

Replication of

Incidental Haptic Sensations Influence Social Judgments and Decisions

by Ackerman, J. M. / Nocera, C. C. / Bargh, J. A. (2010)

in: *Science*, 328, pp. 1712–1715

Replication Authors:

Nick Buttrick, Anup Gampa, Lilian Hummer, and Brian Nosek

In a priming experiment with passersby, Ackerman, Nocera, and Bargh (2010) randomly assigned participants to one of two conditions (“heavy”, “light”) before evaluating job candidates’ resumes for overall suitability, interest in the position, and likeliness of “getting along”. In the heavy-clipboard condition, participants evaluated a resume on a clipboard weighing 2041.2g, and in the light-clipboard condition participants evaluated a resume on a clipboard weighing 340.2g. They found that participants holding heavy clipboards evaluated the job candidates as better overall and as showing more interest in the position. They did not find a difference on the rating for “getting along”, presumably because social likability is irrelevant to the “heavy” metaphor. Also, participants in the heavy condition, rated their own accuracy on the task as more important, but did not self-report devoting more effort to the task. The paper included 6 experimental studies testing the nonconscious influence of weight (1 and 2), texture (3 and 4), and hardness (5 and 6) on social impression formation and decision-making. Experiment 1 is the first experiment in the paper. Following selection guidelines for this replication project, Experiment 1 was therefore chosen for replication.

Hypothesis to replicate and bet on:

Participants that evaluate a resume while using a heavier clipboard will rate the resume as better overall compared to the participants that evaluate the resume while using a lighter clipboard. The original study used F -test for a two condition comparison, $p < 0.05$.

Original test statistics: Heavy Condition: $N = 26$, $M = 5.80$, $SD = 0.76$; Light Condition: $N = 28$, $M = 5.38$, $SD = 0.79$. $F(1, 52) = 4.08$, $p = 0.049$. If there were no covariates in the model, we will convert the F to t for comparison with the replication tests.

Power Analysis and Criteria for Replication: First Data Collection

The original sample was 54 observations (after three were removed because the participants sat down to complete the survey), and the standardized effect size was $r = 0.270$. To

have 90% power to detect 75% of the original effect size, a sample size of 259 is required. The criteria for replication is an effect in the same direction as the original study and a p -value < 0.05 (two-sided t -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 334 additional individuals will be carried out so that the total sample size is 593. 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 593 is required; i.e. a sample size of 334 in the second data collection to have a sample size of 593 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

Following the original study, the sample in the first data collection will be 259 passersby in the Charlottesville community. Participants will be recruited on grounds at the University of Virginia or from popular community locations in Charlottesville area such as the downtown pedestrian mall. Participants will be paid \$5 for ten minutes. 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 334 additional passersby from the area will be carried out so that the total sample size is 593.

Materials

We will use the original survey instructions, candidate evaluation, questionnaire, and clipboards of the same weight (light = 340.2g and heavy = 2041.2g). Additionally, we will use the same resume was used in the original study for evaluation by the participants, with

the exception of updated dates for the experience and activities portions. The experiment will be in English as in the original study.

Procedure

We will follow the procedure described in the original publication, as well as the “Experiment 1: Weight Impression” section of the supporting online materials. Additionally, we will closely follow the specifications provided through our direct communication with the original authors. The experiment will encompass one stage in total where the participant will first read through the resume and continue by filling out a scaled response evaluation sheet. No filler tasks will be utilized.

To begin the experiment, adult passersby on the street will be approached near the University of Virginia and the downtown Charlottesville, VA area and will be asked if they’d like to complete a short study in exchange for \$5. They will be told that the study aims to compare the judgments of non-expert (participant) evaluations to that of expert (professor) evaluations. After reading the experiment instructions, participants will proceed to evaluate a real academic job candidate’s resume and will be informed that even minimal information can lead to the accurate appraisal of a candidate. The resume of the first author of the original study will be provided with redacted identifying information and updated dates. The resume will be attached to either a heavy (2,041.2 grams) or light (340.2 grams) top-opening storage clipboard. The participants will be given 1–2 minutes to evaluate the resume, and then evaluate the candidate on the strength of the resume for the position, likely future job performance, social compatibility with future colleagues, seriousness of interest in the position, impression of the application materials, overall impression of the candidate, likelihood of interviewing the can-

didate, and likelihood of hiring the candidate. Participants will also rate how important it was for them to make the correct evaluation, and the degree of effort they perceived devoting to the evaluation. All responses will be made on 7-point Likert scales. Data for the degree of effort participants perceived devoting to the evaluation task itself will be collected orally by the experimenter at the end of the study (before debriefing), as clarified by the original authors.

Analysis

The analysis will be performed as in the original article. First, any participant that sits down or otherwise places the clipboard on another surface instead of holding it in their hands will be excluded. No other exclusion rules were reported in the original study or by the original authors. As such, we will include all participants that respond to one or more of the dependent measures.

In the original study, a subset of the Likert-scale items was combined to create an “overall job candidate suitability” measure as the dependent variable. The items to be included in the aggregate dependent variable are strength of applicant for the position, likely future job performance, impression of the application materials, overall impression of the candidate, likelihood of interviewing the candidate, and likelihood of hiring the candidate. All responses will be made on 1-7 (negative-positive) Likert-type scales (Supplemental Online Material). Following the supplemental materials (Additional Results: Experiment 1) of the original study, we will average these items to form a composite job candidate rating for analysis. The difference in average responses by conditions of “heavy” and “light” will be tested for statistical significance via a t-test. As clarified by the original authors, in addition to creating a composite

of overall candidate ratings, the items “seriously interested in the position” and “will get along well with future colleagues” were tested separately as a function of the factor analysis results. We will adhere to this clarification during our replication analysis.

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

Differences from Original Study

The replication procedure will be the same as that of the original study, with some unavoidable deviations. The replication will be performed in Charlottesville between September 2016 and September 2017, whereas the original study was carried out in Boston in Fall of 2008. As such, as in all replications, the sample, recruiting, and setting are different from the original study. In addition, participants in this replication will be compensated with a payment of \$5, directly from the experimenter, whereas candy was used in the original study. There are no claims in the original article that suggest that these deviations are material for the tested effects. Nevertheless, we sought review from the original authors before conducting the replication to confirm and incorporate their feedback into our report.

Through communication with Ackerman et al., we have obtained the materials used in the original study, including the resume, scripts

used to engage with the participants, and analysis scripts, which will help in adhering to the original manuscript's process closely and avoiding deviations if possible. Further, during our correspondence, the authors clarified that when the original experiment was conducted in Fall of 2008, data was not collected during periods of inclement weather (i.e. heavy rain or snow). When replicating, we will adhere to this method during our data collection.

The original paper contains six studies: for the replication the focus is only on study 1 following the project protocol to select the first study in the paper reporting treatment effects.

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

Ackerman, J. M., / Nocera, C. C., / Bargh, J. A., (2010): *Incidental Haptic Sensations Influence Social Judgments and Decisions*, Science, 328, pp. 1712–1715.