Meta-Psychology, 2019, vol 3, MP.2018.843,
https://doi.org/10.15626/MP.2018.843
Article type: Original Article
Published under the CC-BY4.0 license

Open data: Not relevant
Open materials: Not relevant
Open and reproducible analysis: Not relevant
Open reviews and editorial process: Yes
Preregistration: Not relevant

Edited by: Rickard Carlsson
Reviewed by: Nuijten, M. & Schimmack, U.
All supplementary files can be accessed at the OSF project page:
https://doi.org/10.17605/OSF.IO/Q56E8

A Brief Guide to Evaluate Replications

Etienne P. LeBel
KU Leuven

Wolf Vanpaemel
KU Leuven

Irene Cheung
Huron University College

Lorne Campbell
Western University

The importance of replication is becoming increasingly appreciated, however, considerably less consensus exists about how to evaluate the design and results of replications. We make concrete recommendations on how to evaluate replications with more nuance than what is typically done currently in the literature. We highlight six study characteristics that are crucial for evaluating replications: replication method similarity, replication differences, investigator independence, method/data transparency, analytic result reproducibility, and auxiliary hypotheses' plausibility evidence. We also recommend a more nuanced approach to statistically interpret replication results at the individual-study and meta-analytic levels, and propose clearer language to communicate replication results.

Keywords: *transparency*, *replicability*, *direct replication*, *evaluating replications*, *reproducibility*

There is growing consensus in the psychology community regarding the fundamental scientific value and importance of replication. Considerably less consensus, however, exists about how to evaluate the design and results of replication studies. In this article, we make concrete recommendations on how to evaluate replications with more nuance than what is typically done currently in the literature. These recommendations are made to maximize the likelihood that replication results are interpreted in a fair and principled manner.

We thank the editor Rickard Carlsson and reviewers Michèle Nuijten and Ulrich Schimmack for valuable feedback on an earlier version of this article. We also thank Chiel Mues for copyediting our manuscript.

Correspondence concerning this article should be addressed to Etienne P. LeBel, Quantitative Psychology and Individual Differences Unit, KU Leuven, Tiensestraat 102 - Box 3713, Leuven, Belgium, 3000. Email: etienne.lebel@gmail.com

We propose a two-stage approach. The first one involves considering and evaluating six crucial study characteristics (the first three specific to replication studies with the last three relevant for any study): (1) replication method similarity, (2) replication differences, (3) investigator independence, (4) method/data transparency, (5) analytic result reproducibility, and (6) auxiliary hypotheses' plausibility evidence. Second, and assuming sound study characteristics, we recommend more nuanced ways to interpret replication results at the individual-study and meta-analytic levels. Finally, we propose the use of clearer and less ambiguous language to more effectively communicate the results of replication studies.

These recommendations are directly based on curating N = 1,127 replications (as of August 2018) available at Curate Science (CurateScience.org), a web platform that organizes and tracks the transparency and replications of published findings in the social sciences (LeBel, McCarthy, Earp, Elson, & Vanpaemel, 2018). This is the largest known meta-scientific effort to evaluate and interpret replication

results of studies across a wide and heterogeneous set of study types, designs, and methodologies.

Replication-Specific Study Characteristics

When evaluating replication studies, the following three study characteristics are of crucial importance:

1. Methodological similarity.

A first aspect is whether a replication study employed a sufficiently similar methodology to the original study (i.e., at minimum, used the same operationalizations for the independent and dependent variables, as in "close replications"; LeBel et al., 2018). This is required because only such replications can cast doubt upon an original hypothesis (assuming sound auxiliary hypotheses, see section below), and hence in principle, falsify a hypothesis (LeBel, Berger, Campbell, & Loving, 2017; Pashler & Harris, 2012). Studies that are not sufficiently similar can only speak to the generalizability -- but not replicability -- of a phenomenon under study, and should therefore be treated as "generalizability studies" rather than "replication studies". Such studies are sometimes called "conceptual replications", but this is a misnomer given that it is more accurate to conceptualize such studies as "extensions" rather than replications (LeBel et al., 2017; Zwaan, Etz, Lucas, & Donnellan, 2017).

2. Replication differences.

A second aspect to carefully consider is whether there are any study design characteristics that differed from the comparison original study. These are important to consider whether the differences were within or beyond a researcher's control (LeBel et al., 2018). Such differences are critical to consider because they help the community begin to understand the replicability and generalizability of an effect. Consistent positive replication evidence across replications with minor design differences suggests an effect is likely robust across those design differences. On the other hand, for inconsistent replication evidence, such differences may provide initial clues regarding potential boundary conditions of an effect.

3. Investigator independence.

A final important consideration is the degree of independence between the replication investigators and researchers who conducted the original study. This is important to consider to mitigate against the problem of "correlated investigators" (Rosenthal, 1991) whereby non-independent investigators may be more susceptible to confirmation biases given vested interest in an effect (although preregistration and other transparent practices can alleviate these issues; see next section).

General Study Characteristics

When evaluating studies in general, the following three study characteristics are important to consider.

1. Study transparency.

Sufficient transparency is required to allow comprehensive scrutiny of how any study was conducted. Sufficient transparency means posting the experimental materials and underlying data in a readable format (e.g., with a codebook) on a public repository (criteria for earning open materials and open data badges, respectively; Kidwell et al., 2016) and following the relevant reporting standards for the type of study and methodology used (e.g., CONSORT reporting standard for experimental studies; Schulz, Altman, & Moher, 2010). If a study is not reported with sufficient transparency, it cannot be properly scrutinized. The findings from such a study are consequently of little value because the target hypothesis was not tested in a sufficiently falsifiable manner. Preregistering a study (which publicly commits data collection, processing, and analysis plans prior to data collection) offers even more transparency and limits researcher degrees of freedom (assuming that the preregistered procedure was actually followed).

2. Analytic result reproducibility.

For any study, it is also important to consider whether a study's primary result (or set of results) is analytically reproducible. That is, whether a study's primary result can be successfully reproduced (within a certain margin of error) from the raw or transformed data (this is contingent of course on the fact that the data are actually available, whether publicly, as in the case of "open data", or otherwise).

If analytic reproducibility is confirmed, then our confidence in a study's reported results is boosted (and ideally results can also be confirmed to be robust across alternative justifiable data-analytic choices; Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). If analytic reproducibility is not confirmed and/or if discrepancies are detected, then our confidence should be reduced and this should be taken into account when interpreting a study's results.

3. Auxiliary hypotheses.

Finally, for any study, researchers should consider all available evidence regarding how plausible it is that the relevant auxiliary hypotheses, needed to test the substantive hypothesis at hand, were true (LeBel et al., 2018). Auxiliary hypotheses include, for example, the psychometric validity of the measuring instruments, and the sound realizations of experimental conditions (Meehl, 1990). This can be done by examining reported evidence of positive controls or evidence that a replication sample had the ability to detect some effect (e.g., replicating a past known effect; manipulation check evidence). These considerations are particularly crucial when interpreting null results so that one can rule out more mundane reasons for not having detected a signal (e.g., fatal experimenter or data processing errors; though such fatal errors can also sometimes cause false positive results).

Nuanced Statistical Interpretation and Language

Once these six study characteristics have been evaluated and taken into account, we recommend statistical approaches to interpret the results of a replication study at the individual-study and meta-analytic levels that are more nuanced than what is currently typically done. We then propose the use of clearer language to communicate replication results.

Statistical interpretation: Individual-study level.

At the individual-study level, we recommend that the following three distinct statistical aspects of a replication result are considered: (1) whether a signal was detected, (2) consistency of the replication effect size (ES) relative to the original study ES, and (3) the relative precision of the replication ES estimate relative to the original study. Such considerations yield the following replication outcome categories for the situation where an original study detected a signal (see Figure 1, Panel A, for visual depictions of these distinct scenarios)[1]:

1. Signal – consistent: replication ES 95% confidence interval (CI) excludes 0 and includes original ES point estimate (Panel A replication scenario #1; e.g., Chartier's, 2015, Reproducibility Project: Psychology [RPP] #31 replication result of McCrea's, 2008 Study 5; see Table 1 in the Appendix for details of Chartier's, 2015 RPP #31 replication and subsequently cited replication examples).
2. Signal – inconsistent: replication ES 95% CI excludes 0 but also excludes original ES point estimate. Three sub categorizations exist within this outcome category:
    a. Signal – inconsistent, larger (same direction): replication ES is larger and in same direction as original ES (Panel A replication scenario #2; e.g., Veer et al.'s, 2015, RPP #36 replication result of Armor et al.'s, 2008 Study 1).
    b. Signal – inconsistent, smaller (same direction): replication ES is smaller and in same direction as original ES (Panel A replication scenario #3; e.g., Ratliff's, 2015, RPP #26 replication result of Fischer et al.'s, 2008 Study 4).
    c. Signal – inconsistent, opposite direction/pattern: replication ES is in opposite direction (or reflects an inconsistent pattern) relative to the original ES direction/pattern (Panel A replication scenario #4; e.g., Earp et al.'s,

[1] The ES estimate precision of an original study is not currently accounted for because the vast majority of legacy literature original studies don't report 95% CIs (and CIs most often cannot be calculated because insufficient information is reported). In rare cases that CIs are reported, they are typically so wide (given the underpowered nature of the legacy literature) that ES estimates are not statistically falsifiable in practical terms. Once it becomes the norm in the field to report highly precise ES estimates, however, it will become possible and desirable to account for original study ES estimate precision when statistically interpreting replication results.

2014 Study 3 replication result of Zhong & Liljenquist's, 2006 Study 2).

3. No signal – consistent: replication ES 95% CI includes 0 but also includes original ES point estimate (Panel A replication scenario #5; e.g., Hull et al.'s, 2002 Study 1b replication result of Bargh et al.'s, 1996 Study 2a).
4. No signal – inconsistent: replication ES 95% CI includes 0 but excludes original ES point estimate (Panel A replication scenario #6; e.g., LeBel & Campbell's, 2013 Study 1 replication result of Vess', 2012 Study 1).

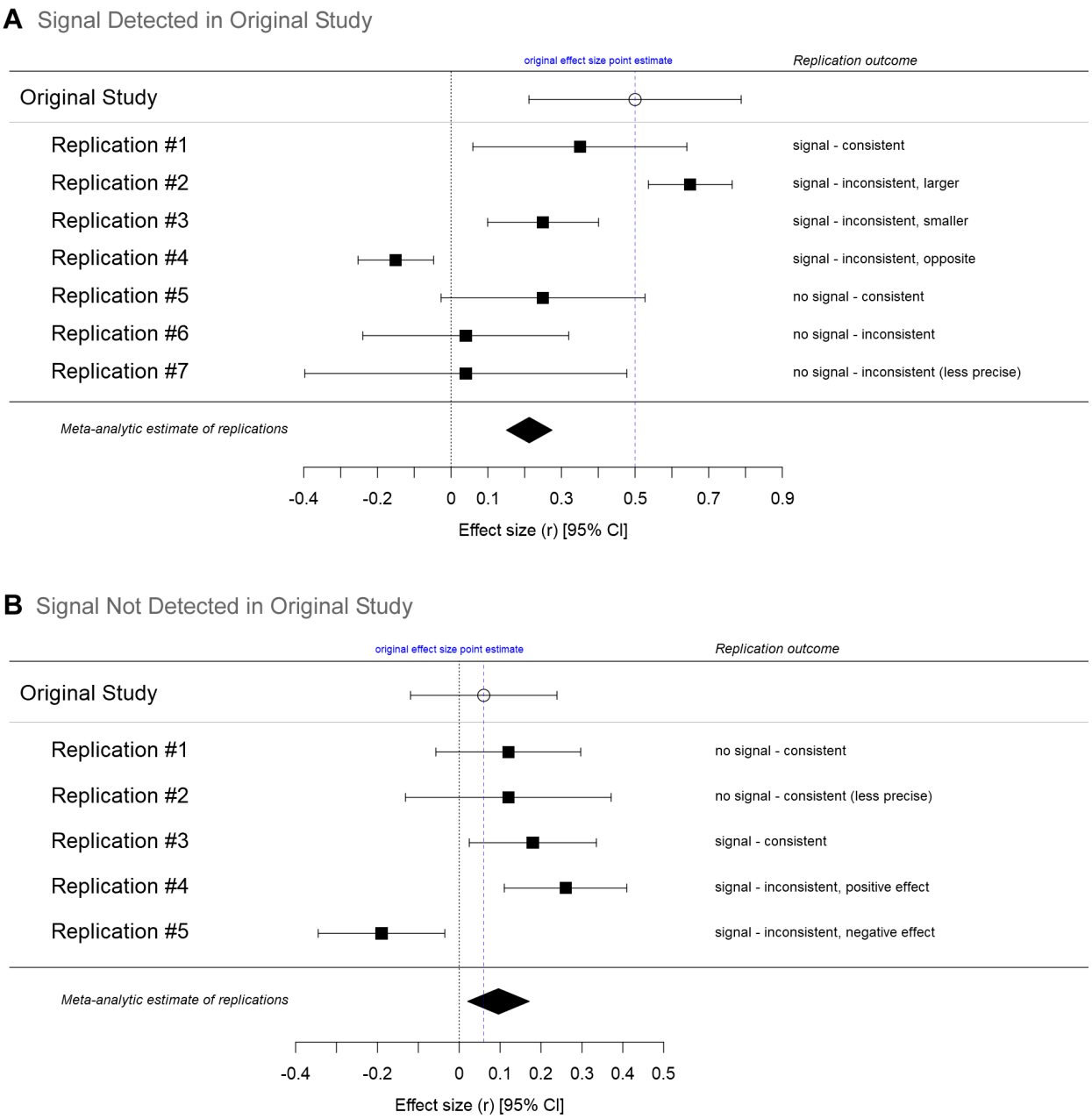

Figure 1. Distinct hypothetical outcomes of a replication study based on considering three statistical aspects of a replication result: (1) whether a signal was detected, (2) consistency of replication effect size (ES) relative to an original study, and (3) the precision of replication ES estimate relative to ES estimate precision in an original study. Outcomes are separated for situations where an original study detected a signal (Panel A) versus did not detect a signal (Panel B).

In cases where a replication effect size estimate was less precise than the original (i.e., the replication ES confidence interval is wider than the original), which can occur when a replication uses a smaller sample size and/or when the replication sample exhibits higher variability, we propose the label "less precise" be used to warn readers that such replication result should only be interpreted meta-analytically (Panel A replication scenario #7; e.g., Schuler & Wanke's, 2016 Study 2 replication result of Caruso et al.'s, 2013 Study 2).

In the situation where an original study did not detect a signal, such considerations yield the following replication outcome categories (see Figure 1, Panel B, for visual depictions of these distinct scenarios):

1. No signal – consistent: replication ES 95% confidence interval (CI) includes 0 and includes original ES point estimate (Panel B replication scenario #1; e.g., Selterman et al.'s, 2015, RPP #29 replication result of Eastwick & Finkel's, 2008 Study 1).
2. No signal – consistent (less precise): replication ES 95% confidence interval (CI) includes 0 and includes original ES point estimate, but replication ES estimate is less precise than in original study (Panel B replication scenario #2; no replication is yet known to fall under this scenario).
3. Signal – consistent: replication ES 95% confidence interval (CI) excludes 0 but includes original ES point estimate (Panel B replication scenario #3; Roebke & Penna's 2015, RPP #76 replication result of Couture et al.'s, 2008 Study 1).
4. Signal – inconsistent: replication ES 95% confidence interval (CI) excludes 0 and excludes original ES point estimate. Two sub categorizations exist within this outcome category:
    a. Signal – inconsistent, positive effect: replication ES involves a positive effect (Panel B replication scenario #4; e.g., Cohn's, 2015, RPP #45 replication result of Ranganath & Nosek's, 2008 Study 1).
    b. Signal – inconsistent, negative effect: replication ES involves a negative effect (Panel B replication scenario #5; e.g., no replication is yet known to fall under this scenario).

From this perspective, the proposed improved language to describe a replication study under replication scenario #6 would be: "We report a replication study of effect X. No signal was detected and the effect size was inconsistent with the original one." This terminology contrasts favorably with several ambiguous or unclear replication-related terminologies that are currently commonly used to describe replication results (e.g., "unsuccessful", "failed", "failure to replicate", "non-replication"). The terms "unsuccessful" or "failed" (or "failure to replicate") are ambiguous: was it the replication methodology or the replication result that was unsuccessful or failed (with similar logic applied to the ambiguous term "non-replication")? The terms "unsuccessful" or "failed" are also problematic because of the implicit message conveyed that something was "wrong" with the replication. For example, though the "small telescope approach" (Simonsohn, 2015) was an improvement over the prior simplistic standard of considering a replication $p < .05$ as "successful" and $p > .05$ as "unsuccessful", the approach nonetheless uses ambiguous language that does not actually describe a replication result (e.g., "uninformative" vs. "informative failure to replicate"). Instead, the terminology we propose offers unambiguous and descriptively accurate language, stating both whether a signal was detected and the consistency of the replication ES estimate relative to the original study. The proposed nuanced approach to statistically interpreting replication evidence improves the clarity of the language to describe and communicate replication results.

Statistical interpretation: Meta-analytic level.

Interpreting the outcomes of a set of replication studies can proceed in two ways: an informal approach, when only a few replications are available, and a more quantitative meta-analytic approach when several replications are available for a specific operationalization of an effect. The first one considers whether replications can consistently detect a signal, each of which is consistent (i.e., of similar magnitude) with the ES point estimate from the original study (Panel A replication scenario #1). Under this situation, one could informally say that an effect is "replicable." When several replications are available, a more quantitative meta-analytic approach can be taken: an effect can be considered "replicable" when the meta-analytic ES estimate excludes zero and is consistent with the original ES point estimate (also replication scenario #1, see

Panel A Figure 1; see also Mathur & VanderWeele, 2018).

Conclusion

It is important to note that replicability should be seen as a minimum requirement for scientific progress rather than an arbiter of truth. Replicability ensures that a research community avoids going down blind alleys chasing after anomalous results that emerged due to chance, noise, or other unknown errors. However, when adjudicating the replicability of an effect, it is important to keep in mind that an effect that does not appear to be replicable does not necessarily mean the tested hypothesis is false: It is always possible that an effect is replicable via alternative methods or operationalizations and/or that there were problems with some of the auxiliary hypotheses (e.g., invalid measurement, or unclear instructions, etc.). This possibility, however, should not be exploited: eventually one must consider the value of continued testing of a hypothesis across different operationalizations and contexts. Conversely, an effect that appears replicable does not necessarily mean the tested hypothesis is true: A replicable effect may not necessarily reflect a valid and/or generalizable effect (e.g., a replicable effect may simply reflect a measurement artifact and/or may not generalize to other methods, populations, or contexts).

The recommendations advocated in this article are based on curating over one thousand replications at Curate Science (as of August 2018). These recommendations have been applied to each of the replication in its database, including employing our suggested language to describe the outcome of each of its curated replication. It is expected, however, that these recommendations will evolve over time as additional replications, from an even wider set of studies, are curated and evaluated (indeed, as of September 2018, approximately 1,800 replications are in the queue to be curated at Curate Science). Consequently, these recommendations should be seen as a starting point for the research community to more accurately evaluate replication results, as we gradually learn more sophisticated approaches to interpret replication results. We hope, however, that our proposed recommendations will be a stepping stone in this direction and consequently accelerate psychology's path on becoming a more cumulative and valid science.

Appendix

Table 1. Known published replication results that fall under the distinct hypothetical replication outcomes depicted in Figure 1 (when available).

Target effect	Original study	Original ES estimate (± 95% CI)	Replication ES estimate (± 95% CI)	Replication study	Replication outcome
Signal detected in original study					
math self-handicapping effect	McCrea (2008) Study 5	r = .34 ± .35	r = .29 ± .24	Chartier (2015, RPP #31)	signal – consistent
prescribed optimism effect	Armor, Massey et al. (2008) Study 1	r = .68 ± .10	r = .76 ± .06	Veer et al. (2015, RPP #36)	signal - inconsistent, larger
selective exposure information quantity effect	Fischer, Schulz-Hardt et al. (2008) Study 4	r = .50 ± .21	r = .22 ± .16	Ratliff (2015, RPP #26)	signal - inconsistent, smaller
Macbeth effect	Zhong & Liljenquist (2006) Study 2	r = .45 ± .31	r = -.11 ± .11	Earp et al. (2014) Study 3	signal - inconsistent, opposite
elderly priming effect	Bargh et al. (1996) Study 2a	d = 1.02 ± .76	d = .53 ± .63	Hull et al. (2002) Study 1b	no signal - consistent
anxious attachment warm food effect	Vess (2012) Study 1	d = .60 ± .55	d = .03 ± .27	LeBel & Campbell (2013) Study 1	no signal - inconsistent
money priming effect	Caruso et al. (2013) Study 2	d = .43 ± .30	d = −.09 ± .39	Schuler & Wänke (2016) Study 2	no signal - inconsistent (less precise)
No signal detected in original study					
generalized earning prospect predicts romantic interest effect	Eastwick & Finkel (2008) Study 1	r = .14 ± .16	r = .03 ± .11	Selterman, Chagnon et al. (2015, RPP #29)	no signal - consistent
Hebb repetition effect revisited	Couture, Lafond, & Tremblay (2008) Study 1	r = .35 ± .38	r = .27 ± .24	Roebke & Penna (2015, RPP #76)	signal - consistent
implicit attitude generalization occurs immediately effect	Ranganath & Nosek (2008) Study 1	r = .00 ± .08	r = .11 ± .04	Cohn (2015, RPP #45)	signal - inconsistent, larger

References

Armor, D. A., Massey, C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science*, 19, 329-331. doi:10.1111/j.1467-9280.2008.02089.x

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2), 230-244. doi:10.1037/0022-3514.71.2.230

Chartier, C. R., & Perna, O. (2015). Replication of "Self-handicapping, excuse making, and counterfactual thinking: Consequences for self-esteem and future motivation." by SM McCrea (2008, Journal of Personality and Social Psychology). Retrieved from https://osf.io/ytxgr/ (Reproducibility Project: Psychology Study #31)

Cohn, M. A. (2015). Replication of "Implicit Attitude Generalization Occurs Immediately; Explicit Attitude Generalization Takes Time" (Ranganath & Nosek, 2008). Retrieved from: https://osf.io/9xt25/ (Reproducibility Project: Psychology Study #45)

Caruso, E. M., Vohs, K. D., Baxter, B., & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. *Journal of Experimental Psychology: General*, 142, 301-306. doi:10.1037/a0029288

Couture, M., Lafond, D., & Tremblay, S. (2008). Learning correct responses and errors in the hebb repetition effect: Two faces of the same coin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 524-532. doi:10.1037/0278-7393.34.3.524

Earp, B. D., Everett, J. A. C., Madva, E. N., & Hamlin, J. K. (2014). Out, damned spot: Can the "macbeth effect" be replicated? *Basic and Applied Social Psychology*, 36, 91-98. doi:10.1080/01973533.2013.856792

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology*, 94, 245-264. doi:10.1037/0022-3514.94.2.245

Fischer, P., Schulz-Hardt, S., & Frey, D. (2008). Selective exposure and information quantity: How different information quantities moderate decision makers' preference for consistent and inconsistent information. *Journal of Personality and Social Psychology*, 94, 231-244. doi:10.1037/0022-3514.94.2.94.2.231

Hull, J., Slone, L., Meteyer, K., & Matthews, A. (2002). The nonconsciousness of self-consciousness. *Journal of Personality and Social Psychology*, 83, 406-424. doi:10.1037//0022-3514.83.2.406

Kidwell, M., Lazarevic, L., Baranski, E., Hardwicke, T., Piechowski, S., Falkenberg, L., . . . Nosek, B. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *Plos Biology*, 14, e1002456. doi:10.1371/journal.pbio.1002456

LeBel, E. P., & Campbell, L. (2013). Heightened sensitivity to temperature cues in individuals with high anxious attachment: Real or elusive phenomenon? *Psychological Science*, 24, 2128-2130. doi:10.1177/0956797613486983

LeBel, E., Berger, D., Campbell, L., & Loving, T. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology*, 113, 696-696. doi:10.1037/pspi0000117

LeBel, E. P., McCarthy, R., Earp, B., Elson, M. & Vanpaemel, W. (2018). A Unified Framework to Quantify the Credibility of Scientific Findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389-402.

Mathur & VanderWeele (2018, May 7). Preprint: "New statistical metrics for multisite replication projects". https://doi.org/10.31219/osf.io/w89s5

McCrea, S. M. (2008). Self-handicapping, excuse making, and counterfactual thinking: Consequences for self-esteem and future motivation. *Journal of Personality and Social Psychology*, 95, 274-292. http://dx.doi.org/10.1037/0022-3514.95.2.274

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195-244. doi:10.2466/PRO.66.1.195-244

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531–536. doi:10.1177/1745691612463401

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, 19, 249-254. doi:10.1111/j.1467-9280.2008.02076.x

Ratliff, K. A. (2015). Replication of Fischer, Schulz-Hardt, and Frey (2008). Retrieved from: https://osf.io/5afur/ (Reproducibility Project: Psychology Study #26)

Roebke, M., & Penna, N. D. (2015). Replication of "Learning correct responses and errors in the Hebb repetition effect: two faces of the same coin" by M Couture, D Lafond, S Tremblay (2008, Journal of Experimental Psychology: Learning, Memory, and Cognition). Retrieved from: https://osf.io/qm5n6/ (Reproducibility Project: Psychology Study #76)

Rosenthal, R. (1991). *Applied Social Research Methods: Meta-analytic procedures for social research*. Thousand Oaks, CA: SAGE. doi: 10.4135/9781412984997

Schuler, J., & Wänke, M. (2016). A fresh look on money priming: Feeling privileged or not makes a difference. *Social Psychological and Personality Science*, 7, 366-373. doi:10.1177/1948550616628608

Schulz, K. F., Altman, D. G., Moher, D., CONSORT Group, & for the CONSORT Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, 698-702. doi:10.1136/bmj.c332

Selterman, D. F., Chagnon, E., & Mackinnon, S. (2015). Replication of: Sex Differences in Mate Preferences Revisited: Do People Know What They Initially Desire in a Romantic Partner? by Paul Eastwick & Eli Finkel (2008, Journal of Personality and Social Psychology). Retrieved from: https://osf.io/5pjsn/ (Reproducibility Project: Psychology Study #29)

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*, 559–569. http://dx.doi .org/10.1177/0956797614567341

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702-712. doi:10.1177/1745691616658637

Veer, A. vt., Lassetter, B., Brandt, M. J., & Mehta, P. H. (2015). The Reproducibility of Psychological Science The Open Science Collaboration Replication of Prescribed Optimism: Is it Right to Be Wrong About the Future? by David A. Armor, Cade Massey & Aaron M. Sackett (2008, Psychological Science). Retrieved from: https://osf.io/8u5v2/ (Reproducibility Project: Psychology Study #36)

Vess, M. (2012). Warm thoughts: Attachment anxiety and sensitivity to temperature cues. *Psychological Science*, *23*, 472-474. doi:10.1177/0956797611435919

Zhong, C., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*, 1451-1452. doi:10.1126/science.1130726

Zwaan, R., Etz, A., Lucas, R., & Donnellan, M. (2017). Making replication mainstream. *Behavioral and Brain Sciences*, 1-50. doi:10.1017/S0140525X17001972