42*(1), 11–26. <https://doi.org/10.1037/0012-1649.42.1.11>

Kimble, C. E., & Hirt, E. R. (2005). Self-focus, gender, and habitual self-handicapping: Do they make a difference in behavioral self-handicapping? *Social Behavior and Personality*, *33*(1), 43–56.

Lemieux, T., MacLeod, B. W., & Parent, D. (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics*, *CXXIV*(1), 1–49.

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 23–34. <https://doi.org/10.1081/E-EWS>

Lucas, J. W., & Lovaglia, M. J. (2005). Self-handicapping: Gender, race, and status. *Current Research in Social Psychology*, *10*(15), 234–249.

Lundeberg, M. A., Fox, P. W., & Puncochaf, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*(1), 114–121.

Marcinkowska, U. M., Kozlov, M. V., Cai, H., Contreras-Garduño, J., Dixson, B. J., Oana, G. A., … Rantala, M. J. (2014). Cross-cultural variation in men’s preference for sexual dimorphism in women’s faces. *Biology Letters*, *10*(4), 4–7. <https://doi.org/10.1098/rsbl.2013.0850>

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>

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. <https://doi.org/10.1037/a0025655>

Mccrea, S. M., Hirt, E. R., Hendrix, K. L., Milner, B. J., & Steele, N. L. (2008a). The worker scale: Developing a measure to explain gender differences in behavioral self-handicapping. *Journal of Research in Personality*, *42*, 949–970. <https://doi.org/10.1016/j.jrp.2007.12.005>

Mccrea, S. M., Hirt, E. R., & Milner, B. J. (2008b). She works hard for the money: Valuing effort underlies gender differences in behavioral self-handicapping. *Journal of Experimental Social Psychology*, *44*, 292–311. <https://doi.org/10.1016/j.jesp.2007.05.006>

McGee, A., McGee, P., & Pan, J. (2015). Performance pay, competitiveness, and the gender wage gap: Evidence from the United States. *Economic Letters*, *128*, 35–38.

Miller, A. R., Petrie, R., & Segal, C. (2019). Does workplace competition increase labor supply? Evidence from a field experiment. *Melbourne Institute Working Paper*, (14).

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. *NBER Working Paper*.

Murayama, K., & Elliot, A. J. (2012). The competition-performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, *138*(6), 1035–1070. <https://doi.org/10.1037/a0028324>

Murphy, K. J. (1999). Executive compensation. In *Handbook of labor economics* (pp. 2485–2563). Retrieved from <papers3://publication/uuid/3DC06196-20D7-49BE-9E07-5779C52B93A7>

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, BBeyond western, educated, industrial, rich, and democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, *31*(6), 678–701. <https://doi.org/10.1177/0956797620916782>

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. <https://doi.org/10.1287/mnsc.1120.1602>

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. Retrieved from [https://web.stanford.edu/{~}niederle/Niederle.Vesterlund.QJE.2007.pdf](https://web.stanford.edu/%7B~%7Dniederle/Niederle.Vesterlund.QJE.2007.pdf)

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, *3*, 601–630. <https://doi.org/10.1016/j.labeco.2009.08.002>

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*(1), 44–59. <https://doi.org/10.1037//0022-3514.83.1.44>

Oney, E., & Oksuzoglu-Guven, G. (2015). Confidence: A critical review of the literature and an alternative perspective for general and specific self-confidence. *Psychological Reports*, *116*(1), 149–163. <https://doi.org/10.2466/07.PR0.116k14w0>

Paulus, M. (2015). Children’s inequity aversion depends on culture: A cross-cultural comparison. *Journal of Experimental Child Psychology*, *132*, 240–246. <https://doi.org/10.1016/j.jecp.2014.12.007>

Pisanski, K., & Feinberg, D. R. (2013). Cross-cultural variation in mate preferences for averageness, symmetry, body size, and masculinity. *Cross-Cultural Research*, *47*(2), 162–197. <https://doi.org/10.1177/1069397112471806>

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*, 172–179. <https://doi.org/10.1016/j.jtbi.2011.03.004>

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In *Psychological experiments on the internet* (pp. 89–117). <https://doi.org/10.1016/b978-012099980-4/50005-8>

Reuben, E., Sapienza, P., & Zingales, L. (2015). Taste for competition and the gender gap among young business professionals. *NBER WORKING PAPER SERIES*.

Saccardo, S., Pietrasz, A., & Gneezy, U. (2018). On the Size of the Gender Difference in Competitiveness. *Management Science*, *64*(4), 1541–1554.

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

Samek, A. (2019). Gender differences in job entry decisions: A university- wide field experiment. *Management Science*, *65*(7), 3272–3281.

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.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype Threat. *Annual Review of Psychology*, *67*(1), 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women’s math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>

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. <https://doi.org/10.1287/mnsc.2014.1981>

Sutter, M., Glätzle-Rützler, D., Balafoutas, L., & Czermak, S. (2016). Cancelling out early age gender differences in competition: an analysis of policy interventions. *Experimental Economics*, *19*(2), 412–432. <https://doi.org/10.1007/s10683-015-9447-y>

Sutter, M., & Rutzler, D. (2010). Gender differences in competition emerge early in life. *Working Papers in Economics and Statistics*.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*(1), 21–36.

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, *78*(4), 751–796. <https://doi.org/10.3102/0034654308321456>

Veldhuizen, R. van. (2017). Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness? *Dicussion Paper*, *14*.

Zhang, Y. J. (2012). Can experimental economics explain competitive behavior outside the lab? *Unpublished Manuscript*, 1–45. <https://doi.org/10.2139/ssrn.2292929>

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the Web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. <https://doi.org/10.1037/pspa0000056>

1. All prior studies were pre-registered on [Open Science Framework](https://osf.io/q39a5/) [↑](#footnote-ref-2)