

Replication of

The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults

by Kovacs, A. M. / Teglas, E. / Endress, A. D. (2010)

in: Science, 330, pp. 1830–1834

Replication Authors:

Teck-Hua Ho and Hang Wu

In a series of visual object detection task experiments, Kovacs et al. (2010) show that participants' beliefs about the beliefs of an agent (whose beliefs are actually irrelevant to performing the task) modulated the participants' reaction times.

We include Experiment 1 of the original study, which uses adult participants and which conveys the key findings of the article. Participants performed a visual detection task while watching 40 animated movies. The movies started with an agent placing a ball on a table in front of an occluder. Then the ball rolled behind the occluder. After this, the movies would continue in one of four ways depending on the experimental condition. The critical manipulations involved the participant's beliefs about the ball's presence and the beliefs of the agent such that, the agent, the participant, both, or neither could believe that the ball was behind the occluder. This was achieved by varying (i) the final location of the ball and (ii) the time at which the agent left the scene. Specifically, (i) participants saw the ball either staying behind the occluder or leaving the scene and (ii) the agent left the scene either before or after the ball had reached its final location (leading to a true/false belief). That is, the agent had a true belief about the ball's location if s/he left the scene after the ball had reached its final location; if s/he left the scene before the ball reached its final location, his belief was false. At the end of each movie, the agent re-entered the scene and the occluder was lowered. The four conditions were paired with two outcomes, in which the ball was either present or absent behind the occluder. Participants were instructed to press a button as soon as they detected the ball. Notably, the agent's beliefs were never mentioned and were irrelevant to the task.

The critical comparison involved the baseline condition (neither the participant nor the agent believed the ball to be behind the occluder), and the condition where only the agent believed that the ball was behind the occluder. The key finding was that in the latter condition, participants' reaction times were faster. This suggests that participants projected the agents' beliefs and that these beliefs influenced participant behavior even though it was inconsistent with participants' own beliefs.

Hypothesis to replicate and bet on:

Participants automatically project agents' beliefs and store them in a way similar to that of their own representation about the environment. A comparison of the mean reaction time between the "P-A- treatment" and the "P-A+ treatment" in Study 1 (within subject variation), shows that reaction time is shorter in the P-A+ treatment; results show that $t(23) = 2.42$, $p\text{-value} = 0.02$ (exact $p = 0.0238$).

Power Analysis and Criteria for Replication: First Data Collection

The original sample size was 24 individuals, and the standardized effect size measured as the correlation coefficient (r) was 0.450. To have 90% power to detect 75% of the original effect size, a sample size of 89 individuals 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 of 120 additional individuals will be carried out so that the total sample size is 209 individuals. If a second data collection is carried out, we will test whether the original result replicates in the pooled sample of the first and second data collections.

In order to have 90% power to detect 50% of the original effect size, a sample size of 209 individuals is required. In the second data collection, a sample size of 120 individuals is needed in order to have a total sample size of 209 individuals for the pooled first and second data collections. The criteria for replication is an effect in the same direction as the original and a p -value < 0.05 (in a two-sided test) in the pooled data.

Sample

The sample size in the first data collection consists of 89 individuals (adults) from the National University of Singapore (NUS).

If the original result is not replicated in the first data collection (two-sided paired t -test p -value < 0.05 in the original direction), a second data collection consists of 120 additional individuals from NUS will be carried out so that the total number of individual is 209.

Materials

We use the same videos as used in the original experiment (8 animated movies, each 18.4 seconds long, generated using Maya 3D software). The original authors sent us the movies used for Experiment 1 in their study. However, they no longer have the original Psyscope X program used to measure reaction times. Therefore, we will develop a new computer program to measure reaction times. We will send the program to the original authors to check that it matches the original program and the purpose of the experiment, before conducting the replication project.

Procedure

We follow the procedure of the original article. Subjects will be recruited through recruiting advertisements posted in the NUS campus, as well as e-mail invitations sent to an existing voluntary database of undergraduate students maintained by the Centre for Behavioural Economics (CBE) at the National University of Singapore. The adult participants taking part in the original experiments received monetary compensation corresponding to about €8.00 per hour. In the replication experiment, the participant will receive a monetary compensation of S\$10.00 (in Singapore Dollars). ~~The following summary of the experimental procedure is based on section 1.1 (pp. 2–4) of the Supplementary Information.~~

Participants watched forty 18.4s long animated movies, generated using Maya 3D software. The movies involved an agent, a ball and an occluder placed on a table. There were four belief conditions, each paired with 2 different outcomes (ball present/ball absent), resulting in 8 different movies, each seen 5 times. Our critical manipulations involved the beliefs of the participant about the ball's presence and the "beliefs" of the agent. This was achieved by varying the final location of the

ball before the occluder was lowered, and the time at which the agent left the scene. Specifically, the agent left the scene either before the ball reached its final location or afterwards. In the last scene of the movie, the agent returned and the occluder was lowered; participants were instructed to press a button as soon as they detected the ball after the occluder was lowered. We recorded their reaction times.

All movies in all experimental conditions had four phases. The first phase and the last phase of the movies (columns 1 and 4 in Fig. 1) were physically identical in each of the four conditions, and had durations of 5s and 3s, respectively. In the last phase of all conditions the agent came back to the scene at 16s. The second and the third phases, in contrast, differed across the four conditions. The following descriptions refer to these two phases.

In the P-A- condition, the ball emerged from behind the occluder without leaving the scene, then rolled back behind the occluder, and finally left the scene (**ball last seen at 12s**), all in the agent's presence. **The agent left the scene at 14s**. Thus, neither the participant nor the agent believed the ball to be behind the occluder (P-A-).

In P-A+ condition, we reversed the order of when the ball and the agent left the scene, relative to the P-A- condition. Thus, the agent left the scene at 8s. The ball then emerged from behind the occluder without leaving the scene, rolled back behind the occluder, and finally left the scene (ball last seen at 14s), all in the agent's absence. Thus, only the agent but not the participant believed the ball to be behind the occluder (P-A+).

As a result of the design (true belief/false belief), the experimental conditions involved differences in the ordering of the events, as well as differences in the relative times when the agent left the scene and when the ball was last seen. The ordering differences re-

sulted from our experimental design, in which some conditions required **the agent to leave the scene before the ball reached its final location (resulting in a false belief)**, and other conditions required the agent to leave after the ball reached its final location (resulting in a true belief).

To control for the timing differences, we used pairs of conditions matched for their timing properties. (It was not possible to use 4 conditions with exactly the same timing properties as a result of the design.) Specifically, in the true belief conditions (P-A-), participants last saw the ball at 12s after the start of the movie. In the false belief conditions (P-A+), participants last saw the ball at 14s. We controlled in the same way for the time-point at which the ball was in the vicinity of the occluder, operationally defined as the last moment at which the ball crossed the midline between the right edge of the occluder and the right edge of the computer screen. In the two true belief conditions, the ball was in the vicinity of the occluder at 11s, while the ball was in the vicinity of the occluder at 13s in the two false belief conditions.

In all conditions, the ball covered the same total visible distance, and rolled twice behind the occluder. In 50% of the trials in all four conditions (20 trials in total), a ball was revealed behind the occluder when the occluder was lowered at the end of the movie.

Participants were not informed about the purpose of the studies in advance; rather, they were simply told to perform a visual detection task. Participants were instructed to press a button with their right hand as soon as they detected the ball when the occluder was lowered. To make sure that they paid attention to the entire movie (and not just to the outcome), they were also instructed to press a button with their left hand when the agent left the scene.

Participants were tested individually in a

sound-attenuated booth using Psyscope X (<http://psy.ck.sissa.it/>) on an Apple PowerBook. Responses were collected on a [button box](#). Before starting the experiment, participants were given four practice trials with feedback; these trials were not included in the analysis.

Analysis

The analysis will be performed exactly as in the original study. [In the original study \$t\$ -test was used to test the equality of response times \(RTs\)](#) between the condition where only the agent believed that the ball was behind the occluder (P-A+) and the condition where neither the participant nor the agent believed the ball to be behind the occluder (P-A-), where $t = 2.42$ and $p = 0.0238$. The same test will be used in the replication.

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 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 collections to test if the original result replicated (two-sided p -value < 0.05 in the same direction as the original study).

The experiment will be conducted in English.

Differences from Original Study

The replication procedure is the same as that of the original study, with some unavoidable deviations. The replication will be carried out in NUS between September 2016 and September 2017, whereas the original study was carried out at the Language, Cognition

and Development Lab at SISSA in Trieste, Italy, from 2006 to 2010. The original experiment was conducted in Italian, while the replication study will be conducted in English.

The original experiment used a Psyscope X program to measure the reaction times, but the original program is no longer available. Therefore, in the replication, we will use a computer program that we will develop ourselves to measure reaction times.

The original paper contains seven experiments (Studies 1–3 used adult subjects, whereas Studies 4–7 used infant subjects). For the replication, the focus is only on Study 1 and on the P-A+ and the P-A- conditions. In the original experiments the adult participants received a monetary compensation corresponding to about €8.00 per hour. In the replication experiment, the participant will receive a monetary compensation of S\$10.00 in Singapore Dollars.

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

In total 95 adult subjects participated in the first data collection (49 female students and 46 male students), which is slightly more than the planned 89 subjects. In our analyses, for each of the 95 participants the data should include 10 reaction times (5 reaction times for the P-A+ treatment and 5 reaction times for the P-A- treatment). However, for the data of 8 participants one reaction time is either missing or invalid. This is because the participant pressed the button before the earliest possible ball detection time at 18.16 seconds (hence the reaction time is negative and invalid), or because the participant forgot to press the button within the given reaction time (so the reaction time record is missing). For 3 participants, two reaction times are either invalid or missing for the same reasons. Another 2 participants experienced program

breakdown during their tests and their reaction time data are incomplete (for one participant 2 reaction times are missing, and for the other participant 4 reaction times are missing). Our analyses are based on all 95 participants' data with their mean reaction times calculated using all the valid reaction times.

Our results show that the mean reaction time for the “P-A- treatment” is 423.86 milliseconds and the mean reaction time for the “P-A+ treatment” is 378.43 milliseconds. As in the original study, we find in the replication experiment that the reaction time is shorter in the P-A+ treatment; $t(94) = 7.0152$ ($p < 0.0001$; exact $3.497e^{-10}$). This effect is significant using the same test as in the original study. Thus we replicate the result of the original study.

If we conduct the same analyses based on the valid reaction time data of the 90 participants with no more than one missing or invalid reaction times, the result of the test becomes $t(89) = 6.5182$ ($p < 0.0001$; exact $4.130e^{-9}$), and the conclusion of the replication result does not change. Based on the test using data of all 95 participants, the standardized effect size measured as the correlation coefficient (r) in the replication experiment equals 0.586. In comparison to the standardized effect size of 0.450 in the original study, the relative effect size of the replication experiment equals 130.22% ($0.586/0.450$).

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

Not applicable because the original result is replicated in the first data collection.

Unplanned Protocol Deviations

The planned sample size in the first data collection consists of 89 subjects; the sample size of the replication experiment is 95 subjects. In the original study the tests were conducted in a 3m×3m sound-attenuated booth using Psyscope X on an Apple PowerBook. Due to difficulties in finding exactly the same experimental venue and materials, in the replication experiment the tests were carried out in a 3m×4m breakout room with participants wearing a sound-attenuated device (earmuffs with a noise reduction rating of 31 decibels) using Psyscope X on a 13-inch Apple MacBook Air. Apart from these deviations, the replication tests have been conducted exactly as outlined above, without further deviations from the protocol.

Discussion

Given the criteria and procedure outlined above, the hypothesis of interest has been replicated at a significance level of $\alpha < 5\%$ for the first data collection of the replication experiment. The relative standardized effect size equals 130.22% and the p -value of the hypothesis test is smaller than 0.0001.

References

Kovacs, A. M. / Teglas E. / Endress, A. D. (2010): *The social sense: susceptibility to others' beliefs in human infants and adults*, Science, 330, pp. 1830–1834.