Replication of

Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions

by Aviezer, H. / Trope, Y. / Todorov, A. (2012)
in: Science, 338, pp. 1225–1229

Replication Authors:

Felix Holzmeister, Jürgen Huber, Michael Kirchler, and Julia Rose

Aviezer et al. (2012) test whether shared mechanisms across positive and negative emotions occur in real-life facial expressions. The findings suggest that during peak intensities of emotion, positive and negative situations are successfully discriminated from isolated bodies but not faces. Though, views tend to perceive illusory positivity or negativity in the non-diagnostic faces when seen with bodies, highlighting the role of the body in expressing and perceiving emotions.

> **Hypothesis to replicate and bet on:**
>
> The body context is diagnostic for the affective valence of the situation during peak intensity moments (tests the hypothesis of a higher mean valence rating of winning bodies versus losing bodies in the "body treatment" in Experiment 1 (within subjects variation, paired t-test; $t(14) = 13.07$, $p < 0.0001$, p. 1226 and Fig. 1c).

Power Analysis and Criteria for Replication: First Data Collection

The original sample size is 15 participants and the standardized effect size measured as the correlation coefficient (r) is 0.961. To have 90% power to detect 75% of the original effect size, a sample size of 14 is required. The criteria for replication are 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 is 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 41 is required, i.e., a sample size of 27 in the second data collection is required. The criteria for replication are 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 14 students from the University of Innsbruck. 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 27 additional students from the University of Innsbruck will be carried out such

that the pooled sample size is 41. There are no exclusion criteria.

Materials

We use the pictures of the original experiment along with the original instructions which have been made available by the authors. While the original experiment was conducted using the software package E-Prime, the replication experiments will be programmed in oTree (Chen et al, 2016). As the replication study is conducted in German, all materials from the original study are translated from English to German.

Procedure

We follow the procedure of the original study, with only slight but unavoidable deviations as out-lined below. The following summary of the experimental procedure is therefore based on the section "Materials and Methods" (pp. 2–3) of the Supplementary Information.

Participants for the replication experiment are recruited via the subject pool of the Innsbruck EconLab. While subjects were invited one by one in the original experiment, sessions in the replication experiments are run in groups up to 24 participants completing the task in private cubicles.

To test the hypothesis, peak expressive reactions to winning and losing points in professional high-stakes tennis matches that typically evoke strong affective reactions are shown to participants of the experiment. The experimental task for the subjects is to rate the affective valence of the body alone, with the face cropped out from the picture.[1]

In the valence rating task, participants rate the body valence of reactions to winning and losing points of professional tennis players shown on 176 (88 winning point and 88 losing point) pictures using a bipolar scale ranging from 1 to 9 with 5 as a neutral midpoint. Images are randomized and pictures appear on screen until a response is made.

Analysis

The analysis will be performed exactly as in the original study. That is, a paired-sample t-test is conducted to test for the difference in the mean valence rating of winning bodies versus losing bodies in the "body treatment" in Experiment 1. The results of the valence rating task are converted from the original scale, ranging from 1 (most negative) to 9 (most positive), with 5 serving as a neutral midpoint, to a scale ranging from -4 (most negative) to +4 (most positive), with 0 as the neutral mid-point.

The original study reports a mean valence rating for winning bodies of 1.91 ($SEM = 0.188$), compared to a mean valence rating for losing bodies of -0.88 ($SEM = 0.113$) for the "body treatment" (Fig. 1c). Based on an paired t-test, the difference between mean valence ratings for winning and losing bodies is statistically significant with a p-value < 0.0001 ($t(14) = 13.07$). 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 collec-

---

[1]In the original study, the task for the subjects was to rate the affective valence *and* intensity of the body alone. The intensity rating block was always administered after the valence rating block and participants were not informed that the study consists of two block at the beginning of the experiment. As only the hypothesis on differences in the valence rating is replicated, the intensity rating block will not be part of the replication experiment.

tion 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).

Differences from Original Study

The replication procedure is identical to that of the original study, with some unavoidable deviations. The replication will be performed at the University of Innsbruck between September 2016 and September 2017, while the original data was gathered at Princeton University in 2011. The experiment will be conducted in German rather than in English (as the original study).

Participants in the original study received course credits to participate in the experiment. In contrast, a monetary show-up fee for participating in the experiment will be used in the replication. At the Innsbruck EconLab it is common practice to compensate participants with about €15.00 per hour (on average) for participating in experiments, with a minimum payment of €5.00 for participating in an experiment. Participation in the original study, on average, took about 10–15 minutes for the valence rating block. In order to comply with the terms of use of the Innsbruck EconLab, subjects will be paid a show-up fee of €5.00 for participating in the replication experiment.

The original study contains four experiments: for the replication, the focus is only on testing the mean valence rating of winning bodies versus losing bodies in the "body treatment" of Experiment 1. As only the hypothesis of a higher mean valence rating of winning bodies versus losing bodies in the "body treatment" is replicated, the intensity rating task -– which was part of Experiment 1 in the orig-inal study as well — will not be replicated.

While the experiments in the original study have been administered one-by-one, sessions in the replication experiment will be run in groups of up to 24 participants in private cubicles. The replication experiments will be programmed in oTree rather than E-Prime as in the original study.

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

Aviezer, H. / Trope, Y. / and Todorov, A. (2012): *Body Cues, Not Facial Expressions, Discriminate Be-tween Intense Positive and Negative Emotions*, Science, 338, pp. 1225–1229.

Chen, D. L. / Schonger, M. / Wickens, C. (2016): *oTree – An open-source platform for laboratory, online, and field experiments*, Journal of Behavioral and Experimental Finance, 9, pp. 88–97.