### Methods

## Participants

Participants were recruited on Amazon Mechanical Turk using the same screening criteria as all previous studies in Chapter 1. Like the last study of Chapter 1, we used Qualtrics’ fraud detection software to filter out responses that were suspicious either because they were likely 1) bots and/or 2) duplicate responses using the same exclusion criteria from before. These exclusions were applied for all main analyses reported in the results section.

The final dataset consists of 3980 participants (57.36% women), with an average age of 41.3 (*SD* = 13.2) years. Of the final sample, 75 participants (30.67% women) dropped out of the study before finishing and 192 participants were flagged by Qualtrics’ fraud detection software as suspicious based on the aforementioned criteria. We include analyses for the full sample in the appendix and all results are unchanged (INSERT DOUBLE CHECK).

## Procedures

Participants included in the study were told they would be completing a multiplication task. Notably, we aimed to recruit a larger sample to grant us enough power to detect our anticipated interaction effects, and decided to shorten the task from two minutes to 30 seconds. Otherwise, the task used was identical to the ones used in previous studies.

Like the studies in Chapter 1, participants were first told about the rules for the multiplication task and were required to pass the same comprehension questions used in the previous studies before moving onto the main manipulation of payment scheme.

### Manipulation of payment scheme

Unlike previous studies, participants were not able to choose a payment scheme. Instead, they were told about their random assignment to one of two payment schemes: the non-competitive piece-rate payment scheme or a competitive tournament payment scheme. Men and women were evenly assigned to both conditions. If they were assigned to the piece-rate payment scheme, they were paid $.10 per problem solved correctly. If they were assigned to the tournament payment scheme, they were randomly matched with another participant that was also assigned to that payment scheme and received $.20 per problem if they solved more problems than the other participant. Otherwise, they received nothing.

Again, we checked that condition was assigned evenly across participants (control= 50.21%) and genders included in the study. 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. We also assessed condition-dependent attrition by identifying the number of participants that dropped out during/after learning about condition and find that a relatively small proportion of participants out of the total sample dropped out after learning about their respective condition (*N* = 42). Given the small sample that dropped out relative to the total number of participants in the study, we are not concerned that condition-dependent attrition is driving any of the effects found in this study.

### Main dependent variables of interest: Measures of preparation and perceptions of relative preparation

After they were informed of their payment scheme, all participants were given the opportunity to spend unlimited time preparing before completing the paid multiplication task. The nature of the unlimited preparation was identical to that used in Study 3 of Chapter 1, where participants who chose to prepare were shown 10 multiplication problems that were created randomly by drawing from the pool of numbers used in the main multiplication task. Unlike the last study in Chapter 1, participants were not asked to explicitly indicate whether they would like to study the times tables. Instead, they were shown the times table right after the practice problems directly on the practice problems page and told they could check their answers using the table as desired. By including the option to check their answers, we hoped to make the practice itself more useful by providing participants a way to receive feedback on their responses. At the bottom of each practice page, participants were asked if they would like to continue practicing multiplication problems, with the option to continue as many times as desired or opt out at any point. The amount of time (in seconds) participants spent on each practice page was also recorded. Thus, like the previous studies, we have multiple measures of preparation by design: 1) the actual decision to practice problems (before knowing what the practice entails), 2) among participants who chose to practice problems, the number of practice problems participants attempted (quantified as number of practice problems not left blank, irrespective of accuracy), 3) among participants who chose to practice problems, the amount of time they spent across all practice rounds they completed, and 4) the number of extra practice rounds participants completed after having completed the first round of practice. Since the practice structure in this study is identical to that of Study 3 in Chapter, the number of extra practice variable was encoded in the same way as that study.

After completing the practicing/studying, participants guessed how much their amount of practicing for the task compared to all other participants that completed the task by indicating the decile of their practice relative to other participants. We also asked participants to indicate their anticipated decile when their amount of practicing was compared to that of all participants who identified as men and women, respectively.

We used these the decile questions to create the perceived practice deviation variables as follows: self-rated decile (either based on the question about practicing relative to all other participants, relative to only men, or relative to only women) - actual percentile based on number of practice problems completed. Therefore, negative values for this variable indicate a participant expected to have practiced less, relative to other participants, than they actually did, and vice versa for positive values. A value of zero, therefore, indicates that a given participant was completely accurate in their guess of relative practicing.

### Paid multiplication task and post-task measures

After practicing, participants completed the paid multiplication task, received feedback about their absolute (but not relative) performance, and completed many of the same follow-up questions used across Chapter 1, including risk attitudes, confidence, and perceptions of gender differences in preparation, competitiveness, and performance. One of the perceptions of gender differences questions deviated slightly from the previous studies, which was edited for the sake of clarity. Instead of asking participants to indicate “Do you think men or women in this study chose the tournament payment option more often?”, they were asked “If given the opportunity to choose between the two payment schemes (Piece Rate or Tournament), do you think men in this study would choose the piece rate or the tournament payment scheme more often?”, with the options to indicate: “Men would choose tournament more often than piece rate”, “Men would choose piece rate more often than tournament”, or “Men would choose each payment scheme equally”. This question was repeated with respect to women in the study.

We paid participants to answer the questions about their confidence and perceptions of gender differences correctly at the same rate as previous studies. Finally, they completed the same demographic questions from Chapter 1 and provided feedback on the study before being paid for their participation.

### Results

## Describing main variables of interest

First, 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, , 95% CI , , , , 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 (that is, rounds of practice after having seen what the practice looks like).

Like all studies in the first chapter, we replicate the effect of gender on risk, , 95% CI , , , , and confidence, , 95% CI , , , , 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, Mwomen=10.45, SD=4.47; Mmen= 12.29, SD =7.28, even when including risk, confidence, and an interaction between gender and condition in the model, , 95% CI , , , (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.

## 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, , 95% CI , , , , where 50.74% of women chose to prepare relative to 37.62% of men (see Figure @ref(fig:s400)). However, we do not find an interaction between gender and condition, , 95% CI , , , , 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, , 95% CI , , , .

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, , 95% CI , , , , more rounds of extra practice , 95% CI , , , , and spent more time completing practice problems, , 95% CI , , , .

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.

## 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, , 95% CI , , , , 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, , 95% CI , , , , Mwomen=23.56, SD=56.11; Mmen= 39.69, sd=54.87 (see Figure @ref(fig:s401)). 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 have felt 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, , 95% CI , , , .

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 (as a proxy for more precise estimate of amount of practicing) in a linear regression. We find that, regardless of the number of practice problems attempted, 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, , 95% CI , , , .

## 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%), , .

Similarly, participants were significantly more likely to say that women prepare more than men in general (68.28% of participants), , , 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, , , 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%), , . 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%), , .

### Discussion

## Main hypotheses:

The most notable finding in this study comprising Chapter 2 is that we replicate findings from the studies in Chapter 1 that women choose to prepare more than men. Interestingly, women chose to prepare more regardless of the payment scheme they were randomly assigned to. Also, we did not find evidence that assignment to either a tournament or piece-rate payment scheme significantly predicted the binary choice to practice problems, after including gender and the interaction between gender and condition in the model. Although we did not pre-register a hypothesis that condition would be a significant predictor of the choice to practice, it is nonetheless important to note that gender explains participants’ decision to practice problems over and above any effect of condition.

We also pre-registered other means of quantifying preparation (i.e., amount of time spent on the pages with practice problems and study tables, number of practice problems completed, and rounds of extra practice problems completed) to test the robustness of the gender effect, and find evidence across our multiple measures of preparation that women tended to prepare more than men.

One important consideration when interpreting the effect of gender on the choice to prepare before the task is that we find a significant effect of gender on task score even while controlling for risk and confidence, unlike two out of the three studies in the last chapter. It is possible that shortening the task contributed to this effect - especially considering the evidence suggesting that women’s performance may suffer under more competitive pressure [cites]. There may be less pressure to perform well during a two-minute task (used in all of the studies of Chapter 1) relative to a 30-second task (used in the study in this Chapter). In support of this possibility, @Shurchkov2012 shows that womens’ performance significantly improves, to the extent to that they outperform men, in a low time pressure competition.

We also found evidence for the hypothesized main effect of gender in our other primary pre-registered analysis, where women were more likely to assume they practice less than others compared to men. We found this effect held when using our pre-registered version of the analysis using the perceived practice deviation variable, representing the accuracy of participants’ guess of how much their level of practicing compared to others participants’ level of practicing, both among the full set of participants and among the subset of participants that chose to practice. We also wanted to test the robustness of the effect using a slightly different way of quantifying our relationship of interest, where we included participants’ raw self-reported practice decile as the main dependent variable of interest (instead of perceived practice decile) with gender and number of practice problems attempted as predictors. We replicate the aforementioned effect, where women tended to think they practice less than others, regardless of how many practice problems they actually attempted - although this effect only held among the full set of participants and not among the subset of participants who chose to practice.

Like the first primary hypothesis in our pre-registered analysis, we did not find the hypothesized interaction between gender and condition on perceived practice deviation - suggesting that, like actual decisions to practice, women’s tendency to perceive they are practicing less than others is not strongly affected by whether they are competing or not. We explore possible implications of the null interaction between gender and condition further in the subsequent general discussion summarizing results across all studies of the dissertation.

## Perceptions of gender differences in performance, competition, and preparation

With respect to the questions asking participants to indicate their perceptions of gender differences about our main behavioral variables of interest, we replicate findings from all three studies in Chapter 1. Even though participants expected that women and men would not have a significant difference in task scores, they expected men to prefer the tournament payment scheme over the piece-rate payment scheme, while expecting women to both i) prefer the piece-rate payment scheme over the tournament payment scheme and ii) prepare more, both before completing the multiplication task used in this study and in general before most tasks. Again, with the exception of the general gender difference in practice questions, all of the other perception questions were incentivized for accuracy to reduce socially desirable responding. Our exploratory analysis of the new set of questions about perceptions of relative practicing compared to each gender included in the study of this Chapter support these general perceptions of gender differences in preparation. Given the targeted nature of the questions, we were able to test how participants’ responses changed based on whether they were asking to compare their level of practicing in the study to only participants that identified as women or only participants that identified as men, and find that participants were significantly more likely to indicate that they practiced less relative to women than relative to men.

## Summary of takeaways

Overall, our results for the study in Chapter 2 suggest that women prepare more than men, regardless of whether they were assigned to a competitive tournament or non-competitive piece-rate payment scheme, and despite thinking they practice relatively less than men for the multiplication task used in the study. It is possible that gender stereotypes are driving these gender differences in behaviors and perceptions, given our replication of the findings from all three studies in Chapter 1 that participants expected women to prepare more both before the specific task used in the study and in general, along with the finding that participants’ tended to rate their relative practicing significantly lower when comparing themselves to women than men.