The role of meta-analysis and preregistration in assessing the evidence for cleansing effects

Ross, Robert M.; van Aert, Robbie C. M.; van den Akker, Olmo R.; van Elk, Michiel

The role of meta-analysis and preregistration in assessing the evidence for cleansing effects:

A commentary in Lee and Schwarz (2020)

Robert M. Ross

Department of Philosophy, Macquarie University, Australia

https://researchers.mq.edu.au/en/persons/robert-ross

robross46@gmail.com

Robbie C. M. van Aert

Department of Methodology & Statistics, Tilburg University, The Netherlands

http://www.robbievanaert.com

r.c.m.vanaert@tilburguniversity.edu

Olmo R. van den Akker

Department of Methodology & Statistics, Tilburg University, The Netherlands

https://www.ovdakker.com

ovdakker@gmail.com

Michiel van Elk

Department of Psychology, University of Amsterdam, The Netherlands

https://www.uva.nl/profiel/e/l/m.vanelk/m.vanelk.html

m.van.elk@fsw.leidenuniv.nl

Abstract

Lee and Schwarz interpret meta-analytic research and replication studies as providing evidence for the robustness of cleansing effects. We argue that the currently available evidence is unconvincing because (a) publication bias and the opportunistic use of researcher degrees of freedom appear to have inflated meta-analytic effect size estimates, and (b) preregistered replications failed to find any evidence of cleansing effects.

The role of meta-analysis and preregistration in assessing the evidence for cleansing effects

Lee and Schwarz (2020; henceforth "L&S") present a "theory of grounded procedures" that aims to account for empirical findings relating to cleansing and other physical actions (henceforth "cleansing effects"). In section 1.2., they report two forms of evidence that they argue indicate that cleansing effects are robust: (a) meta-analytic research and (b) replication studies. While we applaud their consideration of robustness issues, we argue that they have not provided convincing evidence for the existence of cleansing effects.

L&S summarize the results of meta-analysis (currently unpublished and data unavailable) of experimental studies of cleansing effects (Lee, Chen, Ma, & Hoang, 2020) that estimates the overall effect size to be "in the small-to-medium range and highly significant" (p. 11). Moreover, they claim that converging evidence from fail-safe n, trim-and-fill, and normal quantile plots shows that "publication bias alone was unlikely to account for the existence of cleansing effects" (p. 11). However, we agree with Ropovik et al. (this issue) that this conclusion is unwarranted because these bias detection methods rely on untestable assumptions and have been superseded by more sophisticated methods. In addition, we note that these methods are particularly inappropriate for assessing this literature because, as L&S note, effect sizes are "highly heterogeneous" (p. 11). Fail-safe n does not take heterogeneity in effect sizes into account at all (Iyengar & Greenhouse, 1988), while trim-and-fill provides misleading results when heterogeneity is present (Peters, Sutton, Jones, Abrams, & Rushton, 2007; Terrin, Schmid, Lau, & Olkin, 2003). Removing large positive effects identified in a normal-quantile plot is also inappropriate because these large effects may be genuine if the studies are heterogeneous. Consequently, we encourage Lee and colleagues to re-examine the evidence for publication bias in their upcoming meta-analysis using state-of-the-art methods like Bayesian fill-in meta-analysis (Du, Liu, & Wang, 2017),

PET-PEESE (Stanley & Doucouliagos, 2014), and p-uniform* (van Aert & van Assen, 2020).

Another serious concern is that the *p*-curve analysis conducted by Ropovik et al. (this issue) indicates that the statistically significant replication effects reported in the target article contain no evidential value and that the large proportion of p-values just below .05 may have been caused by the opportunistic use of researcher degrees of freedom (Simonsohn, Nelson, & Simmons, 2014). Consequently, we argue that the evaluation of evidence for cleansing effects should be largely focused on preregistered studies. Preregistration is an effective approach for restricting researcher degrees of freedom and, thus, has an important role to play in resolving the replication crisis in psychology (Lakens, 2019; Nosek, Ebersole, DeHaven, & Mellor, 2018). Among other things, a high-quality preregistration includes a specification of a target sample size that prevents optional stopping, a description of primary and secondary outcomes that prevents outcome switching, and an analysis plan that constrains the use of other researcher degrees of freedom (Bakker et al., 2020; Wicherts et al., 2016). By contrast, meta-analytic methods that aim to correct for biases necessarily rely on untestable assumptions about the processes that generate biases and the magnitudes of these biases, which means we cannot be confident that biases have been corrected (Carter, Schonbrodt, Gervais, & Hilgard, 2019). In other words, meta-analysis is no substitute for preregistered replications (van Elk et al., 2015).

We identified 22 replication studies (that reported results) in the target article and found that only four of them (from two publications) were preregistered (Camerer et al., 2018; Johnson, Cheung, & Donnellan, 2014; see https://osf.io/7ehr8). Notably, each of these preregistered studies had much larger samples ($N = 219$, $N = 132$, $N = 123$, and $N = 286$) than the studies they attempted to replicate (all $N = 40$) (Lee & Schwarz, 2010; Schnall, Benton, & Harvey, 2008) and none of them found any evidence for the cleansing effects reported in the

original studies. In fact, in all four studies the point-estimate for the effect size was very close to zero ($d = -0.01$, $d = 0.01$, $r = -0.07$, and $r = -0.05$). In addition, we have identified a large multisite replication project ($N = 7,001$) not cited by L&S that included a test of a cleansing effect (Klein, 2018). This study attempted to replicate Study 2 of Zhong and Liljenquist (2006) (N = 27) across 50 sites and found no evidence for the predicted effect ($d = 0.00$). This fits a general pattern in the psychology literature: preregistered replication studies fail to replicate at a much higher rate than one would expect given the large effect sizes reported in original studies (Camerer et al., 2018; Open Science Collaboration, 2015), including for effects that had been supported by meta-analyses of studies that were not preregistered (Kvarven, Stromland, & Johannesson, 2020).

Because researcher degrees of freedom are curtailed in preregistered studies (if not entirely absent, see Bakker et al., 2020; Claesen, Gomes, Tuerlinckx, & Vanpaemel, 2020) we suggest that Lee and colleagues could enhance the informativeness of their upcoming meta-analysis of cleansing effects by supplementing it with a targeted meta-analysis that includes only those studies that were preregistered. Finding meta-analytic evidence for cleansing effects in preregistered studies would considerably strengthen the case for cleansing effects being robust phenomena, while a failure to find evidence would be cause for concern. A meta-analysis of the money priming effect provides an interesting example of the extent to which results can diverge (Lodder, Ong, Grasman, & Wicherts, 2019). The full meta-analysis of 246 money priming studies estimated an overall effect size of small to medium magnitude ($g = 0.31$; see Figure 1 (top-left plot), p. 701). By contrast, the targeted meta-analysis of the 47 preregistered studies found an average effect size that was non-significant ($g = 0.01$; see Figure 1 (middle-right plot), p. 701).

In summary, we have argued that a scientific assessment of the evidence for cleansing effects requires the application of state-of-the-art publication bias methods and a meta-

analysis of preregistered studies. As things stand, the empirical foundation for the theory of grounded procedures is tenuous.

Author Notes

References

Bakker, M., Veldkamp, C. L. S., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., . . . Wicherts, J. M. (2020, July 8). Ensuring the quality and specificity of preregistrations. https://doi.org/10.31234/osf.io/cdgyh

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior, 2*(9), 637-644. doi:10.1038/s41562-018-0399-z

Carter, E. C., Schonbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science, 2*(2), 115-144. https://doi.org/10.1177/2515245919847196

Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2020, June 25). Preregistration: Comparing dream to reality. https://doi.org/10.31234/osf.io/d8wex

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Du, H., Liu, F., & Wang, L. (2017). A Bayesian "fill-in" method for correcting for publication bias in meta-analysis. *Psychological Methods, 22*(4), 799-817. doi:10.1037/met0000164

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science, 3*(1), 109-135.

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? *Social Psychology, 45*(3), 209-215. doi:10.1027/1864-9335/a000186

Klein, R. A., Fasselman, F., Adams, B. G., Adamsn,R. B., Alper, S., Aveyard, M., ... Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across sample and setting. *Advances in Methods and Practices in Psychological Science, 1*(4), 443-490. https://doi.org/10.1177/2515245918810225

Kvarven, A., Stromland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*, 423-434. doi:10.1038/s41562-019-0787-z

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review, 62*(3), 221-230.

Lee, S. W. S., Chen, K., Ma, C., & Hoang, J. (2020). Psychological antecedents and consequences of physical cleansing: A meta-analytic review [Manuscript in preparation].

Lee, S. W. S., & Schwarz, N. (2010). Washing away postdecisional dissonance. *Science, 328*, 709. https://doi.org/10.1126/science.1186799

Lee, S. W., & Schwarz, N. (2020). Grounded procedures: A proximate mechanism for the psychology of cleansing and other physical actions. *Behavioral and Brain Sciences, 44*, E1. https://doi.org/10.1017/S0140525X20000308

Lodder, P., Ong, H. H., Grasman, R. P. P. P., & Wicherts, J. M. (2019). A comprehensive meta-analysis of money priming. *Journal of Experimental Psychology: General, 148*(4), 688-712. doi:10.1037/xge0000570

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606. doi:10.1073/pnas.1708274114

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Stat Med, 26*(25), 4544-4562. doi:10.1002/sim.2889

Schnall, S., Benton, J., & Harvey, S. (2008). With a Clean Conscience Cleanliness Reduces the Severity of Moral Judgments. *Psychological Science*, 19(12), 1219-1222. doi: 10.1111/j.1467-9280.2008.02227.x

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666-681. dio: 10.1177/1745691614553988

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60-78. doi:10.1002/jrsm.1095

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Stat Med, 22*(13), 2113-2126. doi:10.1002/sim.1461

van Aert, R. C. M., & van Assen, M. A. L. M. (2020, March 11). Correcting for publication bias in a meta-analysis with the P-uniform* method. https://doi.org/10.31222/osf.io/zqjr9

van Elk, M., Matzke, D., Gronau, Q. F., Guana, M., Vandekerckhove, J., & Wagenmakers, E. J. (2015). Meta-analyses are no substitute for registered replications: a skeptical perspective on religious priming. *Frontiers in Psychology, 6*, 1-7. doi:10.3389/fpsyg.2015.01365

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7*, 1-12. doi:10.3389/fpsyg.2016.01832

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