

The replicability crisis and public trust in psychological science

Citation for published version (APA):

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, 3(3), 266-286. <https://doi.org/10.1080/23743603.2019.1684822>

DOI:

[10.1080/23743603.2019.1684822](https://doi.org/10.1080/23743603.2019.1684822)

Document status and date:

Published: 02/09/2018

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

The replicability crisis and public trust in psychological science

Farid Anvari & Daniël Lakens

To cite this article: Farid Anvari & Daniël Lakens (2018) The replicability crisis and public trust in psychological science, *Comprehensive Results in Social Psychology*, 3:3, 266-286, DOI: [10.1080/23743603.2019.1684822](https://doi.org/10.1080/23743603.2019.1684822)

To link to this article: <https://doi.org/10.1080/23743603.2019.1684822>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 19 Nov 2019.



[Submit your article to this journal](#)



Article views: 749



[View related articles](#)



[View Crossmark data](#)

The replicability crisis and public trust in psychological science

Farid Anvari ^{a,b} and Daniël Lakens^a

^aHuman-Technology and Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands; ^bStrategic Organization Design, Institute of Marketing and Management, University of Southern Denmark, Odense, Denmark

ABSTRACT

Replication failures of past findings in several scientific disciplines, including psychology, medicine, and experimental economics, have created a “crisis of confidence” among scientists. Psychological science has been at the forefront of tackling these issues, with discussions about replication failures and scientific self-criticisms of questionable research practices (QRPs) increasingly taking place in public forums. How this replicability crisis impacts the public’s trust is a question yet to be answered by research. Whereas some researchers believe that the public’s trust will be positively impacted or maintained, others believe trust will be diminished. Because it is our field of expertise, we focus on trust in psychological science. We performed a study testing how public trust in past and future psychological research would be impacted by being informed about (i) replication failures (replications group), (ii) replication failures and criticisms of QRPs (QRPs group), and (iii) replication failures, criticisms of QRPs, and proposed reforms (reforms group). Results from a mostly European sample ($N = 1129$) showed that, compared to a control group, people in the replications, QRPs, and reforms groups self-reported less trust in past research. Regarding trust in future research, the replications and QRPs groups did not significantly differ from the control group. Surprisingly, the reforms group had less trust in future research than the control group. Nevertheless, people in the replications, QRPs, and reforms groups did not significantly differ from the control group in how much they believed future research in psychological science should be supported by public funding. Potential explanations are discussed.

ARTICLE HISTORY


Received 20 July 2018
Accepted 27 September 2019

KEYWORDS

Replicability crisis; trust in science; reproducibility crisis; crisis of confidence; open science

There is a social contract between science and society, where in return for the public’s support, science is required to transparently produce reliable knowledge about how the world operates (Gibbons, 1999). No single person has the time to become an expert in all scientific fields, and therefore everyone must have some level of trust in scientific findings, scientists, or the scientific community as a whole (Hendriks, Kienhues, & Bromme, 2015; Pittinsky, 2015). For these reasons, if we want science to continue to play an important role in knowledge generation for society, to continue to be supported by and to receive public funding, we must be concerned with how science is perceived by the general public.

CONTACT Farid Anvari  faridanvari.phd@gmail.com  Human-Technology and Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands

 Supplemental data for this article can be accessed [here](#).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Science is generally considered in high esteem, and the public trusts science and places a high level of confidence in scientists (e.g., Jonge, 2015; Lamberts, 2017; Lindholm, Bergman, & Gustav, 2018; National Science Board, 2016; Scheufele, 2013; German Science Barometer, 2017). However, a crisis of confidence has taken place in some scientific disciplines, caused by concerns about the replicability of past findings (e.g., Baker, 2016; Pashler & Wagenmakers, 2012). This crisis could potentially impact the public's trust in science (e.g., Rutjens, Heine, Sutton, & van Harreveld, 2017). The ongoing scientific discussion on how to improve the way we do science has resulted in a surge of criticisms against questionable research practices that produce irreproducible and misleading results. Criticism on research practices has always been part of the scientific process (e.g., Hull, 1988). What is new is that these discussions and criticisms are now increasingly taking place in blogs, on Twitter, and in Facebook discussion groups, which makes these criticisms more accessible to journalists and the general public (Brumfiel, 2009). People who lack training in the scientific process (e.g., Hallman, 2017) may misperceive legitimate disagreements as something out of the ordinary (see also Pittinsky, 2015). So far, there has been no systematic study to test the impact of different aspects of the replicability crisis (i.e., failures to replicate, criticisms of questionable research practices, and proposed reforms) on the public's trust in science.

Some researchers are concerned that the public will lose trust in scientific fields that have been criticized, such as psychological science. In the extreme case, there is the concern that public criticisms of research practices and issues regarding replicability can fuel anti-science movements (e.g., Ioannidis, 2017; Pickett & Roche, 2018). For example, legitimate calls for increased scientific transparency have been used as the basis for a political attempt to severely restrict research that the US Environmental Protection Agency can use in writing regulations (e.g., Gebellhof, 2018). Others have suggested that public criticisms and attempts to reform research practices may be perceived as science being self-correcting, thereby demonstrating scientists' commitment to improving the way they do science. This could positively impact public trust, or at least help maintain trust in science in the long run (e.g., Rutjens et al., 2017; Srivastava, 2017; Vazire, 2016). Because of the importance of trust in science among the general public (who in democratic societies vote for parties that subsequently determine science policy) and the disagreement about whether the public discussions around reproducibility will reduce or increase public trust in psychological science, we aim to experimentally examine the impact of being informed about psychology's replicability crisis, and attempts to self-correct the way psychological science is practiced.

The purpose of the present study is to provide empirical data to inform discussions that have thus far been largely anecdotal or hypothetical. Although it seems plausible that learning about replication failures will reduce trust in psychological science (especially in past findings), it is unclear whether additionally hearing about publicly expressed criticism about questionable research practices will further reduce trust or be seen as a sign of a healthy field of science and thus increase trust. Finally, a novel question in the proposed study is whether learning about current reforms in psychological science will increase trust in psychological science enough to compensate for the lower trust after hearing about replication failures. This study will allow us to examine whether the current developments in psychological science taken together, replication failures, criticisms,

and proposed reforms, can be expected to decrease, increase, or keep stable, the trust in psychological science.

The replicability crisis and questionable research practices

Several fields of scientific inquiry have encountered what has been widely referred to as a replicability (or reproducibility) crisis. In psychology, the Open Science Collaboration (2015) attempted to replicate 100 studies from three major psychology journals. Of those 100, 64 did not produce statistically significant results – that is, they failed to support the hypothesis of the original study. In cancer biology, of 53 landmark studies investigated, 89% (47) could not be confirmed (Begley & Ellis, 2012; see also Davis et al., 2017). In experimental economics, an attempt to replicate 18 findings from two major journals was unable to confirm the original results in 7 (39%) cases (Camerer et al., 2016). And, in experimental philosophy, approximately 30% of 40 original findings failed to replicate (Cova et al., 2018). Other large-scale attempts to replicate important psychological findings have also yielded null-effects (e.g., Hagger et al., 2016; O'Donnell et al., 2018; Wagenmakers et al., 2016). The replicability crisis has gained considerable news coverage (e.g., Connor, 2015; Engber, 2017; Feilden, 2017; Yong, 2015) and a sizable Wikipedia entry entitled, “Replication crisis”.

Questionable research practices (QRPs), whether intentional or not, have been discussed as one possible reason underlying the current replicability crisis by leading to a misrepresentation of the data (Banks et al., 2016; John, Loewenstein, & Prelec, 2012; Pickett & Roche, 2018; Simmons, Nelson, & Simonsohn, 2011; Wicherts, 2011). QRPs consist of data-analytic choices driven by their utility in producing more favourable statistical results, which are not transparently reported in methods sections (such as optional stopping during data collection, or selectively excluding outliers). They are problematic because they inflate Type 1 error rates (incorrectly rejecting the null hypothesis, Simmons et al., 2011) and thus the percentage of false positives in the literature. Many scientists admit to having engaged in QRPs (for psychology see, Fiedler & Schwarz, 2016; John et al., 2012; for management see Banks et al., 2016; for ecology and evolution see; Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018; for science more generally see, Martinson, Anderson, & De Vries, 2005). Many psychological scientists have started educating researchers about QRPs, as well as pointing out QRPs in the literature, perceiving them as an important underlying cause of the replicability crisis (in addition to other problems, such as publication bias) (e.g., Cumming, 2016; Neuroskeptic, 2015).

Effects of the replicability crisis on trust in psychological science

How might the published failures to replicate, and the public discussion of QRPs, impact the public's trust of psychological science? One perspective suggests that transparency about problems and improvements in scientific practice may maintain the public's trust and confidence in the long term. A compelling argument for this view is that, given QRPs are perceived to be morally unacceptable and deserving of punishment (Pickett & Roche, 2018), scientific criticisms against them may signal to the public that such behaviours are taken seriously by the psychological science community, and that psychological scientists are attempting to address these problems. As Pittinsky (2015) argues, scientists can

inspire the public's faith in science by acknowledging the seriousness of scientific misconduct, and being transparent about discussions among scientists for how to improve research practices.

Vazire (2017) posits that transparency increases the accountability of researchers, incentivising more reliable scientific practices that produce better quality works, and thus reducing public uncertainty about the quality of research findings which, in turn, increases trust in science. Although her arguments specifically referred to transparency about data and methods, the argument can be extended to publicly expressed criticisms against QRPs, for this too would incentivise more reliable practices. Hence, although awareness of replication failures may negatively affect public perceptions about the quality of past works (i.e., the published literature), perceptions about quality in the future of the scientific field, and thus trust in its future findings, may be increased (or maintained) when there is also awareness about scientific criticisms of QRPs and initiatives aimed at reform.

Hendriks, Kienhues, and Bromme (2016) present findings that support this view. Participants in their study read what was described as a blog entry by an expert science blogger, followed by a commentary said to be written either by the blogger or by another person criticising the original blog post. They found that self-criticism of the blog post, compared to the external criticism, resulted in the expert science blogger being rated higher on integrity and benevolence (Hendriks et al., 2016). Integrity and benevolence have been shown to represent two dimensions of epistemic trust (Hendriks et al., 2015, 2016; Mayer & Davis, 1999). In another scenario-based study conducted by Fetterman and Sassenberg (2015), scientists tended to overestimate how negatively they would be perceived by other scientists following failed replications of their work. In addition, they found that the scientific reputation of a researcher who admitted to being "wrong about the effect" was higher than the researcher who questioned the replication. At the group level, one could argue that because psychological scientists themselves are testing the reliability of past research findings and criticising QRPs, their publicly expressed criticisms can have beneficial effects on the scientific reputation and the perceived benevolence and integrity of, and thus trust in, the collective (i.e., the psychological science community).

In contrast to the preceding view that transparency instils trust in the scientific process, others are concerned that the replicability crisis and public discussion about QRPs will negatively influence public perceptions of psychological science (e.g., Fanelli, 2018). This concern is expressed clearly by Klaus Fiedler in an interview, "... I believe that the way [the debate] unfolded over the last decade was counterproductive. It damaged psychology's public image and undermined the self-confidence of our young scientists and students" (Genschow & Crusias, 2018). The enhanced status of scientific knowledge is derived from its procedural norms, including replication and peer review (Gluckman, 2014). When large replication attempts of research in psychological science fail, and publicly expressed criticisms suggest that these failures are a result of QRPs, public perceptions of psychological science may be negatively affected, diminishing the field's status, and leading to less trust in psychological science.

Although it is possible that concerns about the negative impact of failed replications and criticisms of QRPs on public trust may be somewhat overestimated (e.g., Fetterman & Sassenberg, 2015), there is some research that supports these concerns. Pickett and Roche (2018) randomly allocated participants to be presented with either a definition of data

falsification (i.e., fraud; $n = 415$) or selective reporting (i.e., QRPs; $n = 406$). Participants then judged how morally unacceptable they believed these behaviours to be, and the action that should be taken against scientists engaging in them. Although a greater percentage of participants presented with data fraud said that it was morally unacceptable (96%), a majority of those presented with QRPs also judged these as morally unacceptable (71%). Of the participants responding to QRPs, 73% thought that scientists who engaged in them should be banned from receiving funding, 63% thought they should be fired, and 37% thought QRPs should be a crime. If members of the public consider QRPs to be such a severe moral transgression deserving of punishment, then it is likely that publicly expressed criticisms that connect the replicability crisis with scientists using QRPs will have a negative impact on perceptions of the scientific field that is being criticised. Relatedly, in a representative sample of over 1000 Australians, Critchley (2008) found that the perceived benevolence of scientists (including ethical research methods and honesty about results) was positively related to trust in scientists ($r = .41$). Together, these findings would suggest that learning about QRPs, and criticisms on their use by psychological scientists, could reduce the public's trust in the field.

The most direct support for this position comes from a recent study conducted by Chopik, Bremner, Defever, and Keller (2018). They used a 1-hour lecture to educate undergraduate students about the replicability crisis in psychology, the potential causes for it (including QRPs), and good research practices such as sharing data and materials and designing studies with high power. In the pre-post within-subjects design, after having taken part in the lecture, students had less trust in "the results of studies done by psychologists".

It seems sensible that learning about the replicability crisis in psychology reduces trust in studies that have been done (i.e., trust in past research). But Chopik et al.'s (2018) study does not inform us about trust in future research done by psychologists nor does it inform us about how the various additive aspects of the replicability crisis (i.e., replication failures, criticisms of QRPs, and proposed reforms) work together to impact trust. Will public criticism of QRPs have a positive or negative effect on trust in future research in psychological science? And to what extent can learning about proposed reforms in psychological science improve trust, even after learning about replication failures?

Overview of present study

In the present study, we experimentally manipulated whether participants were presented with information about (1) replication failures, or (2) replication failures and criticisms of QRPs, or (3) replication failures, criticisms of QRPs, and suggestions for reform, or (4) a control condition with general information about psychology, but no information about any of the issues discussed in the other experimental conditions. We investigated the effects of being exposed to information about replication failures, criticisms of QRPs, and suggestions for reforms on three dependent variables: (1) trust in past research within psychological science, (2) trust in future research within psychological science, and (3) support for future research in psychological science. We compared the four experimental conditions on the three outcome variables. We predicted that knowledge about replication failures would reduce trust, especially in past research, but it might also reduce trust in future research. Learning about criticisms of QRP's might further reduce trust in both past and future research in psychological science. However, an alternative perspective would suggest that self-criticism would reduce

trust in past research but improve trust in *future* research. Finally, we expected knowledge about reforms to improve trust in future research in psychological science, but an important question was whether the increase in trust when hearing about reforms would be enough to counteract the predicted negative effect of hearing about replication failures. Would knowledge about replication failures outweigh knowledge about reforms, or might learning about reforms in psychology maintain or even lead to an improvement in trust in psychological science, despite the failures to replicate past studies?

We were interested in three comparisons:

- (1) First, compared to the control group, we expected that people informed about replication failures would have less trust in past and future research in psychological science.
- (2) Our second interest was in whether and how learning about criticisms of QRP's in addition to learning about failures to replicate would impact trust in past and future science. It seemed plausible that if learning about QRP's had an effect on past research, it would be negative. However, with respect to future research, we did not have a directional prediction. Trust in future research could decrease (if people believe QRP's would also affect future studies) but trust in future research could also increase (if people see criticisms on QRP's as healthy self-criticism of a field trying to improve).
- (3) Finally, we were interested in whether, compared to a control condition, learning about all information about the reproducibility crisis (replication failures, criticisms on QRP's, *and* proposed reforms) would overall reduce, maintain, or even improve people's trust in future research in psychological science. We had no directional prediction, but we believed that this comparison addressed an important question about the extent to which the discussion as it has thus far unfolded would impact peoples' trust in the future of psychological science.

Non-preregistered pilot study

In a pilot study, performed in response to the first round of reviews, we aimed to examine three main questions. First, we tested the comprehension checks to estimate what percentage of participants would remember key points in the presented information. As one reviewer noted, including comprehension checks might increase a demand effect, so if the percentage of participants who pass the comprehension checks is high enough, we can ignore them in the main study to reduce demand effects. Second, we more closely examined a composite measure of trust in the psychological science community (which had too poor reliability to be used in the main study), and, third, we collected initial data (as the pilot study was largely similar in terms of independent and dependent variables as the proposed study). These materials, data, and analysis script are available on the OSF (<https://osf.io/sftz2/>).

Participants

We recruited 201 participants using Prolific Academic (98 females and 103 males; mean age 31.9 years ranging from 18 to 67; 81.6% residing in Europe, 9.9% in North America, 6% in

Central/South America, 2.5% other). We pre-screened participants to be fluent in English and used the pre-screening criteria to gather demographic variables including subjective socioeconomic status (SES) using the MacArthur Ladder (Adler, Epel, Castellazzo, & Ickovics, 2000), the highest level of education level (no formal qualifications, secondary school/GCSE, college/A levels, undergraduate degree BA/BSc/other, graduate degree MA/MSc/MPhil/other, doctorate degree PhD/MD/other), personal income (<£10,000, £10,000–£19,999, £20,000–£29,999, £30,000–£39,999, £40,000–£49,999, £50,000–£59,999, £60,000–£69,999, £70,000–£79,999, £80,000–£89,999, £90,000–£99,999), student status (student vs. not student), sex, and country of residence.

Procedure

Participants were randomly allocated to the control, replication failures, criticisms of QRPs, or the reforms condition (see Materials and Measures section of the preregistered study). After reading the information for their respective condition, participants responded to two multiple-choice comprehension checks (see Supplementary Materials) to ensure they had read and understood the critical pieces of information they were provided with.

Next, participants were asked how much they trust past research in psychological science, how much they trust future research in psychological science, and how much they agree that public funding should be used to support future research in psychological science (from 1 = *not at all*, to 10 = *completely*; see Materials and Measures section below). In this pilot study we also included a scale to measure trust in the psychological science community from a previous pilot test, minus one item due to poor fit statistics, but the scale showed poor psychometric properties. We do not report the scale here (but the data are available on OSF, <https://osf.io/sftz2/>, where data are also available from another pilot study designed to develop a measure for trust in the science community). We decided not to use the scale measuring trust in the psychological science community in the final study. All trust measures were presented in random order on a separate page. After responding to the trust measures, participants were asked, “Before taking this survey, how informed were you about a replicability or reproducibility crisis in psychology (from 1 = *never heard about it*, to 10 = *very well-informed*)?”. Finally, before being debriefed and paid, participants provided their age and a single-item measure of political self-identification, “Please indicate how you politically self-identify, from very liberal/left-wing to very conservative/right-wing”, on a 7-point scale (the farthest left point was labelled “very liberal/left-wing”, the farthest right point was labelled “very conservative/right-wing”, and the midpoint was labelled “Moderate/Centre”). This item was included so as to test the construct validity of the trust in the psychological science community scale (i.e., the further right/conservative people political self-identify the less trust in the science community they would be expected to have).

Results

Out of the 201 participants in the pilot study, 184 (91.5%) passed both comprehension checks. We deem this an acceptable pass-rate and, to avoid potential demand effects

introduced by the comprehension checks, we decided not to use comprehension checks in the final study.

Results showed that how informed people were about the replicability crisis was not strongly correlated with subjective SES ($r = .12, p = .104, CI_{95\%} [-.02; .25]$), level of education ($r = .10, p = .154, CI_{95\%} [-.04; .24]$), personal income ($r = .04, p = .648, CI_{95\%} [-.12; .18]$), or political self-identification ($r = -.01, p = .876, CI_{95\%} [-.15; .13]$). Exploratory analyses suggested that being informed about the replicability crisis was significantly negatively correlated with age ($r = -.21, p = .003, CI_{95\%} [-.07; -.34]$) such that younger people tended to be more informed. Moreover, participants who were current students were significantly more informed ($M = 4.14, SD = 2.60$) than participants who were not currently studying ($M = 3.04, SD = 2.27$), $t_{(113.73)} = 2.91, p = .004$, Cohen's $d = 0.46, CI_{95\%} [0.16; 0.76]$. Although being informed about the replicability crisis was not significantly correlated with trust in past research ($r = .09, p = .181, CI_{95\%} [-.04; .23]$), trust in future research ($r = .13, p = .058, CI_{95\%} [-.004; .27]$), or support for future research in psychological science ($r = .11, p = .126, CI_{95\%} [-.03; .24]$), we could not rule out small-medium sized correlations in the positive direction due to the small sample size.

We further explored differences between naïve vs. more informed participants by dichotomising the informed variable so that we could compare those who had never heard about the replicability crisis (i.e., selected 1 on the rating scale, labelled "Never heard about it") with participants who had (i.e., selected 2–10 on the rating scale). From the 201 participants in the study, 60 (29.9%) had never heard about the replicability crisis. There were non-significant differences (that should be interpreted with special caution given the small sample size in the group who answered '1') between those who had never heard about the replicability crisis and those who had, in subjective SES ($p = .672$), level of education ($p = .126$), income level ($p = .762$), political self-identification ($p = .380$), age ($p = .118$), trust in past research ($p = .789$), trust in future research ($p = .641$), and support for future research ($p = .678$). [Table 1](#) presents the descriptive statistics. Although we find no reason to assume strong effects, due to the small sample size we cannot rule out small or medium effects.

This pilot study allowed us to collect data very similar to that planned in the preregistered (albeit with smaller sample sizes). We examined the effect of being informed about the various elements of the replicability crisis on trust and support for research in psychological science. Our focus was on three contrasts we were particularly interested in (see Analysis Plan). The descriptive statistics (M s and SD s) are presented in [Table 2](#). Almost none of the comparisons were statistically significant with 50 participants in each between-subject condition, but the

Table 1. Descriptive statistics, means and (standard deviations), on the outcome variables as a function of naivety.

Outcome	Naivety M (SD)		Effect Size
	Naïve	Not-Naïve	
SES	5.43 (1.77)	5.55 (1.61)	$d = 0.07, CI_{95\%} [-0.24; 0.37]$
Education	2.18 (1.19)	2.46 (1.17)	$d = 0.24, CI_{95\%} [-0.07; 0.54]$
Income	2.29 (1.70)	2.20 (1.65)	$d = 0.05, CI_{95\%} [-0.28; 0.38]$
Political ID	3.58 (1.37)	3.40 (1.37)	$d = 0.14, CI_{95\%} [-0.17; 0.44]$
Trust (past)	5.82 (1.75)	5.89 (1.53)	$d = 0.04, CI_{95\%} [-0.26; 0.35]$
Trust (future)	7.13 (1.59)	7.02 (1.48)	$d = 0.07, CI_{95\%} [-0.23; 0.38]$
Support (future)	6.95 (2.20)	7.09 (1.88)	$d = 0.07, CI_{95\%} [-0.24; 0.37]$

Table 2. Descriptive statistics, means and (standard deviations), on the outcome variables as a function of experimental condition (all participants).

Outcome	Condition			
	Control (<i>n</i> = 50)	Replication failures (<i>n</i> = 50)	Criticisms (<i>n</i> = 51)	Reforms (<i>n</i> = 50)
Trust past	6.28 (1.26)	5.78 (1.48)	5.53 (1.98)	5.88 (1.52)
Trust future	7.06 (1.39)	6.84 (1.35)	7.08 (1.68)	7.24 (1.60)
Support future	7.28 (1.86)	6.66 (1.90)	6.92 (2.26)	7.32 (1.81)

pattern of effect sizes is in line with expectations. Participants who read information about replication failures, compared to the control group, had lower trust in past (Cohen's $d = 0.36$, $CI_{95\%} [-0.04, 0.76]$) and future research (Cohen's $d = 0.16$, $CI_{95\%} [-0.24, 0.56]$) and supported future research less (Cohen's $d = 0.33$, $CI_{95\%} [-0.07, 0.73]$), $ps = .072, .424$, and $.103$, respectively. Compared to the control group, participants who read information about criticisms of QRPs had significantly lower trust in past research (Cohen's $d = 0.45$, $CI_{95\%} [0.05, 0.85]$, $p = .026$). There was basically no difference in trust in future research (Cohen's $d = -0.01$, $CI_{95\%} [-0.41, 0.38]$, $p = .952$) or support for future research (Cohen's $d = 0.17$, $CI_{95\%} [-0.22, 0.57]$, $p = .386$). Although participants who read information about reforms in psychological science, compared to the control group, had lower trust in past research (Cohen's $d = 0.29$, $CI_{95\%} [-0.11, 0.69]$), their trust in future research was slightly higher (Cohen's $d = -0.12$, $CI_{95\%} [-0.52, 0.28]$), with equal support for future research (Cohen's $d = -0.02$, $CI_{95\%} [-0.42, 0.38]$), $ps = .156, .549$, and $.914$, respectively.

We note that these preliminary tests are underpowered for effect sizes we still consider interesting (see Participants section below for power calculations and target sample size for the final study). Nevertheless, the descriptive statistics (and effect size confidence intervals) seem to suggest that although being informed about the various aspects of the replicability crisis might reduce trust in past research, trust in future research may be maintained (e.g., when there are criticisms of research practices or when participants are informed about proposed reforms). We performed the main study to examine whether this predicted pattern of means would emerge when collecting data from a larger sample.

In response to a concern raised by a reviewer that selecting participants who have not heard of the replication crisis might lead to selection effects, because these participants differ from participants who have heard of the replication crisis, we tested whether the manipulations had differential effects on trust for those who had heard about the replication crisis in psychology as compared to those who had not. To do this, we dichotomised participants into those who had never heard about the replicability crisis in psychology (i.e., those who selected option 1 on the item asking participants how informed they are about the replicability crisis) and those who had at least heard about it (i.e., those who selected options 2–10). Entering this dichotomised naivety variable as a factor together with experimental manipulation in an ANOVA, showed statistically non-significant interaction effects on trust in past research ($p = .167$), trust in future research ($p = .622$), and support for future research ($p = .717$). Moreover, the main effects of naivety were also not statistically significant, providing no evidence for differences between those who had never heard about the replicability crisis compared to those who had ($ps = .776, .632$, and $.659$ for trust in past research, trust in future research, and support for future research, respectively). Given the small sample size, these tests are underpowered.

Table 3. Descriptive statistics, means and (standard deviations), on the outcome variables as a function of experimental condition (naïve participants only).

Outcome	Condition			
	Control (<i>n</i> = 13)	Replication failures (<i>n</i> = 13)	Criticisms (<i>n</i> = 18)	Reforms (<i>n</i> = 16)
Trust past	6.77 (1.42)	5.85 (1.72)	5.17 (1.79)	5.75 (1.77)
Trust future	7.31 (1.49)	6.92 (1.50)	7.06 (1.73)	7.25 (1.69)
Support future	7.31 (1.84)	6.92 (2.56)	6.33 (2.22)	7.38 (2.16)

Nevertheless, it can be seen that the pattern of descriptive statistics for participants who were naïve about the replicability crisis, presented in Table 3, is very similar to the pattern of results for the entire sample examined as a whole (i.e., Table 2).

Preregistered main study

Participants

We were interested in small effects because, whereas in the real world people are likely to be repeatedly presented with stories about the replicability crisis, our manipulation involved a single reading of short statements. A single exposure with a small effect may translate to larger effects with repeated exposure over a longer period of time. We had difficulty in setting a smallest effect size of interest. In our pilot study, the average effect size estimate for decreased trust in past research, compared to the control group, due to being informed about each aspect of the replicability crisis was Cohen's $d = 0.37$ (0.36 for replication failures, 0.45 for criticisms of research practices, and 0.29 for reforms). We therefore decided to set the smallest effect size of interest to what we believed would be an effect small enough for an initial test of the concerns about the impact of the replicability crisis: Cohen's $d = 0.3$. Our main contrasts of interest (described in more detail in the Analysis Plan further below) were tested using one- and two-tailed t -tests. We decided on a Type 1 error rate based on Good's (1988) proposal to adjust the alpha level as a function of the sample size (adjusting the p -value to $p\sqrt{N/100}$, or adjusting the alpha level to $0.05/\sqrt{N/100}$), which intends to prevent Lindley's paradox when collecting larger samples. Using the TOSTER package in R (Lakens, 2017) we calculated that for alpha = .0307, and 90% power to exclude an effect size outside the equivalence range of $d = -0.3$ and $d = 0.3$, we require a sample size of 275 per condition (Lakens, Scheel, & Isager, 2018). We therefore aimed to recruit a total of 1100 participants (275 participants across 4 conditions) using Prolific Academic, without allowing those who took part in the pilot study to participate in the main study. We pre-screened participants to be fluent in English. Participant recruitment continued until the number of participants who had completed the survey reached the sample size requirement ($N = 1100$).

Our sampling process resulted in 1129 complete responses on our preregistered dependent variables (the three main outcome measures) because some completed all dependent variables, but not some of the other questions. These participants (528 female, 568 male, 4 other/non-binary, 29 missing values) had mean age of 30.56 years (range: 18–75 years) and mostly resided in Europe (91.6%). For the exploratory analyses involving variables included later in the survey, sample size varies and includes only those with complete responses on the variables included in the analysis.

Procedure

We randomly allocated (using the Survey Monkey randomiser) participants to one of four groups. As in the pilot study, in one (control) group, participants read three paragraphs of information about psychological research (see the Materials and Measures section below). In a second group, participants read the same first two paragraphs as the control group, but the last paragraph was replaced with some information about failed replications in psychological science. For the third group of participants the second and third paragraphs of the control group were respectively replaced with information about replication failures and information about researchers criticising the use of QRPs as a cause of the failed replications. A fourth group of participants read the two paragraphs with information about replication failures and criticisms of QRPs, plus another paragraph in which the criticisms are followed by suggestions and implementation of reforms to improve the reliability of research. Participants were asked to indicate their trust in past research, trust in future research, and support for future research in psychological science. Participants were then asked to explain why they had those levels of trust in past and future research in an open answer format (answers to this question are not reported in this paper), and extending the pilot, followed by two measures of ambivalence (mixed feelings) about the events described in the vignettes (a subjective measure and an objective measure). Next, participants were asked, "Before taking this survey, how informed were you about a replicability or reproducibility crisis in psychology?" (from 1 = *Never heard about it*, to 10 = *Very well-informed*). Finally, participants provided demographic details (age, gender; and the same demographic details as those collected in the pilot study were obtained from Prolific's pre-screening criteria), were debriefed, thanked for their time, and paid.

Materials and measures

The information provided to each group of participants was designed to be comparable in length (160 words) so as to control for how much information each group received regarding psychological science.

Information for the control group

Psychology was originally studied as a philosophical pursuit with insights coming from introspection and observation. During the Enlightenment in Europe, psychology became more popular and shortly afterwards it was studied using experiments. Now, many universities around the world study psychology as a scientific pursuit.

Psychological science includes several different subfields. Cognitive psychologists study mental processes such as perception, attention, reasoning, memory, and learning. Social psychologists study how humans think about and relate to each other, including the influence of others on one's behaviour, beliefs, attitudes, and stereotypes. Personality psychologists study enduring patterns of emotion, thought, and behaviour, in individuals.

Research in psychological science includes many different methodologies, such as controlled experiments in laboratory settings, field experiments, and large surveys, that require the use of statistical analyses. For example, psychological scientists use statistics to draw conclusions about the effects of new behavioural interventions or policy changes. Much of the conclusions drawn from research in psychological science often rely heavily on statistics.

Information for the replication failures group

Psychology was originally studied as a philosophical pursuit with insights coming from introspection and observation. During the Enlightenment in Europe, psychology became more popular and shortly afterwards it was studied using experiments. Now, many universities around the world study psychology as a scientific pursuit.

Psychological science includes several different subfields. Cognitive psychologists study mental processes such as perception, attention, reasoning, memory, and learning. Social psychologists study how humans think about and relate to each other, including the influence of others on one's behaviour, beliefs, attitudes, and stereotypes. Personality psychologists study enduring patterns of emotion, thought, and behaviour, in individuals.

Recently, some psychological scientists tested the reliability of research in their field by repeating the procedures of past studies to see if they could produce the same results. In 2015, one large collaborative project found that from the 100 original studies examined, 65 didn't produce the same results. In other words, researchers were able to successfully replicate 35% of the studies.

Information for the criticisms of QRPs group

Psychology was originally studied as a philosophical pursuit with insights coming from introspection and observation. During the Enlightenment in Europe, psychology became more popular and shortly afterwards it was studied using experiments. Now, many universities around the world study psychology as a scientific pursuit.

Recently, some psychological scientists tested the reliability of research in their field by repeating the procedures of past studies to see if they could produce the same results. In 2015, one large collaborative project found that from the 100 original studies examined, 65 didn't produce the same results. In other words, researchers were able to successfully replicate 35% of the studies.

Some psychological scientists have criticised research practices, stating that a major reason for failures to replicate past studies is that the original findings were based on poor and un-transparent research practices. These include things like selective reporting, where many analyses are performed but only favourable ones reported. They argue that such practices produce false findings.

Information for the suggestions for reforms group

Recently, some psychological scientists tested the reliability of research in their field by repeating the procedures of past studies to see if they could produce the same results. In 2015, one large collaborative project found that from the 100 original studies examined, 65 didn't produce the same results. In other words, researchers were able to successfully replicate 35% of the studies.

Some psychological scientists have criticised research practices, stating that a major reason for failures to replicate past studies is that the original findings were based on poor and un-transparent research practices. These include things like selective reporting, where many analyses are performed but only favourable ones reported. They argue that such practices produce false findings.

Because of this, many psychological scientists have started initiatives to increase the reliability of research. They are pushing for greater transparency in research practices and increased use of registered reports where statistical analyses are specified before data collection and articles published regardless of results.

In the real world, news stories may frame the replicability crisis in a negative way, with headlines such as "Study reveals that a lot of psychology research really is just 'psycho-babble'" (Connor, 2015), or provide content that is more nuanced and somewhat positive,

“Housecleaning is a crucial corrective in science, and psychology has led by example.” (Carey, 2018). The effects of such framing in media outlets is beyond the scope of this study, the purpose of which was to investigate how merely being informed about the various aspects of the replicability crisis impacts the public’s trust.

Outcome measures. Participants responded to the following questions by giving ratings on 10-point Likert-type scales. The scales were labelled: 1 = *not at all*, to 10 = *completely*.

We measured trust in past research by asking: “How much do you trust past research in psychological science?”. We measured trust in future research by asking: “How much do you trust future research in psychological science?”. We measured support for future research by asking: “How much do you agree that public funding should be used to support future research in psychological science?”. The above three items were presented in randomised order each on a different page.

Additional measures. To examine potential reasons for how people’s trust might be impacted by the replicability crisis, we included some additional measures. Participants were asked an open question: “explain why you have these levels of trust in past and future research in psychological science”. Following this open question, we measured both subjective and objective ambivalence (Schneider & Schwarz, 2017; Schneider, Veenstra, van Harreveld, Schwarz, & Koole, 2016). First, we measured subjective ambivalence by asking participants to “Think about the events you have just read about in psychological science. To what extent do you have mixed thoughts or feelings about the events?”. Second, we measured what is called objective ambivalence using two items, one each for the positive feelings and negative feelings:

Think about the events you just read about in psychological science. When you think about the positive [negative] aspects of the events described, while ignoring the negative [positive] aspects, how positive [negative] is your evaluation of the events in psychological science?.

Both subjective and objective ambivalence items were measured with Likert-type scales (from 1 = *not at all*, to 9 = *very much*). To calculate objective ambivalence, we use the formula $((P + N)/2) - |P - N|$, where P is the positive rating and N is the negative rating. This measure can range from -3 to 9, with scores below 1 reflecting univalence, and from 1 to 9 reflecting increasing ambivalence (a score of 1 reflects neutrality; see Schneider et al., 2016 for a thorough explanation).

Results and discussion

Preregistered analysis plan

We were particularly interested in planned contrasts that tested our main hypotheses, for which we planned to use Welch’s *t*-tests against 0 (which does not assume homogeneity of variance), and equivalence tests with an equivalence range from $d = -0.3$ to $d = 0.3$. Welch’s *t*-test is quite robust to violations of normality as long as the sample size is sufficiently large (Delacre, Lakens, & Leys, 2017). We planned that if a visual inspection suggested strongly asymmetric distributions, we would also perform a Mann–Whitney *U* test and rely on the

more conservative of the two p -values. A visual inspection of the distribution of the three main outcome variables did not suggest strong asymmetric distributions.

We believe that trust in past research and trust in future research are related to different questions. Being informed about each of the different elements of the replicability crisis may affect trust in past research similarly (particularly given that all elements involve failures to replicate past studies). On the other hand, how they affect trust in future research may differ from each other, given that each element may signal different things about what future research may involve (e.g., reforms to research practices). Therefore, in correcting for multiple comparisons, we considered all tests that involve trust in past research as the dependent variable to be a family of tests, and all tests that involve trust in future research to be another family of tests. We used the Holm–Bonferroni correction to adjust our alpha levels. There are two comparisons for trust in past research (i.e., control vs. replication failures, and control vs. criticisms of QRPs): The lowest of the two p -values from these comparisons was tested against an alpha of $.0307/2$ (.0154), and the other against an alpha of $.0307$. All three comparisons examine trust in future research: The lowest of the three p -values was tested against an alpha of $0.0307/3$ (.0102), the second lowest against an alpha of $0.0307/2$ (.0154), and the last test at the 0.0307 level.

We included a question asking participants about their support for future research in psychological science specifically. We expected the same pattern of results as for trust in future research (i.e., compared to control, learning about (i) replication failures will reduce support, (ii) criticisms of QRPs may either reduce or increase support, and (iii) all aspects of the replication crisis may either reduce or increase support). Although this variable may be taken as belonging to the same family of tests as trust for future research, we examined support for future research as something more exploratory and treated it separately. Hence, we corrected the alpha for multiple comparisons, treating support for future research as a separate family of tests. Support for future research was thus involved in all three comparisons and had the same corrected alpha levels as trust in future research.

Preregistered analyses

Descriptive statistics (M , SD) of the three preregistered outcome measures for each of the four groups are presented in Table 4 and the data visualized in Figure 1.

Replication failures and trust

First, we tested whether, compared to the control group, the group informed only about replication failures had less trust in past and future research in psychological science,

Table 4. Descriptive statistics, means and (standard deviations), on the outcome variables as a function of experimental condition (all participants).

Outcome	Condition			
	Control ($n = 257$)	Replication failures ($n = 277$)	Criticisms ($n = 310$)	Reforms ($n = 285$)
Trust past	6.12 (1.66)	5.63 (1.76)	5.59 (1.95)	5.63 (1.74)
Trust future	7.34 (1.44)	7.20 (1.62)	7.07 (1.60)	6.97 (1.53)
Support future	7.39 (1.86)	7.32 (1.80)	7.35 (1.97)	7.38 (1.89)

SurveyMonkey's randomisation process did not allow full control of the number of participants allocated to each condition; we could only indicate that we wanted participants allocated evenly. Therefore, the number of participants in each condition varies.

using one-sided Welch's t -tests. Being informed about replication failures significantly reduced trust in past research, $t_{(531.91)} = 3.31, p < .001$ ($d = 0.29, CI_{95\%} [0.11; 0.46]$); but not trust in future research, $t_{(531.01)} = 1.06, p = .146$ ($d = 0.09, CI_{95\%} [-0.08; 0.26]$).¹ The equivalence test (with lower bound = $-\infty$, upper bound = 0.3) was not significant for trust in past research, $t_{(531.91)} = 0.16, p = .436$; but it was significant for trust in future research, $t_{(531.01)} = 2.42, p = .008$. Therefore, we do not find support for the idea that trust in future research is reduced by learning about replication failures and we can conclude that the effect is not larger than 0.3. Hence, in line with previous research (e.g., Chopik et al., 2018; Wingen, Berkessel, & Englich, 2019), and as expected, we found that learning about replication failures reduces trust in past research. However, extending on the previous works, we found no evidence that trust in *future* research is affected. This suggests that attempting to replicate past findings may act as a signal that researchers want to get things right and that psychological science is self-correcting.

QRPs and trust

Second, we tested how trust in past and future research would be affected by learning about criticisms of QRPs in addition to learning about replication failures. Expecting that learning about criticisms of QRPs would decrease trust in past research, we used a one-sided Welch's t -test: The group who learned about replication failures *and* criticisms of QRPs had significantly less trust in past research than the control group, $t_{(564.63)} = 3.48, p < .001$ ($d = 0.29, CI_{95\%} [0.12; 0.46]$), and the equivalence test (with lower bound = $-\infty$, upper bound = 0.3) was not significant, $t_{(564.63)} = 0.10, p = .461$. Given that for trust in future research we did not have a directional prediction, we used a two-sided Welch's t -test: Trust in future research was not significantly different between the control group and the group who learned about replication failures *and* criticisms of QRPs, $t_{(560.91)} = 2.12, p = .034$ ($d = 0.18, CI_{95\%} [0.01; 0.34]$); the Holm-Bonferroni corrected alpha for this test is .0154. And the equivalence test (with lower bound = -0.3 , upper bound = 0.3) was not significant, $t_{(560.91)} = 1.45, p = .074$. The results of this comparison are thus inconclusive. Similar to learning only about replication failures, when people learn about replication failures and criticisms of QRPs they lose trust in past research from psychological science. This is in line with previous work cited above. However, the findings are not so clear cut regarding trust in future research. There was a non-significant reduction in trust for future research compared to the control group, but the equivalence test was inconclusive. Further research is needed to disambiguate these findings. It is possible that the replication attempts signal that psychological science is self-correcting, and criticisms of QRPs indicate that researchers take these issues seriously, but it is also possible that learning about QRPs undermines these signals.

Reforms and trust

Next, we examined whether and how trust in future research would be affected if the general public was informed about the three core aspects of the replicability crisis, which include replication failures, criticisms of QRPs, and suggested reforms. This comparison informs us about whether receiving brief but complete information regarding current discussions in the field of psychology will affect trust in future research from psychological science. For this, we used a two-sided Welch's t -test and found that compared to the

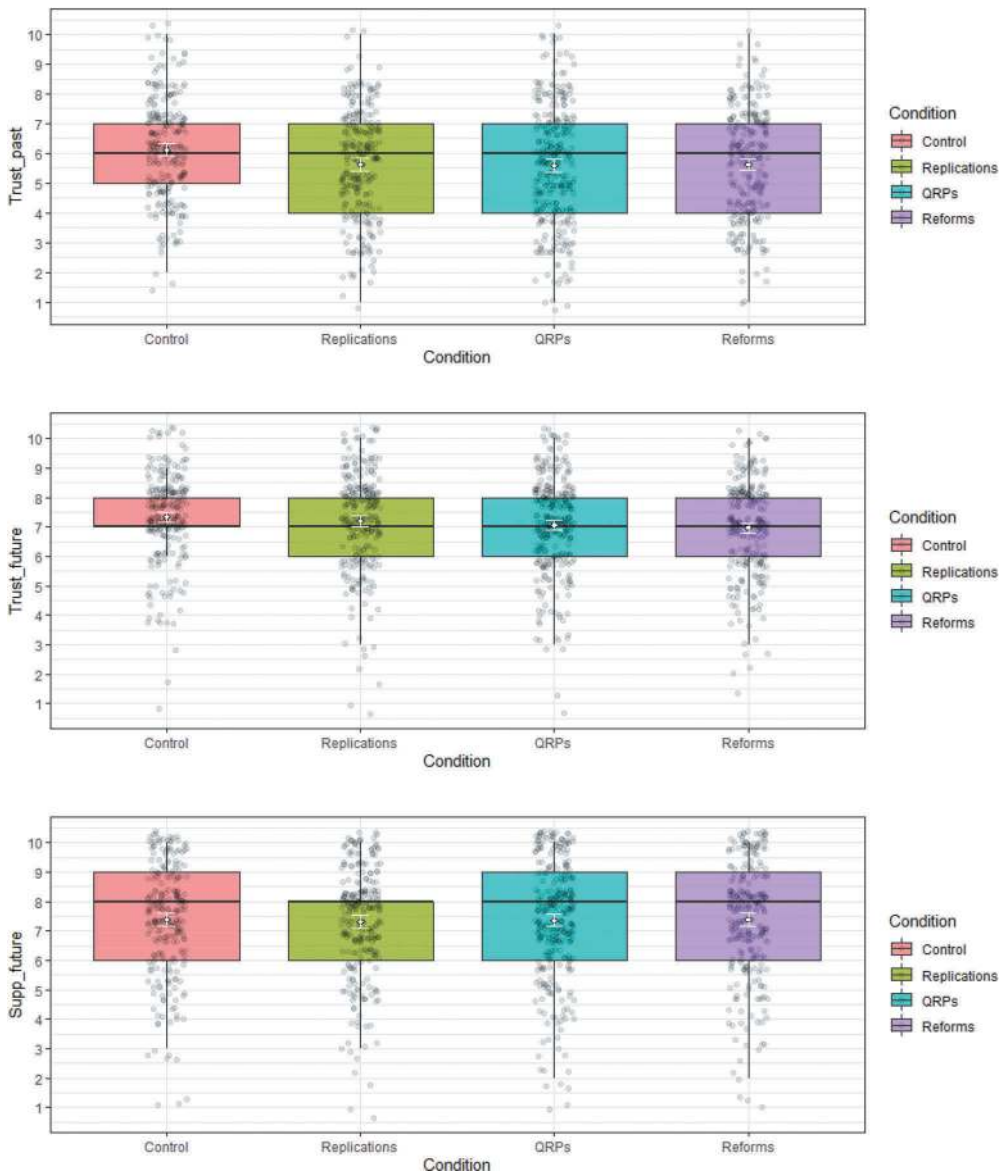


Figure 1. The plots display the means (white diamonds) with 95% confidence intervals (white error bars) and median boxplots for the preregistered outcome measures. Trust in past research, trust in future research, and support for future research in psychological science are plotted in the top, middle, and bottom panels, respectively. The plots were created using ggplot2 function in R (Wickham, 2016).

control group, the group of participants who learned about replication failures, criticisms of QRPs, and suggested reforms had significantly less trust in future research, $t_{(538.86)} = 2.88$, $p = .004$ ($d = 0.25$, $CI_{95\%} [0.08; 0.42]$); the Holm-Bonferroni corrected alpha for this test is .0102. The equivalence test (with lower bound = -0.3 , upper bound = 0.3) was not significant for trust in future research, $t_{(538.86)} = 0.62$, $p = .269$. This suggests that learning about all three aspects of the replicability crisis (replication failures, criticisms of QRPs, and

suggested reforms) reduces trust in future research from psychological science. This result is unexpected, given that learning about replication failures did not significantly affect trust in future research, nor did learning also about criticisms of QRPs.

Why would additionally learning about reforms reduce trust? One explanation might be that people, upon learning that new reforms such as increased transparency in research practices, might be surprised to find out that transparency is not already part of the process, resulting in a negative reaction. An alternative explanation is a methodological confound. The three other conditions began with the paragraph describing the history of how psychology has been studied, whereas the reforms condition started with the paragraph on replication failures (i.e., immediately started on a negative note). This may have caused people to react more negatively to the reforms condition. However, results of exploratory robustness checks for this finding (see “Sensitivity Analyses” in supplementary materials) showed that there were non-significant differences in trust of future research, between the control and reforms groups, unless we included people who indicated that they were somewhat well-informed (i.e., a rating of 8 or higher on that item) in the analyses. As such, we would warn against over interpreting this unexpected result. Yet another explanation may be that people do not believe that the suggested reforms are extreme enough or that they will not have a significant impact soon enough to restore their trust in future research. Further research is needed to test these alternate explanations.

Support for future research

We also pre-registered our intention to examine support for future research in psychological science. Compared to the control group, support for future research was not significantly different for the groups learning about replication failures ($t_{(525.91)} = 0.45$, $p = .326$, $d = 0.04$, $CI_{95\%} [-0.13; 0.21]$), criticisms of QRPs ($t_{(555.83)} = 0.21$, $p = .832$, $d = 0.02$, $CI_{95\%} [-0.15; 0.18]$), or suggested reforms ($t_{(535.78)} = 0.04$, $p = .967$, $d = 0.004$, $CI_{95\%} [-0.17; 0.17]$). The equivalence tests (with lower bound = -0.3 , upper bound = 0.3) for all three comparisons to the control group were significant: replication failures, $t_{(525.91)} = 3.01$, $p = .001$; criticisms of QRPs, $t_{(555.83)} = 3.35$, $p < .001$; and reforms, $t_{(535.78)} = 3.45$, $p < .001$. Therefore, we can conclude that for each of these comparisons the effect is not larger than 0.3. People thus support future research in psychological science regardless of whether they have heard about any aspects of the replicability crisis. One explanation may be that people understand that psychological science is a young science and is thus likely to make mistakes along the way. Moreover, the replicability crisis, and each of the core components that we have presented (replication failures, criticisms of QRPs, and reforms), may act as a signal to the public showing that we care about getting it right, that we will identify and correct mistakes. People’s views about the process of science may be analogous to how researchers view science, in so far as it is self-correcting and that it advances through periods of crises (e.g., Kuhn, 1970; Vazire, 2018). Thus, wanting to further the advancement of knowledge about humans, people desire for psychology to continue to be supported by public funding, even if they may trust its future research outputs a little less, because in the long-run the research may become more trustworthy. Of course, these explanations are speculative and should be examined by future research.

Exploratory analyses

In exploratory analyses, we examined the measures of negative and positive feelings (designed to be used to calculate an objective measure of ambivalence) people had in response to the information they were given. A one-way ANOVA revealed no statistically significant differences between groups on positive feelings, $F_{(3, 1094)} = 2.51, p = .058$ ($M_{\text{control}} = 6.67, SD_{\text{control}} = 1.20$; $M_{\text{rep}} = 6.54, SD_{\text{rep}} = 1.39$; $M_{\text{qrp}} = 6.38, SD_{\text{qrp}} = 1.38$; $M_{\text{reform}} = 6.44, SD_{\text{reform}} = 1.23$). However, there were statistically significant differences between groups on negative feelings, $F_{(3, 1096)} = 4.84, p = .002$. Bonferroni corrected *post hoc* comparisons (corrected alpha .0083) showed that the reforms group ($M = 5.01, SD = 1.83$) had significantly more negative feelings than the control group ($M = 4.49, SD = 1.69$; $p < .001$, Cohen's $d = 0.29, CI_{95\%} [0.12; 0.46]$), and the replications group ($M = 4.60, SD = 1.78, p = .009$, Cohen's $d = 0.22, CI_{95\%} [0.06; 0.39]$), but not the QRPs group ($M = 4.87, SD = 1.77, p = .348$, Cohen's $d = 0.08, CI_{95\%} [-0.08; 0.24]$). The other group differences were not statistically significant. Therefore, people in the reforms group had more negative feelings about the events they read than people in the control group and the replication failures group. This supports our speculation that the unexpected reduction in trust for future research that resulted from reading about reforms may have been due to a methodological confound (i.e., beginning on a negative note). But this finding is also in line with the alternative explanation that people may have reacted negatively to learning about the reforms because they may have been previously unaware that the suggested reforms (e.g., transparency) were not already part of the process. Moreover, this analysis was exploratory and our conclusions are thus speculative. Future research can examine these speculative explanations.

Conclusion

Psychology's replication crisis has sparked many debates about whether public discussions about replication failures and questionable research practices will cause people to lose trust in psychological science. We examined whether informing people about three major aspects of the replication crisis (i.e., replication failures, criticisms of questionable research practices, and reforms) affects how much people trust psychological research, distinguishing trust in past research from trust in future research. Our results, using a Prolific Academic sample, suggest that being informed about replication failures and criticisms of questionable research practices may reduce trust in past research but not trust in future research. The results were a little less clear regarding what happens when people are also informed about reforms, such as increased transparency, which caused people to trust future research less. Nevertheless, people tended to indicate that future research in psychological science should be supported by public funding. We hope that these results will inform the debates about the impact of psychology's replication crisis and that future research will examine these findings in more representative samples, testing also the alternative explanations of the findings. Moreover, future research should examine the impact of these events on different samples of people (e.g., science funders may respond more favourably to reforms; students may respond negatively to the crisis and be less attracted to study past research), and the impact of different variations of presenting the information.

Note

1. Cohen's d effect sizes were calculated using the "effsize" package in R (Torchiano, 2018).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Australian Endeavour Postdoctoral Fellowship [Australia Awards-Endeavour Fellowships]; Nederlandse Organisatie voor Wetenschappelijk Onderzoek [452-17-013].

ORCID

Farid Anvari  <http://orcid.org/0000-0002-5806-5654>

References

- Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology, 19*(6), 586.
- Baker, M. (2016). Is there a reproducibility crisis? *Nature, 452*–454(26). doi:10.1038/533452a
- Banks, G. C., O'Boyle, E. H., Jr, Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., ... Adkins, C. L. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management, 32*(1), 5-20. doi:10.1177/0149206315619011
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*(7391), 531.
- Brumfiel, G. (2009). Science journalism: Supplanting the old media? *Nature News, 458*(7236), 274–277.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436.
- Carey, B. (2018, July 16). Psychology itself is under scrutiny. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/07/16/health/psychology-studies-stanford-prison.html>
- 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* (2), 158–163.
- Connor, S. (2015, August 27). Study reveals that a lot of psychology research really is just 'psycho-babble'. *Independent*. Retrieved from <http://www.independent.co.uk/news/science/study-reveals-that-a-lot-of-psychology-research-really-is-just-psycho-babble-10474646.html>
- Cova, F., Strickland, B., Abatista, A. G., Allard, A., Andow, J., Attie, M., ... Zhou, X. (2018, April 9). Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology, 14*, 1–36.
- Critchley, C. R. (2008). Public opinion and trust in scientists: The role of the research context, and the perceived motivation of stem cell researchers. *Public Understanding of Science, 17*(3), 309–327.
- Cumming, G. (2016, October 6). One reason so many scientific studies may be wrong. *The Conversation*. Retrieved from <https://theconversation.com/one-reason-so-many-scientific-studies-may-be-wrong-66384>
- Davis, C., Naci, H., Gurpinar, E., Poplavska, E., Pinto, A., & Aggarwal, A. (2017). Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European medicines agency: Retrospective cohort study of drug approvals 2009–13. *BMJ, 359*, j4530.

- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology*, 30, 1.
- Engber, D. (2017, August 21). Is science broken? *Slate*. Retrieved from http://www.slate.com/articles/health_and_science/science/2017/08/science_is_not_self_correcting_science_is_broken.html
- 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(11), 2628–2631.
- Feilden, T. (2017, February 22). Most scientists 'can't replicate studies by their peers'. *British Broadcasting Corporation*. Retrieved from <http://www.bbc.com/news/science-environment-39054778>
- Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *PLoS One*, 10(12), e0143723.
- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, 7(1), 45–52.
- Fraser, H., Parker, T. H., Nakagawa, S., Barnett, A., & Fidler, F. (2018, March 21). Questionable research practices in ecology and evolution. *PLoS One*, 13(7), e0200303.
- Gebellhof, R. (2018, April 6). Thanks a lot, Scott Pruitt. You've made transparency a bad word. *The Washington Post*. Retrieved from https://www.washingtonpost.com/blogs/post-partisan/wp/2018/04/06/thanks-a-lot-scott-pruitt-youve-made-transparency-a-bad-word/?utm_term=.4e4c24fb1e1c
- Genschow, O., & Crusias, J. (2018, February 11). "We are ready to move!" An interview with Daniel Lakens and Klaus Fiedler on the current challenges in the field of psychological research. *The inMind blog*. Retrieved from <http://www.in-mind.org/blog/post/we-are-ready-to-move-an-interview-with-daniel-lakens-and-klaus-fiedler-on-the-current>
- Gibbons, M. (1999). Nature. *Science's New Social Contract with Society*, 402(6761supp), C81–C84.
- Gluckman, P. (2014). The art of science advice to government. *Nature*, 507(7491), 163–165.
- Good, I. J. (1988). The interface between statistics and philosophy of science. *Statistical Science*, 3(4), 386–397.
- Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., ... Calvillo, D. P. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573.
- Hallman, W. K. (2017). What the public thinks and knows about science—And why it matters. In K. H. Jamieson, D. Kahan, & D. A. Scheufele (Eds.), *The oxford handbook of the science of science communication* (pp. 61–72). New York, NY: Oxford University Press.
- Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: The Muenster Epistemic Trustworthiness Inventory (METI). *PLoS One*, 10(10), e0139309.
- Hendriks, F., Kienhues, D., & Bromme, R. (2016). Disclose your flaws! Admission positively affects the perceived trustworthiness of an expert science blogger. *Studies in Communication Sciences*, 16(2), 124–131.
- Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.
- Ioannidis, J. P. A. (2017). Statistical biases in science communication: What we know about them and how they can be addressed. In K. H. Jamieson, D. Kahan, & D. A. Scheufele (Eds.), *The oxford handbook of the science of science communication* (pp. 61–72). New York, NY: Oxford University Press.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.
- Jonge, J. D. (2015). *Trust in Science in the Netherlands 2015*. Den Haag: Rathenau Instituut.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- 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(2), 259–269.
- Lamberts, R. (2017). *The Australian beliefs and attitudes towards science survey*. Canberra, Australia: The Australian National University.
- Lindholm, M., Bergman, M., & Gustav, B., translated by Garrison, H., graphics by Isaksson, P. (2018). Vetenskap & Allmänhet (VA) Barometer 2017/2018. Retrieved from https://v-a.se/downloads/varapport2017_3_eng.pdf
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737.

- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*(1), 123.
- National Science Board. (2016). *Science and engineering indicators. 2016 (NSB-2016-1)*. Arlington, VA: National Science Foundation.
- Neuroskeptic (2015, May 18). P-hacking: A talk and further thoughts. *Discover Magazine*. Retrieved from <http://blogs.discovermagazine.com/neuroskeptic/2015/05/18/p-hacking-a-talk-and-further-thoughts/#.W0WtHfZuK2w>
- O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., ... Balatekin, N. (2018). Registered replication report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science, 13*(2), 268–294.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science, 7*(6), 528–530.
- Pickett, J. T., & Roche, S. P. (2018). Questionable, objectionable or criminal? Public opinion on data fraud and selective reporting in science. *Science and Engineering Ethics, 24*(1), 151–171.
- Pittinsky, T. L. (2015). America's crisis of faith in science. *Science, 348*(6234), 511–512.
- Rutjens, B. T., Heine, S. J., Sutton, R., & van Harreveld, F. (2017). Attitudes towards science. *Advances in Experimental Social Psychology, 57*, 2.
- Scheufele, D. A. (2013). Communicating science in social settings. *Proceedings of the National Academy of Sciences, 110*(Supplement 3), 14040–14047.
- Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: The case of ambivalence. *Current Opinion in Behavioral Sciences, 15*, 39–45.
- Schneider, I. K., Veenstra, L., van Harreveld, F., Schwarz, N., & Koole, S. L. (2016). Let's not be indifferent about neutrality: Neutral ratings in the international affective picture system (IAPS) mask mixed affective responses. *Emotion, 16*(4), 426.
- 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*(11), 1359–1366.
- Srivastava, S. (2017, April 6). Re: Learning exactly the wrong lesson [Web log message]. Retrieved from <https://hardsci.wordpress.com/2017/04/06/learning-exactly-the-wrong-lesson/>
- Torchiano, M. (2018). Effsize: Efficient effect size computation. R package version 0.7.4, Retrieved from <https://CRAN.R-project.org/package=effsize>.
- Vazire, S. (2016, March). Re: Is this what it sounds like when the doves cry? [Web log message]. Retrieved from <http://sometimesimwrong.typepad.com/wrong/2016/03/doves-cry.html>
- Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology, 3*, 1.
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417.
- Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Jr, ... Bulnes, L. C. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science, 11*(6), 917–928.
- Wicherts, J. M. (2011). Psychology must learn a lesson from fraud case. *Nature News, 480*(7375), 7.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
- Wingen, T., Berkessel, J., & Englich, B. (2019, February 22). No replication, no trust? *How Low Replicability Influences Trust in Psychology*. doi:10.31219/osf.io/4ukq5
- Wissenschaft im Dialog/Kantar Emnid. (2017). German science barometer. Retrieved from https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Projekte/Wissenschaftsbarometer/Dokumente_17/Einzelgrafiken/Sciencebarometer2017_brochure_web.pdf
- Yong, E. (2015, August 27). How reliable are psychology studies? A new study shows that the field suffers from a reproducibility problem, but the extent of the issue is still hard to nail down. *The Atlantic*. Retrieved from <https://www.theatlantic.com/science/archive/2015/08/psychology-studies-reliability-reproducibility-nosek/402466/>