

1

Chapter 2: INSERT TITLE

1.1 Introduction

1.2 Rising use of performance pay, especially competitive pay over time - and how it may contribute to the gender wage gap/other gender diff in labor market outcomes

Compensation packages based on performance pay, such as bonuses, commissions, and piece-rate payments, have risen in popularity relative to hourly/salaried pay, especially among workers in the highest tiers of occupations [**Hall1998**, **Murphy1999**, **Cunat2005**, **Lemieux2009**]. There is evidence that the increasing use of performance pay lends itself to wage inequality. **Lemieux2009** 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 [**Bell2010**, **Bell2014**, **Benabou2016**]. Importantly, performance pay may contribute to the gender wage gap too. Using data from the National Longitudinal Surveys of Youth, **McGee2015** 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.

1.3 Gender differences in competitiveness as possible contributor to labor market gaps

Since competition is relevant to labor market outcomes, researchers began to focus on how a person's gender affects their competitiveness, both in terms of willingness to enter competitions and response when required to enter a competition. To date, most of the research on gender differences in competitiveness has focused on either i) explaining the sources of the gender difference [**Veldhuizen2017**] or ii) designing interventions to encourage women to compete more [**Balafoutas2012, Sutter2016, Cassar2016, Brandts2015, Niederle2013, Brandts2015, Healy2011, Alan2018**]. Less consideration has been paid to how competitions may differentially, and perhaps negatively, impact women.

The introduction of Chapter 1 provides an overview of the literature on gender differences in willingness to compete, so we will only review the literature on gender differences in response to entering competitive environments here. There are three major time points at which competition may affect men and women differently: before, during, and after competition. The majority of previous studies in this space have examined gender differences in response to competition during and after performance, which we will briefly review here.

NOTE TO COREN: what do you think of these sections breaking down competition into 3 time points and reviewing some of the literature in these areas? Is this a useful way to break it down or do you prefer another way?

1.4 Gender differences in response to competitive environments

NOTE TO COREN: for the following sections, if included, do you have any recommendations on citations to include outside of what is listed below?

1.4.1 During competition

Gender differences in performance during competition: Although competitions are generally motivating and designed to improve performance by increasing effort [Connelly2014a, Murayama2012, Miller2019a], previous 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 [Gneezy2003, Gneezy2004, Gunther2010, Samak2013]. Gneezy2003 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, with males performing better. Gunther2010 replicate the effect of competition on gender differences in performance for a male-typed task, but find no gender differences in performance during competition for female-typed or gender-neutral tasks.

INSERT OTHER CITES: - Evidence that women's performance doesn't strongly respond to competition compared to men: <https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12417> - Suggests that women may not respond well to competitive pressure (aka when stress is kept to a minimum, there are no gender differences in performance, but when certain knock-out rules are applied, a difference emerges): https://www.sciencedirect.com/science/article/pii/S0167268121001785?casa_token=1G3VrTCCNu8AAAAA:ds0sjejPKHnunOTRSqkEHU-odJMjDPhHUBXy-JPX4KqAqrH4bihs5riR7gypyza2Rko_vg - there is a growing literature showing that women are less willing to guess on exams [Pekkarinen2015, Baldiga2014, Iriberri2021], which in turn negatively impacts performance on said exams [Pekkarinen2015, Baldiga2014] - which they argue may driven by women being

less confident in their probability of answering correctly or being more risk averse. **Riener2018a** suggests this phenomenon starts at an early age, with girls as young as 8 years of age being significantly less willing to guess on exams relative to men - **Paserman2007**: “Data on serve speed, on first serve percentages and on rally length suggest that women play a more conservative and less aggressive strategy as points become more important.”

1.4.2 After competition

Gender differences in response to losing: 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 [**Gill2014**]. Other research suggests women stop competing altogether after losing if given the choice. **Buser2019**, who examine the effects of losing while competing in the Dutch Math Olympiad on the choice to compete in subsequent years, show that men are just as likely to compete even if they lost the previous year, while women are less likely to compete again if they lost before. Overall, this body of literature suggests that competitions may differentially impact women and men, both during and after the competition.

INSERT OTHER CITES - negative feedback increases women’s likelihood of dropping out of their major: **Astorne-Figari2018** - <https://drive.google.com/file/d/1eMZJpkqa0QvDhcf76r2U8bkuelVa1Byt/view>: “We find that, among assistant professors, a flat rejection reduces the confidence in publishing the paper in any leading journal to a significantly greater extent for women than it does for men. We find no gender differences among associate and full professors, likely due to survivorship bias.” - https://www.nber.org/system/files/working_papers/w29382/w29382.pdf: “We find that, holding fixed performance and decisions before feedback, women update their beliefs and choices more negatively than men do after bad news.”

1.4.3 Before competition

As mentioned previously, little research has examined how competitions may affect gender differences in behavior during arguably the most critical period: before an individual enters a competition, where they have the most control of their subsequent performance in the competition. Given previous research suggesting that women and men may respond differently during and after competitions, we expect that they will also employ different behaviors and have different perceptions of themselves and others in advance of a competition.

We only know of a few studies that explore this open question: insert possible cites if relevant

Preparation as a coping strategy before competition & possible mechanisms

Preparing for a competition, through either practicing or studying, is a potential response to entering a competition. Since competitions, by definition, compare the performance among two or more individuals, they naturally lead to self-evaluation and comparative judgments of self with others - 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, which naturally invoke self-other assessments. Thus, less confident individuals may prepare more. Moreover, they may prepare more in order to reduce the negative feelings caused by low confidence independent of any ambitions to win, since mastery is an important driver of confidence [Gist1992, Usher2008]. 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, at minimum, an equal desire for self-improvement [Apicella2017a]. Similarly, given the inherent risk of competitive payment schemes relative to non-competitive payment schemes, it is possible that

the aforementioned gender differences in risk attitudes may also lead women to be more likely to cope by preparing before performing in a competition relative to men.

Gender stereotypes as a possible mechanism

A novel prediction deriving from the results showing robust perceptions of gender differences in preparation across all studies in Chapter 1 is that gender differences in preparing may be driven by persistent stereotypes men and women's tendencies to prepare before performance.

Prominence and characteristics of gender stereotypes

- Gender stereotypes derive from observers' automatic tendency to make correspondent inferences about men and women's dispositions [**Gilbert1995, Ross1977, Jones1967, Gawronski2004**]. These correspondent inferences have led to prominent gender stereotypes that exist across cultures [**Williams1990, Williams1982, Steinmetz2014, Fiske2017**].
- Stereotypes involve prescriptive, proscriptive, and descriptive components [**Prentice2002**], where prescriptive and proscriptive stereotypes reflect cognitive representations of the characteristics women and men should and should not have, respectively, while descriptive stereotypes are representations of the typical man and woman [**Burgess1999**].
- Gender stereotypes encompass a variety of attributes. For instance, there are physical (e.g., women are dainty), cognitive (e.g., men are analytical), and personality-based (e.g., women are nurturing) stereotypes [**Cejka1999, Deaux1984**].

Implications of gender stereotypes for behavior

- Importantly, there is evidence that gender stereotypes can affect behavior [INSERT cites]. For instance, **Coffman2014a** show that both men and women are less likely to contribute ideas to a group decision in gender-incongruent decision-making domains (e.g., women contributing ideas to a decision on

sports), even when the group would have made a better decision with their contribution. [INSERT other example showing stereotypes affecting behavior].

Given the power of gender stereotypes in affecting subsequent behavior, we expect that our findings of robust perceptions of gender differences in preparation likely contributes to gender differences in actual preparation behavior.

1.5 The current experiment

Overall, women may engage in more coping strategies than men, such as preparation, before entering competitions because they tend to be more risk-averse [Croson2009, Dohmen2011b, Eckel2008, Bertrand2010a] and less confident [Bertrand2010, Lundeberg1994, Mobius2011, Barber2001, Croson2009], and/or may be adhering to gender stereotypes [insert coffman etc cites]. In support of this possibility, in Chapter 1 of this dissertation, we found evidence of a sizable gender difference in preparation, where women were more likely than men to choose to prepare before completing a multiplication task.

Here, we focus on how women and men differentially respond to competition through preparation. We expect to see both gender differences in actual preparation behavior, along with gender differences in perceptions of relative preparation when men and women are required to compete (relative to non-competitive environments). More specifically, in the study included in this chapter, we tested whether competition exacerbates previously established gender differences in preparation by manipulating participants' assigned payment scheme (i.e., competitive or non-competitive). We hypothesize that women will choose to practice problems at a higher rate than men, especially when assigned to the competitive tournament payment scheme (i.e., we anticipate a main effect of gender on practice, and an interaction between gender and condition, such that women will practice more than men in both conditions, but the difference-in-differences between practicing rates across genders will be greater in the competition condition). To be clear, there was no interaction between gender and choice to compete on the choice to prepare in any

of the previous studies from Chapter 1. That is, women prepared more than men regardless of which payment scheme they had chosen. However, it is not possible to draw conclusions from this because i) we did not manipulate the payment scheme, so there could have been selection effects on one's choice to prepare across payment schemes, such that those who were more likely to choose to compete may have been less likely to prepare, and ii) there was little power to detect any possible interaction effects. For instance, in the first study, only 11% [DOUBLE CHECK] of women chose to compete, and in the second, INSERT% of women made this decision. Through the proposed experiment, we intend to expand upon the studies in Chapter 1 by directly manipulating participants' payment scheme and recruiting a large sample to provide power to detect small effects.

We also tested whether gender predicts participants' perceptions of their relative amount of preparation, given our hypothesis based on Study 3 of Chapter 1 that women may be especially susceptible to feelings of underpreparation relative to others when they have unlimited time to prepare. More concretely, we expect women will be more likely to assume they practice less than others compared to men (that is, the effect of gender on perceived practice deviation will be negative), especially when assigned to the competitive tournament payment scheme (such that women in general will think that they practice less than other participants than men, but this difference will be exacerbated in the competition condition).

The research design, hypotheses, measures and analyses for this chapter were pre-registered on OSF and all analyses were conducted in R statistical software (version 4.0.4).

1.5.1 Methods

note to self: pulled from nsf app 01_project-description.Rmd

2

Study 1: Does competition elicit gender differences in effort?

Participants

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.

Given the difficulty of powering interaction effects [Simonsohn2014, Giner-Sorolla2018], we conducted a power analysis to determine an adequate sample size for the main hypothesized interaction effect in the primary analysis [Hughes2017a]. 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).

Will also note number of ppts that dropped out during/after learning about condition to asses condition-dependent attrition: `clean %>% filter(Finished == "FALSE" & !is.na(condition)) %>% select(Finished, condition)`

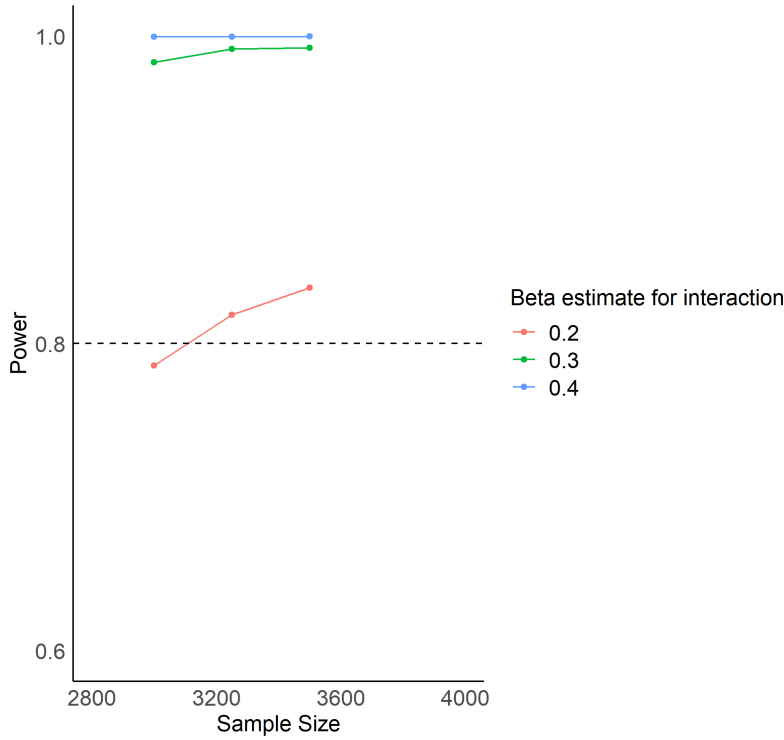


Figure 3. Plot of simulation output used to determine necessary sample size for at least 80% power in both proposed studies.

Manipulation

Participants will be randomly assigned to follow either a competitive or non-competitive 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 carryover effects. 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 carryover effects 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 prepare for the task by completing unlimited practice 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. We chose a multiplication task because we expect participants will improve with practice. Indeed, research suggests that rehearsing and recalling associative memories can speed up retrieval of those memories [Rundus1971]. Moreover, we have already established a robust gender difference in both the choice to prepare and compete using this task (Richards et al., in prep). For participants who agree to practice, they will be able to practice for as long as they want, with the option to pause in case of any unexpected interruptions, such as children coming into the room. Also, participants will have the option to exit the preparation and move onto the task at any point via an “Exit” button in the bottom right corner of the survey screen. The dependent variable will be quantified as the total number of seconds of preparation, excluding the amount of time participants paused during the practice.

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 payment scheme will not be told whether they won, since this serves as an indicator of relative performance.

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. 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 [**Niederle2007**], but instead of asking participants to indicate whether they won against a randomly selected opponent, we ask them 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 [**Oney2015**], 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?” [**Dohmen2011b**]. There is evidence that risky behavior (i.e., lottery choices) is strongly associated with the risk measure included in the current proposal [**Dohmen2011b**]. 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 predictive of behavior [**Dohmen2011b**]. These measures are

included after completing the task largely because the confidence measure requires participants to state their perceived relative performance on the task.

make sure to mention self-rated decile variable and how perceived practice deviation variable will be created from that.

include extra_practice_count var description: number of times continuing to practice after having seen what the practicing/studying looks like & gone through it at least once. - if they didn't choose to practice at all = 0 - if they chose to practice but did not practice thereafter = 0 - if they said they want to practice again after having said yes the first round = 1

2.0.1 Results

2.1 Describing main variables of interest

First, we confirmed that condition was assigned evenly across participants (control= 50.21%). Of the men who completed the study, 50.38% were assigned to the control condition and of the women who completed the study, 49.72% were assigned to the control condition.

Like all studies in the first chapter, we replicate the effect of gender on risk, $b = -0.86$, 95% CI $[-1.01, -0.70]$, $t(3920) = -10.71$, $p < .001$, and confidence, $b = -8.46$, 95% CI $[-10.12, -6.79]$, $t(3978) = -9.96$, $p < .001$, such that women were more risk averse and less confident relative to men.

Contrary to the majority of studies in the first chapter, we find a significant effect of gender on task score, $M_{\text{women}}=10.45$, $SD=4.47$; $M_{\text{men}}= 12.29$, $sd=7.28$, even when including risk, confidence, and an interaction between gender and condition in the model, $b = -1.38$, 95% CI $[-1.89, -0.86]$, $t(3916) = -5.20$, $p < .001$ (NOTE: interesting that the sd for men appears to be higher than for women - aka men may have been significantly more likely to fall into both extremes of the performance distribution - Emily is this worth mentioning?). We explore this finding further in the discussion section for this study.

Next we explored the characteristics of the main practice variables in the dataset. 45.51% of all participants chose to practice, with 48.22% choosing to practice in the piece-rate payment condition and 51.78% choosing to practice in the tournament payment condition. This difference in the choice to practice across conditions is significant when condition is included as a predictor alone, $b = 0.15$, 95% CI [0.02, 0.28], $z = 2.35$, $p = .019$, but in the subsequent section we explain how the effect changes when including other predictors in the model. Participants spent an average of 29.12 seconds practicing across all rounds of practice and of those who chose to practice, completed 0.14 total rounds of extra practice problems (COME BACK TO THIS ONCE DECIDED).

2.2 Effects of gender and condition on both practicing and perceptions of one's relative practicing

Through a logistic regression with gender, condition, and the interaction between the two predicting the binary choice to practice problems, we replicate the effect of gender on the choice to practice found in Chapter 1, $b = 0.51$, 95% CI [0.33, 0.69], $z = 5.49$, $p < .001$, where 0.51% of women chose to prepare relative to 0.38% of men (see Figure 2.1). However, we do not find an interaction between gender and condition, $b = 0.08$, 95% CI [-0.18, 0.34], $z = 0.60$, $p = .547$, contrary to our hypothesis that the gender difference in the choice to prepare would be exacerbated under the tournament payment scheme relative to the piece-rate payment scheme. Additionally, the aforementioned effect of condition on the choice to practice is no longer significant in the model including these additional predictors, $b = 0.10$, 95% CI [-0.09, 0.30], $z = 1.03$, $p = .302$.

As part of our pre-registered analyses, we also explored other measures of practice to test the robustness of the effect of gender on practicing. We find that women, relative to men, completed a significantly higher number of practice problems, $b = 0.39$, 95% CI [0.35, 0.43], $z = 18.54$, $p < .001$, more rounds of extra practice

$b = 0.62$, 95% CI [0.35, 0.91], $z = 4.31$, $p < .001$, and spent more time completing practice problems, $b = 13.12$, 95% CI [7.95, 18.28], $t(3959) = 4.98$, $p < .001$.

Based on previous literature on risk aversion and confidence affecting competition entry, we expected participants with especially high levels of risk aversion and/or low levels of confidence would be especially likely to choose to practice before entering a competition, and that this effect may interact with gender. Thus, we tested possible three-way interactions between gender, condition, and risk aversion or confidence on the choice to practice problems through two pre-registered logistic regressions, but did not find evidence that risk aversion, INSERT, nor confidence, INSERT, interacted with gender and condition.

2.3 Accuracy of levels of practicing based on participant gender

Next, we ran a linear regression with gender, condition, and the interaction between those two variables predicting the aforementioned perceived practice deviation variable (that is, subtracting each participants' percentile based on number of practice problems completed from their self-rated decile) to test our second hypothesis that women would be more likely to assume they practice less than others compared to men, especially under the competitive tournament payment scheme. Though we did not find evidence of the anticipated interaction effect, $b = -3.25$, 95% CI [-10.26, 3.76], $t(3959) = -0.91$, $p = .364$, we find a significant effect of gender on perceived practice deviation, such that women (relative to men) were significantly less likely to assume they practice more than others, $b = -14.49$, 95% CI [-19.43, -9.55], $t(3959) = -5.75$, $p < .001$, $M_{\text{women}}=23.56$, $SD=56.11$; $M_{\text{men}}= 39.69$, $sd=54.87$ (see Figure 2.2). We performed a more targeted exploratory analysis to see if the effect held when participants have actually practiced (and as a result, the question about their relative practicing may feel more relevant), and find that among this subset of the data, women (again, relative to men) were still

significantly less likely to believe that they practiced more than others, $b = -3.25$, 95% CI $[-10.26, 3.76]$, $t(3959) = -0.91$, $p = .364$.

Since this is the first time we have used the perceived practice deviation variable and are not able to attest to its robustness, we also explored another way of testing this hypothesized effect by using participants' self-rated decile as the dependent variable instead of perceived practice deviation and then controlling for number of practice problems attempted in a linear regression. We find that, regardless of the number of problems practiced, women are significantly less likely to say they practice more than others compared to men, INSERT, although this effect does not hold when focusing on the subset of participants who chose to practice, INSERT.

On top of the differences in how much women and men in this study perceived they practiced relative to others, we also pre-registered tests of men and women's accuracy of their relative practice through a series of t-tests comparing the perceived practice deviation variable to 0 (which would represent a participant guessing their exact decile correctly). Across the full dataset, most participants tended to overestimate how much they practiced relative to others, INSERT. After honing in on each gender included in the study, we find that this effect holds among both women, INSERT, and men, INSERT. Notably, participants who chose to practice significantly underestimated their relative practice, both among women, INSERT, and men, INSERT.

We also explored how self-rated decile changes based on whether participants were asked to compare their amount of practicing to men or women in the study specifically, and find that participants' perceptions of how much they practiced relative to women in the study are significantly lower than perceptions of much they practiced relative to men, $M_d = -8.09$, 95% CI $[-9.03, -7.14]$, $t(3,979) = -16.85$, $p < .001$.

2.4 Perceptions of gender differences in behavior

Across all measures of perceptions of gender differences in behavior, we replicate effects found in the previous studies. First, the majority of participants (59.57%)

said that women would be more likely to practice/study for the task, which was significantly higher than the proportion of participants that said men would be more likely to practice/study than women (4.73%) and the proportion of participants that said there was no difference in the likelihood that men and women would practice/study (35.7%), $\chi^2(2, n = 4172) = 1,782.43, p < .001$.

Similarly, participants were significantly more likely to say that women prepare more than men in general (68.28% of participants), $\chi^2(2, n = 4172) = 2,464.02, p < .001$, relative to the proportion of participants that said men prepare more than women (4.41% of participants) or that there is no difference in how much men and women prepare (27.31% of participants).

Yet, participants did not expect a gender difference in performance on the main multiplication task used, $\chi^2(2, n = 4172) = 781.11, p < .001$, where 54.17% of participants said that there was no difference in how many multiplication problems men and women correctly solved, while 20.56% said men correctly solved more multiplication problems than women and 25.27% said women had a performance advantage over men.

Finally, 64.24% of participants expected women would be more likely to choose the piece-rate payment scheme than the tournament payment scheme, which was a significantly higher proportion of participants than those who expected women would choose each payment scheme equally 20.9% and than those who expected women would choose tournament more often than piece rate, 14.86%, $\chi^2(2, n = 4172) = 1,707.40, p < .001$. On the contrary, when asked about how much men in the study would compete, a significant majority of participants (63.5%) expected men to be more likely to choose the tournament payment scheme over the piece-rate payment scheme, relative to the proportion of participants who said men would choose each payment scheme equally (15.8%) and the proportion who said men would choose piece rate more often than tournament (20.7%), $\chi^2(2, n = 4172) = 1,620.33, p < .001$.

2.5 Figures

40 -

deviation

30

40

20

