# Project summary

## Overview

Competitions are increasingly prevalent in the global labor market (Lavy, 2004; Lemieux, MacLeod, & Parent, 2009) and the winners of competitions are disproportionately rewarded (Frank & Cook, 2010). Understanding individual differences in response to competitive situations may help address economic disparities across groups, like persistent gender differences in labor market outcomes (Altonji & Blank, 1999; Blau & Kahn, 2017). Our proposed work will examine how much effort women choose to expend before feeling comfortable entering a competition. Study 1 will examine how tournament and piece-rate payment schemes for performance on a multiplication task affect how much time men and women spend practicing multiplication. We hypothesize that women will spend more time, on average, preparing relative to men, especially under a tournament payment scheme. In Study 2, we will explore the boundary conditions of the findings from Study 1 by exploring how manipulating beliefs about gender differences in performance on a gender-neutral matching task affects gender differences in the decision to prepare. We expect women will spend significantly more time preparing for the matching task when participants believe men outperform women, but the gender difference in preparation will be reduced when participants believe women outperform men on the task.

## Intellectual merit

The proposed work will advance knowledge by providing the foundation for a fruitful line of work focused on gender differences in the choice to prepare before a competition and its possible economic ramifications for women. Additionally, many standard economic models do not account for gender differences in behavioral traits as explanations for gender differences in economic outcomes (Niederle, 2017). The current proposal will demonstrate the need to include gender-based behavior in future economic models (e.g., Akerlof & Kranton, 2000), and provide data to lay the foundation for building these models. Also, gender differences in willingness to prepare before competition may serve as a possible explanation for robust gender differences in the choice to compete, which has not yet been fully explained in previous research (Niederle & Vesterlund, 2011). Specifically, gender differences in the choice to compete may result from women’s desire to spend more time preparing before willingly entering a competition. Thus, the proposed research will identify a robust phenomenon that previous researchers have not yet considered in explaining gender differences in the choice to compete.

## Broader impacts

Much of the research on gender differences in competitiveness is focused on designing interventions to increase women’s competitiveness, with less attention paid to potential downstream consequences of these interventions. Yet, understanding how women and men respond to competitions may also be crucial for addressing gender disparities in labor market outcomes. If competitions exacerbate gender differences in the amount of effort exerted (e.g., preparing or studying) before performance, this may affect women’s labor output, career advancement, the ability to achieve a satisfying work-life balance, and even the decision to enter or stay in competitive environments. Additionally, there are opportunity costs for expending more effort than necessary on only one task or skill, especially when an individual’s overall performance is evaluated based on their success across many domains. These effects are especially relevant to professional development in STEM fields, where gender disparities are arguably driven by gender differences in confidence, increasing the likelihood women will overprepare before competition (Cheryan, Ziegler, Montoya, & Jiang, 2017). Finally, gender differences in effort before entry into competitions would have important implications for many of the competitions we observe in the modern labor market, where small differences in effort and/or ability can lead to massive disparities in economic outcomes. Thus, understanding gender differences in the choice to expend effort before competitions may be one avenue to explain persistent gender gaps in economic outcomes.

# Project description

## Prevalence of competition in the modern labor market

As competition in the labor market increases, compensation packages based on performance pay (including bonuses, piece-rate, and commission) have become more popular compared to hourly/salaried pay, especially among the higher tiers of an organization (Cuñat & Guadalupe, 2005; Hall & Liebman, 1998; Lemieux et al., 2009; Murphy, 1999). There is evidence that the increasing use of performance pay lends itself to wage inequality. For instance, Lemieux et al. (2009) show that an increased dependence on performance pay during the late 1970’s (1976-1979) and early 1990’s (1990-1993) accounted for 21% of the observed growth in variance of male wages. Bonuses and commissions, arguably the most competitive compensation schemes, have been especially important in driving the large disparity between the highest and lowest percentile earners within an organization (Bell & Van Reenen, 2010, 2014; Bénabou & Tirole, 2016).

There is evidence suggesting the prevalence of performance pay across labor markets contributes to the gender wage gap. Using data from the National Longitudinal Surveys of Youth in 1979 and 1997, McGee, McGee, & Pan (2015) show that women are less likely to be employed in occupations that receive the most competitive form of performance pay (i.e., bonuses) and simultaneously are more likely to receive piece-rate pay, arguably the least competitive form of performance pay.

## Gender differences in competitiveness, confidence, risk aversion, and effort

### Competitiveness

Since competition is relevant to labor market outcomes (Buser, Niederle, & Oosterbeek, 2014; Reuben, Sapienza, & Zingales, 2015; Zhang, 2012), researchers have started 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 (Niederle & Vesterlund, 2011). Seminal work on gender differences in competitiveness operationalized competitiveness as the choice to follow a payment scheme that reaped higher rewards but required beating an opponent and found women are more likely to avoid competitions when given the choice (Niederle & Vesterlund, 2007). Follow-up research has replicated the effect of gender on the choice to opt into competitions, and simultaneously shown that women who are required to compete perform more poorly in tournaments and stop competing after losing, especially in mixed-sex tournaments and male-typed tasks (Apicella & Dreber, 2015; Buser & Yuan, 2019; Dreber, Essen, & Ranehill, 2014; Gill & Prowse, 2014; Gneezy, Niederle, & Rustichini, 2003; Niederle, 2017; Niederle & Vesterlund, 2007, 2011). Given the robust nature of gender differences in competitiveness and its implications for labor market outcomes (Niederle & Vesterlund, 2011), many researchers have sought to explain gender differences in competitiveness over the past several years. Most notably, confidence and risk aversion have been identified as two of the most robust explanations for gender differences in competitiveness (for review, see Niederle & Vesterlund, 2011).

### Confidence

In the steadily growing literature on gender differences in competitiveness, confidence has repeatedly been shown across contexts to explain gender differences in competitiveness (Niederle & Vesterlund, 2011). Women tend to be less (over)confident on average than men (Bertrand, 2010; Beyer, 1990; Beyer & Bowden, 1997; Croson & Gneezy, 2009; Lundeberg, Fox, & Puncochaf, 1994; Mobius, Niederle, Niehaus, & Rosenblat, 2011; Niederle & Vesterlund, 2007, 2011), especially in certain domains, like math (Beyer & Bowden, 1997; Grosse & Riener, 2010). Confidence is conceptualized as the accuracy of one’s perceived performance or ability on a task (Beyer & Bowden, 1997), and as such, affects the decision to compete. Within the literature on the gender gap in competitiveness, confidence is operationalized as beliefs 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 performance (Niederle & Vesterlund, 2007). Gender differences in confidence consistently explain part of the gender gap in competitiveness, although the gap typically remains significant even after controlling for confidence, suggesting unexplained factors are driving the gender gap in competitiveness (Niederle & Vesterlund, 2011).

### Risk aversion

Another variable that has been identified as a possible explanation for gender differences in competitiveness is risk aversion, 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 more risky 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). Since risk aversion is relevant to the decision to compete and women tend to be more risk-averse than men on average (Bertrand, 2010; Croson & Gneezy, 2009), researchers have examined whether risk aversion is sufficient to fully or partially explain the gender gap in competitiveness. Risk aversion is typically measured using one’s self-reported “willingness to take risk in general” on a 10-point scale, which has been confirmed as a valid measure of risk aversion that maps onto risk-taking behavior (Dohmen et al., 2011). Using this measure, there is evidence that risk aversion explains part of the gender gap in competitiveness, although the role of risk aversion is less robust and smaller than the role of confidence (Niederle & Vesterlund, 2011).

### Effort

Outside of the choice to enter competitions, another behavior that is likely affected by these established gender differences in confidence and risk aversion is the choice to exert effort before or during performance. Specifically, competitive environments are inherently more risky than non-competitive environments, which may lead individuals to reduce the risk of missing a reward by choosing to exert more effort before competition. Simultaneously, competitions may reduce one’s confidence by drawing attention to their relative performance, which may lead people to engage in several self-regulatory behaviors before and while performing in competitive environments. For instance, they may exert more effort before the competition and try to master relevant skills, which is one of the most important drivers of confidence (for review, see Gist & Mitchell, 1992; Usher & Pajares, 2008). Since women tend to be more risk-averse and less confident to begin with (Croson & Gneezy, 2009; Lenney, 1977; Maccoby & Jacklin, 1974; Niederle, 2014; Niederle & Vesterlund, 2011), they may respond more negatively to competitive environments than men and as a result, engage in more coping strategies, like preparing more before entering the competition. In fact, previous evidence suggests women value effort and 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, Hirt, & Milner, 2008). For instance, Hirt, Mccrea, & Boris (2003) find that women perceive effort as the norm and any type of effort withdrawal is unacceptable. Similarly, more feminine individuals rated the importance of studying higher compared to masculine individuals (Grabill et al., 2005). Observing gender differences in effort likely leads people to perceive effort as feminine, which is supported by evidence that people more readily identify a student as female when they are told the individual has exerted effort to improve their grades (Grabill et al., 2005; Power, Whitty, Edwards, & Wigfall, 1998).

#### Preliminary data

Our previous work substantiates gender differences in effort, where women were more likely to choose to prepare before performing on a multiplication task, even though they were less likely to compete (Richards et al., in prep). All previous studies were pre-registered on Open Science Framework. The first preliminary study in this line of work manipulated participants’ (*N* = 1010) 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 the chance to prepare as much as they wanted 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, where men were 40.8% less likely to choose to prepare for a multiplication task compared to women when offered the opportunity (OR = .59; 95% CI for odds ratio [.46, .76], p < .001). At the end of the experiment, participants were incentivized to state which gender they believed would be more likely to prepare. Both men (77.61%) and women (88.35%) believed that women would spend more time preparing for the task (χ2 = 449.78, *p* < .001), with similar results when asked which gender prepares more in general, where 85.9% of men and 92.4% of women believed women generally prepare more (χ2 = 10.38, *p* < .01). 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 exert effort by choosing to prepare more often. In fact, women chose to prepare more often than men while controlling for their choice to compete (95% CI for odds ratio [1.35, 2.34], *p* < .01) (see Figure 1). Notably, there was no interaction between gender and choice to compete on the choice to prepare (*p* = .97).

In a follow-up study examining the role of forced preparation on the decision to compete with 1026 participants from MTurk, we replicated the effect of gender on preparation, where 42% of women and 35.7% of men chose to complete the optional preparation (*p* < .05). In fact, even in the condition where participants were forced to prepare for 12 rounds lasting on average 2 minutes, there was a marginally significant effect of gender on the choice to prepare (OR = 1.41; 95% CI for odds ratio [.97, 2.04], p = .0706). Again, we find that these behaviors align with participants’ expectations that women will prepare more for the task (χ2 = 391.77, *p* < .01). Also, we find that women prepare more while controlling for their choice to compete (95% CI for odds ratio [1.04, 1.82], *p* < .05) (see Figure 2). 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, despite competing less often. Although we do not find that women who choose to compete are more likely to prepare, 1) we did not manipulate the payment scheme, so there were clear selection effects on one’s choice to prepare across payment schemes and 2) only 11% of women within the study chose to compete, so there was little power to detect any possible interaction effect. 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.

## Open questions

We expect gender differences in the choice to prepare before performing will be exacerbated when individuals are required to compete, given women’s greater tendency to avoid competitions entirely, perform more poorly in competitions, and stop competing after failure (Buser & Yuan, 2019; Gill & Prowse, 2014; Gneezy et al., 2003; Niederle, 2017; Niederle & Vesterlund, 2007, 2011). If women cannot avoid ubiquitous competitions in the labor market, they may try to cope by preparing more, which may increase their confidence or reduce the perceived risk of the competition. Our proposed Study 1 will manipulate participants’ payment scheme (i.e., competitive or non-competitive) to examine whether competition exacerbates previously established gender differences in effort.

Previous evidence suggests that the effect of gender on competitiveness varies by task type, where women are far more likely to choose to enter a tournament when they are completing a verbal task compared to a math task (Apicella & Dreber, 2015; Dreber et al., 2014; Grosse & Riener, 2010), suggesting an important boundary condition of the gender gap in competitiveness driven by gender stereotypes about ability. Beliefs about gender differences in ability likely contribute to persistent occupational segregation. For instance, Cheryan et al. (2017) argue that gender stereotypes about ability affect differences in the representation of women across STEM fields. Occupational gender segregation explains gender differences in wages (Blau & Kahn, 2017; Levanon, England, & Allison, 2009), so it is important to identify how beliefs about gender differences in ability shape decisions to enter competitive environments. To this end, Study 2 will manipulate participants’ beliefs about gender differences on the task under a competitive payment scheme, where participants will be under the impression that women outperformed men, men outperformed women, or, in the control condition, will not be provided information about gender differences in performance. Like Study 1, we expect an interaction between gender and competition choice, where women will spend more time preparing than men when they believe men performed better during a previous iteration of the study and in the control condition, but this effect will be reduced when they believe women performed better. Since previous research suggests confidence and risk aversion are relevant factors in one’s decision to compete (fo review, see Niederle & Vesterlund, 2011), we will include exploratory analyses testing whether confidence or risk aversion interaction with gender and condition.

## Proposed research

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

#### Procedure

Participants (*N* = 3250; see sample size justification below) 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 (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.

**Manipulation:** Participants will be randomly assigned to follow either a competitive or noncompetitive payment scheme for one round (2 minutes) of multiplication problems conditional on their indicated gender (to guarantee women and men are represented at similar rates in each condition). 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 allow us to have more power in detecting the proposed interaction effect between condition and gender in predicting decision to prepare, we anticipate interpreting the effect of the interaction on time spent preparing for whichever payment scheme is presented second (since the payment schemes would be counterbalanced) would be difficult considering the number of factors that could affect the decision (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). Thus, we would only be able to confidently interpret the results for whichever condition were presented first. For this reason, we have opted to use a between-subjects design.

**Dependent variable:** Participants will have the option to complete unlimited “preparation” problems, which they will be told might 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. We will also include a measure of the number of problems participants complete in the amount of time they spent preparing (including both correct and incorrect responses, since completing any problems, regardless of one’s accuracy, is considered preparation) as a secondary dependent variable. Replicating the effects using the secondary dependent variable will serve as a robustness check for the slider scale measure, in case participants do not actually prepare as much as they anticipate they will. 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 choosing their unlimited practice time, participants will complete the task for payment based on their designated pay scheme. 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. Participants will be presented with their absolute (but not relative) performance (i.e., the number of questions they answered correctly). We do not include information about their relative performance since we ask them to guess their relative performance in the confidence measure.

**Post-manipulation measures:** Participants will complete a series of measures after completing the task, which will be used for exploratory analyses. All questions will be counterbalanced. The confidence measure will incentivize ($.25 = 10% of their guaranteed earnings) 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. 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 relative performance is much more relevant for competitive contexts than noncompetitive contexts in the labor market. 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 on the task used within the context of competition. We will also measure risk aversion 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). Notably, 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

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 3 costless strategies (e.g., 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. On top of 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 β1\*Condition + β2\*Gender + β3\*Gender\*Condition 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). We also note that our pilot data where participants were required to complete one round of a task under each type of payment scheme suggests that a small proportion of participants (6%) drop out during the study at all (Richards et al., in prep). For participants who did not finish the pilot study, all participants who performed under the piece-rate payment scheme (which was presented first) also performed under the tournament payment scheme. Thus, our previous work suggests that condition-dependent attrition is unlikely, with the caveat that the participants in the pilot study saw both payment schemes, which is slightly different from the proposed research. To account for the possibility that attrition rates in previous studies do not allow us to infer attrition rates for the proposed research, we take the aforementioned steps to protect the integrity of the study from condition-dependent attrition.

#### Hypotheses and analyses

We will be using two-tailed tests during all hypothesis testing (*p* < .05) and all analyses will be conducted using *R* (version 3.6.3). To control the false-discovery rate during exploratory analyses, we will apply the Benjamini-Hochberg correction to all exploratory analyses.

**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: (Log) Time spent preparing = β1\*Gender + β2\*Condition + β3\*Gender\*Condition, 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 (β3) 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. We will also run a separate linear regression with number of problems completed (including both correct and incorrect responses) as a secondary dependent variable, to check that the results agree. 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., Condition = β1\*Age). 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:** Given the literature suggesting confidence is an important driver of the gender gap in competitiveness (Niederle & Vesterlund, 2011), 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 post-manipulation confidence (operationalized as participants’ projected decile rating) 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: (Log) Time spent preparing = β1\*Gender + β2\*Condition + β3\*Confidence + β4\*Gender\*Condition + β5\*Confidence\*Condition + β6\*Gender\*Confidence + β7\*Gender\*Condition\*Confidence. 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. Given the sample size, a three-way interaction may be underpowered, so the analysis of the three-way interaction will serve as the foundation for future work using the effect sizes found.

**Exploratory analysis 2:** Finally, we will explore whether risk aversion 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 aversion will replace confidence.

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

#### Procedure

Study 2 will follow nearly identical procedures as those employed in Study 1, where participants will see the manipulation, complete the main task, and answer post-manipulation measures, including risk aversion, confidence, and a manipulation check. There are a few notable changes that will be implemented in Study 2. Instead of manipulating the payment scheme, all participants will be required to submit their performance to a tournament, following the same rate of pay as Study 1 (i.e., 4 cents per problem if the participant outperforms a randomly assigned partner). We do not manipulate payment scheme here to reserve power for the main interaction effect of interest between gender and condition on time spent preparing. Stereotypes about gender differences in performance on the task will be manipulated through a statement about gender differences in performance supposedly found in a previous study, where men outperformed women or women outperformed men on the task (Fryer, Levitt, & List, 2008). Notably, 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. On top of the two manipulated conditions, we will have a control condition where participants are not told about any gender differences in performance on the task. If we do not find a significant difference between the conditions that establish gender differences in performance, we will be able to use the control condition to identify whether manipulating gender differences in performance has any effect on time spent preparing, while simultaneously using it as a baseline point of comparison for determining how female-typed and male-typed tasks may have differential effects on the choice to prepare. For the paid task, participants will 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 pilot study online with MTurkers. Since our pilot data suggest 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 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, the pilot study showed that 79.69% of participants believed their score on the matching task would have improved with practice if they had been given the chance (χ2 = 112.81, p < .001). 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 pilot study using the matching task with MTurkers, there was no significant gender difference in performance on the task (95% CI for t [-1.95, 5.86], *p* = .33), and we did not find a gender difference in competitiveness, χ2 (1, n = 546)=0.32, *p*=.574. 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 (χ2 = 63.01, *p* < .001). 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.

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.

After completing the task, participants will complete the measures of risk aversion and confidence from Study 1, along with a manipulation check, where participants are asked to identify whether, on average, our previous study showed that a) men performed better on the task, b) women performed better on the task, or c) there were no gender differences in performance on the task. The presentation of these options will be counterbalanced across participants. Participants will be incentivized to answer all post-manipulation measures at the same rate (i.e., $.25).

#### 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, with a few changes. First, the reference group for the condition variable will be the control 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. For the interaction between gender and the female-typed task condition, we do not have strong a priori predictions about the direction of the effect. It is possible women may still be motivated to prepare since they would not want to perform worse than other women, in which case we would expect the interaction term between gender and the female-typed task condition would be positive, albeit weaker than interaction effect between gender and the male-typed task condition. However, if women feel less motivated to prepare after learning the task is female-typed, the coefficient would be close to zero and nonsignificant, suggesting that believing women perform better on the task will lead women and men to spend relatively similar amounts of time practicing. 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). As in the previous study, we will run a separate linear regression with the number of correct and incorrect responses within the amount of time spent preparing as a secondary dependent variable.

**Exploratory analyses 1 and 2:** Exploratory analyses 1 and 2 will be similar to the analyses in Study 1, with the control 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: (Log) Time spent preparing = β1\*Gender+ β2\*Manipulation check + β3\*Gender\*Manipulation check, 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.

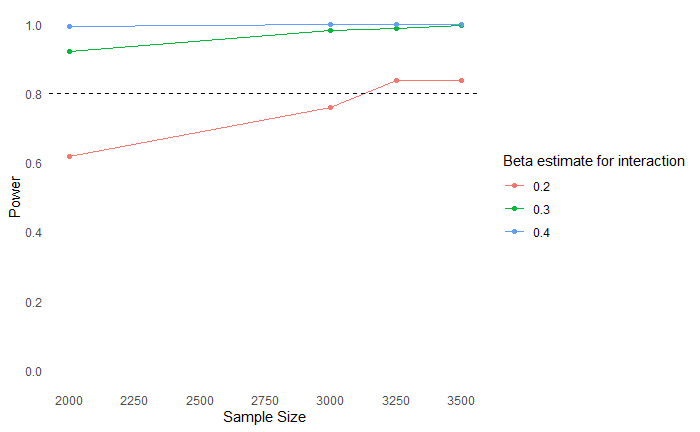
## Future directions

A follow-up study based on these experiments can explore whether the anticipated gender difference in choice to prepare during competition persists even when there is a monetary cost for preparing. 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. Also, if our hypotheses are supported, our research raises the question: Which gender is preparing more (or less) than needed? This should be addressed in research by testing whether gender and time chosen to prepare interact to affect a participants’ probability of winning a competition (see Niederle & Vesterlund, 2007).

## Sample size justification

We estimate power for the hypothesized interaction effect in the primary analyses (see [Simonsohn](http://datacolada.org/17) and [Giner-Sorolla](https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/)).

We ran 1000 simulations while varying the sample size (*N* = 2000, 3000, 3250, 3500) and the effect size for the interaction effect (*b* = .2, .3, .4) (see below). Based on these simulated estimates, we will recruit 3250 participants to achieve 83.7% power for a relatively small effect (*b* = .2).



# Facilities, equipment, and other resources

# Facilities, equipment, and other resources

Keana Richards has support from several knowledgeable mentors, including her adviser, Dr. Apicella, who has extensive experience conducting research in behavioral economics, especially with regards to gender and competition (Apicella et al., 2017a, 2017b, 2011; Apicella, Demiral, & Mollerstrom, 2020; Apicella & Dreber, 2015). On top of that, Dr. Apicella is co-director of the Social and Behavioral Science Initiative (SBSI), which fosters interdisciplinary discussion by hosting weekly brown bags and has opportunities for funding graduate research. Throughout her dissertation, Richards will receive feedback from an interdisciplinary committee, including Dr. Apicella, Associate Professor of Psychology, Dr. Gideon Nave, Assistant Professor of Marketing, Dr. Emily Falk, Professor of Communication, Psychology, and Marketing, and Dr. Geoffrey Goodwin, Associate Professor of Psychology.

All studies will be completed on Qualtrics survey software, which both Dr. Apicella and Richards have access to, along with 24/7 access to the Qualtrics support team. To mediate the recruitment process on Amazon Mechanical Turk, Richards has a Turkprime account, which has features to prevent repeated participation (e.g., checks for duplicate IP addresses), prevent bots from contaminating data collection (e.g., checks for duplicate geolocations), and verify worker information. Additionally, Dr. Apicella and Keana Richards will use the Benjamin Franklin Laboratory and the William Penn Laboratory for participant recruitment, which include 34 computers with Microsoft Office, Adobe Reader Pro, Internet access (Chrome and Firefox) and removable privacy shields, along with 20 noise canceling headphones. Through Dr. Apicella’s affiliation with SBSI, there are two available participant pools for recruitment: the psychology student pool who will receive credit for participating and the paid participant pool maintained by SBSI. Finally, Dr. Apicella and Richards have access to additional support, if needed, from 3 grants managers, 3 IT support specialists, 1 business office manager, 1 undergraduate coordinator, and 6 administrative coordinators.

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

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market. In *Handbook of labor economics* (pp. 3143–3259).

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., Demiral, E. E., & Mollerstrom, J. (2020). Compete with others? No, thanks. With myself? Yes, please ! *Economic Letters*, *187*, 1–3.

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., Dreber, A., Gray, P. B., Hoffman, M., Little, A. C., & Campbell, B. C. (2011). Androgens and competitiveness in men. *Journal of Neuroscience, Psychology, and Economics*, *4*(1), 54–62. <https://doi.org/10.1037/a0021979>

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>

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

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.

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>

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

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., & 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.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some stem fields more gender balanced than others? *Psychological Bulletin*, *143*(1), 1–35.

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

Frank, R. H., & Cook, P. J. (2010). *The winner-take-all society: Why the few at the top get so much more than the rest of us*.

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>

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

Grabill, K., Lasane, T., Povitsky, W., Saxe, P., Munro, G., Phelps, L., & Straub, J. (2005). Gender and study behavior: How social perception, social norm adherence, and structured academic behavior are predicted by gender. *North American Journal of Psychology*, *7*(1), 7–24.

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

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

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.

Hirt, E. R., Mccrea, S. M., & Boris, H. I. (2003). "I know you self-handicapped last exam": Gender differences in reactions to self-handicapping. *Journal of Personality and Social Psychology*, *84*(1), 177–193. <https://doi.org/10.1037/0022-3514.84.1.177>

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.

Lavy, V. (2004). Performance pay and teachers’ effort, productivity and grading ethics. *NBER WORKING PAPER SERIES*.

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

Lenney, E. (1977). Women’s self-confidence in achievement settings. *Psychological Bulletin*, *84*(1), 1–13. <https://doi.org/10.1037/0033-2909.84.1.1>

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>

Levanon, A., England, P., & Allison, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950-2000 U.S. Census data. *Social Forces*, *88*(2), 865–891. <https://doi.org/10.1353/sof.0.0264>

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.

Maccoby, E., & Jacklin, C. (1974). Myth, reality and shades of gray-what we know and don’t know about sex differences. *Psychology Today*, 109–112.

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>

Mccrea, S. M., Hirt, E. R., & Milner, B. J. (2008). She works hard for the money: Valuing effort underlies gender differences in behavioral self-handicapping q. *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.

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

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

Niederle, M. (2014). Gender. *Handbook of Experimental Economics*, *2*, 1–120. Retrieved from [http://web.stanford.edu/{~}niederle/Gender.Niederle.pdf](http://web.stanford.edu/%7B~%7Dniederle/Gender.Niederle.pdf)

Niederle, M. (2017). A gender agenda: A progress report on competitiveness. *American Economic Review*, *107*(5), 115–119. <https://doi.org/10.1257/aer.p20171066>

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>

Power, S., Whitty, G., Edwards, T., & Wigfall, V. (1998). Schoolboys and schoolwork: gender identification and academic achievement. *International Journal of Inclusive Education*, *2*(2), 135–153.

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*.

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>

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>