Of Two Minds: a registered replication

Tobias Heycke[1], Frederik Aust[1], Mahzarin R. Banaji[2], John G. Conway[3], Pieter Van Dessel[4], Xiaoqing Hu[5], Congjiao Jiang[3], Benedek Kurdi[2], Robert Rydell[6], Lisa Spitzer[1], Christoph Stahl[1], Christine A. Vitiello[3], & Jan De Houwer[4]

[1] University of Cologne
[2] Hardvard University
[3] University of Florida
[4] Ghent University
[5] The University of Hong Kong
[6] Indiana University

Study Information

Title

Of Two Minds: a registered replication

Note

Note that parts of this registration are copied from a previous registration by Heycke et al.: https://osf.io/cm2sj/ and the corresponding paper: https://osf.io/tsvkd/

Research questions

One possible way of acquiring preferences is through evaluative conditioning (EC). In general, the EC-effect is defined as a change in evaluation, that is due to the pairing of a negative or positive valenced stimulus (called unconditioned stimulus, US) with a target object (referred to as conditioned stimulus, CS) (De Houwer, 2007).

Multiple models have been proposed in an effort to answer the question of the underlying mechanisms in evaluative conditioning effects. These models can be allocated to one of two conceptually different assumptions. On the one hand, single propositional process theories claim that there is only one path that leads to preference acquisition. The propositional assumption states that evaluative conditioning works through the conscious formation of a proposition about the relation between a CS-US pair. Therefore, contingency awareness of the connection between the CS and US would be necessary to establish an EC-Effect (e.g. Mitchell, De Houwer, & Lovibond, 2009).

Dual-process theories, on the other hand, propose two distinct systems that leads to preference acquisition. Apart from the conscious, propositional system, they describe a

second associative system, where a link between the stimuli is formed automatically without contingency awareness. Thus, in contrast to the single propositional view, an EC effect could occur through an associative system even when participants have no contingency awareness for the link between CS and US. An influential example for such a theoretical account is the associative-propositional evaluation (APE) model, developed by Gawronski and Bodenhausen (2006). Along the division between propositional and associative processes, they distinguish between implicit and explicit attitudes. Thereby, associations as CS-US pairings have a direct effect on implicit evaluations, whereas propositions form the basis for explicit attitude change (Gawronski & Bodenhausen, 2011).

One testable difference between dual-and single-process theories is the question of contingency awareness as a necessity for an EC effect to occur. One way to experimentally assess the effect of contingency awareness on EC is subliminal conditioning: If one stimulus is presented subliminally, contingency awareness between CS and US can be ruled out. Following the APE model, it should be possible to form an unconscious association between CS and US through repeated pairing, since awareness of the relationship between stimuli is not needed for association formation. These pairings would ultimately result in an EC-effect, which one should be able to detect with implicit evaluation measures (Gawronski & Bodenhausen, 2011). Finding an EC effect after a subliminal conditioning procedure would be a strong argument for dual-process theories of conditioning. If no EC occurred after subliminal conditioning, it would be an argument for propositional single-process theories (Mitchell et al., 2009).

In fact, there are several studies that found evidence for subliminal EC-effects (e.g. De Houwer, Baeyens, & Eelen, 1994; De Houwer, Hendrickx, & Baeyens, 1997; Field & Moore, 2005; Fulcher & Hammerl, 2001). Most of these studies however, can be criticized for methodological issues (see Sweldens, Corneille, & Yzerbyt, 2014 for an overview). Therefore, some authors still hold the opinion that an automatic link formation is not possible (Lovibond & Shanks, 2002; Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009). Certainly, finding an EC-effect in a well executed subliminal conditioning procedure would be a valuable contribution to the ongoing controversy, providing evidence against the necessity of contingency awareness.

One study showing a selective influence of subliminally presented USs is the study of Rydell, McConnell, Mackie, and Strain (2006). In this experiment, the CS was a picture of a character named “Bob”. On each trial, after a fixation cross, the picture of Bob was preceded by a prime word of either positive or negative valence that was presented subliminally. Then, participants saw the picture of Bob, and a few moments later behavioral information about Bob was presented alongside the image. This information was again either positive or negative. The participants were then asked to judge whether this behavior was characteristic of Bob or not. They subsequently received feedback about whether their judgment was correct or false. The experiment consisted of two experimental blocks with 100 trials in each block: If the primes in the first block were negative, Bob was characterized by positive behavior and vice versa. Moreover, if the primes in the first block were negative, then they were positive in the second block and Bob was described by negative behavior. The participants therefore received deviating implicit and explicit information in each experimental block. After each block, participants’ implicit and explicit attitudes with regard to “Bob” were assessed.

The main finding of the study is that explicit attitudes were affected by behavioral

information, whereas implicit attitudes reflected the valence of subliminal primes. In line with the behavioral feedback information, there was a significant difference between the explicit attitudes measured after block one compared to the results of the explicit attitude measure after the second block. Independently, in line with the valence of subliminally presented primes, there was a significant difference between the two implicit attitude measures. Rydell et al. (2006) conclude that their participants simultaneously held opposite implicit and explicit attitudes at the same time.

In a recent replication attempt the results of Rydell and colleagues could not be replicated (Heycke, Gehrmann, Haaf, & Stahl, in press). Even though the same stimulus material was used in this replication attempt, differences between the procedure by Heycke et al. (in press) and the original study might be the reason why the original result could not be replicated: Specifically in Experiment 1 the stimulus material was translated to German and small changes to some words were made and in Experiment 2 the original material was used but the presentation time of the primes was reduced (16 ms compared to 25 ms in the original study).

Therefore, it seems necessary to conduct a replication of the original study which reproduces the original method as close as possible. If the original results can be reproduced further studies could investigate why Heycke et al. (in press) were unable to reproduce the results (e.g., investigate the question whether the prime words were presented truly subliminally).

Hypotheses

Explicit Evaluation. Under both single-propositional and dual-process theories we would expect a more favorable rating of the target stimulus (i.e., Bob) when positive behavior is characteristic of Bob, than when negative behavior is characteristic of him (Rydell et al., 2006). We thus expect a significant difference between the results of the explicit attitude measure at time 1 compared to time 2, in the direction described above.

IAT effect. Based on the findings by Rydell et al. (2006) we would expect an IAT effect based on the valence of primes. We would therefore expect a significant difference between the results of the implicit attitude measure at time 1 compared to time 2. This result corresponds to the assumptions of a dual-process account. According to such an account, we would thus expect that the implicit measure contradicts the explicit measure in each experimental block.

Based on the findings by Heycke and colleagues (in press) we would expect that the IAT effect reflects the supraliminal behavioral information of Bob, as does the explicit rating.

Sampling Plan

Existing data

Registration prior to creation of data. As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.

Data collection procedures

We plan to conduct our study at Ghent University (Belgium), Harvard University (USA), and the University of Cologne (Germany). Participants will receive 8 € or partial

course credit for their participation. At each location participants will be seated behind a computer as the entire task is computerized.

If participants already took part in a previous replication attempt of this study, they are not allowed to participate in the study.

Sample size

We will collect the data of 50 participants in each lab. Participants who signed up for the study are always allowed to participate, even when $N = 50$ has been reached before, which could result in a slightly larger sample size.

Sample size rationale

We will collect the same number of participants as used in the original study (N = 50) in each lab. An a priori power analysis with the smallest effect ($f = 0.32$) from the original study and $\alpha = \beta = .05$ yielded a required sample of 34 participants. Pooling the data from all three labs (N = 150) would allow us to detect effects as small as $d_z = 0.27$ (with $\alpha = \beta = .05$).

Stopping rule

We will collect the data of 50 participants in each lab. When participants are excluded we will continue the data collection until 50 valid data sets are collected. Participants who signed up for the study are always allowed to participate, even when $N = 50$ has been reached before, which could result in a slightly larger sample size.

Variables

Manipulated variables

Our main variable of interest is the valence of the primes and the valence of the (characteristic) behaviors. Depending on whether participants receive positive explicit and negative implicit information first or vice versa, we should observe different responses of the measurement of implicit and explicit attitudes at time 1 and at time 2 (see Hypotheses).

We additionally manipulate the order of dependent measures to ensure that those are not responsible for any effects. Half of the participants first respond to the implicit measure followed by the explicit attitudes measures, with the other half of participants having the opposite order. As in the original study, we counterbalanced the order of the critical IAT blocks over participants (Combination 1: “Bob” or negative; “not Bob” or positive and Combination 2: “Bob” or positive; “not Bob” or negative). Additionally, it is counterbalanced over participants whether in the IAT positive words have to be reacted to with the left or right key (and negative words with the other key). As in the original study, we use different images for the target person. Specifically, one out of six images is randomly assign to represent Bob and the other five images are used in the IAT as “Not Bob”.

Measured variables

Explicit Attitude. The explicit attitude toward the target person will be measures after the first and after the second learning block. We will use the same measures as in the original study, which will be administered on a computer and not (as in the original study) with a paper and pencil task:

In the Explicit Attitude Measure participants are asked to judge how likable the target person is, on a 9-point Likert scale, ranging from 1 (very unlikable) to 9 (very likable). After the participant indicated his/her response, five 9-point semantic differential scales, are presented together on a computer screen and participants are asked to judge the target person. These scales are as follows:

1. good - bad
2. pleasant – mean
3. agreeable – disagreeable
4. caring – uncaring
5. kind – cruel

Subsequently the participants are requested to evaluate the target person on a feeling thermometer that ranges from 0° to 100°, with higher values indicating a more positive evaluation of Bob.

Implicit Attitude. We additionally measure the implicit attitude, by means of reaction times in an IAT, at time 1 and time 2. As in Rydell et al. (2006), we used a modified IAT version reported by McConnell and Leibold (2001) consisting of seven 20-trial blocks. Participants had to press the left key ("d") or right key ("k") as part of a classification task. In the first block, images of the target character as well as images of other persons had to be categorized as depicting the target character or other persons (counterbalanced over participants whether the target character was on the left or the right key). The target character name and "not" plus the target character name were displayed on the left or right bottom of the screen as category labels. As in the original study we selected one image—out of the six images in the original material at random, for each participant anew, to represent the target character (in the learning phase and in the IAT). The remaining five other images were then used in the IAT to represent other individuals that were not the target character. In the second block, positive and negative words had to be categorized as positive or negative (with the position of the corresponding key for positive/negative counterbalanced across participants). Now the labels "positive" and "negative" were displayed on the left and right bottom of the screen. In the third and fourth block, the first and second block were combined: Half of the participants had the target character together with negative words on one key and all other images and positive words on the other key (Combination 1), while the other half of participants classified the target character together with positive words on one key and all other images and negative words on the other key (Combination 2). The category labels from block 1 and block 2 were combined in this block (i.e., Bob + positive). In the fifth block, a classification of the target character vs. other persons was again performed, but the location of the key for classifying the target character (vs. other persons) was switched compared to block 1 and block 3/4. In the final two blocks (i.e., blocks six and seven), participants again worked on a combined classification task (i.e., a

combination of blocks 2 and 5): Participants who worked on Combination 1 in blocks 3 and 4 now worked on Combination 2 (and vice versa). The order of trials in the critical blocks 3/4 and 6/7 was randomized for each participant anew, controlling that there were no more than 5 repetition trials (same key pressed as in previous trial) and controlling that there were approximately the same amount of response-switch (switching between the d and k key) and response-repetition trials.

During the learning phase we additionally measure the reaction time when judging a behavior to be characteristic or uncharacteristic of the target person. We also measure the amount of correct primes identified in the memory task and ask demographic questions (Age, Study/Job, Gender, Suspected goal of study, Comments).

Indices

Explicit Evaluation. In the original study, seven items were used to assess participants' attitudes towards Bob. After the items are z-standardized, an average score will be calculated for each participant at each of the two measurements.

Implicit Evaluation. An IAT effect is calculated by subtracting the mean log RT of the block in which the target character and positive words (Combination 2) were presented from the mean log RT of the block in which the target character and negative words (Combination 1) were presented. The IAT effect is then z-standardized (across both IAT measures and all participants). Therefore a larger IAT effect indicates a more positive evaluation of the target character.

Design Plan

Study type

Experiment. A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

Participants are not informed about the experimental manipulation.

Study design

As Rydell et al. (2006) we will conduct a modified version of an attitude-learning paradigm developed by Kerpelman and Himmelfarb (1971).

Participants will be seated at a computer and receive the instructions that they are going to learn something about a person, who is called Bob. Each trial starts with a blank screen for 1000 ms. Afterwards a fixation cross is displayed in the middle of the screen for 200 ms, which is replaced by either a positive or negative word (depending on the condition and block, see above) in white font on a black background, which is presented for 23.5 ms (Harvard) or 26.6 ms (Ghent and Cologne). There are 10 positive and 10 negative primes, each presented 10 times, resulting in 200 trials altogether. Half of the participant will receive 10×10 positive Primes in the first block and 10×10 negative Primes in the second block. The other half will receive negative primes first and positives primes in the

second block. The prime is immediately followed by an image of the target person, which is shown without any information for 247 ms (Harvard) or 253 ms (Ghent and Cologne). Then negative or positive behavior written in white font appears underneath the image, until the participant presses either the "c" key or the "u" key to indicate whether he/she thinks that the behavior is characteristic or uncharacteristic for the target person. In the first block of the learning phase, 50 positive behaviors and 50 negative behaviors are randomly drawn out of a list of 100 positive and 100 negative behaviors respectively, and presented in random order for each participant. In the second block, the remaining 50 positive and 50 negative behaviors are presented. The valence of the primes is always opposite to the valence of the characteristic behaviors. The order of the 100 primes within each block is random for each participant. After the judgment whether the behavior is characteristic of the target person or not, the participant receives feedback for 5000 ms about whether their judgment was correct or incorrect, by presenting the corresponding word on the screen. After the first block, participants perform the first attitude measure consisting of explicit evaluations of the target person and an IAT. The order of these measures is counterbalanced across participants, with the same order after block 1 and block 2.

To assess implicit associations that the participants may hold towards Bob, an IAT is administered. As Rydell et al. (2006) we used a modified IAT version of the task by McConnell and Leibold (2001). As in those studies, we counterbalance which side is associated with positive/negative words and also counterbalance the order of critical blocks.

After the second experimental block and the second measure of dependent variables, a recognition memory test of the subliminally presented primes is administered. 40 words are presented in a random order on the computer screen simultaneously. 20 of them are original primes, the other 20 are words were not used as primes. Participants are asked to select 20 words from which they think that they occurred during the learning phase.

Randomization

See above.

Analysis Plan

Statistical models and hypotheses

How to evaluate the success of a replication attempt statistically is subject of current debate (e.g., Fabrigar & Wegener, 2016; Simonsohn, 2015; Verhagen & Wagenmakers, 2014). Whether a pattern of results has been reproduced is especially difficult to diagnose if the to-be-replicated pattern consists of more than two cells of a factorial design.

It is, however, possible to instantiate specific patterns of mean differences, predicted by a theory or observed in a study, as a statistical model and evaluate the predictive accuracy of this model based on the results observed in the study (e.g., Rouder, Haaf, & Aust, 2017).

In this project we have two competing predictions:

(1) Rydell et al. (2006) report that across both measurement times explicit ratings are congruent with behavioral descriptions of Bob. In contrast, IAT scores were incongruent with behavioral descriptions.

(2) In their replication, Heycke et al. (in press) observed the same pattern for explicit ratings but the opposite for IAT scores. Like explicit ratings, IAT scores were congruent with behavioral descriptions.

In a replication we could, thus, expect that the manipulation influences ratings and IAT scores in the same direction ($\mathcal{H}_{\text{same direction}}$), or in opposite directions ($\mathcal{H}_{\text{opposite directions}}$). In a Bayesian analysis we would directly compare which of the two patterns is a better a priori-description of the replication outcome. Hence, the analysis would directly inform the question which of the two patterns was replicated.

Additionally, it is advisable to add two theoretically less interesting exploratory models to the analysis: one model that assumes no effect of the manipulation ($\mathcal{H}_{\text{no effect}}$) and one model that is compatible with any pattern ($\mathcal{H}_{\text{unconstrained effect}}$). If the latter model provides the best a priori-description of the data, the observed data would be inconsistent with both previously reported outcomes.

$$
\begin{aligned}
\mathcal{H}_{\text{no effect}} &: \delta_{\text{rating}} = 0 \\
&\quad\ \delta_{\text{IAT}} = 0 \\
\mathcal{H}_{\text{same direction}} &: \delta_{\text{rating}} \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
&\quad\ \delta_{\text{IAT}} \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
\mathcal{H}_{\text{opposite directions}} &: \delta_{\text{rating}} \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2) \\
&\quad\ \delta_{\text{IAT}} \sim \text{Negative-Half-Cauchy}(r = \sqrt{2}/2) \\
\mathcal{H}_{\text{unconstrained effect}} &: \delta_{\text{rating}} \sim \text{Cauchy}(r = \sqrt{2}/2) \\
&\quad\ \delta_{\text{IAT}} \sim \text{Cauchy}(r = \sqrt{2}/2)
\end{aligned}
$$

We will analyse target evaluations using Bayesian ANOVA with default (multivariate) Cauchy prior ($r = \sqrt{2}/2$; Rouder, Morey, Speckman, & Province, 2012) and recognition memory performance using a one-tailed Bayesian t test with default Cauchy prior ($r = \sqrt{2}/2$; Rouder, Speckman, Sun, Morey, & Iverson, 2009).

$$
\begin{aligned}
\mathcal{H}_{\text{guessing}} &: \delta_{\text{recognition}} = 0 \\
\mathcal{H}_{\text{memory}} &: \delta_{\text{recognition}} \sim \text{Positive-Half-Cauchy}(r = \sqrt{2}/2)
\end{aligned}
$$

Furthermore, we will replicate the traditional frequentist ANOVA analyses as described by Rydell et al. (2006). However, this traditional ANOVA approach ignores systematic trial-to-trial variability in IAT response latencies due to stimuli, which can lead to inflated test statistics and underestimated confidence intervals. Hence, we will supplement the ANOVA analysis of IAT scores by a mixed model analysis to ensure that our conclusion are not affected by stimulus effects (Wolsiefer, Westfall, & Judd, 2017).

Transformations

We will transform the data as described in the original paper (see Indices).

Follow-up analyses

We will check if there are any differences in the results depending in the location of data collection.

Inference criteria

For the frequentist analysis we use p-values $p < .05$ as inference criteria. For the Bayesian analyses we will use Bayes factors. We consider a Bayes factor > 10 or $< 1/10$ to be convincing evidence for an effect, or the absence of an effect, respectively. BF_{10} or BF_{01} smaller than 3 will only be considered to be anecdotal evidence (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015).

Data exclusion

If participants already took part in a previous replication attempt of this study, they are not allowed to participate in the study. Data from participants who abort the study before completing it will be excluded. Additionally, data will be excluded if participants let us know that they did not follow the instructions properly.

Missing data

If we encounter incomplete or missing data, the participant will be excluded from any analyses.

Other

Differences to the original study by Rydell et al. (2006)

We aimed to replicate the study of Rydell et al. (2006) as close as possible, but the following changes were made in accordance with the first author of the original study:

We translated all instructions and materials to German / Dutch for the experiment in Cologne/ Ghent to ensure that valent primes carry the intended affect for the participants and to ensure that participants understood all parts of the study. For efficiency reasons, we decided to conduct the experiment solely at the computer and therefore computerized the paper-pencil tasks (i.e., the explicit evaluation and the memory test). Additionally, we will present the subliminal prime for 23.5 ms / 26.6 ms instead of 25 ms, due to a slight difference in CRT monitors (i.e., 75/ 85 Hz instead of 80 Hz in the original study).

References

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(02), 230–241. doi:10.1017/S1138741600006491

De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, *32*(6), 629–633. doi:10.1016/0005-7967(94)90017-5

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative Learning with "Subliminally" Presented Stimuli. *Consciousness and Cognition*, *6*(1), 87–107.

doi:10.1006/ccog.1996.0281

Fabrigar, L. R., & Wegener, D. T. (2016). Conceptualizing and evaluating the replication of research results. *Journal of Experimental Social Psychology*, *66*, 68–80. doi:10.1016/j.jesp.2015.07.009

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion*, *19*(2), 217–243. doi:10.1080/02699930441000292

Fulcher, E. P., & Hammerl, M. (2001). When All Is Revealed: A Dissociation between Evaluative Learning and Contingency Awareness. *Consciousness and Cognition*, *10*(4), 524–549. doi:10.1006/ccog.2001.0525

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative–Propositional Evaluation Model. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 59–127). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B9780123855220000020

Heycke, T., Gehrmann, S., Haaf, J., & Stahl, C. (in press). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition & Emotion*. doi:10.1080/02699931.2018.1429389

Kerpelman, J. P., & Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, *19*(3), 301–305. doi:10.1037/h0031447

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*(1), 3–26. doi:10.1037//0097-7403.28.1.3

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. doi:10.1006/jesp.2000.1470

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183. doi:10.1017/S0140525X09000855

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (Dis)Liking: Item-Based Analyses Reveal That Valence Acquisition via Evaluative Conditioning Emerges Only When There Is Contingency Awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 130–144. doi:10.1037/0278-7393.33.1.130

Rouder, J. N., Haaf, J. M., & Aust, F. (2017). From theories to models to predictions: A Bayesian model comparison approach. *Communication Monographs*.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi:10.3758/PBR.16.2.225

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological*

Science, *17*(11), 954–958. doi:10.1111/j.1467-9280.2006.01811.x

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing with Bayes Factors: Efficiently Testing Mean Differences. *SSRN Electronic Journal*. doi:10.2139/ssrn.2604513

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559–569. doi:10.1177/0956797614567341

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*(3), 404–420. doi:10.1037/a0016196

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. doi:10.1177/1088868314527832

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4), 1457–1475. doi:10.1037/a0036731

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods*, *49*(4), 1193–1209. doi:10.3758/s13428-016-0779-0