### Methods

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

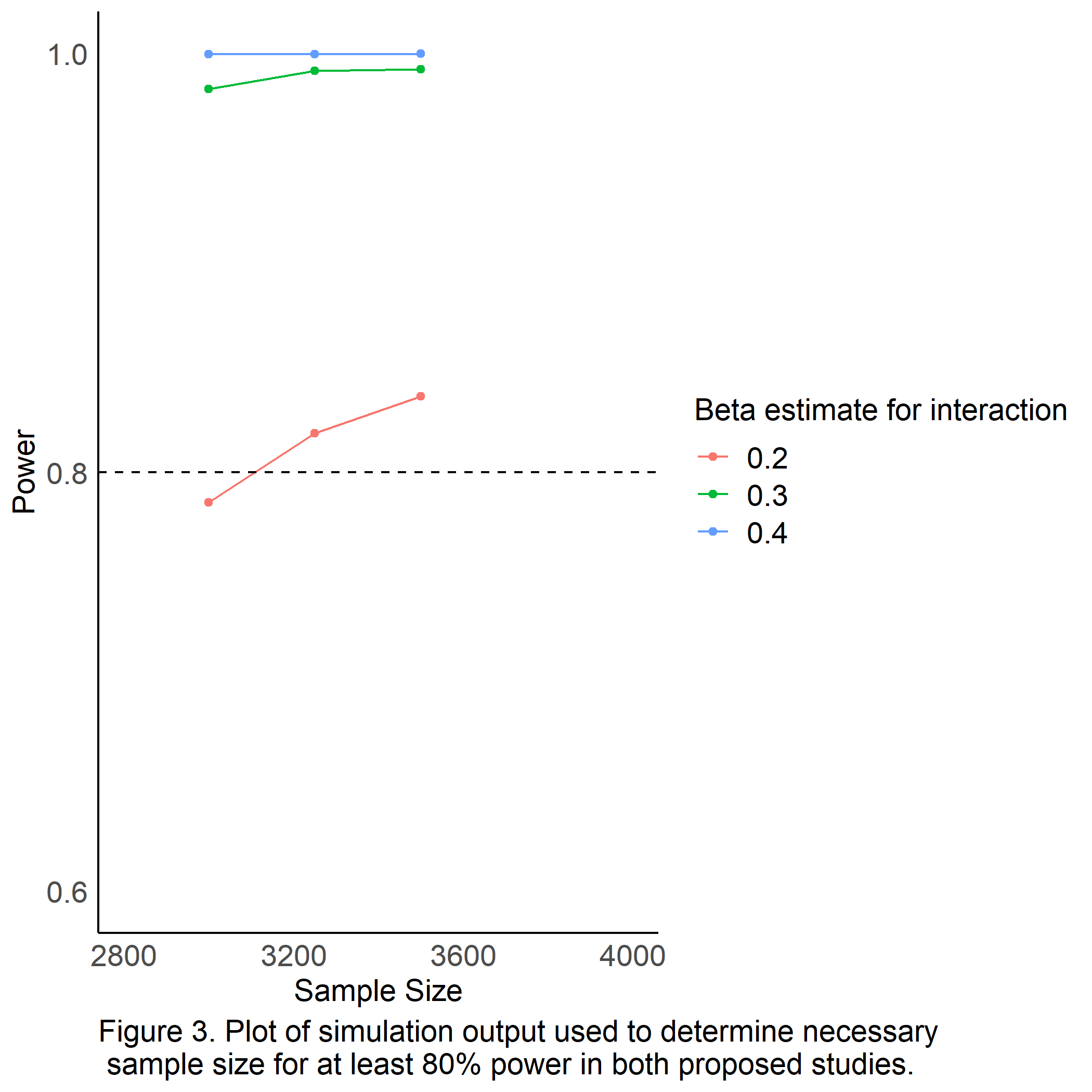
# Study 1: Does competition elicit gender differences in effort?

*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 [see @Simonsohn2014; @Giner-Sorolla2018], we conducted a power analysis to determine an adequate sample size for the main hypothesized interaction effect in the primary analysis [simulations modeled after code from @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)



*Manipulation*

Participants will be randomly assigned to follow either a competitive or noncompetitive payment scheme for one round (2 minutes) of multiplication problems. The payment scheme will be manipulated between subjects, where participants in the competition (tournament) condition will be paid 4 cents per problem on the task, but only if they beat another randomly assigned MTurker, while participants assigned to the noncompetitive (piece-rate) payment scheme will be paid 2 cents per problem. Although a within-subjects design would provide more power in detecting the hypothesized interaction effect, we opted to use a between-subjects design to avoid 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 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 [see @Oney2015 for review], the confidence measure assesses participants’ beliefs within the context of the task used. We will also measure risk attitude by asking participants to indicate on a 0-10 scale “How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?” [@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.

### Results

## 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, , 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 , , , .

Next we explored the characteristics of the main practice variables in the dataset. 45.51% of participants chose to practice, with 0.48% choosing to practice in the piece-rate payment condition and 0.52% choosing to practice in the tournament payment condition. 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).

## Effects of gender and condition on both practicing and perceptions of one’s relative practicing

Through a logistic regression, we replicate the effect of gender on the choice to practice found in Chapter 1, , 95% CI , , , , where 0.51% of women chose to prepare relative to 0.38% of men (see Figure @ref(fig:INSERT)). 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.

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, INSERT exploratory4a, more rounds of extra practice (INSERT, COME BACK TO THIS ONCE DECIDED), and spent more time completing practice problems, INSERT.

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 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 (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 compared to men, , 95% CI , , , , Mwomen=23.56, SD=56.11; Mmen= 39.69, sd=54.87 (see Figure @ref(fig:INSERT)). 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, , 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 in a linear regression. Again, we find that 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 tested 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 were significantly more likely to underestimate 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 in general (68.28% of participants) than men (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%), , .