Replication of "The fluency of social hierarchy: the ease with which hierarchical relationships are seen, remembered, learned, and liked"

Zitek & Tiedens (2012, Journal of Personality and Social Psychology)

Erica J. Yoon
ejyoon@stanford.edu

Introduction

Zitek and Tiedens (2012) probed whether social hierarchies are processed more easily, and thus are liked more, than less hierarchical stimuli. In five studies, they showed that hierarchical relationships are identified faster, learned more easily, and preferred as social relationships become more complex. The target finding of this replication project is their Study 1, which suggested that participants identified pairs of people in a hierarchy (i.e., dominant-submissive pairs) faster than pairs of people equal in their facial dominance (i.e., dominant-dominant, or submissive-submissive pairs).

Methods

Sample Size Determination

We adopted the standard suggested by Simonsohn (2015), that sample size be 2.5 times larger than the original (N=235), providing 80% power (Simonsohn, 2013)

Planned Sample

A total of 235 adult participants will be recruited on Mechanical Turk (http://www.mturk.com), and only those whose IP address is in the U.S. or Canada will be recruited. Collection of the data will be terminated when the sample size reaches 235 after the data has been screened based on the exclusion criteria.

Materials

The current experiment will closely follow what is reported in the original experiment:

> *Four sets of pairs of pictures were used as stimuli in the following four within-subjects conditions: equality (submissive), equality (dominant), hierarchy, and animal. In the equality (submissive) condition, faces of two submissive-looking people were paired together, and in the equality (dominant) condition, faces of two dominant-looking people were paired together. In the hierarchy condition, a dominant and a submissive face were paired together. Finally, in the animal condition, one animal face and one human face were paired together. The participants' task was to decide whether the pictures represented two humans or one animal and one human, and the amount of time required to make this judgment was measured ...*
>
> *The human pictures were black and white photographs taken from the 1950 West Point*

yearbook (Howitzer, 1950). All of the people in the pictures were White men. These pictures had been rated for facial dominance on a 7-point scale (1 □- very submissive, 4 □- neutral or undecided, 7 □ very dominant) by large groups of judges in previous research (Mazur et al., 1984; Mueller & Mazur, 1996), and the median rating for each picture was calculated ...

We used the 16 pictures with a median score of 2 on the submissive–dominant scale as our submissive stimuli, and we used 16 of the pictures with a median score of 6 on the submissive– dominant scale as our dominant stimuli. For the equality (submissive) condition we made 20 different pairs out of the submissive faces, and for the equality (dominant) condition we made 20 different pairs out of the dominant faces. We also paired the same submissive and dominant faces to make 20 different pairs for the hierarchy condition. In half of the pairs in the hierarchy condition, the dominant face was on the left, and in the other half, the dominant face was on the right. Finally, we used pictures of 16 different animal faces (e.g., squirrel, elephant), altered to be similar in size and color to the human faces, and paired each with one picture of a cadet who had been rated a 3 on the submissive– dominant scale and with one picture of a cadet who had been rated a 5, making 32 different pairs in our animal condition. In half of the trials, the animal was on the right, and in the other half of the trials, the animal was on the left.

The original author will provide the pictorial stimuli for all of the four conditions.

Procedure

In the original experiment, participants

were seated at a computer cubicle … [and] were told that they would be viewing pairs of pictures on the screen and that these pairs of pictures would be either of one animal and one human or of two humans. They were told to hit the A key if they saw one animal and one human and to hit the H key if they saw two humans. The participants were instructed to try to make this decision as quickly as possible while still being accurate. The two pictures in the pair were centered in the middle of the screen, and they were each about 250 pixels wide and 370 pixels high with a small space in between them. After the participants hit the appropriate key to make their choice about whether it was a human or animal trial, the screen said "get ready," and then after 500 ms, the next set of stimuli appeared. Participants were presented with five practice trials before the actual stimuli were presented. There were then 92 trials in which participants saw all of the equality (submissive), equality (dominant), hierarchy, and animal–person pictures pairs. The stimuli were presented in random order.

The procedure will be followed precisely except that the task will be administered by an online identification reaction time paradigm on Mechanical Turk.

Analysis Plan

In the original experiment, the following data were excluded: 1) data from participants who made errors that are 3 SDs above the mean errors of the remaining participants; 2) data from participants who had average reaction times that are 4 SDs higher than the mean of the remaining participants; 3) each participant's incorrect responses; and 4) responses that took response time that is greater than 3 SDs above the overall mean response time on trials from the same condition. The current work will adopt similar but slightly modified exclusion criteria: 1) data from participants who make errors that are 3 SDs above the mean errors of all participants; 2) data from participants who had average reaction times that are 4 SDs higher than the mean of all participants; 3) each participant's incorrect responses; and 4) responses that took response time that is greater than 3 SDs above the overall mean response time on all trials that were exempt from the previously mentioned exclusion criteria. The remaining trials for each participant will be used to calculate an adjusted mean response time for each condition.

Differences from Original Study

The major difference of the current replication project from the original experiment is the site of identification task: whereas the original task was an in-lab experiment, the replication study will be an online paradigm where participants will perform the task on their own computers in a browser window running JavaScript code. This method has been shown to yield reliable reaction-time measurements in previous validation studies (Crump et al., 2013).

(Post Data Collection) Methods Addendum

Actual Sample

253 Mechanical Turk workers participated in the current study. Three participants participated in the experiment twice, and the data from their second participation was excluded from the analysis. Data from seven participants who failed to pass the data inclusion criteria as outlined in the analysis plan were also excluded from further analysis. Thus, data from 243 participants total were included in the planned analysis.

Differences from pre-data collection methods plan

The planned sample size was 235, but the number of participants run in data collection was slightly increased as expected from the data exclusion rate in the original study (10%). However, the exclusion rate was lower for the current replication sample.

Results

Data preparation

The number of participants who were excluded based on the criteria was as follows: Five participants had error rates > 3 SDs above the mean error rate for the remaining participants; Two participants had average reaction times 4 SDs higher than the mean of the remaining participants. We also excluded each participant's incorrect responses; and responses with RT greater than 3 SDs above the overall mean response time for all participants. The remaining trials for each participant were used to calculate an adjusted mean response time for each

condition.

Confirmatory analysis

The mean response time for each condition was calculated for each participant, then the mean response rate for each condition across all participants was calculated. The visual comparison between the data from the current study and the original study is shown in Figure 1.

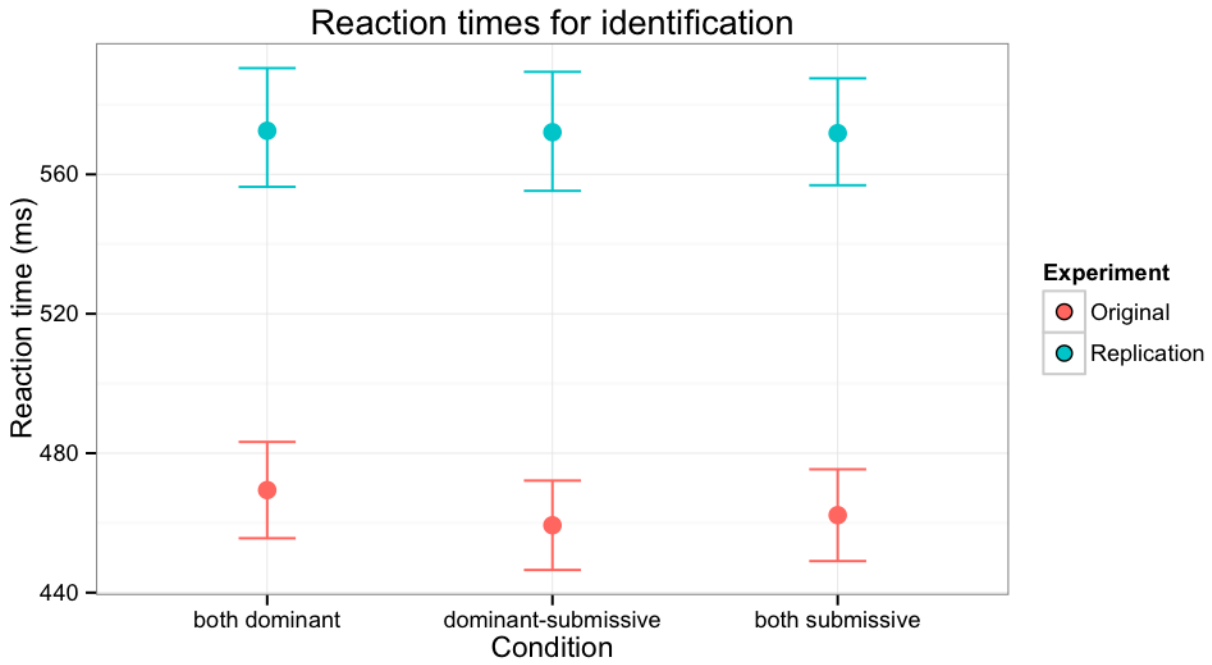

Figure 1. Comparison between the mean reaction times to identify faces across three conditions. Error bars indicate 95% confidence intervals.

The RTs in the current replication were overall slower compared the RTs in the original results. The means for the three conditions of interest (both human faces), dominant-submissive condition (M = 572.1, SD = 179.3), both-dominant condition (M = 572.5, SD = 181.9) and both-submissive condition (M = 571.8, SD = 177.6), were close to each other and did not show the expected pattern of differences, namely faster responses for dominant-submissive condition compared to the other two conditions. One-way repeated-measures analysis of variance (ANOVA) with condition (both dominant, dominant-submissive, and both submissive) as a factor revealed no significant effect of condition on reaction times: $F(2, 484) = 0.13$, $p > .87$. Thus, the expected difference across the three conditions was not shown in the current study.

The original study performed an analysis using means computed over the reciprocal (inverse) of reaction time, and reported a significant difference between the dominant-submissive condition and the other two conditions. The means and standard deviations of these reciprocals were not reported in the original study, however. For the current study, the ANOVA on means of reciprocal reaction times did not reveal an effect of condition:

$F(2, 484) = 0.39$, $p > .68$.

Exploratory analyses

To further explore the effect of condition (which could potentially be obscured by subject and item effects), we fit two linear mixed-effects models. These models predicted reaction times and reciprocal reaction times on the basis of condition. We included in the models random slopes and intercepts for participants random intercepts for items (a “maximal” random-effects structure following the recommendation of Barr et al., 2013). The model predicting raw reaction times failed to converge, and thus we resorted to the alternative random-effects structure without by-participant random slopes. Neither the model using the raw reaction times or the model using the reciprocal reaction times found significant effects of condition (refer to Table 1 and 2).

Table 1. Coefficient estimates from mixed-effects models predicting reaction times in the replication study.

Predictor	Estimate	Value (SE)	t-value
Intercept (dominant-submissive)	576.557	8.857	65.1
Dominant-dominant	-1.118	5.389	-0.20
Submissive-submissive	0.168	5.393	0.00

Table 2. Coefficient estimates from mixed-effects models predicting reciprocal reaction times in the replication study.

Predictor	Estimate	Value (SE)	t-value
Intercept (dominant-submissive)	1.88e-03	2.41e-05	76.8
Dominant-dominant	7.47e-07	1.48e-05	0.10
Submissive-submissive	-5.28e-06	1.48e-05	-0.30

Discussion

Summary of Replication Attempt

Confirmatory analyses of the current study failed to reproduce the findings of the original study. The original study reported a significant effect of condition on reciprocal reaction times, such that responses to faces in dominant-sensitive condition seemed to be faster than responses in both-dominant or both-submissive conditions. However, the current study did not find an effect of condition on either reciprocal or raw reaction times. In particular, mean reaction time in the dominant-sensitive condition was not different from the mean reaction times in the two other conditions.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PloS one, 8(3), e57410.

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. Psychological Science.

Zitek, E. M., & Tiedens, L. Z. (2012). The fluency of social hierarchy: the ease with which hierarchical relationships are seen, remembered, learned, and liked.Journal of personality and social psychology, 102(1), 98.