Replication of

Google Effects on Memory:
Cognitive Consequences of Having Information at Our Fingertips

by Sparrow, B. / Liu, J. / Wegner, D. M. (2011)
in: Science, 333, pp. 776–778

Replication Authors:

Colin Camerer, Felix Holzmeister, Taisuke Imai, Dylan Manfredi, Gideon Nave, and Julia Rose

Sparrow et al. (2011) investigate whether the Internet has become an external memory system, that is primed by the need to acquire information. They do so by testing whether once information has been accessed, participants' internal encodings of where the information is stored are enhanced, as opposed to the information itself. The results of four studies suggest that (1) when faced with difficult questions, people are primed to think about computers, and (2) when people expect to have future access to information, they have lower recall rates for the information itself, and instead have enhanced recall for where to access it.

Hypothesis to replicate and bet on:

Computer terms are more accessible than general words after answering a block of hard trivia questions; measured as longer color-naming reaction times in a Modified Stroop Task after priming with computer terms compared to priming with non-computer terms (paired t-test, within subject variation); $t(45) = 3.26$, $p = 0.0021$, study 1, p. 776, and Fig. 1).

Power Analysis and Criteria for Replication: First Data Collection

The original sample size is 46 participants and the standardized effect size, measured as the correlation coefficient (r) is 0.437. To have 90% power to detect 75% of the original effect size, a sample size of 93 is required. The criteria for replication is an effect in the same direction as the original study and a p-value < 0.05 (in a two-sided test).

Power Analysis and Criteria for Replication: Second Data Collection

If the original result is not replicated in the first data collection, a second data collection will be carried out. To have 90% power to detect 50% of the original effect size in the pooled sample (first and second data collection), a sample size of 218 is required, i.e., a sample size of 125 in the second data collection is required. The criteria for replication is an effect in the same direction as in the original study and a p-value < 0.05 (in a two-sided test) in the pooled data.

Sample

The sample in the first data collection consists of 93 participants from the Wharton behavioral lab subject pool. If the original result is not replicated in the first data collection (two-sided p-value < 0.05 in the same direction as the original study), a second data

collection consisting of 125 additional individuals from the same population will be carried out such that the pooled sample size is 218. There are no exclusion criteria.

Materials

As the original authors did not respond to our request asking for the original material, the study will be replicated as close as possible based on the information available in the paper and Supporting Online Material. That is, we will use the 32 trivia questions listed in Appendix A and B of the Supporting Online Material (pp. 7–8) as well as the 8 target words and the 16 unrelated words used in the Modified Stroop Task of the original experiment (mentioned on p. 2 of the Supporting Online Material). As the instructions of the original study are not available either, we will compose instructions for the replication experiments to the best of our knowledge.

The original study was conducted using the software Direct RT. The replication experiments will make use of a similar, suitable software solution capable of measuring reaction times precisely in milliseconds.

Procedure

We follow the procedure of the original study as closely as possible given the information available in the paper and Supporting Online Material (the procedure is described on pages 2–3 of the Supporting Online Material). The experiment will be conducted in English as in the original study.

When participants arrive at the lab for their assigned session they will be seated by a computer to complete the study. Participants will first answer a block of either 16 easy or 16 hard (difficult) yes/no trivia questions, and then perform a Modified Stroop Task. After that, participants answer the second block of questions, such that if the first block was easy, the second block would be hard, and vice versa. Then, participants will take a second Modified Stroop Task. It is randomly determined if a participant start with the block of easy or the block of hard trivia questions.

In the Modified Stroop task, participants will be presented with words in either blue or red, and asked to press a key corresponding with the correct color (the color of the words is randomly determined). At the same time, they are to hold a 6-digit number in memory, creating cognitive load. In total, there will be two blocks of 24 words each (8 target and 16 unrelated words, in random order). The 8 target words are related to computers and search engines Google, Yahoo, screen, browser, modem, keys, internet, computer, and the 16 unrelated (control) words are Target, Nike, Coca Cola, Yoplait, table, telephone, book, hammer, nails, chair, piano, pencil, paper, eraser, laser, television. There will be no explicit mention of neither the difficulty of the questions, nor the types of words that are included in the Modified Stroop Task. The outcome measure is the reaction time for naming the color of the word. It is hypothesized that the reaction time will be longer for the target words than for the unrelated words after answering the block of hard trivia questions.

In the original study it was not mentioned how many trials were used in the Modified Stroop Task. The description of the procedure in the Supporting Online Material suggests that at least 24 trials were conducted (including each of the 8 target words once and each of the 16 unrelated words once). But it is unclear if more trials than that were conducted. We decided to conduct 48 trials in the Modified Stroop Task so that each of the 8 target words and each of the 16 unrelated words are included twice.

For the cognitive load manipulation we will randomly generate one 6-digit number used for all participants in the Modified Stroop

Task that follows the easy question block and we will randomly generate a second 6-digit number used by all participants in the Modified Stroop Task that follows the hard question block. Before the participants start the Modified Stroop Task they will be told to try and remember the 6 digit number while they complete the Modified Stroop Task and they will be told that they will be asked for the number when the task is completed. After the Modified Stroop Task is completed the participants will be asked about the 6 diigit number. In the original study it was not mentioned exactly how the cognitive load manipulation was implemented. It was only mentioned that participants were told to hold a 6-digit number in memory to create cognitive load while completing the Modified Stroop Task.

Participants will complete the study in the lab, which is a large room with isolated computer stations. A research assistant will always be in the room in order to answer any questions, and to ensure that the room remains quiet throughout the session.

Analysis

The analysis will be performed as in the original study. That is, we will conduct a paired-sample *t*-test for differences in color-naming reaction times, between the target (computer) and unrelated words, after the block of hard trivia questions. We will exclude mistakes when we estimate the average reaction times; i.e. exclude responses where the participants click on the key for the wrong color. It is unclear whether mistakes were excluded from the analysis in the original study or not.

In the original study, the color-naming re-action time in the Modified Stroop Task was higher for target (computer) words than for unrelated words after participants had encountered the block of hard trivia questions, with a mean color-naming reaction time of 712ms ($SD = 413$ms) for computer words compared to a mean reaction time of 591ms ($SD = 204$ ms) for unrelated words. A paired-sample *t*-test revealed that this difference was statistically significant with $t(45) = 3.26$, $p = 0.0021$[1]. The same test will be used in the replication study.

The results will first be estimated based on the first data collection. If the original result is replicated in the first data collection (a two-sided *p*-value < 0.05 in the same direction as in the original study), the second data collection will not be carried out. If the original result is not replicated in the first data collection, a second data collection will be carried out. The above statistical test will then be estimated for the pooled sample of the first and second data collection, to test if the original result replicated (a two-sided *p*-value < 0.05 in the same direction as in the original study).

The main replication test will be the above test based on all the 48 trials of the Modified Stroop Task. But for completeness we will also report results only using the first 24 trials of the Modified Stroop Task (using each of the 8 target words and each of the 16 unrelated words once), as it is unclear from the original study if more than 24 trials were implemented.

Differences from Original Study

As the original authors did not respond to questions on details of the experimental design employed in the original study, the repli-

---

[1]In the original study, the result is referred to as *t*-test with 68 degrees of freedom rather than 45. However, in an initial LinkedIn conversation, the first author confirmed that the sample size was $n = 46$ and that the degrees of freedom were wrongly specified in the paper.

cation will be administered as similar as possible, given the information in the paper and Supporting Online Material.

The replication will be performed at Wharton behavioral lab between September 2016 and September 2017, while the original data was gathered at Harvard University, year unknown.

From the original paper and Supporting Online Material, it was not clear how many trials of the Modified Stroop Task were conducted. The description of the task suggests that they used the 8 target words and the 16 unrelated words at least once, resulting in at least 24 trials. In the replication we will conduct 48 trials using all of the 24 words twice.

From the original paper and Supporting Online Material it is also not clear if mistakes (clicking on the key for the wrong color) in the Modified Stroop Task were excluded or not. We will exclude mistakes when we estimate the average reaction times; i.e. exclude responses where the participants click on the key for the wrong color. From the original paper and Supporting Online Material it is furthermore unclear exactly how the cognitive load manipulation was implemented, so this may differ between the original study and the replication.

Participants in the original study were incentivized by partial course credits or payments for participating. In the replication all participants will receive a show-up fee of $20. The original study contains four experiments: for the replication, the focus is only on testing for differences in color-naming reaction times to computer and unrelated (non-computer) words after the block of hard trivia questions in Experiment 1.

Replication Results for the First Data Collection (90% power to detect 75% of the original effect size)

[To be added when replication experiments have been completed.]

Replication Results for the First and Second Data Collection Pooled (90% power to detect 50% of the original effect size)

[To be added when replication experiments have been completed.]

Unplanned Protocol Deviations

[To be added when replication experiments have been completed.]

Discussion

[To be added when replication experiments have been completed.]

References

Sparrow, B. / Liu, J. / Wegner, D.M. (2011): *Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips*, Science, 333, pp. 776–778.