Pilot study methods

Keana Richards

6/9/2020

We recruited workers on Amazon Mechanical Turk for a study on decision-making. The workers who opted into the study had to pass several screening questions to be included as participants in the paid portion of the study. Specifically, participants included in the study had to (i) identify their nationality as American and live in the United States, (ii) identify as male or female, (iii) be using a computer (rather than a phone or tablet), and (iV) pass three comprehension questions for the task they would be completing. If they did not meet these criteria, they did not proceed to the paid portion of the study. Additionally, upon reviewing the data, we found that certain participants had duplicate IP addresses, MTurkID, and gender, so we excluded the second response from those participants. Based on these criteria, 206 participants were excluded from the analyses. Thus, the final sample consisted of 337 participants (56.97% women), with an average age of 37.4 (*SD* = 11.67) years. However, 17 participants (76.47% women) chose to drop out of the study before finishing (*M* = 41 years of age, *SD* = 13.57).

Participants were told they would complete three rounds of a key-entry task where they tried to solve as many problems as they could within one minute and they would be paid based on their performance. The task required participants to enter letters that correspond to a series of 2-digit numbers shown on the screen using a legend. The legend consisted of five numbers with their corresponding letters below them. The letters were randomly drawn, and the legend was the same across all participants and rounds. For instance, if the letter “C” was associated with the number 1 and “R” was associated with the number 2, and participants were presented with the number 12, they would have to enter “CR” into a corresponding text box. Before the paid rounds, participants were shown an example problem with the correct answer. Following this, they were required to pass 3 practice problems to test their comprehension. Each of these problems was identical in structure to the problems they would receive during the incentivized rounds.

In the first and second round of the task, participants were incentivized according to a piece-rate and tournament payment scheme, respectively [@Niederle2007]. Under the piece-rate scheme, participants were told that they would earn $.05 for each correct answer. Under the tournament payment scheme, participants were told that they would earn $.10 for each correct answer, but were only paid if their score was greater than the second-round performance of another anonymous, randomly matched opponent who was doing the same experiment with the same sequence of numbers as them. In the third round, participants could choose between the two above payment schemes. We denote this decision as “willingness to compete” [@Niederle2007].

After the third round, participants completed a series of follow-up questions which measured confidence, perceptions of gender differences in performance on the task, and risk aversion.

The confidence measures required participants to (i) indicate whether they thought their round 2 score was higher or lower than the person they had competed against and (ii, iii) predict which decile their score might fall into relative to all men (women) who completed the task during round 2. We also asked if they thought their score would have improved if they practiced the task beforehand. Finally, we asked if they would they have taken the opportunity to practice the task, if offered. If they responded yes, they were asked how long they would have practiced (in minutes) if they were given unlimited time to practice. Participants were told that if they correctly answered one of the confidence and perceptions of gender differences questions correctly they could receive a bonus. Specifically, we randomly selected one of their responses and if the selected prediction was correct, they received a bonus of $.10.

Finally, we measured risk aversion by asking participants “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] on a 10 point scale, where 0 indicates participants are “Not at all willing to take risks,” while 10 indicates participants are “very willing to take risks.” All hypotheses were pre-registered (<https://osf.io/q39a5/>) unless otherwise stated and all analyses were conducted in R.