

## **Participants**

All study measures described below are publicly available on OSF both as a .pdf and .qsf. 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.

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 aimed to recruit 4000 participants to achieve at least 80% power for a relatively small effect ( $b = .2$ ) (see Figure @ref(fig:power-analysis-study5)). 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 (we use their data when available) and 192 participants were flagged by Qualtrics' fraud detection software as suspicious based on the aforementioned criteria.

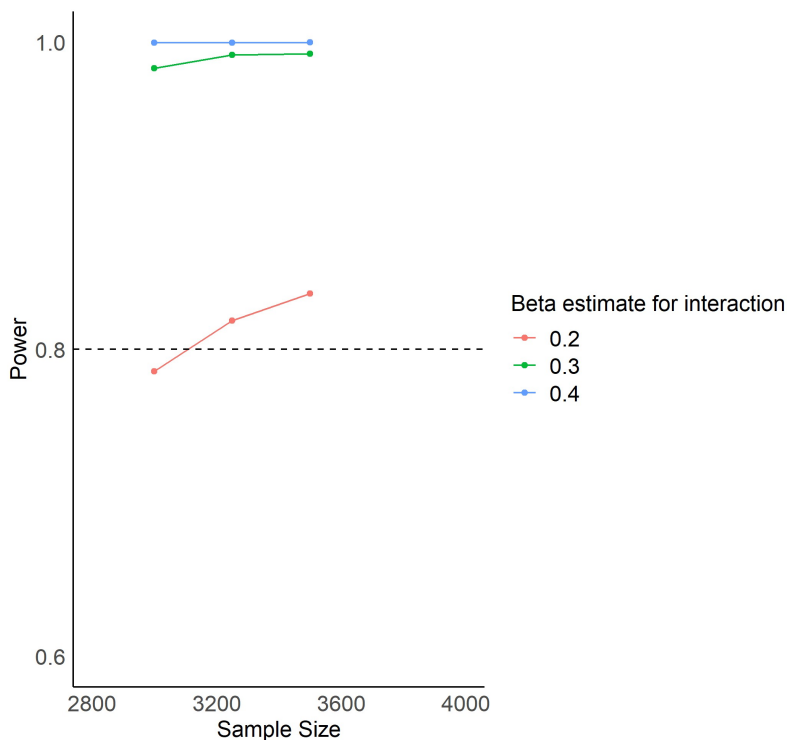


Figure 1: Plot of simulation output used to determine necessary sample size for at least 80 percent power.

### Demographics Across Conditions

| Characteristic  | Control condition, N =<br>1990 | Competition condition, N =<br>1973 | p-<br>value |
|---|--------------------------------|------------------------------------|-------------|
| <b>Age</b>  | 41 (13)                        | 41 (13)                            | 0.6         |
| Unknown   | 17                             | 24                                 |             |
| <b>Gender</b>   |                                |                                    | 0.7         |
| Man   | 855 (43%)                      | 837 (42%)                          |             |
| Woman   | 1,135 (57%)                    | 1,136 (58%)                        |             |
| <b>Education</b>  |                                |                                    | 0.9         |
| No formal education   | 0 (0%)                         | 0 (0%)                             |             |
| Less than high school   | 8 (0.4%)                       | 10 (0.5%)                          |             |
| High school graduate (diploma)                                  | 162 (8.2%)                     | 179 (9.2%)                         |             |
| High school graduate (GED)                                      | 58 (2.9%)                      | 49 (2.5%)                          |             |
| Some college (1-4 years, no degree)                             | 402 (20%)                      | 403 (21%)                          |             |
| Associate's degree (including occupational or academic degrees) | 223 (11%)                      | 207 (11%)                          |             |
| Bachelor's degree (BA, BS, etc)                                 | 777 (39%)                      | 761 (39%)                          |             |
| Master's degree (MA, MS, MENG, MSW, etc)                        | 274 (14%)                      | 276 (14%)                          |             |
| Professional school degree (MD, DDC, JD, etc)                   | 35 (1.8%)                      | 38 (1.9%)                          |             |
| Doctorate degree (PhD, EdD, etc)                                | 34 (1.7%)                      | 26 (1.3%)                          |             |
| Unknown   | 17                             | 24                                 |             |

Table 1: Size of sample with corresponding percentage listed for gender and education, with p-values derived from Fisher's exact test. Mean with corresponding standard deviation listed for age, with p-values derived from Kruskal-Wallis test. If a participant did not respond to a given question, we list their response as 'Unknown'. Note: we did not include questions about race/ethnicity nor income in this study.

| <b>Demographics Based on Participant Gender</b>                 |                       |                         |                |
|---|-----------------------|-------------------------|----------------|
| <b>Characteristic</b>   | <b>Man, N = 1,697</b> | <b>Woman, N = 2,283</b> | <b>p-value</b> |
| <b>Age</b>  | 40 (13)               | 42 (13)                 | <0.001         |
| Unknown   | 19                    | 39                      |                |
| <b>Education</b>  |                       |                         | 0.005          |
| No formal education   | 0 (0%)                | 0 (0%)                  |                |
| Less than high school   | 6 (0.4%)              | 12 (0.5%)               |                |
| High school graduate (diploma)                                  | 154 (9.2%)            | 187 (8.3%)              |                |
| High school graduate (GED)                                      | 49 (2.9%)             | 58 (2.6%)               |                |
| Some college (1-4 years, no degree)                             | 330 (20%)             | 475 (21%)               |                |
| Associate's degree (including occupational or academic degrees) | 152 (9.1%)            | 278 (12%)               |                |
| Bachelor's degree (BA, BS, etc)                                 | 687 (41%)             | 851 (38%)               |                |
| Master's degree (MA, MS, MENG, MSW, etc)                        | 229 (14%)             | 321 (14%)               |                |
| Professional school degree (MD, DDC, JD, etc)                   | 39 (2.3%)             | 34 (1.5%)               |                |
| Doctorate degree (PhD, EdD, etc)                                | 32 (1.9%)             | 28 (1.2%)               |                |
| Unknown   | 19                    | 39                      |                |

Table 2: Size of sample with corresponding percentage listed for education, with p-values derived from Fisher's exact test. Mean with corresponding standard deviation listed for age, with p-values derived from Kruskal-Wallis test. If a participant did not respond to a given question, we list their response as 'Unknown'. Note: we did not include questions about race/ethnicity nor income in this study.

## Procedures

Participants included in the study were told they would be completing a multiplication task. Notably, we aimed to recruit a larger sample to provide enough power to detect our anticipated interaction effects, and shortened the task from two minutes (as in Chapter 1) to 30 seconds (in the present study). 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 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 for their bonus payment.

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, 50.53% were assigned to the control condition and

of the women who completed the study, 49.98% were assigned to the control condition,  $\chi^2(1, n = 3980) = 0.10$ ,  $p = .754$ . We also assessed condition-dependent attrition by identifying the number of participants that dropped out during/after learning about condition and found that a relatively small proportion of participants out of the total sample dropped out after learning about their respective condition ( $N = 42$ ; 0.35% of men dropped out after learning about assigned condition versus 0.7% of women dropped out after learning about assigned condition,  $\chi^2(1, n = 3980) = 1.14$ ,  $p = .285$ ). 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) the number of practice problems participants attempted (quantified as number of practice problems not left blank, irrespective of accuracy - with participants who did not opt into the practice having a value of zero), 3) the amount of time participants spent across all practice rounds they completed (where those who chose not to practice had a value of zero for this variable), and 4) the number of practice rounds participants completed. Since the practice structure in this study is identical to that of Study 3 in Chapter 1, the number of practice rounds variable was encoded in the same way as that study ( $M = 0.59$ ,  $SD = 0.98$ ).

After completing the practicing/studying, participants guessed how much their amount of practicing for the task compared to all other participants who 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 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. We subtracted percentile from decile here because it provides more variation in the variable, thus allowing us to be more precise in our estimates of the effects of various predictors on this variable. We chose to ask participants to indicate their decile rather than percentile because it would be cumbersome for participants and it is unlikely they would be able to provide concise responses. Overall, 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.

**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. Like Study 3 of Chapter 1, all questions had three response options (e.g., men are more likely to compete than women, women are more likely to compete than men, or there are no differences in how much men or women would choose to compete). 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.