Replication of Thomas & Jacoby (2013)

Alan Jern and the Students of SV 472*
Rose-Hulman Institute of Technology
February 15, 2017

1 Introduction

We attempted to perform a nearly exact replication of Experiment 1 from Thomas and Jacoby (2013) (henceforth referred to as TJ). In the study, participants were shown the answers to a series of general knowledge questions. Later, they were asked to answer general knowledge questions, including some they had seen previously (old questions), and some they had not (new questions). After each question, participants estimated what percentage of their peers would know the answer to each question. TJ found reduced egocentrism for estimates of old questions compared to new questions. That is, their estimates for old questions were less correlated with their own performance than their estimates for new questions.

All materials, data, and analyses are available at https://osf.io/hsve8.

2 Methods

2.1 Participants

Participants were undergraduate students of Rose-Hulman Institute of Technology. We collected data from as many participants as we could during a two week period. By the end of the two weeks, we had data from 43 participants. The participants were paid for participating.

2.2 Materials

100 general knowledge questions were randomly selected from Nelson and Narens (1980) to create two lists of 50 questions each—List 1 and List 2. After random selection, several test questions were swapped between the two lists so that the recall probability for the questions in each list, according to the norms in Nelson and Narens (1980), had approximately equal means and standard deviations.

*Dalal Bima, Chase Bishop, Lizzy Bohnet, Taelar Burriss, Zhuohao Chen, Bailey Craig, Izzy Cuasay, Connor Cutler, Timmy Darvillo, Austin Derrow-Pinion, Ben Dreikosen, Kyle Eriksson, Austin Fahsl, Cobi Ferrario, Tim Furman, Jalen Gutierrez, Eric Haug, Liz Klaas, Amanda Krauss, Kevin Kum, Tressa Lauer, Amanda McKean, Bill Metcalf, Kajun Miller, Ryan Murphy, Daniel Myers, Christian Nunnally, Ester Perez, Matt Pokoj, Noah Randall, David Saadatnezhad, Kice Sanders, Maria Schaefer, Nate Scharpenberg, Billy Scheid, Logan Schiessle, Cameron Sheehan, Wyatt Smith, Isaiah Smith, Sarthak Suri, Tasha Tepley, Carrie Utter, Nick Wallien, Jessica Wells, Michael Wojcik
The mean recall probability for List 1 was 0.44 ($SD = 0.29$) and the mean recall probability for List 2 was 0.44 ($SD = 0.27$). TJ reported a mean recall probability of 0.49 ($SD = 0.28$) for both of their lists.

For the fluency task that took place during the filled interval between the preexposure and test phases, we selected 13 categories from [Van Overschelde, Rawson, & Dunlowsky, 2004].

2.3 Procedure

We followed the procedure of TJ’s original experiment as exactly as possible. We implemented the experiment using PsychoPy software. Our experiment files are available in the OSF repository.

During the preexposure phase, participants saw 50 questions and answers for 6 seconds, followed by a 1-second blank screen. Approximately half of participants saw List 1 and the remaining participants saw List 2. The order of the questions was randomized for each participant. At the end of this phase, participants estimated what percentage of the questions they knew the answers to before the experiment started, and at that moment.

Participants then completed a fluency task for 10 minutes, in which they were presented with the names of categories and were instructed to write down as many examples of the category as they could think of. Each of the 13 category names was visible for 36 seconds followed by 6 seconds of blank screen. The order of the categories was randomized for each participant.

Participants then completed the test phase, in which they answered 100 general knowledge questions (combining Lists 1 and 2). 50 of the questions were “old” (the same questions from the preexposure phase) and the remaining questions were “new”. If participants did not answer a question within 20 seconds, the experiment advanced. After answering each question, participants were told to “Estimate what percentage of your peers (Rose-Hulman students) would know the correct answer to this question.” Participants provided their answers using a slider that ranged from 0% to 100%. The order of the 100 questions was randomized for each participant.

2.4 Differences from the original experiment

We aimed for a nearly exact replication of the original methods. However, we made several minor modifications.

- TJ used a total of 160 general knowledge questions (split into two sets of 80). We used 100 questions (split into sets of 50) to shorten the duration of the experiment after initial tests revealed that the experiment was taking longer than our allotted 50 minute time slots.

- For the peer accuracy estimate, TJ instructed participants that peers were “other students at Washington University.” Analogously, we instructed participants that peers were Rose-Hulman students.

- TJ had participants type their responses for the fluency task. We had participants write their responses down on paper because it was not straightforward to collect free response input using PsychoPy.

- TJ measured response time during the peer accuracy estimates starting from the moment that participants first clicked on the sliders. We measured response times starting from the moment the slider appeared on screen. We did this because it was not straightforward to measure response time like TJ did in PsychoPy.
We did not use exactly the same questions nor the same fluency task categories but we drew them from the same lists (Nelson & Narens, 1980; Van Overschelde et al., 2004) that TJ did.

3 Results

Data from two participants were excluded from analysis because they did not answer any of the questions during the test phase. Therefore, all analyses are based on \( N = 41 \) participants. The statistical tests reported in this section were performed using JASP software. The JASP output file is posted in the OSF repository.

3.1 Accuracy

TJ did not specify how they computed participants’ accuracy scores in the test phase. Therefore, for computational simplicity, we used a strict scoring method that only counted an answer as correct if the participant submitted a string that matched exactly the correct answer to the question (ignoring case and leading or trailing spaces). For example, if the correct answer was “Einstein”, the following answers would be marked as correct: “einstein”, “einSTEIN”. The following answers would be marked as incorrect: “Einstien”, “Albert Einstein”\(^1\) The strict correctness scores were computed using an R script, posted in the OSF repository.

3.2 Participants’ own accuracy versus their estimates of peer accuracy

Accuracy scores and peer accuracy estimates are shown in Figure[1]. As predicted by TJ, the average increase in estimated peer accuracy from new to old questions was smaller (0.09) than the average increase in participant accuracy from new to old questions (0.27). Consistent with TJ, we found that the interaction between question type (old or new) and performance type (participant’s accuracy or participant’s estimated peer accuracy) was significant, \( F(1, 40) = 88.40, p < .001, \eta^2_p = 0.69 \).

3.3 Egocentrism in peer estimates

Next, we measured how correlated each participant’s performance was with their peer estimates. TJ argued that if participants’ own accuracy is less correlated with peer estimates for old questions than for new questions, this is evidence of diminished egocentrism for old questions. We computed gamma correlations between each participant’s own accuracy and their peer accuracy estimates using the rococo R package. The results are shown in Figure[2]. Consistent with TJ, the mean gamma correlation for old questions (\( M = 0.63, SD = 0.11 \)) was significantly different from the mean gamma correlation for new questions (\( M = 0.72, SD = 0.22 \)), \( F(1, 40) = 7.79, p = 0.008, \eta^2_p = 0.16 \).

3.4 Speed of estimating peer accuracy

Finally, we compared participants’ mean response time when making peer accuracy estimates for old and new questions. TJ argued that if participants make peer accuracy estimates through

\(^1\)As a class exercise, we also computed accuracy scores using a more lenient scoring method. For this method, student scored correctness by hand, counting clear misspellings as correct answers. Although not reported here, the results of statistical tests using this lenient scoring method were essentially the same as the results reported here using the strict scoring method.
Figure 1: Comparison between participants’ own accuracy and their estimates of peer accuracy for new and old questions. Colored bars indicate means. Error bars are 95% confidence intervals. Dots indicate individual data points.
anchoring and adjustment, they should take longer to respond for old questions because participants will have to make larger adjustments for old questions. The mean response time for old questions ($M = 4.63$ seconds, $SD = 1.63$) was faster than for new questions ($M = 4.74$ seconds, $SD = 1.60$), but this difference was not statistically significant, $F(1, 40) = 2.43, p = 0.127$.

4 Conclusion

We successfully replicated the primary result of TJ’s Experiment 1. Our results support their hypothesis that people exhibit reduced egocentrism in estimating what others would know when they are aware of the fact that they had only recently learned a fact. We were unable to replicate the result that participants were significantly slower to make peer accuracy estimates for old questions than new questions. Therefore, our results are inconclusive about the method by which participants reduced their egocentrism. However, as noted above, we measured response time slightly differently than TJ did. Almost certainly for this reason, our mean response times were about 9 times longer than the mean response times found by TJ.

References
