

Registered Reports

A Method to Increase the Credibility of Published Results

Brian A. Nosek¹ and Daniël Lakens²

¹University of Virginia and Center for Open Science, Charlottesville, VA, USA,
²Eindhoven University of Technology, The Netherlands

Ignoring Replications and Negative Results Is Bad for Science

The published journal article is the primary means of communicating scientific ideas, methods, and empirical data. Not all ideas and data get published. In the present scientific culture, novel and positive results are considered more publishable than replications and negative results. This creates incentives to avoid or ignore replications and negative results, even at the expense of accuracy (Giner-Sorolla, 2012; Nosek, Spies, & Motyl, 2012). As a consequence, replications (Makel, Plucker, & Hegarty, 2012) and negative results (Fanelli, 2010; Sterling, 1959) are rare in the published literature. This insight is not new, but the culture is resistant to change. This article introduces the first known journal issue in any discipline consisting exclusively of pre-registered replication studies. It demonstrates that replications have substantial value, and that incentives can be changed.

There are a number of advantages of performing direct replications, and publishing the results irrespective of the outcome. First, direct replications add data to increase precision of the effect size estimate via meta-analysis. Under some circumstances, this can lead to the identification of false positive research findings. Without direct replication, there is no way to confidently identify false positives. Conceptual replications have been more popular than direct replications because they abstract a phenomenon from its original operationalization and contribute to our theoretical understanding of an effect. However, conceptual replication are not best suited to clarify the truth of any particular effect because nonsignificant findings are attributable to changes in the research design, and rarely lead researchers to question the phenomenon (LeBel & Peters, 2011; Nosek, Spies, & Motyl, 2012).

Second, direct replication can establish generalizability of effects. There is no such thing as an exact replication. Any replication will differ in innumerable ways from the

original. A direct replication is the attempt to duplicate the conditions and procedure that existing theory and evidence anticipate as necessary for obtaining the effect (Open Science Collaboration, 2012, 2013; Schmidt, 2009). Successful replication bolsters evidence that all of the sample, setting, and procedural differences presumed to be irrelevant are, in fact, irrelevant.

Third, direct replications that produce negative results facilitate the identification of boundary conditions for real effects. If existing theory anticipates the same result should occur and, with a high-powered test, it does not, then something in the presumed irrelevant differences between original and replication could be the basis for identifying constraints on the effect. In other words, understanding any effect requires knowing when it does and does not occur. Therefore, replications and negative results are consequential for theory development.

Registered Reports Are a Partial Solution

Despite their theoretical and empirical value, the existing scientific culture provides few incentives for researchers to conduct replications or report negative results (Greenwald, 1975; Koole & Lakens, 2012). Editors and reviewers of psychology journals often recommend against the publication of replications (Neuliep & Crandall, 1990, 1993). If journals will not publish replications, why would researchers bother doing them?

This special issue of *Social Psychology* presents 15 articles with replications of important results in social psychology. Moreover, these articles demonstrate a novel publishing format – Registered Reports. By reviewing and accepting preregistered proposals prior to data collection, Registered Reports are an efficient way to change incentive

structures for conducting replications and reporting results irrespective of their statistical significance.

In 2013, the guest editors issued calls for submissions of proposals to replicate published studies in social psychology (Nosek & Lakens, 2013). Prospective authors proposed a study or studies for replication and articulated (1) why the result is important to replicate, and (2) the design and analysis plan for a high-powered replication effort. Proposals that passed initial editorial review went out for peer review. Reviewers evaluated the importance of conducting a replication and the quality of the methodology. At least one author of the original article was invited to be a reviewer if any were still alive. Most invited original authors provided a review. Authors incorporated feedback from peer review in their designs and, if the proposal had not been accepted initially, resubmitted for review and acceptance (or rejection) based on reviewer feedback.

We received 36 pre-proposals of which 24 were encouraged to submit full proposals. Ultimately 14 proposals were accepted. A 15th article (Moon & Roeder, 2014) was solicited as a second replication of one of the peer reviewed, accepted proposals (Gibson, Losee, & Vitiello, 2014) because reviewers suggest that the effect may not occur among Asian women at southern US universities (Gibson et al.'s sample).

Accepted proposals were registered at the Open Science Framework (OSF; <http://osf.io/>) prior to data collection along with the study materials. Authors proceeded with the data collection with assurance that the results would be published irrespective of the outcome, as long they followed the registered plans or provided reasonable, explicit justifications for deviating from the plan. The infrequent deviations were assessed by the action editor as whether they sacrificed integrity of the confirmatory plan before acceptance. For example, in two cases, the sample size was far short of the registered plan. The editors required additional data collection prior to acceptance. In the published articles, authors report results according to the registered confirmatory analysis plan, disclose any deviation from the plan, and sometimes provide additional exploratory analyses – clearly designated as such.

Successful proposals were designs that peer reviewers considered to be high-powered, high-quality, faithful replication designs. Peer review prior to data collection lowered the barrier to conduct replications because authors received editorial feedback about publication likelihood before much of the work was done. Furthermore, authors could focus on reporting a confirmatory analysis (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012), without the need to hunt for positive and clean results (Simmons, Nelson, & Simonsohn, 2011).

Registered reports also shift the incentives for reviewers. When the results are known, evaluation of quality is likely influenced by preexisting beliefs (Bastardi, Uhlmann, & Ross, 2011). Motivated reasoning makes it easy to generate stories for why results differed from expectations. Following Kerr's (1998) observation of hypothesizing about one's own research outcomes post facto, this might be termed, CARKing, critiquing after the results are known.

When reviewing a study proposal, only the design is available as a basis for critique. Reviewers' motivation is to make sure that the design provides a fair test. Reviewers could insist that there are innumerable conditions and moderating influences that must be met. However, each of these constrains the scope of the original effect and risks trivializing the result. So, reviewers may have competing interests – just as they do in theorizing – providing just enough constraint to ensure a fair test, but not so much to make the effect uninteresting or inapplicable.

In sum, review prior to data collection focused researchers and reviewers to evaluate the methodological quality of the research, rather than the results.

What Registered Reports Do Not Do

Preregistration and peer review in advance of data collection or analysis do not lead to definitive results. Even highly powered designs – like those in this issue – leave room for Type 1 and Type 2 errors. Furthermore, when registered reports are used for replication studies, different results between original and replication research could mean that there are unknown moderators or boundary conditions that differentiate the two studies. As such, the replication can raise more questions than it answers. At the same time, effects size estimates in small samples – common in original research – can vary considerably and are more likely to elicit an exaggerated effect size than results from larger sample sizes (Schönbrodt & Perugini, 2013). Therefore, *not finding* a predicted effect in a *large* study may indicate more about the likelihood that an effect is true than *finding* a predicted effect in a *small study*, because the former is statistically less likely if an effect is true (Button et al., 2013; Lakens & Evers, in press).

Registered Reports do not prevent or discourage exploratory analysis. Rather, they make clear the distinction between confirmatory and exploratory analysis. This applies to registered reports whether they are conducted for replications or original research. Confirmatory results follow a preregistered analysis plan and thereby ensure interpretability of the reported *p*-values (Wagenmakers et al., 2012). In exploratory analysis, *p*-values lose their meaning due to an unknown inflation of the alpha-level. That does not mean that exploratory analysis is not valuable; it is just more tentative.

Open Science Practices in This Special Issue

The articles published in this special issue adopted transparency practices that further enhance the credibility of the published results. These practices make explicit how the research was conducted and make all the relevant materials and data available to facilitate reanalysis, reuse, and replication. The practices include:

- For all articles, original proposals, anonymized data, and study materials are registered and available at the Open Science Framework (OSF; <http://osf.io/>). Each article earned badges acknowledging preregistration, open data, and open materials (Miguel et al., 2014) that are maintained by the Open Science Collaboration (<https://osf.io/tvyxz/>). Badges and links to the OSF projects appear in the acknowledgments section of each article.
- Some OSF projects have additional material such as photos or video simulations of the procedures.
- All articles specify the contributions of each author.
- All articles specify funding sources.
- All articles disclosed whether authors had conflicts of interest (Greenwald, 2009).
- All articles make explicit all conditions, measures, data exclusions, and how samples sizes were determined (LeBel et al., 2013; Simmons, Nelson, & Simonsohn, 2012). This disclosure standard has been introduced at *Psychological Science* starting in January 2014 as an expectation for all reviewed submissions (Eich, 2013).

than suggested by the published findings (Blanken, Van de Ven, Zeelenberg, & Meijers, 2014).

Finally, Many Labs replication project (Klein et al., 2014) was a large international collaboration that amassed 36 samples and 6,344 participants to assess variation in replicability across samples and settings of 13 effects. It revealed relatively little variation in effect sizes across samples and settings, and demonstrated that crowdsourcing offers a feasible way to collect very large sample sizes and gain substantial knowledge about replicability.

No single replication provides the definitive word for or against the reality of an effect, just as no original study provides definitive evidence for it. Original and replication research each provides a piece of accumulating evidence for understanding an effect and the conditions necessary to obtain it. Following this special issue, *Social Psychology* will publish some commentaries and responses by original and replication authors of their reflections on the inferences from the accumulated data, and questions that could be addressed in follow-up research.

Closing

Registered Reports are new model for publishing that incorporates preregistration of designs and peer review before data collection. The approach nudges incentives for research accuracy to be more aligned with research success. As a result, the model may increase the credibility of the published results. Some pioneering journals in psychology and neuroscience have adopted Registered Reports offering substantial opportunity to evaluate and improve this publishing format (e.g., Chambers, 2013; Simons & Holcombe, 2013; Wolfe, 2013). Further, through the OSF (<http://osf.io/>), the Center for Open Science (<http://cos.io/>) provides free services to researchers and journals to facilitate Registered Reports and other transparency practices including badges, disclosure standards, and private or public archiving of research materials and data.

This special issue shows that the incentive structures to perform and publish replication studies and negative results can change. However, it is just a demonstration. Many cultural barriers remain. For example, when judging the importance of replication proposals, some reviewers judged a replication as unimportant because as expert “insiders” they already knew that the original result was not robust, even though this knowledge is not shared in the scientific literature. The irreproducibility of certain effects may be informally communicated among particular insiders, but never become common knowledge. Knowledge accumulation will be much more efficient if insider knowledge is accessible and discoverable. Registered Reports are just one step for addressing that challenge.

Two central values of science are openness and reproducibility. In principle, the evidence supporting scientific knowledge can be reproduced by following the original methodologies. This differentiates science from other ways of knowing – confidence in claims is not based on trusting the source, but in evaluating the evidence itself.

This Special Issue

The articles in this special issue demonstrate a variety of ways in which published findings can be important enough to replicate. For one, every discipline has a number of classic, textbook studies that exemplify a research area. These studies are worth revisiting, both to assure their robustness and sometimes to analyze the data with modern statistical techniques. This special issue contains several replications of textbook studies, sometimes with surprising results (Nauts, Langner, Huijismans, Vonk, & Wigboldus, 2014; Sinclair, Hood, & Wright, 2014; Vermeulen, Batenburg, Beukeboom, & Smits, 2014; Wessellmann et al., 2014).

Second, several teams (Brandt, IJzerman, & Blanken, 2014; Calin-Jageman & Caldwell, 2014; Johnson, Cheung, & Donnellan, 2014; Lynott et al., 2014) replicated recent work that has received substantial attention and citation. Given their high impact on contemporary research trajectories, it is important to investigate these effects and the conditions necessary to elicit them to ensure efficient development of theory, evidence, and implications.

Third, replication studies might provide a way to validate results when previous research lines have reached opposite conclusions (Žeželj & Joki, 2014), or provide more certainty about the presence and mechanisms of the original effect by performing direct replications while simultaneously testing theorized moderators (Gibson, Losee, & Foxwell, 2014; Moon & Roeder, 2014; Müller & Rothermund, 2014).

Fourth, replications can reveal boundary conditions, for example by showing how sex differences in distress from infidelity is reliably observed in a young sample, but not in an older sample (IJzerman et al., 2014). Performing direct replications can be especially insightful when a previous meta-analysis suggests the effect is much smaller

Acknowledgments

This article was supported by the Center for Open Science. B. N. and D. L. both conceived and wrote the paper. The authors declare no conflict of interest in the preparation of this article.

References

- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking, belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, *22*, 731–732.
- Blanken, I., Van de Ven, N., Zeelenberg, M., & Meijers, M. H. C. (2014). Three attempts to replicate the moral licensing effect. *Social Psychology*, *45*, 232–238. doi: 10.1027/1864-9335/a000189
- Brandt, M. J., IJzerman, H., & Blanken, I. (2014). Does recalling moral behavior change the perception of brightness? A replication and meta-analysis of Banerjee, Chatterjee, and Sinha (2012). *Social Psychology*, *45*, 246–252. doi: 10.1027/1864-9335/a000191
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 1–12. doi: 10.1038/nrn3475
- Calin-Jageman, R. J., & Caldwell, T. L. (2014). Replication of the superstition, performance study by Damisch, Stoberock, Mussweiler (2010). *Social Psychology*, *45*, 239–245. doi: 10.1027/1864-9335/a000190
- Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*, 609–610.
- Eich, E. (2013). Business not as usual. *Psychological Science*, *25*, 3–6. doi: 10.1177/0956797613512465
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS One*, *5*, e10068.
- Gibson, C. E., Losee, J., & Foxwell, C. (2014). A replication of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999): Identity salience and shifts in quantitative performance. *Social Psychology*, *45*, 194–198. doi: 10.1027/1864-9335/a000184
- Gibson, C. E., Losee, J., & Vitiello, C. (2014). A replication attempt of stereotype susceptibility: Identity salience and shifts in quantitative performance. *Social Psychology*, *45*, 1XX–1XX. doi: 10.1027/1864-9335/a000184
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*, 562–571.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1–20. doi: 10.1037/h0076157
- Greenwald, A. G. (2009). What (and where) is the ethical code concerning researcher conflict of interest? *Perspectives on Psychological Science*, *4*, 32–35.
- IJzerman, H., Blanken, I., Brandt, M. J., Oerlemans, J. M., Van den Hoogenhof, M. M. W., Franken, S. J. M., & Oerlemans, M. W. G. (2014). Sex differences in distress from infidelity in early adulthood and in later life: A replication and meta-analysis of Shackelford et al. (2004). *Social Psychology*, *45*, 202–208. doi: 10.1027/1864-9335/a000185
- Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). *Social Psychology*, *45*, 209–215. doi: 10.1027/1864-9335/a000186
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.
- Klein, R. A., Ratliff, K., Vianello, M., Adams, A. B. Jr., Bahník, S., Bernstein, N. B., ... Nosek, B. A. (2014). Investigating variation in replicability: A “Many Labs” Replication Project. *Social Psychology*, *45*, 142–152. doi: 10.1027/1864-9335/a000178
- Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, *7*, 608–614. doi: 10.1177/1745691612462586
- Lakens, D., & Evers, E. (in press). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*.
- LeBel, E., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K., Ratliff, K., & Smith, C. (2013). Psychdisclosure.Org: Grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, *8*, 424–432.
- LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem’s (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, *15*, 371–379. doi: 10.1037/a0025172
- Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. J., Lucas, R. E., & O’Brien, K. (2014). Replication of “Experiencing physical warmth promotes interpersonal warmth” by Williams and Bargh (2008). *Social Psychology*, *45*, 216–222. doi: 10.1027/1864-9335/a000187
- Makel, M., Plucker, J., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives in Psychological Science*, *7*, 537–542. doi: 10.1177/1745691612460688
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, *343*, 30–31. doi: 10.1126/science.1245317
- Moon, A., & Roeder, S. S. (2014). A secondary replication attempt of stereotype susceptibility (Shih, Pittinsky, & Ambady, 1999). *Social Psychology*, *45*, 199–201. doi: 10.1027/1864-9335/a000193
- Müller, F., & Rothermund, K. (2014). Replication of stereotype activation (Banaji & Hardin, 1996; Blair & Banaji, 1996). *Social Psychology*, *45*, 187–193. doi: 10.1027/1864-9335/a000183
- Nauts, S., Langner, O., Huijmsmans, I., Vonk, R., & Wigboldus, D. H. J. (2014). Forming impressions of personality: A replication and review of Asch’s (1946) evidence for a primacy-of-warmth effect in impression formation. *Social Psychology*, *45*, 153–163. doi: 10.1027/1864-9335/a000179
- Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, *5*, 85–90.
- Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, *8*, 21–29.
- Nosek, B. A., & Lakens, D. (2013). Call for proposals: Special issue of social psychology on “Replications of important results in social psychology”. *Social Psychology*, *44*, 59–60. doi: 10.1027/1864-9335/a000143
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II Restructuring incentives and practices to promote truth over publishability. *Perspectives in Psychological Science*, *7*, 615–631.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*, 657–660. doi: 10.1177/1745691612462588
- Open Science Collaboration. (2013). The reproducibility project: A model of large-scale collaboration for empirical research on reproducibility. In V. Stodden, F. Leisch, & R. Peng (Eds.),

- Implementing reproducible computational research (a volume in the r series)*. New York, NY: Taylor & Francis.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100. doi: 10.1037/a0015108
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612. doi: 10.1016/j.jrp.2013.05.009
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 word solution*. Retrieved from <http://ssrn.com/abstract=2160588>
- Simons, D., & Holcombe, A. (2013). *Registered replication reports*. Retrieved from <http://www.psychologicalscience.org/index.php/replication>
- Sinclair, H. C., Hood, K., & Wright, B. (2014). Revisiting Romeo and Juliet (Driscoll, Davis, & Lipetz, 1972): Reexamining the links between social network opinions and romantic relationship outcomes. *Social Psychology*, *45*, 170–178. doi: 10.1027/1864-9335/a000181
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – Or vice versa. *Journal of the American Statistical Association*, *54*, 30–34.
- Vermeulen, I., Batenburg, A., Beukeboom, C., & Smits, T. (2014). Breakthrough or one-hit wonder? Replicating effects of single-exposure musical conditioning on choice behavior (Gorn, 1982). *Social Psychology*, *45*, 179–186. doi: 10.1027/1864-9335/a000182
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638.
- Wesselmann, E. D., Williams, K. D., Pryor, J. B., Eichler, F. A., Gill, D., & Hogue, J. D. (2014). Revisiting Schachter's research on rejection, deviance, and communication (1951). *Social Psychology*, *45*, 164–169. doi: 10.1027/1864-9335/a000180
- Wolfe, J. M. (2013). Registered reports and replications in attention, perception, & psychophysics. *Attention, Perception, & Psychophysics*, *75*, 781–783. doi: 10.3758/s13414-013-0502-5
- Žeželj, I. L., & Jokić, B. R. (2014). Replication of experiments evaluating impact of psychological distance on moral judgment (Eyal, Liberman & Trope, 2008; Gong & Medin, 2012). *Social Psychology*, *45*, 223–231. doi: 10.1027/1864-9335/a000188

Published online May 19, 2014

Brian A. Nosek

Department of Psychology
University of Virginia
102 Gilmer Hall, Box 400400
Charlottesville, VA 22904
USA
E-mail nosek@virginia.edu
