Study 1 methods

Keana Richards

6/9/2020

Participants All study measures described below are publicly available on OSF both as a .pdf and .qsf. Participants on Amazon Mechanical Turk who opted into the study had to pass several screening questions. Specifically, participants included in the paid portion of the study had to (i) identify their nationality as American and live in the United States, (ii) identify as a man or a woman, and (iii) be using a computer (rather than a phone or tablet). If they did not meet these criteria, they did not proceed to the paid portion of the study. Additionally, upon reviewing the data, we had reason to suspect that some participants completed the study more than once. Specifically, some participants had the same IP address, MTurk ID, and were of the same gender. When entries matched on all three identifiers, we included only the first entry and excluded all subsequent entries. The final sample consisted of 1056 participants (53.6% women), with an average age of 37.74 ($SD = 13.19$) years.[1] 54 participants (53.7% women) dropped out of the study before finishing and we use their data when available. See Table 1 for all demographic information.

[1]Note: Per our screening criteria for participants, the number of participants who do not identify as men or women is 0 across all studies. Therefore, the proportion of men in the sample is represented by whatever proportion of the full sample does not identify as women

Demographics Across Conditions

Characteristic	Control condition, N = 522	Preparation condition, N = 522	**p-value**
Age	38 (13)	38 (14)	>0.9
Unknown	24	17	
Gender			>0.9
Man	243 (47%)	244 (47%)	
Woman	279 (53%)	278 (53%)	
Race/ethnicity			0.3
Black/African-American	56 (11%)	49 (9.7%)	
East Asian	41 (8.2%)	34 (6.7%)	
Hispanic/Latino	33 (6.6%)	23 (4.6%)	
Middle Eastern	0 (0%)	1 (0.2%)	
Native American	1 (0.2%)	4 (0.8%)	
Pacific Islander	0 (0%)	2 (0.4%)	
Selected more than one option	14 (2.8%)	21 (4.2%)	
Selected other	5 (1.0%)	8 (1.6%)	
White/Caucasian	348 (70%)	363 (72%)	
Unknown	24	17	
Household Income			0.4
Less than $10,000	73 (15%)	69 (14%)	
$10,000 to $20,000	75 (15%)	55 (11%)	
$20,000 to $30,000	71 (14%)	69 (14%)	
$30,000 to $40,000	66 (13%)	71 (14%)	
$40,000 to $50,000	51 (10%)	60 (12%)	
$50,000 to $60,000	48 (9.6%)	51 (10%)	
$60,000 to $70,000	28 (5.6%)	34 (6.7%)	
$70,000 to $80,000	21 (4.2%)	29 (5.7%)	
$80,000 to $90,000	11 (2.2%)	15 (3.0%)	
$90,000 to $100,000	15 (3.0%)	12 (2.4%)	
$100,000 to $200,00	39 (7.8%)	35 (6.9%)	
Over $200,000	0 (0%)	5 (1.0%)	
Unknown	24	17	
Education			0.2
Less than a high school degree	3 (1.0%)	3 (1.0%)	
High School Diploma	65 (21%)	55 (18%)	
Vocational Training	9 (3.0%)	17 (5.5%)	
Some College	167 (55%)	155 (50%)	
Bachelor's degree	0 (0%)	0 (0%)	
Graduate Degree	61 (20%)	78 (25%)	
Unknown	217	214	

Table 1: Size of sample in Study 1 with corresponding percentage listed for gender, race, education, and household income, 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'.

Demographics Based on Participant Gender			
Characteristic	**Man**, N = 490	**Woman**, N = 566	**p-value**
Age	36 (13)	39 (13)	<0.001
Unknown	25	28	
Race/ethnicity			0.092
Black/African-American	43 (9.2%)	62 (12%)	
East Asian	42 (9.0%)	33 (6.1%)	
Hispanic/Latino	34 (7.3%)	22 (4.1%)	
Middle Eastern	1 (0.2%)	0 (0%)	
Native American	2 (0.4%)	3 (0.6%)	
Pacific Islander	1 (0.2%)	1 (0.2%)	
Selected more than one option	12 (2.6%)	23 (4.3%)	
Selected other	5 (1.1%)	8 (1.5%)	
White/Caucasian	325 (70%)	386 (72%)	
Unknown	25	28	
Household Income			0.8
Less than $10,000	63 (14%)	79 (15%)	
$10,000 to $20,000	60 (13%)	70 (13%)	
$20,000 to $30,000	60 (13%)	80 (15%)	
$30,000 to $40,000	62 (13%)	75 (14%)	
$40,000 to $50,000	52 (11%)	59 (11%)	
$50,000 to $60,000	49 (11%)	50 (9.3%)	
$60,000 to $70,000	30 (6.5%)	32 (5.9%)	
$70,000 to $80,000	30 (6.5%)	20 (3.7%)	
$80,000 to $90,000	10 (2.2%)	16 (3.0%)	
$90,000 to $100,000	12 (2.6%)	15 (2.8%)	
$100,000 to $200,00	34 (7.3%)	40 (7.4%)	
Over $200,000	3 (0.6%)	2 (0.4%)	
Unknown	25	28	
Education			0.9
Less than a high school degree	2 (0.7%)	4 (1.2%)	
High School Diploma	61 (21%)	59 (18%)	
Vocational Training	13 (4.5%)	13 (4.0%)	
Some College	153 (52%)	169 (53%)	
Bachelor's degree	0 (0%)	0 (0%)	
Graduate Degree	63 (22%)	76 (24%)	
Unknown	198	245	

Table 2: Size of sample in Study 1 with corresponding percentage listed for race, education, and household income, 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'.

Procedures Participants were told they would be completing a timed multiplication task where they could choose how they would be paid for their performance. We chose a multiplication task because we expected participants' performance to improve with practice. Indeed, research suggests that rehearsing and recalling associative memories can speed up retrieval of those memories [@Rundus1971]. The task involved solving problems from multiplication tables 1-12 as quickly as possible within a two-minute period. They were provided an example of a question with the correct response and had to answer three practice problems correctly to proceed, as a test of their comprehension. After completing the comprehension questions, participants were randomly assigned to either a "knowledge of preparation" condition or a control condition. Participants in the "knowledge of preparation" condition were presented the following text:

"There is an option to practice/study before completing the multiplication task that is available to all participants. If you take this opportunity to practice/study, we will provide you with materials that may help boost your performance in the multiplication task. You will have unlimited time to practice/study before completing the task. You can stop practicing/studying at any point."

Participants assigned to the control condition simply proceeded without seeing this text. An equal number of participants were assigned to both conditions (control= 50%). Of the men who completed the study, 50.1% were randomly assigned to the control condition. Of the women who completed the study, 49.91% were randomly assigned to the control condition, $\chi^2(1, n = 1056) = 0.00$, $p > .999$.

Then, all participants were asked to choose how they wanted to be paid. They were given two options, either a piece-rate payment scheme or a tournament payment scheme. They first read a description of each payment scheme, and had to correctly answer three comprehension questions before making their selection.

Under the piece-rate scheme participants were told that they would be paid \$.10 for every problem answered correctly. Under the tournament scheme, participants were told that they would be paid \$.20 for every problem they answered correctly, but only if they answered more questions correctly than a randomly assigned competitor. The order of presentation of the tournament and piece-rate payment options was randomized for participants. Participants in the experimental condition were reminded that they had the option to prepare before completing the task. After choosing a payment scheme, participants in both conditions were given an opportunity to prepare before the multiplication task. If they chose not to prepare, they proceeded to the timed multiplication task. If they chose to prepare, participants were presented with each multiplication table, 1 through 12, in sequential order. Each multiplication table provided products of numbers up to 12. Thus, participants could use the tables to study. Additionally, participants were asked if they wanted to complete practice problems for each times table. If they said yes, participants were asked to solve all multiples in that table and could only proceed to the next table if they answered all the questions correctly. Once they completed all practice questions for a given times table, they were shown the multiplication table again and were asked if they would like to continue solving problems from that table or move onto the next multiplication table. This process was repeated for each multiplication table. We originally pre-registered that we would use time spent preparing as a secondary dependent variable of interest outside of the binary choice to prepare. However, upon reflection, we decided to use a different measure, "number of practice rounds completed" ($M = 1.11$, $SD = 2.22$), which was collected in the Qualtrics survey instead of the time preparing variable. This decision was made prior to conducting any analyses and was incorporated into the preregistration plans for Studies 2 and 3. We decided against using time spent preparing as a secondary dependent variable because of concerns that we were not able to monitor whether participants were actually practicing during that time or were doing other activities, making interpretation of any significant gender differences in the results difficult. For instance, gender differences may arise if women (or men) on MTurk are more likely to be interrupted by other family members while they are working on the computer in general, rather than because of gender differences in desire to prepare. With the number of practice rounds completed

variable, participants had to make the conscious decision to continue preparing by clicking a "Yes" button every time they wanted to complete a new round of preparation. This variable was encoded as follows: if participants chose not to practice at all, they were assigned a "0" whereas participants who said they wanted to practice were assigned a value of at least "1", and this number increased incrementally by one for each additional round of practice completed.

Overall, we had two measures of preparation behavior: 1) binary choice to practice, and 2) number of practice rounds completed. The decision to practice measure conceptually captures a participants' baseline willingness to prepare, before they know what the preparation will involve. The number of practice rounds serves as a way to quantify the number of times participants continue to practice after having made the initial decision to prepare, and having seen what the practice rounds look like, which we imagine may reflect different underlying decision processes.

Following the preparation portion of the study, participants moved on to the paid portion of the study. They were required to solve as many problems as possible in two minutes. Participants' scores on the task were quantified as the number of questions correct within the two-minute time frame allotted, without any penalties for incorrect responses. After completion, participants were told how many problems 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 were not told whether they won, since this serves as an indicator of relative performance. We employ this design across all studies in this dissertation.

Then, they completed a series of incentivized follow-up questions, including measures of confidence and perceptions of gender differences. For these measures, participants were told one of these questions would be selected for a possible bonus payment, and if they answered the selected question correctly, they would earn a bonus of $.10.

For the measure of confidence, participants were asked to correctly predict their relative performance compared to all other participants completing the task by indicating the decile of their score. We used 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 compete given competition inherently requires a comparison of one's performance to the performance of one's competitors. The confidence measure draws from previous research [@Niederle2007], but instead of asking participants to indicate whether they won against a randomly selected opponent, we asked 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 was phrased so participants did not need to understand the word "decile," but were asked instead: "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 multiplication task used.

Participants were also asked to correctly predict whether men or women 1) correctly solved more problems 2) spent more time practicing before completing the multiplication task, and 3) chose the tournament payment option more. An additional question about perceptions of general gender differences in willingness to prepare that was not incentivized was included after participants responded to the incentivized questions: "For most tasks, do you think men or women generally prepare (i.e., practice and/or study) more?"

Finally, participants completed a measure of risk attitudes, where they answered if they generally are willing to take risks or try to avoid taking risks [@Dohmen2011b] on a 10-point scale with 0 meaning participants are "Not at all willing to take risks" and 10 indicating participants are "Very willing to take risks." There is evidence that risky behavior (i.e., lottery choices) is strongly associated with the risk measure used in this study [@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]. To determine whether participants used calculators to improve their performance on the task and whether there were gender differences in the use of calculators, we also asked participants about their use of calculators and perceptions of calculator use on the multiplication task. Neither of these measures was incentivized.