

No Replication, no Trust? How Low Replicability Influences Trust in Psychology

Tobias Wingen

University of Cologne

Jana B. Berkessel

University of Mannheim

Birte Englich

University of Cologne

This paper is currently in press in *Social Psychological and Personality Science* (SPPS). The present post-print may not replicate the final version published in this SAGE journal. It is not the copy of record and may differ from the final version.

Please find the published version at:

<https://journals.sagepub.com/doi/10.1177/1948550619877412>

Author Note.

Tobias Wingen is a PhD candidate at the University of Cologne. His research focuses on how novel research methods can contribute to answering classical social-psychological questions. Moreover, he is interested in the open science movement, replicability, social hierarchies, and implicit theories.

Jana Berkessel is a PhD candidate at the Mannheim Centre for European Social Research at the University of Mannheim. Her work revolves around cross-cultural social and personality psychology, specifically culture fit, social class, and well-being. She is further interested in the social-psychological aspects of replicability and open science.

Birte Englich is a full professor at the University of Cologne. Her research, which currently focuses on topics including heuristics and biases, bias correction, social power, expertise, and indecisiveness, combines an applied perspective with research on the relevant underlying socio-cognitive processes.

We thank Nicolas Alef, Amelie Conrad, Elisabeth Jackson, and Estella Umbach for their support with the preparation of materials. We thank Alexandra Fleischmann and Oscar Lecuona for their helpful comments. Finally, we thank the Graduate School of the Faculty of Human Sciences at the University of Cologne for providing a travel grant to present this work at an international conference. We would like to dedicate this paper to the memory and friendship of Prof. Birte Englich (co-author of this article), who passed away in September 2019, shortly after the manuscript had been accepted for publication. The authors are grateful for her enduring trust and support, during this research project and way beyond.

Correspondence concerning this article should be addressed to Tobias Wingen, Applied Social Psychology and Decision Making, Social Cognition Center Cologne, University of Cologne, Herbert-Lewin-Str. 10, 50931 Köln, Germany. E-Mail: tobias.wingen@uni-koeln.de

Abstract

In the current psychological debate, low replicability of psychological findings is a central topic. While the discussion about the replication crisis has a huge impact on psychological research, we know less about how it impacts public trust in psychology. In this paper, we examine whether low replicability damages public trust and how this damage can be repaired. Studies 1, 2 and 3 provide correlational and experimental evidence that low replicability reduces public trust in psychology. Additionally, Studies 3, 4, and 5 evaluate the effectiveness of commonly used trust-repair strategies, such as information about increased transparency (Study 3), explanations for low replicability (Study 4), or recovered replicability (Study 5). We found no evidence that these strategies significantly repair trust. However, it remains possible that they have small but potentially meaningful effects, which could be detected with larger samples. Overall, our studies highlight the importance of replicability for public trust in psychology.

Keywords: replicability, public trust, replication crisis, open science

No Replication, no Trust? How Low Replicability Influences Trust in Psychology

A trustworthy reputation is crucial for psychology. Researchers in psychology aim to have societal impact and to inform practitioners and policymakers. Additionally, psychological research relies on public funding and participation. In turn, when facing complex scientific questions, decision-makers seek advice from the (psychological) science community (Bromme & Thomm, 2016). If based on robust evidence, these well-informed decisions can lead to improved individual and societal outcomes (Ruggeri et al., 2019). Thus, public trust is not only crucial for psychology itself, but also for the public. What happens to this trust when psychological findings often fail to replicate?

This is a relevant question since many prominent studies indeed suggest a low replicability of psychological findings. For example, the *Reproducibility Project: Psychology* replicated 100 psychological studies and only about one-third to one-half of the original findings were replicated (Open Science Collaboration, 2015). This low replication rate often serves as an illustration for a “replication crisis” (e.g. Anderson & Maxwell, 2017).

Low Replicability Might Damage Public Trust in Psychology

Many researchers see crises positively since they play a central role in the advancement of sciences and show that science self-corrects (Kuhn, 1970; Vazire, 2018). Indeed, in the course of the replication crisis, psychology has gone through major changes. Currently, journals and scientific societies encourage open science practices (e.g. Schönbrodt, Gollwitzer, & Abele-Brehm, 2017) and many researchers implement open science (Lindsay & Nosek, 2018). These changes are often seen as a major advancement of psychological science: Researchers argue that the rate of scientific progress is likely to increase (Vazire, 2018), that the widespread use of preregistration will increase the interpretability and credibility of research findings (Nosek, Ebersole, DeHaven, & Mellor, 2018), and that open science will liberate researchers and foster creativity (Frankenhuis & Nettle, 2018).

Nevertheless, there are reasons to assume that information about low replicability might damage public trust in psychology. While findings regarding the effects of (scientific) uncertainty on audience reactions are rather inconclusive (for a review, see van der Bles et al., 2019) some studies suggest that non-scientists react negatively to scientific uncertainty. For example, non-scientists who perceive scientific evidence as uncertain further perceive the corresponding research field as less valuable (Broomell & Kane, 2017). Likewise, even modest amounts of scientific dissent reduce public support for government policies and lead to disagreement with the scientific consensus (Aklin & Urpelainen, 2014). Similarly, low replicability might also result in reputational damage and diminished public trust (Białek, 2018; Chopik, Bremner, Defever, & Keller, 2018; Fanelli, 2018). We test this hypothesis in the present article.

How can Public Trust be Repaired?

If low replicability damages public trust, an important question for the psychological science community is if and how this damage can be repaired. We tested the following three theory-based and commonly used approaches to repair public trust.

Repairing trust through increased transparency. Transparency signals that there is nothing to hide and thus repairs trust (Bachmann, Gillespie, & Priem, 2015). Indeed, one major response to the replication crisis is the open science movement (Frankenhuis & Nettle, 2018; LeBel, Campbell, & Loving, 2017). Central aspects of this movement, such as preregistrations, open data, and open materials, aim to increase the transparency of psychological research (Miguel et al., 2014; Nosek et al., 2015). Thus, building on the idea that transparency can repair trust, information about the open science movement might help to repair public trust in psychology.

Repairing trust through explanations. The causes and responsibilities of a transgression are often not evident (Bachmann et al., 2015). Explanations of the causes of a

transgression can help to repair trust by establishing a shared understanding of what and why the transgression happened (Bachmann et al., 2015; Dirks, Lewicky, & Zaheer, 2009). If low replicability violates the public expectations of reliable published findings, the public may perceive low replicability as a transgression. In this case, explanations for low replicability could be an effective trust repair strategy.

Considering the replication crisis, two major explanations emerged. Some scholars argue that questionable research practices (QRPs) are the main reason for the replication crisis (e.g. Sijtsma, 2016; Simmons, Nelson, & Simonsohn, 2011). Other scholars attribute failed replications to hidden moderators and the high context-sensitivity of psychological effects (e.g. Stroebe & Strack, 2014; Van Bavel, Mende-Siedlecki, Brady, & Reinero, 2016). While this debate has not been settled, it is an additional open question whether any of those explanations – QRPs vs. hidden moderators – would repair public trust damaged by low replicability.

Repairing trust by restoring the status quo. Trust can further be repaired by restoring the status quo, as norms and expectations are also restored (Dirks et al., 2009). Before the replication crisis, the majority of replications in psychology journals reported similar findings to their original studies (Makel, Plucker, & Hegarty, 2012) and it was thus likely assumed that psychology is highly replicable. To restore this status quo, psychological science would thus need to achieve high levels of replicability. Indeed, many new methodological standards in psychology aim at increasing replicability (Cook, Lloyd, Mellor, Nosek, & Therrien, 2018; Van Bavel et al., 2016). If those standards succeed, the status quo belief that psychology is a highly replicable science might be restored. Eventually, this increase in replicability might also lead to a restoration of public trust.

Overview of Studies

We conducted five studies to test whether low replicability damages public trust and if this damage can be repaired. Study 1 examined whether trust in psychology correlates with expected replicability. Study 2 experimentally tested whether low replicability causes reduced trust, which we replicated in Study 3. Moreover, Studies 3 to 5 tested different commonly used trust-repair strategies: information about increased transparency (Study 3), explanations for low replicability (Study 4), and information about increased replicability (Study 5). Participants who took part in one of our studies were not allowed to participate in subsequent studies. We relied on MTurk workers in all studies, since they are significantly more socio-economically and ethnically diverse, and presumably less likely to have prior knowledge of the replication crisis, compared with a student sample (Casler, Bickel, & Hackett, 2013).

We include all studies we conducted, and report all collected variables and all conditions included in the study designs across all studies. We preregistered all analyses presented in the manuscript (except for specifically highlighted correlations presented in Table 1), and we report all preregistered analyses in either the manuscript or the supplemental materials. We discuss the central preregistered hypotheses when introducing each study. All analyses with a preregistered hypothesis are accompanied by one-sided p -values. All participants who completed our studies were included in the analyses except if they met preregistered exclusion criteria. All materials, data, analyses syntaxes, and preregistrations are shared on https://osf.io/9ba28/?view_only=7f2edfc9b5f143beb5f86dfdc657d73d.

Study 1

Study 1 investigated which replication rate non-scientists assume for psychological studies and whether their expected replication rate correlates with their trust in psychology and their perceived value of psychological science. We expected positive correlations.

Method

Participants and design. Participants completed a short online study on Amazon's Mechanical Turk (MTurk) website for \$0.50. The sample size was set to 266, based on an a priori power analysis for 95% power (one-sided α of .05) to detect a small to moderate effect of $r = .2$, that would be typical for similar social-psychological research (Richard, Bond Jr, & Stokes-Zoota, 2003). The final sample was slightly larger as is often the case in online studies and consisted of 271 participants (54.3% male; age: $M = 33.7$ years, $SD = 8.9$). No participants were excluded from the analyses. A sensitivity analysis showed that our final sample had a high chance ($1 - \beta = 0.80$, one-sided $\alpha = 0.05$) to detect a correlation of $r = .15$ and a very high chance ($1 - \beta = 0.95$, one-sided $\alpha = 0.05$) to detect $r = .20$.

Procedure. Participants read a short, jargon-free description of the *Reproducibility Project: Psychology* (for details, see https://osf.io/9ba28/?view_only=7f2edfc9b5f143beb5f86dfdc657d73d). Participants then guessed how many of these 100 original findings were successfully replicated.

Afterward, participants indicated their trust in psychology with five items (e.g., "I trust the psychological science community to do what is right; 1 = *strongly disagree*, 7 = *strongly agree*; $\alpha = .90$; adapted from Nisbet et al., 2015). Although we conveniently call this measure "trust in psychology" throughout this manuscript, it is important to note that it was designed to measure institutional trust in the (psychological) science community (Nisbet et al., 2015). Alternatively, trust in psychology could for example also be conceptualized as trust in psychological findings (Anvari & Lakens, 2019) or in the scientific methods used by psychologists. However, for non-scientists, the scientific community might be the most vivid aspect of psychology. Moreover, prior research showed that the used "trust in psychology"-measure is affected by dissonant science communication (Nisbet et al., 2015), so it could be particularly suitable to capture potential effects of (expected) low replicability.

Although this measure showed acceptable to excellent reliability in Studies 1-5, it showed a poor confirmatory model fit across most indices and studies. We believed this to likely be due to the reverse-coding of items and found that accounting for this drastically improves the model fit and does not change the pattern of our results (see supplemental materials).

As an additional dependent variable, we measured participants perceived value of psychological science with four items (e.g., “Please rate the societal benefit of research produced by psychological science.”; $\alpha = .80$; 1 = *very low*, 5 = *very high*; adapted from Broomell & Kane, 2017). In all studies, perceived value showed a similar result pattern to trust in psychology. All preregistered analyses regarding perceived value can be found in the supplemental material. Finally, participants indicated whether they knew the results of the *Reproducibility Project: Psychology* and completed demographics.

Results

Eleven participants (4.0%) said they had heard of the *Reproducibility Project: Psychology* but only one participant reported to know the results. This participant, however, indicated an incorrect replication rate of 14 out of 100 studies. On average, participants estimated that 60.9 out of 100 studies could be replicated ($SD = 22.9$). Descriptive statistics for the “trust in psychology”-measure across all studies and conditions are presented in Table 1. As predicted, the higher participants estimated the replication rate, the more they trusted psychology, $r(268) = .329$, one-sided $p < .001$, 95% CI [.218, .431] (see Fig. 1). Perceived value showed similar results to trust in psychology (see supplemental materials).

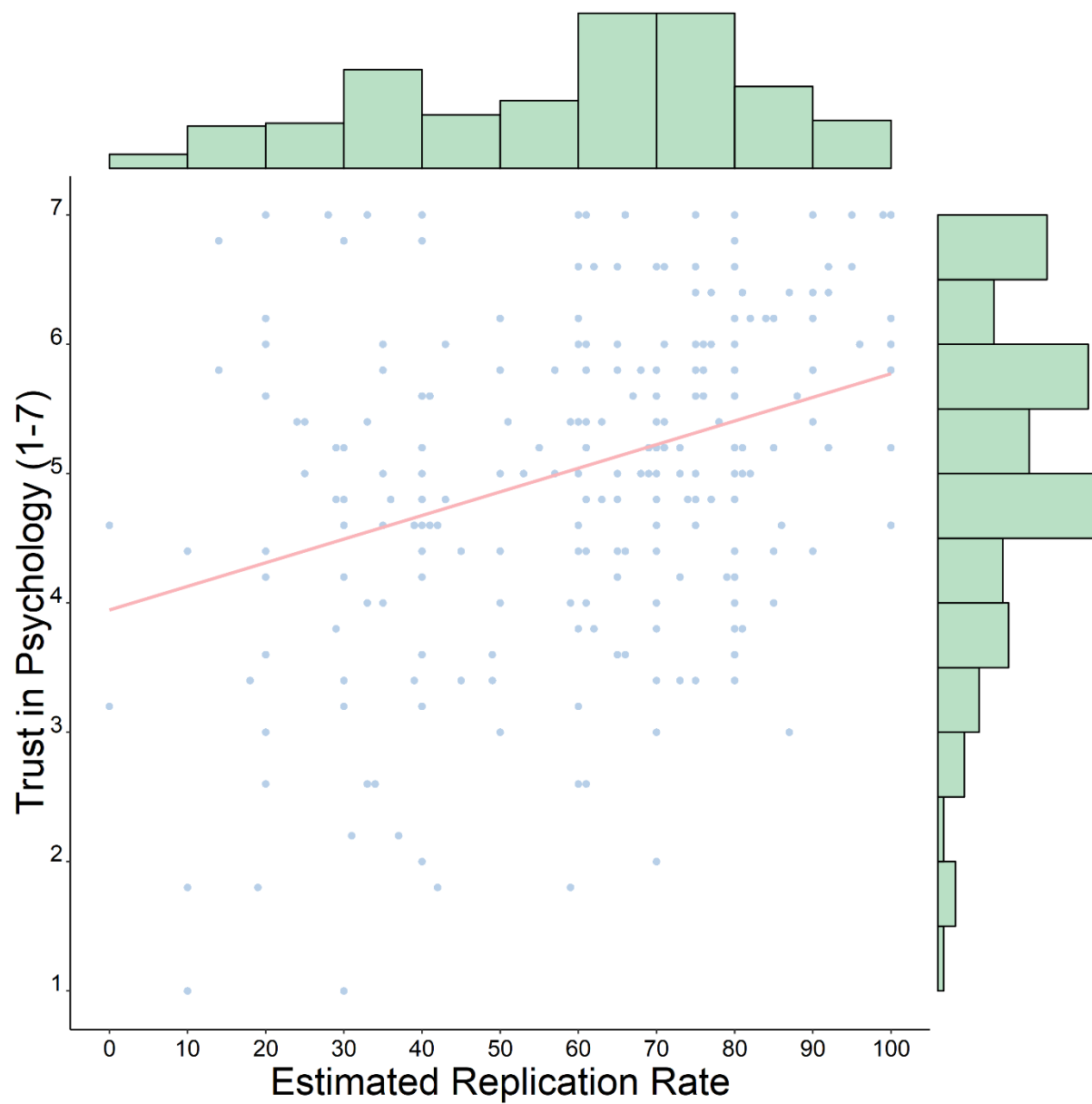


Figure 1. Relationship between the estimated replication rate and trust in psychology in Study

1. Histograms show the distribution of each measure.

Study 2

Study 1 provided correlational evidence that expected replicability is related to public trust in psychology. Building on Study 1, we employed an experimental approach to test causality. We expected low replicability (compared with high replicability) to reduce trust in psychology and to reduce the perceived value of psychological science.

Method

Participants and design. Participants completed a short online study on MTurk for \$0.50. We randomly assigned participants to three conditions (low replicability, medium replicability, high replicability). We set sample size to 264, based on an a priori power analysis for 95% power (one-sided α of .05) to find a moderate effect ($d = 0.5$), that would be typical for similar social psychological research (Richard et al., 2003), requiring 88 participants per cell (the same power analysis was applied to Studies 3, 4, and 5). The final sample consisted of 269 participants (59.9% male; age: $M = 34.59$ years, $SD = 10.74$). No participants were excluded from the analyses. A sensitivity analysis showed that our final sample had a high chance ($1 - \beta = 0.80$, one-sided $\alpha = 0.05$) to detect a difference of $d = 0.37$ between the low replicability and any of the two other conditions and a very high chance ($1 - \beta = 0.95$, one-sided $\alpha = 0.05$) to detect $d = 0.50$.

Procedure. Participants read the same description of the *Reproducibility Project: Psychology* as in Study 1. This time, however, participants received information about the results. Depending on their condition, participants were told that out of the 100 investigated studies, 39 (low replicability condition), 61 (medium replicability condition) or 83 (high replicability condition) could be successfully replicated. These values were based on the estimated replication rates found in Study 1: 61 represents the mean estimated replication rate in Study 1, and 39 and 83 represent the mean plus/minus one standard deviation in Study 1. Afterward, participants responded to three text-understanding items and a manipulation check

(“Psychological research is replicable”; 1 = *strongly disagree*, 7 = *strongly agree*). Then, they filled out the five items from Study 1 to measure trust in psychology ($\alpha = .92$). We also measured participants’ perceived value of psychological science and various individual differences as preregistered potential moderators (beliefs about science, error culture, error attribution style; see supplemental materials for details). Finally, participants completed a brief demographic questionnaire and were debriefed.

Results

The manipulation check suggested that the manipulation was successful (see supplemental materials). A one-way analysis of variance revealed significantly different levels of trust in psychology between the three conditions, $F(2, 265) = 4.86, p = .008, \eta^2 = .04$, 90%-CI [.01, .07], see Figure 2. Participants in the low replicability condition indicated a significantly lower trust in psychology than participants in the high replicability condition, $t(176) = 3.25$, one-sided $p < .001, d = 0.49$, 95%-CI [0.19, 0.79].

Further analyses indicated that participants in the exploratory medium replicability condition tended to be more trustful than the participants in the low replicability condition, $t(176) = 1.48$, one-sided $p = .070, d = 0.22$, 95%-CI [-0.07, 0.52], and less trustful than participants in the high replicability condition, $t(176) = 1.59$, one-sided $p = .057, d = 0.24$, 95% CI [-0.06, 0.53], but these differences were not significant. Perceived value showed similar results to trust in psychology (see supplemental materials).

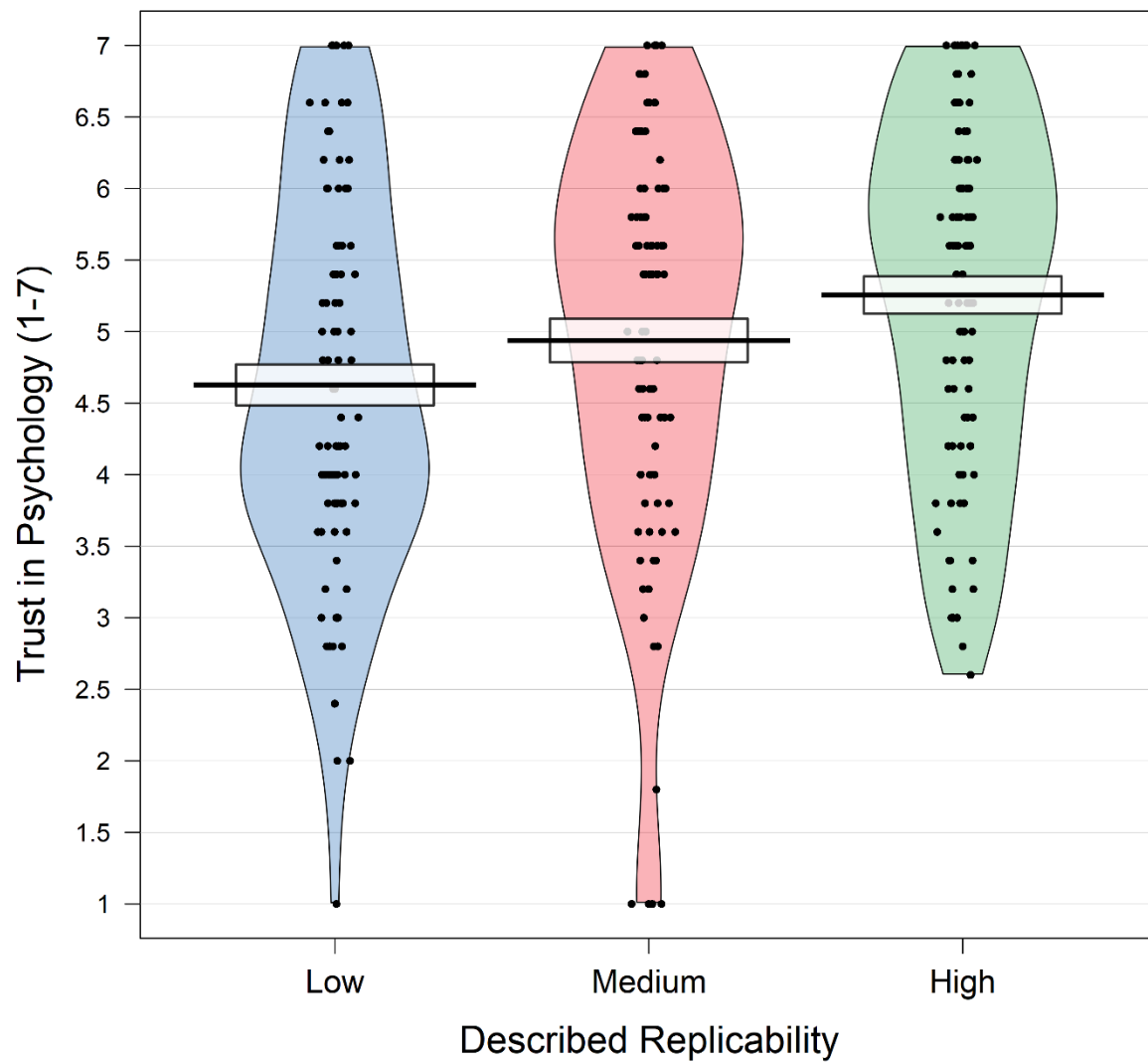


Figure 2. Pirate plot (Phillips, 2017) showing trust in psychology in the different replicability conditions in Study 2. The black dots represent the raw data which is shown with smoothed densities indicating the distributions in each condition. The central tendency is the mean and the intervals represent two standard errors around the mean.

Study 3

As expected, Study 1 and 2 provided evidence that low replicability, compared with high replicability, reduces trust in psychology. In Study 3, we replicated the trust-damaging effect of low replicability and tested whether informing participants about the open science movement and increased transparency of psychological science would repair trust damaged by low replicability. We expected low replicability (compared with high replicability) to reduce trust in psychology (as in Study 2). Crucially, we expected information about increased transparency to repair public trust, compared with information about low replicability only.

Method

Participants and design. Three hundred and four participants were recruited to complete a short online study on MTurk for \$0.60 each. Compared with Study 2, we increased the target sample size to 300 to compensate for potential exclusions. Indeed, seven participants were excluded for meeting the preregistered exclusion criteria (failing more than one text understanding questions). We randomly assigned participants to three conditions (low replicability, low replicability but transparency, high replicability). The final sample consisted of 297 participants (56.9% male; age: $M = 35.7$ years, $SD = 11.6$). A sensitivity analysis showed that our final sample had a high chance ($1 - \beta = 0.80$, one-sided $\alpha = 0.05$) to detect a difference of $d = 0.36$ between the low replicability and the low replicability but transparency condition and a very high chance ($1 - \beta = 0.95$, one-sided $\alpha = 0.05$) to detect $d = 0.47$.

Procedure. Participants read the same description of the *Reproducibility Project: Psychology* as in Study 2. Once again, participants were told that out of the investigated 100 studies, 39 (low replicability condition) or 83 (high replicability condition) could be successfully replicated. In a third condition (low replicability but transparency condition),

participants were also told that 39 studies could be replicated, but that psychology has since then become much more open and transparent. This comprehensible information described major aspects of the open science movement, including preregistration, open data, and open materials, and highlighted that those measures contribute to increased transparency (for details see https://osf.io/9ba28/?view_only=7f2edfc9b5f143beb5f86dfdc657d73d).

Afterward, participants responded to three text-understanding items and two manipulation checks (“Psychological research is replicable.”; “Psychological research is transparent.”; 1 = *strongly disagree*, 7 = *strongly agree*). Then, they filled out the five items from Studies 1 and 2 to measure their trust in psychology ($\alpha = .86$). Participants also completed a brief demographic questionnaire and were debriefed.

Results

The manipulation check suggested that the manipulations were successful (see supplemental materials).

As predicted, and replicating our prior findings, participants in the low replicability condition indicated a significantly lower trust in psychology than participants in the high replicability condition, $t(196) = 3.36$, one-sided $p < .001$, $d = 0.48$, 95%-CI [0.19, 0.76]; see Figure 3. However, contrary to our prediction, participants in the low replicability but transparency condition did not indicate a significantly higher trust in psychology than participants in the low replicability condition, $t(194) = 0.74$, one-sided $p = .231$, $d = 0.11$, 95%-CI [-0.18, 0.39].

Finally, participants in the low replicability but transparency condition indicated a significantly lower trust in psychology compared with participants in the high replicability condition, $t(192) = 2.60$, $p = .010$, $d = 0.37$, 95%-CI [0.09, 0.66].

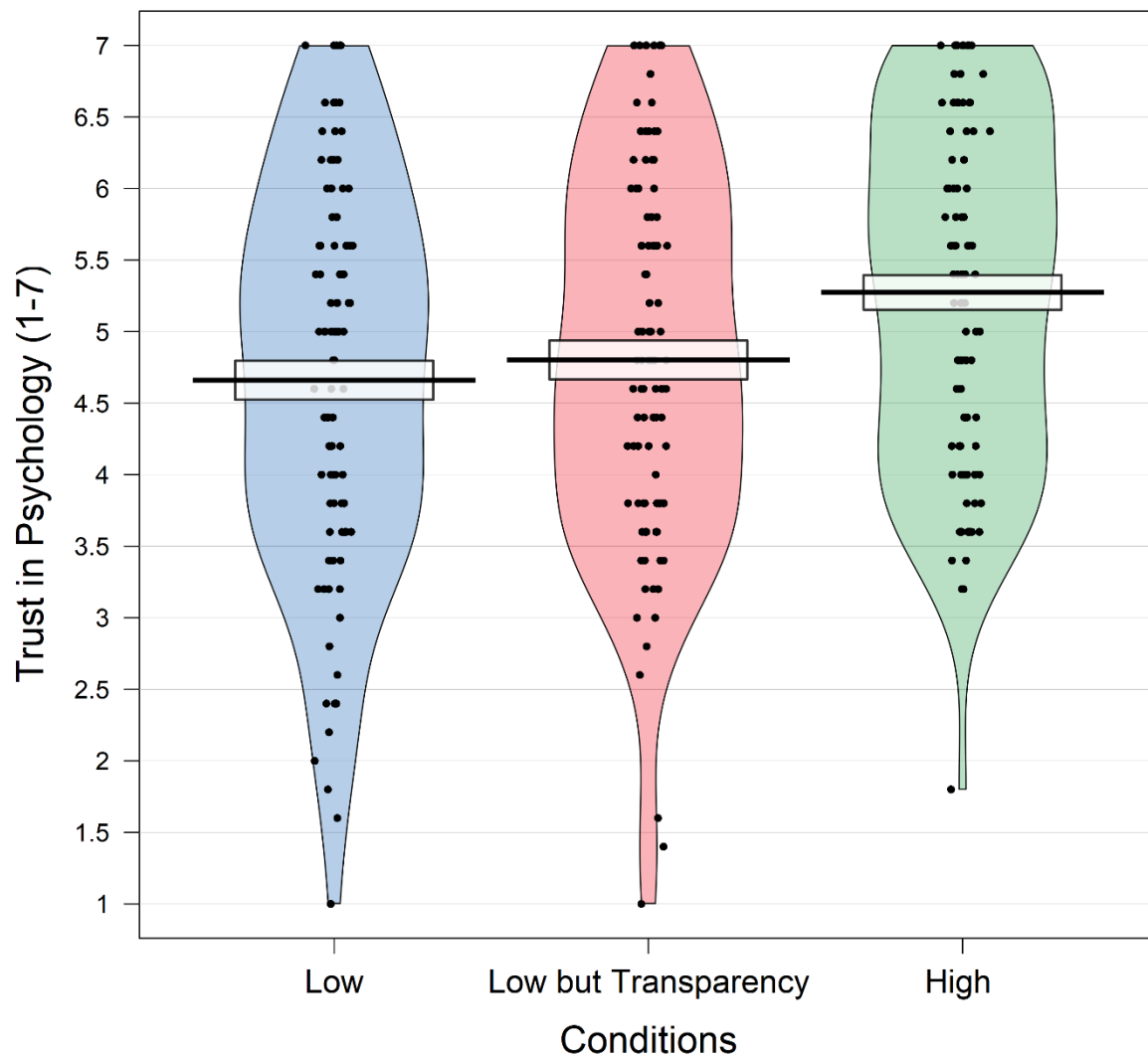


Figure 3. Pirate plot showing trust in psychology in the different replicability and transparency conditions in Study 3.

Study 4

Study 3 found no evidence that increased transparency can repair trust. While this approach focused on the consequences of the replication crisis, another approach to repair public trust might be to explain the causes of low replicability. Thus, in Study 4, we tested the effectiveness of the two most common explanatory strategies: hidden moderators and QRPs.

We expected the hidden moderator explanation to lead to a higher trust in psychology than the QRPs explanation. However, given the little effectiveness of our trust repair strategy in Study 3, we had no clear hypotheses on whether any of the explanations would be able to repair trust compared with providing no explanation, so we preregistered these analyses as exploratory.

Method

Participants and design. Three hundred and three participants were recruited to complete a short online study on MTurk for \$0.60. 20 participants were excluded for meeting the preregistered exclusion criteria (failing more than one text understanding questions). We randomly assigned participants to three conditions (low replicability condition, hidden moderator condition, QRPs condition). The final sample consisted of 283 participants (55.5% male; age: $M = 36.5$ years, $SD = 12.0$). Sensitivity analyses showed that our final sample had a high chance ($1 - \beta = 0.80$, $\alpha = 0.05$) to detect a difference of $d = 0.41$ between the low replicability and any of the two explanation conditions and a very high chance ($1 - \beta = 0.95$, $\alpha = 0.05$) to detect $d = 0.53$.

Procedure. Participants read the same description of the *Reproducibility Project: Psychology* as in Studies 2 and 3. All participants were told that out of the 100 investigated studies, 39 were successfully replicated. Depending on their condition, participants received no explanation (low replicability condition), an explanation stating that QRPs caused the low replication rate (QRPs condition) or an explanation stating that hidden moderators caused the low replication rate (hidden moderator condition). In the QRPs condition, participants read that researchers “primarily look for new and spectacular results which can lead to bad research practices, for example repeating an experiment until a surprising effect emerges. Often researchers only publish the spectacular results, while less spectacular – but potentially more reliable – results end up in a drawer somewhere.” In contrast to that, participants in the hidden moderator condition learned that: “When studying humans, unknown or hidden factors

such as individual differences between participants, participants' current state, or minimal differences in the experimental procedure can affect the results. It is very difficult to always have absolute control over these conditions and keep possible influencing factors constant.”

(for details see https://osf.io/9ba28/?view_only=7f2edfc9b5f143beb5f86dfdc657d73d).

Afterward, participants responded to three text-understanding items and two manipulation checks (1. “Unknown or hidden factors explain the low replication rate”, 2. “questionable research practices explain the low replication rate”; 1 = *strongly disagree*, 7 = *strongly agree*).

Participants did not answer the manipulation checks in the low condition, to avoid highlighting these explanations to control participants. Then, participants filled out the same five items used in Studies 1 - 3 to measure their trust in psychology ($\alpha = .80$). Participants also completed a brief demographic questionnaire and were debriefed.

Results

Manipulation checks indicated that the manipulation was successful (see supplemental materials).

Participants in the QRPs condition showed a significantly lower trust in psychology than participants in the hidden moderator condition, $t(185) = 2.11$, one-sided $p = .018$, $d = 0.31$, 95%-CI [0.02, 0.60]; see Figure 4. However, the low replicability condition, which served as a control condition, did not differ significantly from the hidden moderator condition, $t(188) = 0.20$, $p = .839$, $d = 0.03$, 95%-CI [-0.27, 0.32] nor from the QRPs condition, $t(183) = -1.68$, $p = .094$, $d = 0.25$, 95%-CI [0.04, 0.54], which showed an even lower trust than the low replicability condition.

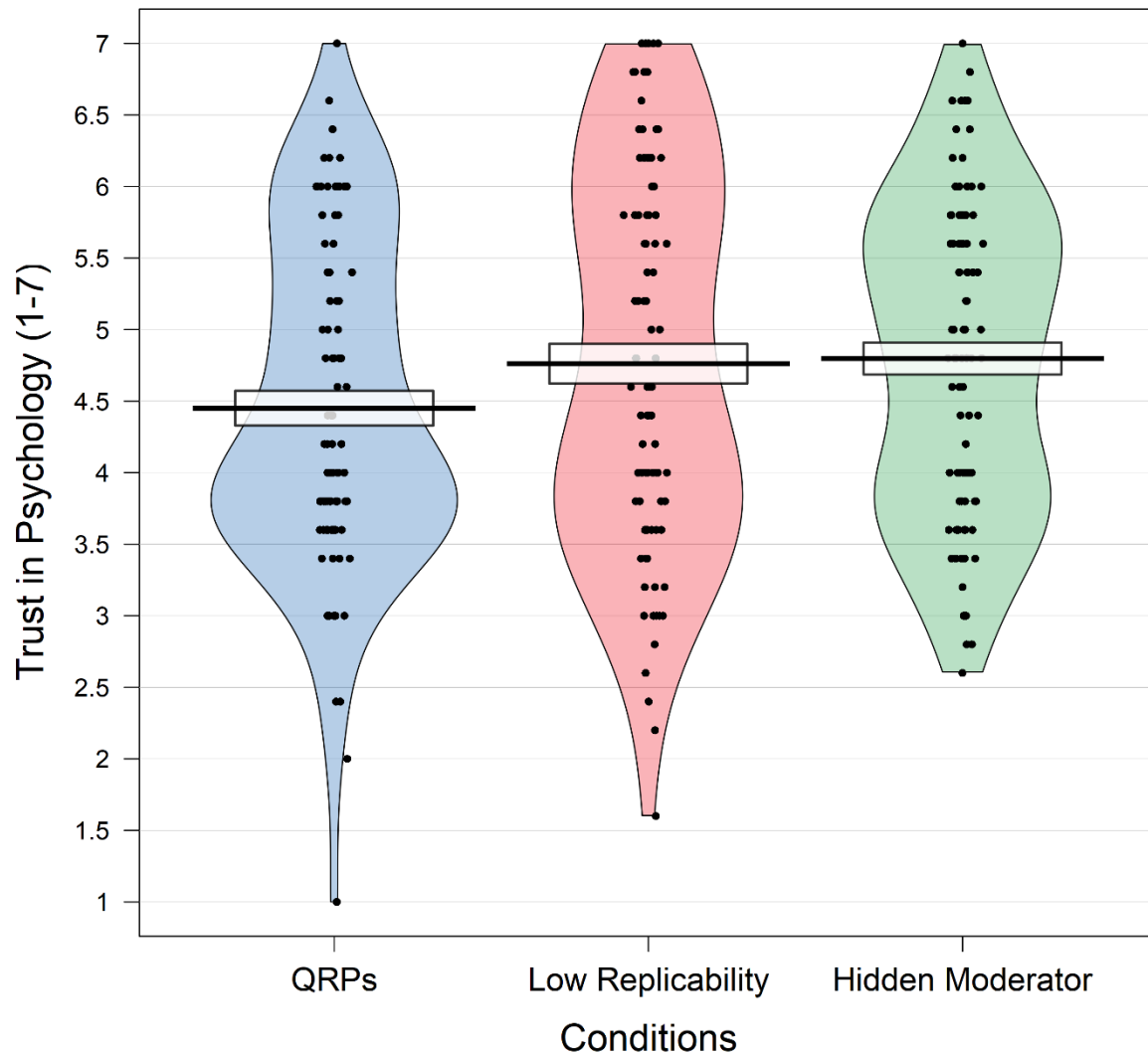


Figure 4. Pirate plot showing trust in psychology in the different explanation conditions in Study 4.

Study 5

Neither increased transparency (Study 3), nor explanations (Study 4) significantly repaired trust. One reason for this might be that we did not provide information about both, the causes and adequate solutions to the crisis, in one study. If non-scientists intuitively do not

believe that nontransparent practices (e.g., QRPs) cause low replicability, increasing transparency would not be a sensible response to low replicability. Thus, it might be necessary to inform non-scientists about both: QRPs as a cause of low replicability and transparency as an adequate solution. Whereas a QRP explanation on its own might even damage public trust (see Study 4), such an explanation could be especially effective when combined with information about increased transparency.

Moreover, we did not provide information about whether increased transparency was indeed effective in increasing replicability. Thus, we conducted Study 5 to address these concerns. In this final study, we tested whether public trust can be repaired by providing participants with both, information about the causes of, and adequate solutions for low replicability, and by informing them that these solutions successfully restored high replicability. We expected successfully restored replicability to lead to increased trust in psychology.

Method

Participants and design. Three hundred and four participants were recruited to complete a short online study on MTurk for \$0.50 each. We again used an increased target sample size of 300 to compensate for potential exclusions. 26 participants were excluded for meeting the preregistered exclusion criteria (failing more than one text understanding questions). We randomly assigned participants to three conditions (low replicability condition, “now high” replicability condition, “still low” replicability condition). The final sample consisted of 278 participants (64.7% male; age: $M = 33.8$ years, $SD = 11.0$). Sensitivity analyses showed that our final sample had a high chance ($1 - \beta = 0.80$, $\alpha = 0.05$) to detect a difference of $d = 0.36$ between the low replicability and any of the two other conditions and a very high chance ($1 - \beta = 0.95$, $\alpha = 0.05$) to detect $d = 0.48$.

Procedure. Participants read the same description of the *Reproducibility Project: Psychology* as in Studies 2, 3, and 4, and additionally learned that the *Reproducibility Project: Psychology* was published in 2015. All participants read that out of the 100 investigated studies, 39 were successfully replicated. In the low replicability condition, participants received no further information. In the “still low” replicability and “now high” replicability conditions, participants received an explanation that QRPs caused the low replication rate, but that this issue was now addressed through the open science movement and the increased transparency of psychological science. In the “still low” condition, which served as an additional control group, participants learned that these measures were not successful. Concretely, they were informed that an (alleged) new systematic replication project in 2018 revealed that out of 100 studies conducted under the new transparency guidelines, still only 41 could be successfully replicated. In contrast, in the “now high” replicability condition, participants learned that those measures were very successful since the alleged new replication project in 2018 revealed that now 83 out of 100 recent studies could be successfully replicated (for details see https://osf.io/9ba28/?view_only=7f2edfc9b5f143beb5f86dfdc657d73d). Afterward, participants responded to three text-understanding items and to the manipulation check (“Psychological research is now more replicable”). Participants did not fill out the manipulation check in the low replicability condition, which received no information about the change in replicability. Then, participants answered the five items from Studies 1 - 4 to measure their trust in psychology ($\alpha = .73$). Participants also completed a brief demographic questionnaire and were debriefed.

Results

According to our manipulation checks, the manipulation was successful (see supplemental materials).

Participants in the “now high” replicability condition did not show significantly higher trust in psychology than participants in the “still low” replicability condition, $t(178) = 1.29$, one-sided $p = .099$, $d = 0.19$, 95%-CI [-0.10, 0.49], or participants in the low replicability condition ($M = 4.35$, $SD = 1.30$), $t(186) = 1.04$, one-sided $p = .149$, $d = 0.15$, 95% CI[-0.14, 0.44]; see Figure 5.

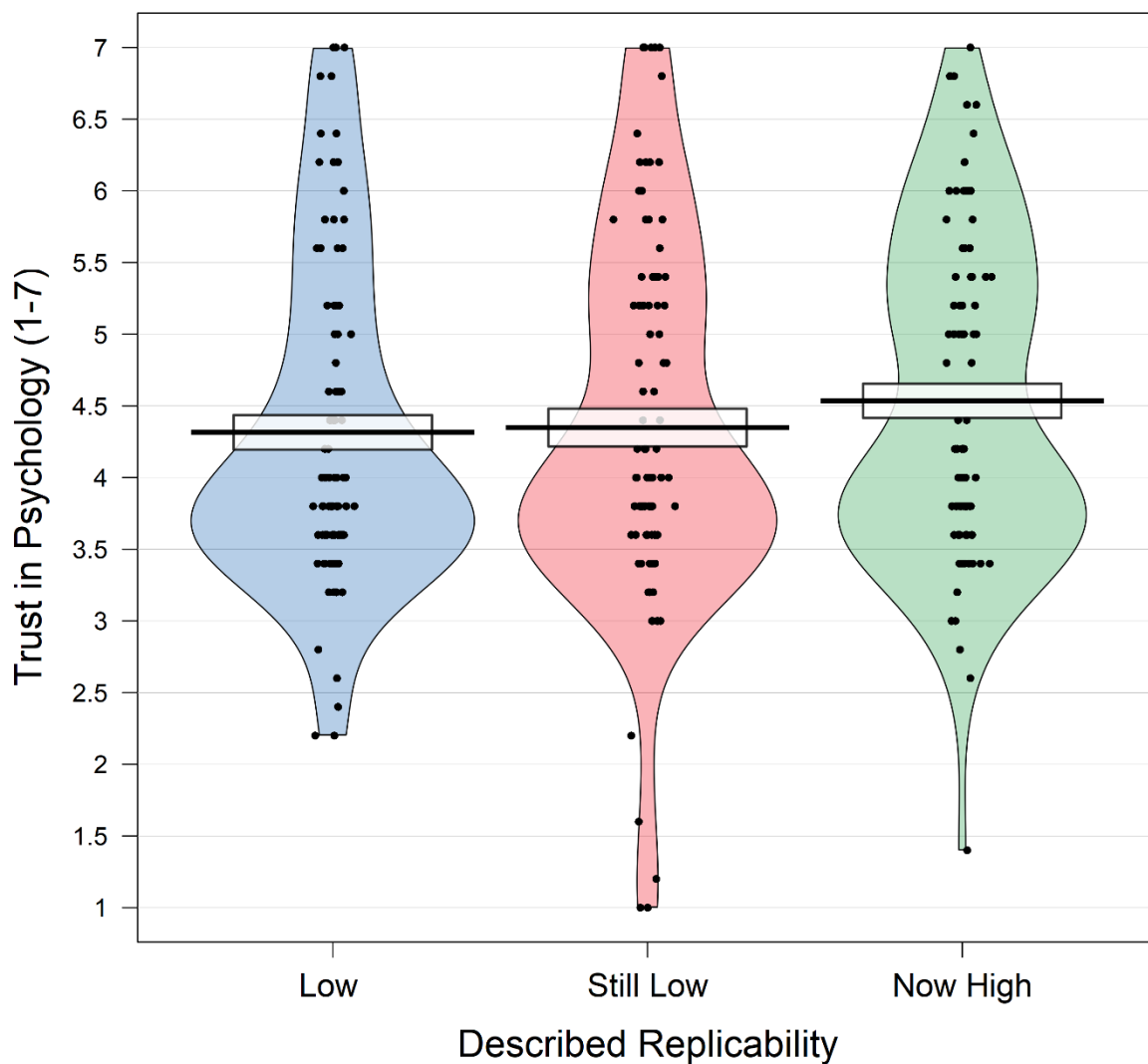


Figure 5. Pirate plot showing trust in psychology in the different replicability conditions in Study 5.

Table 1

Descriptive Statistics for the “Trust in Psychology”-Measure Across Studies

Study	Sample Size ^a	Mean	Standard Deviation	Correlation with Replicability ^b
Study 1	270	5.06	1.27	$r(268) = .329^{***}$
Study 2	268	4.94	1.36	
1. Low Replicability	88	4.63	1.34	$r(86) = .205$
2. Medium Replicability	90	4.94	1.44	$r(88) = .424^{***}$
3. High Replicability	90	5.26	1.24	$r(88) = .239^*$
Study 3	294	4.91	1.32	
1. Low Replicability	100	4.66	1.36	$r(98) = .218^*$
2. Transparency	96	4.80	1.33	$r(94) = .295^{**}$
3. High Replicability	98	5.27	1.20	$r(96) = .373^{***}$
Study 4	281	4.67	1.21	
1. Low Replicability	94	4.76	1.34	-
2. Hidden Moderators	96	4.80	1.09	-
3. QRPs	91	4.45	1.16	-
Study 5	278	4.40	1.20	
1. Low Replicability	98	4.35	1.30	-
2. Still Low Replicability	90	4.32	1.14	-
3. Now High Replicability	90	4.54	1.14	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

^aNumber of participants who completed the “trust in psychology”-measure

^bIn Study 1, this correlation refers to the preregistered correlation of the “trust in psychology”-measure (ranging from 1 to 7) with the estimated replication rate. In Studies 2 and 3 this refers to the not preregistered correlation with the manipulation check “Psychological research is replicable”. This manipulation check was not administered in Studies 4 and 5.

General Discussion

Our results show that concerns about reduced public trust in light of the replication crisis are justified. Across three studies (Studies 1 - 3), we find correlational and experimental evidence that low replicability reduces trust in psychology. Studies 1 and 2 suggest that not only public trust but also the perceived value of psychological science is damaged by low replicability. Moreover, Studies 3 to 5 found no evidence that commonly used trust repair strategies significantly repair this damaged trust in psychology.

So does low replicability damage public trust beyond repair? Although sensitivity analyses showed that it is unlikely that the tested strategies have large trust-repairing effects, they also suggest that we had no sufficient power to rule out small, but potentially meaningful effects, which could only be detected with larger samples (equivalence tests and Bayes factors in line with this argumentation are presented in the supplemental materials). Our findings thus do not allow us to conclude that the tested strategies are certainly ineffective. However, given the non-significant observed effects of trust repair strategies, our findings also do not provide evidence for the effectiveness of the tested strategies on trust in psychology.

Hence, the critical question is: What should psychological researchers do if they encounter low replicability? Considering that replication studies have limitations and that there is often no consensus about their interpretation (Gilbert, King, Pettigrew, & Wilson, 2016), one could potentially argue that psychologists should avoid informing the public about low replicability. However, this non-transparent approach would be ethically problematic and violates, for example, the APA Ethics Code (see APA, 2017, *pp.* 3- 4). Moreover, failed attempts to cover up problematic research findings might reduce public trust even more (Leiserowitz, Maibach, Roser-Renouf, Smith, & Dawson, 2013). Therefore, covering up low replicability is neither an ethical nor an effective way to handle the problem.

A more promising approach to maintaining the public trust might be to substantially improve the replicability of psychological research findings. Although Study 5 remains inconclusive about whether this is an effective strategy to repair the public trust directly after a replication crisis, Studies 1 to 3 provide evidence that high replicability in the first-place results in increased trust in psychology. Thus, if replicability is constantly high, public trust in psychology might rise. Currently, there is considerable debate about whether constantly high replicability is a worthwhile goal for psychological science. For example, Baumeister (2016) discussed whether a strong focus on replicability could potentially reduce the likelihood of discoveries and the progress and influence of the field. Moreover, scholars debate whether conducting direct replications is after all meaningful (Stroebe & Strack, 2014; cf. Simons, 2014). Although we do not directly speak to these arguments, our work suggests that the debate should also consider the reputational benefits associated with high replicability.

However, it is important to note that we communicated information about low replicability in the form of very short texts, inspired by brief news reports. Potentially, an in-depth explanation of the replication crisis and the open science movement might lead to less negative, or even positive, audience reactions. This is especially likely for highly science-interested audiences which would be willing to engage with such a detailed explanation. Indeed, recent research suggests less negative consequences in such a situation: After a 1-hr lecture on the replication crisis, psychology students' attitudes toward psychology remained relatively stable (Chopik et al., 2018).

Moreover, we conceptualized trust in psychology as trust in the psychological science community. Trust in psychology could however also refer to trust in psychological findings. Since low replicability typically refers to past findings, it seems possible that low replicability of past findings does not necessarily damage trust in future findings (Anvari & Lakens, 2019). Likewise, it is possible that the damaged trust in the psychological science community does

not generalize to future generations of psychological researchers educated under new, more rigorous methodological guidelines.

Overall, our studies highlight the crucial importance of replicability for public trust in psychology. Thus, the immense effort of the psychological science community to increase replicability is not only scientifically important but also highly relevant to psychology's public reputation. This is especially important in the current political climate, where the credibility of scientific evidence is questioned and science is threatened by defunding (Fanelli, 2018; Yong, 2017).

References

- Aklin, M., & Urpelainen, J. (2014). Perceptions of scientific dissent undermine public support for environmental policy. *Environmental Science & Policy*, 38, 173–177.
<https://doi.org/10.1016/j.envsci.2013.10.006>
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “Replication Crisis”: Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52, 305–324.
<https://doi.org/10.1080/00273171.2017.1289361>
- Anvari, F., & Lakens, D. (2019, September 9). *The Replicability Crisis and Public Trust in Psychological Science*. <https://doi.org/10.31234/osf.io/vtmpc>
- APA. (2017). *Ethical principles of psychologists and code of conduct*. Washington, DC.
- Bachmann, R., Gillespie, N., & Priem, R. (2015). Repairing trust in organizations and institutions: Toward a conceptual framework. *Organization Studies*, 36, 1123–1142.
- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153–158. <https://doi.org/10.1016/j.jesp.2016.02.003>
- Białek, M. (2018). Replications can cause distorted belief in scientific progress: BBS commentary by Białek on Zwaan et al. *PsyArXiv*.
<https://doi.org/10.31234/osf.io/8a4h6>
- Bromme, R., & Thomm, E. (2016). Knowing who knows: Laypersons’ capabilities to judge experts’ pertinence for science topics. *Cognitive Science*, 40, 241–252.
- Broomell, S. B., & Kane, P. B. (2017). Public perception and communication of scientific uncertainty. *Journal of Experimental Psychology: General*, 146, 286–304.
<https://doi.org/10.1037/xge0000260>

- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156–2160.
- Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, 45, 158–163.
- Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, 85, 104–118.
- Dirks, K. T., Lewicky, R. J., & Zaheer, A. (2009). Repairing relationships within and between organizations: Building a conceptual foundation. *The Academy of Management Review*, 34, 68–84.
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, 115, 2628–2631.
- Frankenhuis, W. E., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science*, 13, 439–447.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science.” *Science*, 351, 1037–1037.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl]). Chicago: University of Chicago Press.
- LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology*, 113, 230–243. <https://doi.org/10.1037/pspi0000049>
- Leiserowitz, A. A., Maibach, E. W., Roser-Renouf, C., Smith, N., & Dawson, E. (2013). Climategate, public opinion, and the loss of trust. *American Behavioral Scientist*, 57, 818–837.

- Lindsay, D. S., & Nosek, B. A. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, 31. Retrieved from <https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537–542. <https://doi.org/10.1177/1745691612460688>
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... Imbens, G. (2014). Promoting transparency in social science research. *Science*, 343, 30–31.
- Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The ANNALS of the American Academy of Political and Social Science*, 658, 36–66. <https://doi.org/10.1177/0002716214555474>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Christensen, G. (2015). Promoting an open research culture. *Science*, 348, 1422–1425.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115, 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
- Phillips, N. (2017). Yarr: A companion to the e-book “yarr!: The pirate’s guide to r.” Retrieved from <https://CRAN.R-project.org/package=Yarr>.
- Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331–363.

Ruggeri, K., Ojinaga-Alfageme, O., Benzerger, A., Berkessel, J., Hlavová, R., Kunz, M., ...

Sampat, B. (2019). Evidence-based policy. In *Behavioral Insights for Public Policy: Concepts and Case* (pp. 17–40). London: Routledge.

Schönbrodt, F., Gollwitzer, M., & Abele-Brehm, A. (2017). *Data management in psychological science: Specification of the DFG guidelines*.

Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, 81, 1–15.

<https://doi.org/10.1007/s11336-015-9446-0>

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.

<https://doi.org/10.1177/0956797611417632>

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76–80.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication.

Perspectives on Psychological Science, 9, 59–71.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual

sensitivity in scientific reproducibility. *Proceedings of the National Academy of*

Sciences, 113, 6454–6459. <https://doi.org/10.1073/pnas.1521897113>

van der Bles, A. M., van der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L.,

& Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, 6(5), 181870.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13, 411–417.

Yong, E. (2017). How the GOP could use science's reform movement against it. *The Atlantic*.

Retrieved from <https://www.theatlantic.com/science/archive/2017/04/reproducibility-science-open-judoflip/521952/>

Supplemental Materials for*No Replication, no Trust? How Low Replicability Influences Trust in Psychology***Analyses with Additional Variables (Perceived Value of Psychological Science and Potential Moderators)****Study 1**

Additional variables and analyses. We measured participants' perceived value of psychological science as an additional dependent variable with four items (e.g., "Please rate the societal benefit of research produced by psychological science."; $\alpha = .80$; 1 = very low, 5 = very high; adapted from Broomell & Kane, 2017). This estimated replication rate was significantly correlated with the perceived value of psychological science, $r(267) = .309$, 95% CI [0.20, 0.41], one-sided $p < .001$. The results regarding the perceived value of psychological science were thus parallel to our results regarding the public trust in the psychological science community.

Study 2

Additional variables. The same four items from Study 1 were used to measure the perceived value of psychological science ($\alpha = .83$) as an additional dependent variable. To explore potential moderators, participants responded to three items to measure beliefs about whether science is an absolute truth or a debate (Rabinovich & Morton, 2012; $\alpha = .72$; e.g., "There may be more than one correct answer to most scientific questions."), and finally two subscales of the error orientation questionnaire (Rybowiak, Garst, Frese, & Batinic, 1999), namely "learning from errors" ($\alpha = .92$; e.g., "My mistakes help me to improve my work.") and "error communication" ($\alpha = .82$; e.g., "When I make a mistake, I tell others about it in order that they do not make the same mistake."). We also created two items to measure

internal (“Mistakes are often caused by internal factors, such as lack of skill or effort.”) vs. external error attribution style (“Mistakes are often caused by external factors, such as other persons or circumstances”). However, these self-created items seem to lack validity, as they correlated positively with each other, $r(264) = .18, p = .003$. Agreement to all items, except for the perceived value of psychological science, was measured on a scale from 1 = “very low” to 7 = “very high”.

Additional analyses. Regarding the perceived value of psychological science, a one-way analysis of variance revealed a significant difference between the three conditions, $F(2, 264) = 3.04, p = .049, \eta^2 = .02, 95\% \text{ CI } [0.0003, 0.41]$.

Participants in the low replicability condition indicated a significantly lower perceived value of psychological science ($M = 3.36, SD = 0.85$) than participants in the high replicability condition ($M = 3.66, SD = 0.72$), $t(175) = 2.51$, one-sided $p = .006, d = 0.38$, $95\% \text{ CI } [0.08, 0.68]$. The results regarding the perceived value of psychological science were thus again parallel to our results regarding the public trust in the psychological science community.

Further analyses indicated that participants in the medium replicability condition ($M = 3.45, SD = 0.90$) perceived psychological science as not significantly more valuable than participants in the low replicability condition, $t(175) = 0.67$, one-sided $p = .25, d = 0.10, 95\% \text{ CI } [-.20, 0.40]$, and as less valuable than participants in the high replicability condition, $t(178) = 1.72$, one-sided $p = .043, d = 0.26, 95\% \text{ CI } [-0.04, 0.55]$.

We conducted multiple linear regression analyses to test whether any of our potential moderator variables moderated the relationship between replicability (low vs. high) and public trust. However, neither participants’ centered beliefs about science, $B = -0.10, t(173) = 0.74, p = .458$, nor their centered error attribution style, $B = -0.02, t(173) = 0.20, p = .839$, their centered score on the learning from errors-subscale, $B = -0.28, t(172) = 1.55, p = .122$ or

their centered score on the error communication-subscale, $B = -0.23$, $t(173) = 1.31$, $p = .193$ significantly moderated this relationship.

The relationship between replicability (low vs. high) and the perceived value of psychological science was also not significantly moderated by participants' centered beliefs about science, $B = -0.16$, $t(173) = 1.94$, $p = .054$, their centered error attribution style, $B = 0.06$, $t(173) = 0.92$, $p = .359$, their centered score on the learning from errors-subscale, $B = 0.04$, $t(172) = 0.31$, $p = .756$, or their centered score on the error communication-subscale, $B = 0.06$, $t(173) = 0.55$, $p = .582$.

Manipulation Checks

Study 2.

Participants in the low replicability condition indicated a significantly lower agreement to the manipulation check "Psychological research is replicable" ($M = 5.02$, $SD = 1.69$) than participants in the high replicability condition ($M = 5.72$, $SD = 1.08$), $t(177) = 3.30$, $p < .001$, $d = 0.49$, 95% CI [0.19, 0.79], so we deemed our manipulation successful.

In the exploratory medium replicability condition participants ($M = 4.99$, $SD = 1.34$) also agreed less to the manipulation check than in the high replicability condition, $t(178) = 4.03$, one-sided $p < .001$, $d = 0.60$, 95% CI [0.30, 0.90]. However, the medium and the low replicability condition did not differ significantly, $t(177) = 0.15$, one-sided $p = .558$, $d = 0.02$, 95% CI [-0.27, 0.32].

Study 3.

Participants in the high replicability condition indicated a significantly higher agreement to the manipulation check "psychological research is replicable" ($M = 5.78$, $SD = 1.00$) than participants in the low replicability condition ($M = 4.92$, $SD = 1.59$), $t(198) = 4.55$, one-sided $p < .001$, $d = 0.64$, 95% CI [0.36, 0.93], and participants in the low replicability but

transparency condition ($M = 5.04$, $SD = 1.55$), $t(194) = 3.96$, one-sided $p < .001$, $d = 0.57$, 95% CI [0.28, 0.85]. Participants in the low replicability but transparency condition indicated a significantly higher agreement to the manipulation check “psychological research is transparent” ($M = 5.05$, $SD = 1.70$) than participants in the low replicability condition ($M = 4.59$, $SD = 1.57$), $t(196) = 1.97$, one-sided $p = .025$, $d = 0.28$, 95% CI [-0.002, 0.56]..

Interestingly, a non-preregistered t -test showed that participants in the low replicability but transparency condition did not differ in their agreement to the transparency manipulation check compared to participants in the high replicability condition ($M = 4.91$, $SD = 1.39$), $t(194) = 0.64$, $p = .521$, $d = 0.09$, 95% CI [-0.19, 0.37].

Study 4.

Participants in the QRPs condition indicated a significantly lower agreement to the manipulation check “Unknown or hidden factors explain the low replication rate.” ($M = 4.14$, $SD = 1.96$) than participants in the hidden moderator condition ($M = 5.94$, $SD = 1.28$), $t(186) = 7.46$, one-sided $p < .001$, $d = 1.09$, 95% CI [0.78, 1.40]. In contrast, participants in the QRPs condition indicated a significantly higher agreement to the manipulation check “Questionable research practices explain the low replication rate.” ($M = 5.95$, $SD = 1.37$) than participants in the hidden moderator condition ($M = 3.30$, $SD = 1.96$), $t(186) = 10.68$, one-sided $p < .001$, $d = 1.56$, 95% CI [1.23, 1.89], indicating that our manipulations were successful.

Study 5.

Participants in the “still low” condition indicated a significantly lower agreement to the statement “Psychological research is now more replicable” ($M = 4.39$, $SD = 2.05$) than participants in the “now high” condition ($M = 6.02$, $SD = 0.97$), $t(178) = 6.82$, one-sided $p < .001$, $d = 1.02$, 95% CI [0.70, 1.33], indicating that our manipulation was successful.

Psychometric Properties of the Trust in Psychology Scale

In all studies, we measured participants' trust in psychology by using a scale for institutional trust in the psychological science community, which contains five items (e.g., "I trust the psychological science community to do what is right; 1 = strongly disagree, 7 = strongly agree; adapted from Nisbet, Cooper, & Garrett, 2015). Importantly, three of the items are reverse coded (i.e., measure distrust in the psychological science community), while only two of the items are positively coded. The scale showed acceptable to excellent reliability across all studies (all α s. $> .73$). However, a closer inspection of the fit indices in a confirmatory factor analysis (CFA) using the R package lavaan (Rosseel, 2012) showed poor fit of a simple one factor solution (s. Table 1).

To address this issue, we examined modification indices (*MI*) and expected parameter changes (*EPC*) to gain insight into how to improve the low fit. Across all 5 studies, the highest modification indices were found for the correlation of residual variances of the second and third item of the scale (Study 1, *MI* = 20.0, *EPC* = 0.249; Study 2, *MI* = 50.7, *EPC* = 0.440; Study 3, *MI* = 137.5, *EPC* = 1.06; Study 4, *MI* = 126.9, *EPC* = 1.09; Study 5, *MI* = 140.5, *EPC* = 1.31). A closer inspection revealed that these two items indeed have something in common that is not captured by the latent variables, namely that both are coded positively while all other items were reverse-coded. To improve the model fit, we thus modified the model to allow the residual variances of the second and third item to be correlated. This dramatically improved the model fit, especially in Studies 3 – 5, as displayed in Table 1.

Table 1.

Fit indices for a simple one-factor solution and for the modified model across studies.

Study	CFI	TLI	RMSEA	SRMR	Lowest standardized loading
Study 1	.97	.94	.14	.03	.73
Modified Model	.99	.98	.08	.02	.75
Study 2	.94	.87	.22	.05	.72
Modified Model	.99	.97	.11	.02	.68
Study 3	.79	.58	.35	.13	.46
Modified Model	.99	.97	.09	.03	.42
Study 4	.78	.55	.35	.16	.28
Modified Model	.99	.98	.08	.02	.26
Study 5	.73	.46	.39	.18	.02
Modified Model	.98	.95	.12	.02	.02

Note. Modified model refers to a model that allows the residual variances of the second and third item to be correlated.

To test whether this modification would affect our results, we conducted a robustness check by rerunning all our central analyses from the manuscript. However, this time we modeled trust in psychology by using factor scores obtained from the modified CFA (i.e., the CFA which allows the residual variances of the second and third item to be correlated) instead of simply calculating the mean.

Robustness Checks Using Factor Scores Instead of Means

Robustness Check: Study 1. In Study 1, a higher estimated replication rate still correlated significantly with a higher trust factor score $r(268) = .321$, one-sided $p < .001$, 95% CI[.210;.425].

Robustness Check: Study 2. Participants in the low replicability condition indicated a significantly lower trust factor score ($M = -0.35$, $SD = 1.39$) than participants in the high

replicability condition ($M = 0.30$, $SD = 1.24$), $t(176) = 3.31$, one-sided $p < .001$, $d = 0.50$, 95% CI [0.20, 0.80].

Robustness Check: Study 3. Again, participants in the low replicability condition indicated a significantly lower trust factor score ($M = -0.21$, $SD = 1.46$) than participants in the high replicability condition ($M = 0.30$, $SD = 1.39$), $t(196) = 2.54$, one-sided $p = .006$, $d = 0.36$, 95% CI [0.08, 0.64].

Participants in the low replicability but transparency condition did not indicate a significantly higher trust factor score ($M = -0.09$, $SD = 1.50$) than participants in the low replicability condition, $t(194) = 0.59$, one-sided $p = .279$, $d = 0.08$, 95% CI [-0.19, 0.37].

Robustness Check: Study 4. Participants in the QRPs condition showed a significantly lower trust factor score ($M = -0.33$, $SD = 1.43$) than participants in the hidden moderator condition ($M = 0.27$, $SD = 1.39$), $t(185) = 2.91$, one-sided $p = .002$, $d = 0.43$, 95% CI [0.13, 0.72].

However, the low replicability condition ($M = 0.03$, $SD = 1.52$), which served as a control condition and only reported on the low replication rate, did not differ significantly from the hidden moderator condition, $t(188) = 1.15$, $p = .254$, $d = 0.17$, 95% CI [-0.12, 0.45] or from the QRPs condition, $t(183) = 1.65$, $p = .100$, $d = 0.24$, 95% CI [-0.05, 0.53].

Robustness Check: Study 5. Participants in the “now high” replicability condition did not show significantly higher trust factor scores ($M = 0.13$, $SD = 1.69$) than participants in the low replicability condition ($M = -0.13$, $SD = 1.75$), $t(186) = 1.01$, one-sided $p = .157$, $d = 0.15$, 95% CI [-0.14, 0.44].

Does low replicability reduce public trust in psychology beyond repair –**Assessing evidence for the null hypothesis in Studies 3 to 5**

Studies 3 to 5 found no evidence that commonly used trust repair strategies significantly repair this damaged trust. So does low replicability damage public trust beyond repair? It is important to note that non-significant results do not allow to conclude that the null hypothesis is true (i.e., that the tested strategies have absolutely no effect). Thus, in addition to the sensitivity analyses reported in the manuscript, we also conducted equivalence tests (Lakens, 2017; Lakens, Scheel, & Isager, 2018) to test whether our observed effect sizes are statistically equivalent to an interval only containing very small effects ($|d| < .2$). Moreover, we computed Bayes factors (BF) to quantify the support the data provides for the null hypothesis vis-a-vis the alternative hypothesis (Jarosz & Wiley, 2014). Results for these analyses are presented in Table 2.

The results from the equivalence tests indicate that our observed effects are not statistically equivalent with the interval covered by the equivalence bounds [$d = -0.2, d = 0.2$]. Following the logic of equivalence tests, we can thus not declare the absence of meaningful effects. This is especially noteworthy since we used a rather high value for what constitutes a meaningful effect ($d = 0.2$). While $d = 0.2$ is conventionally considered a small effect (Cohen, 1988), it could potentially still be a meaningful effect of a trust repair intervention after a replication crisis: Given that such an intervention ideally would affect large parts of the population, even very small effects could be meaningful (Matz, Gladstone, & Stillwell, 2017). However, even using this potentially very liberal criterion of $d = 0.2$, equivalence tests were not significant.

In contrast, the default Bayes factor for an unpaired t -test (calculated using JASP version 0.8.4) shows that the data favors the null hypothesis over the alternative hypothesis in

all 3 studies. However, given that all Bayes factors are < 10 , this would conventionally not be considered as strong support for the null hypothesis (Jarosz & Wiley, 2014).

Table 2.

Observed Effect Sizes, Equivalence Test Results and Bayes Factors for the Tested Trust Repair Strategies in Study 3 to 5.

Study	Observed d ,	Equivalence tests	BF ^b (H0/H1)
Study 3 (Transparency)	$d = 0.11$	$t(194) = 0.671$, $p = .251$	3.27
Study 4 (QRPs)	$d = -0.25^a$	$t(183) = 0.320$, $p = .625$	1.68
Study 4 (Hidden Moderators)	$d = 0.03$	$t(188) = 1.152$, $p = .125$	6.22
Study 5 (Increased Replicability)	$d = 0.15$	$t(186) = 0.308$, $p = .379$	2.26

Note. The presented Bayes factors indicate the likelihood of the obtained data under the null hypothesis, divided by the likelihood of the data under the alternative hypothesis. Alternative hypotheses are one-sided in Studies 3 and 5 and two-sided in Study 4 (in line with our preregistered hypotheses)

^aInformation about QRPs have a negative effect on trust in psychology, indicated by the negative value.

Overall, neither equivalence tests nor Bayes Factors provide conclusive evidence for the null hypothesis. This is in line with our interpretation of the data presented in the manuscript.

Supplemental References

- Broomell, S. B., & Kane, P. B. (2017). Public perception and communication of scientific uncertainty. *Journal of Experimental Psychology: General*, 146, 286–304.
<https://doi.org/10.1037/xge0000260>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Routledge.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7, 2.
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362.
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1, 259–269.
- Matz, S. C., Gladstone, J. J., & Stillwell, D. (2017). In a world of big data, small effects can still matter: A reply to Boyce, Daly, Hounkpatin, and Wood (2017). *Psychological science*, 28, 547–550.
- Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The Partisan Brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The ANNALS of the American Academy of Political and Social Science*, 658, 36–66.
<https://doi.org/10.1177/0002716214555474>
- Rabinovich, A., & Morton, T. A. (2012). Unquestioned answers or unanswered questions: Beliefs about science guide responses to uncertainty in climate change risk communication. *Risk Analysis: An International Journal*, 32, 992–1002.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48, 1–36.

Rybowiak, V., Garst, H., Frese, M., & Batinic, B. (1999). Error orientation questionnaire (EOQ): Reliability, validity, and different language equivalence. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 20, 527–547.