Replication of

Reading Literary Fiction Improves Theory of Mind

by Kidd, D. C. / Castano, E. (2013)
in: Science, 342, pp. 377–380

Replication Authors:

Adam Altmejd, Anna Dreber, Emma Heikensten, Siri Isaksson, and Magnus Johannesson

In a priming experiment on Amazon Mechanical Turk (AMT), Kidd and Castano (2013) investigate whether reading literary fiction improves the ability to identify and understand the emotions of others (affective Theory of Mind (ToM)). They carried out five experiments and following our protocol, we chose the first study that fulfills our criteria for replication, namely experiment 1. In experiment 1 participants were randomly assigned to read a literary fiction text or a non-fiction text. Having read the text, participants completed a test of cognitive ToM and a test of affective ToM (Reading the Mind in the Eyes (RMET)). In the RMET the participants were presented with 36 pictures of actors and a list of four different emotions. The task was to match each photo with one of the four emotions by looking at the feeling expressed in the actor's eyes.

Experiment 1 finds that participants who read a literary fiction text had a higher RMET score than those who read the non-fiction text. The result thus suggests that reading literary fiction improves the ability to understand and identify the emotional state of others. The original paper includes five experiments.

Hypothesis to replicate and bet on:

Reading literary fiction improves affective Theory of Mind (a comparison of the mean Reading the Mind in the Eyes Test (RMET) score between the literary fiction treatment and the nonfiction treatment in experiment 1; ANOVA test, $F(1, 82) = 6.40$ and $p = 0.0133$ (reported as $p = 0.01$, p. 378).

Power Analysis and Criteria for Replication: First Data Collection

The original sample size was 86 observations (after excluding 4 subjects), and the standardized effect size measured as the correlation coefficient (r) was 0.269. To have 90% power to detect 75% of the original effect size a sample size of 263 (after excluding subjects) 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 338 (after excluding subjects) additional individuals will be carried out so that the total sample size is 601 (after excluding subjects).

If a second data collection is carried out, it will be tested if the original result replicates in the pooled sample of the first and second data collection.

To have 90% power to detect 50% of the original effect size a sample size of 601 is required; i.e. a sample size of 338 in the second data collection to have a sample size of 601 in total for the first and second data collection pooled. 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 263 individuals from AMT (after excluding subjects). We will use a new requester account on AMT when recruiting for this study. If the original result is not replicated in the first data collection (two-sided *p*-value < 0.05 in the original direction) a second data collection of 338 additional individuals from AMT will be carried out so that the total sample size is 601 (after excluding subjects).

We will use the same exclusion criteria as in the original study with an adjustment to the reading time exclusion criteria as recommended by the original authors (see below) and add an additional exclusion criteria as recommended by the original authors. At the end of the experiment an additional question regarding whether or not subjects have participated in a similar study will be added. Only subjects who answer "No, I do not think I have participated in a similar study" will be included in the analysis of the replication result.

Materials

We use the same Qualtrics computer program as used in the original article. The consent form, which is the landing page of the web interface, was adjusted to the current study changing researcher info, etc. As mentioned above we also added a question towards the end of the experiment about previous participation in a similar study.

Procedure

We follow the procedure of the original article (with the exception of a higher compensation for completing the HIT). Subjects are recruited using AMT and told they will receive a $3.00 show-up fee (rather than $2.00 as used in the original study) for participating. If necessary to recruit a sufficient number of participants the show-up fee will be increased.

The following summary of the experimental procedure is based on page 378 in the original article and pages 1–3 of the Supplementary Information. After accepting the task, subjects were redirected to a Qualtrics website where they participated in the study.

Experiment 1 consisted of five steps that all participants completed. First each participant reads a treatment specific text. After this different tests were completed. Thus while the text was treatment specific, all participants participated in the same tests. Each step is described in more detail below:

1. Each participant was randomly allocated one out of six short texts. Three of which were literary fiction and three of which were nonfiction.

2. A false belief test measuring cognitive ToM was completed. Participants were asked to indicate the probability that a character would act according to the characters own false belief or the participants true belief.

3. The RMET test measuring affective ToM was completed. In the RMET the

participants are presented with 36 head-shots of actors and a list of four different emotions. The task was to match each photo with one of the four emotions by looking at the feeling expressed in the actor's eyes.

4. The Author Recognition Test was completed. In this test participants were given a list of 130 names out of which 65 were names of authors. The task consisted of identifying the authors they recognized.

5. Participants were asked to complete The Positive and Negative Affect Schedule, a single item assessing current sadness, the Transportation scale which assesses absorption in the text, and a series of demographic questions. An additional measure for social perception was collected.

Analysis

The analysis will be performed as in the original article. The analysis compared the mean score on the RMET for participants who read a literary fiction text and those that read a non-fiction text. The score on the RMET task was computed by summing the number of correctly identified facial expressions in the RMET task. Since the participants were asked to identify 36 facial expressions, the minimum score is 0 and the max score 36. In the original study outliers ($> 3.5SD$ from the mean) on reading time, score of guessing on the Author Recognition Test or on the RMET score were excluded. We will use the same exclusion criteria. According to the original authors there was no pre-established lower threshold for reading time used in Study 1, but they excluded the subject with 0 reading time. In the final 3 experiments in the original study a lower threshold of 30 seconds was used

(i.e. subjects with less than 30 seconds reading time were excluded) and the original authors recommended that we use this exclusion criteria in the replication. We will therefore use this exclusion criteria in the replication (a reading time of less than 30 seconds). The original authors also recommended that we add an additional exclusion criteria to exclude subjects who have participated in a similar study before. We will therefore add a question about previous participation and only subjects answering "No, I do not think I have participated in a similar study" will be included in the analysis of the replication result.

In the original article the mean RMET score was 25.9 ($SD = 4.38$) in the literary fiction treatment and 23.47 ($SD = 5.17$) in the non-fiction treatment. The difference in mean scores across treatments were analyzed with an ANOVA with Condition, Author Recognition Test, Author Recognition Test × Condition as between-participants factors; $F(1, 82) = 6.40$ and $p = 0.0133$ (reported as $p = 0.01$). 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 mentioned 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 the original study).

The result with the above exclusion criteria will be the main replication result. However, for completeness, we will also report the result without the additional exclusion criteria above compared to the original study (thus

using all observations regardless of whether or not they answered "No" on the question about previous participation and only exclude subjects with 0 reading time; the exclusion criteria for outliers will still be used in this additional analysis).

Differences from Original Study

The replication procedure is the same as that of the original study, with some deviations. The replication will be performed at AMT between September 2016 and September 2017, whereas the data in the original study was carried out at AMT in 2012. The consent form was also adjusted to the current study. The participation payment for completing the HIT was $2.00 in the original study and will be $3.00 in the replication, to be able to recruit a sufficient number of participants to the study. If necessary to recruit a sufficient number of participants the $3.00 show-up fee will be increased.

Also an additional exclusion criteria was added (about previous participation) to the original protocol following a recommendation from the original authors and the exclusion criteria for reading time was changed to 30 seconds as recommended by the original authors (only the observation with 0 reading time was excluded in the original study based on the reading time exclusion criteria).

The original paper contains five studies: for the replication the focus is only on 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

Kidd, D. C. / Castano, E. (2013): *Reading literary fiction improves theory of mind*, Science, 342 (6156), pp. 377–380.