Check for updates

Empirical Article

Advances in Methods and Practices in Psychological Science
January-March 2022, Vol. 5, No. 1, pp. 1–15
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459211070559
www.psychologicalscience.org/AMPPS

Caution, Preprint! Brief Explanations Allow Nonscientists to Differentiate Between Preprints and Peer-Reviewed Journal Articles

Tobias Wingen[1], Jana B. Berkessel[2], and Simone Dohle[1]
[1]Department of Psychology, University of Cologne, Cologne, Germany, and [2]Mannheim Centre for European Social Research, University of Mannheim, Mannheim, Germany

Abstract

A growing number of psychological research findings are initially published as preprints. Preprints are not peer reviewed and thus did not undergo the established scientific quality-control process. Many researchers hence worry that these preprints reach nonscientists, such as practitioners, journalists, and policymakers, who might be unable to differentiate them from the peer-reviewed literature. Across five studies in Germany and the United States, we investigated whether this concern is warranted and whether this problem can be solved by providing nonscientists with a brief explanation of preprints and the peer-review process. Studies 1 and 2 showed that without an explanation, nonscientists perceive research findings published as preprints as equally credible as findings published as peer-reviewed articles. However, an explanation of the peer-review process reduces the credibility of preprints (Studies 3 and 4). In Study 5, we developed and tested a shortened version of this explanation, which we recommend adding to preprints. This explanation again allowed nonscientists to differentiate between preprints and the peer-reviewed literature. In sum, our research demonstrates that even a short explanation of the concept of preprints and their lack of peer review allows nonscientists who evaluate scientific findings to adjust their credibility perception accordingly. This would allow harvesting the benefits of preprints, such as faster and more accessible science communication, while reducing concerns about public overconfidence in the presented findings.

Keywords

preprints, peer review, credibility, science communication, publishing

Received 6/18/21; Revision accepted 11/27/21

Scientific findings, in psychology and beyond, are rapidly becoming more open and accessible. As part of this open-science movement, preprints—that is, scientific manuscripts preceding formal peer review and publication—have gained popularity, and their number is growing exponentially (see Fig. 1). This development has been accelerated by the COVID-19 crisis, during which researchers aim to rapidly disseminate their findings instead of going through the traditional peer-review process (Kwon, 2020; Polka et al., 2021; Rahal & Heycke, 2020). Moreover, this development was facilitated by an increasing availability of preprint servers in general (e.g., OSF Preprints) but also for specific disciplines (e.g., PsyArXiv for psychological research).

The fact that preprints are typically not peer reviewed does not seem to be a significant barrier to their success. One reason for this may be that the peer-review process has several drawbacks. First, the peer-review process is time-consuming and contributes to a substantial delay between the discovery and the publication of research findings (Cooke et al., 2016; Huisman & Smits, 2017). Second, peer reviewers are humans, and thus their judgments can be biased and influenced by factors other

Corresponding Author:
Tobias Wingen, Department of Psychology, University of Cologne, Richard-Strauss-Str. 2, 50931 Cologne, Germany
Email: tobias.wingen@uni-koeln.de

Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/), which permits noncommercial use, reproduction, and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

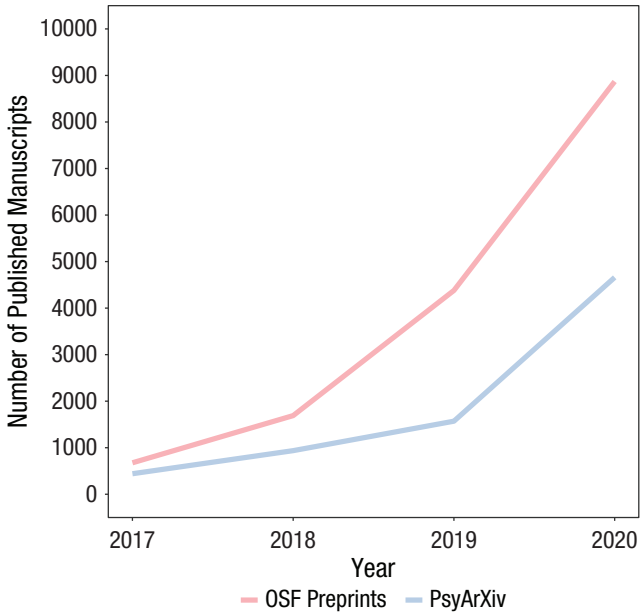

Fig. 1. Development of the number of manuscripts published per year on two major preprint servers for psychology and the social sciences since 2017. Numbers were derived by searching for available preprints on Google Scholar and filtering for each year and server.

than scientific quality (Helmer et al., 2017; Jukola, 2017; Okike et al., 2016). Finally, peer review may further hinder scientific progress because some reviewers oppose unconventional theories, methods, and practices, such as publishing nonsignificant findings or failed replications (Eisenhart, 2002; Elson et al., 2020; French, 2012; Olson et al., 2002). For these reasons, some scholars even argue that peer review is a deeply flawed process and should be abolished (Heesen & Bright, 2021; Smith, 2006).

Nevertheless, peer review is currently the established standard quality-control process for scientific publications (e.g., Elson et al., 2020; Nosek & Bar-Anan, 2012). Indeed, there is empirical evidence that peer-reviewed manuscripts have a higher quality of reporting compared with their non-peer-reviewed version (Carneiro et al., 2019; Cobo et al., 2011; Goodman et al., 1994). Moreover, various studies have shown that peer reviewers usually detect some errors in manuscripts (Godlee et al., 1998; Okike et al., 2016; Schroter et al., 2004). Hence, researchers across disciplines consider peer review as a guiding principle on which work they read and cite. For example, a large international survey found that scientists considered peer review as the most significant factor for determining the quality and trustworthiness of research (Tenopir et al., 2016), and most scientists emphasize that it is important that preprints are ultimately submitted to a peer-reviewed journal (Soderberg et al., 2020).

However, preprints are not available only to scientists (who, in general, can be assumed to know that preprints are not peer reviewed). Instead, because preprints typically are published in open access, they are also openly available to the general public, who might not be aware that preprints are usually not peer reviewed. In fact, especially during the COVID-19 crisis, many preprints became part of the public discourse through traditional and social media (Fraser et al., 2021). For example, a now-retracted preprint that described an "uncanny similarity" between SARS-CoV-2 and HIV spurred discussion on social media on whether SARS-CoV-2 is a genetically engineered bioweapon (Koerber, 2021), which later became one of the leading coronavirus-related conspiracy theories (Imhoff & Lamberty, 2020). Presumably because of this incident, the preprint server bioRxiv, who provided this questionable preprint, added a warning to their website that preprints are preliminary, non-peer-reviewed reports (Forster, 2020). In another example, a preprint on the SARS-CoV-2 viral load in children was disparaged on the title page of the largest German newspaper (Niggemeier, 2020). The newspaper, however, ignored that the work was a preprint and heavily criticized some preliminary analyses. This public debate over a preprint might have damaged trust in science in Germany (Lindner, 2020), which could have had serious consequences for the adherence and adoption of recommended protective behaviors (Dohle et al., 2020). These examples illustrate what many researchers fear: members of the general public treating non-peer-reviewed preprints as established evidence, leading to ill-advised decisions and potentially damaging public trust in science (Fox, 2018; Heimstädt, 2020; Rahal & Heycke, 2020; Sheldon, 2018).

This concern about preprints, which has been described as the most frequent argument against them (Vazire, 2020), goes beyond COVID-19-related research and is highly relevant for all research findings of public interest. Indeed, media outlets and public-science communication blogs also cover preprints on psychological topics such as climate change anxiety (Chow, 2021), personality (Adam, 2019), or even the trustworthiness of psychological research as a whole (Chivers, 2020). Preprints in psychology may be especially likely to catch the public eye because they deal with questions related to human behavior and society. It thus seems likely that some nonscientists even directly seek out psychological preprints because they often address topics highly relevant to their lives.

The central assumption underlying concerns about the public availability of preprints is that nonscientists fail to differentiate between preprints and peer-reviewed literature and thus treat them as equally credible sources. However, this assumption currently lacks empirical

evidence. Because preprints are often presented with no or very little accompanying information (e.g., simply stating that the results stem from a preprint), we believe that in such a situation, nonscientists will indeed fail to incorporate this information in their credibility judgment. This is because they lack the necessary background knowledge that preprints are not peer reviewed. We hypothesize that without an additional explanation of preprints and their lack of peer review, people will perceive research findings from preprints as equally credible compared with research findings from the peer-reviewed literature (Hypothesis 1).

However, recent research suggests that even very brief explanations (e.g., warning labels) allow nonscientists to adjust their credibility ratings (Koch et al., 2021), even for complex scientific topics (Anvari & Lakens, 2018; Hendriks et al., 2020; Wingen et al., 2020). If such a brief explanation of preprints includes that they are not peer reviewed and thus did not undergo the established standard quality-control process for psychological publications, nonscientists might perceive preprints as less credible. Emphasizing increased quality control, for example through consumer reviews or quality-management systems (Adena et al., 2019; Boiral, 2012; Resnick et al., 2006; Silva & Topolinski, 2018), and highlighting adherence to community norms and standards (Bachmann & Inkpen, 2011; Blanchard et al., 2011; Wenegrat et al., 1996) are linked to increased credibility and trustworthiness. We thus hypothesize that after receiving an explanation of preprints and their lack of peer review, nonscientists would perceive preprints as less credible than peer-reviewed articles (Hypothesis 2).

Overview of Studies

We conducted five experimental studies to test whether nonscientists perceive preprints as less credible than peer-reviewed literature and whether this depends on whether they receive an explanation of the peer-review process. We focused on preprints covering research findings from psychology and the social sciences because they seem particularly likely to be comprehensible and interesting to the general public. In the pilot study, we explored whether preprints in psychology and the social sciences typically provide an explanation of preprints and the peer-review process. We coded 200 recent preprints and examined whether they sufficiently explain their lack of peer review. Study 1 (German sample) and Study 2 (U.S. sample) tested whether nonscientists would be able to differentiate between peer-reviewed literature and preprints without an explanation of preprints and the peer-review process. Study 3 (within-subjects design) and Study 4 (between-subjects design) tested whether nonscientists would perceive preprints as less credible than peer-reviewed articles after receiving an explanation of preprints and their lack of peer review. Finally, in Study 5, we developed a shortened version of this explanation and tested whether this very brief explanation allowed nonscientists to differentiate between preprints and peer-reviewed literature. We, moreover, cross-sectionally explored how this explanation may work (mediation) and whether the effect of this explanation depends on education and familiarity with the publication process (moderation).

Preregistration

Studies 1 to 5 and the Supplemental Study 1 are preregistered. All preregistration forms are shared on the OSF (https://osf.io/egkpb). The pilot study, which focused on coding existing data, was not preregistered.

Data, materials, and online resources

All materials, anonymized data sets, and analyses code are shared on the OSF. Statistical analyses were conducted using R (Version 4.0.4; R Development Core Team, 2021), and for the main analyses, we relied on the packages *effsize* (Torchiano, 2020), *lavaan* (Rosseel, 2012), *psych* (Revelle, 2021), *pwr* (Champely et al., 2018), *yarrr* (Phillips, 2017), and *TOSTER* (Lakens, 2017). Details regarding our recruitment strategy and regarding one additional study (see Reporting section) are reported in the Supplemental Material available online.

Reporting

For each study, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

The studies are numbered 1 through 5 for narrative style. Chronologically, the studies were run in the following order: 3, 4, 1, 5, 2. Coding for the pilot study was completed shortly after Study 3. We conducted one further study before Study 5. We found that this study likely contains a high percentage of inattentive respondents (for details, see the Supplemental Material available online), which render the obtained null results largely uninterpretable. We thus refrain from discussing this study in the main text, but to increase transparency, we provide details about this study in the Supplemental Material available online and on the OSF. All analyses with a preregistered hypothesis were tested with one-sided p values. In all studies in which we predicted the absence of an effect, we relied on equivalence tests with preregistered equivalence bounds. This is a commonly recommended frequentist method to provide evidence

Table 1. Information About Peer Review in Recent Preprints on Two Major Preprint Servers

Number of preprints informing their readers that:	OSF Preprints	PsyArXiv	Overall
They are a preprint (or similar)	30.00%	25.00%	27.50%
Are not peer reviewed (or similar)	13.00%	18.00%	15.50%[a]
Peer review is typically part of the scientific publication process	1.00%	0.00%	0.50%
Peer review serves as a quality-control measure	0.00%	0.00%	0.00%
Their findings might be preliminary (or similar)	6.00%	0.00%	3.00%

[a]The overall number includes nine publications mentioning that they are "under review" but not 11 publications mentioning that they have been "submitted for publication" because we believe the latter does not clearly indicate to nonscientists that the work has not yet been peer reviewed.

for the absence of a meaningful effect (Lakens, 2017; Lakens et al., 2018).

All participants who completed our studies were included in the analyses unless they met preregistered exclusion criteria or did not respond to our central dependent variable (i.e., perceived credibility, not explicitly preregistered). Participants were blocked from participating in more than one study to avoid nonnaïveté (Chandler et al., 2015). Sample sizes were preregistered in Studies 1 to 5; however, some deviations occurred because we recruited participants online and thus had limited control over the final sample size (for details regarding sample sizes and deviations, see the Supplemental Material available online). However, in no case was the final sample size determined based on the obtained results.

Ethical approval

All studies were conducted consistently with the Declaration of Helsinki, and all are exempt from institutional review board approval by guidelines of the German Psychological Society (2018).

Pilot Study

Method

For the pilot study, we collected the information presented in the 303 most recent manuscripts (at the time of coding; June 2020) on two popular social science preprint servers, commonly used by psychological scientists. These servers were PsyArXiv (https://psyarxiv.com) and the social and behavioral sciences section at OSF Preprints (https://osf.io/preprints). We first collected general bibliographic information (authors, publication date, language, doi, whether the manuscript was a postprint). We excluded 63 manuscripts from our analyses because they appeared to be accepted versions of articles (postprints) and thus peer reviewed, thereby not meeting our definition of preprints. We furthermore excluded 33 non-English preprints and, finally, seven documents that were not preprints (e.g., supplemental materials, book chapter scans).

Given these necessary exclusions, the coders continued coding (by going back further in time and coding earlier preprints) until eventually 200 suitable manuscripts (100 from each server) were included. We coded whether the authors of the preprint (a) mentioned that it is a preprint, (b) mentioned that it is thus not peer reviewed, (c) explained that peer review serves as a quality-control process, (d) explained that peer review is the standard procedure for scientific publication, (e) and/or added another indication that the findings might be preliminary or less credible.

Results

The results showed that only 27.50% of the preprints explicitly stated that they were preprints. Even fewer preprints (15.50%) contained information that they had not undergone peer review yet. Finally, not a single preprint provided information explaining that peer review serves as a quality-control measure. Detailed results for each preprint server are presented in Table 1.

Study 1

In Study 1, we tested whether participants would evaluate psychological research findings that were published as peer-reviewed articles as equally credible as research findings published as preprints.

Method

Participants and design. Participants were German university students recruited online in exchange for course credits and individuals recruited through postings in public German social media groups for voluntary research participation. The study employed a between-subjects experimental design. We randomly assigned participants to one of two between-subjects conditions (preprint condition, peer-review condition). Sample size considerations

Table 2. Overview of Research Findings Used as Stimuli in Studies 1, 2, 4, and 5

Authors	Short description
Gervais and Norenzayan (2012)	Analytical thinking promotes religious disbelief.
Hauser et al. (2014)	When making collective decisions, people share more common resources for future generations.
Nishi et al. (2015)	Financial inequality between group members remains when people are informed about each member's wealth.
Shah et al. (2012)	Poverty drains people's attention.
Wilson et al. (2014)	People dislike doing nothing and prefer an engaged mind.

were made in relation to Study 4, which chronologically took place before Study 1; compared with Study 4, we aimed to double our sample size. The recruited sample was slightly larger and consisted of 277 participants (after excluding 35 participants who already took part in Study 4, as preregistered), out of which 204 provided responses to all credibility ratings and were therefore included in the main analysis (74.5% female; age: M = 25.41 years, SD = 7.09). Power analyses revealed that the sample size of 204 had a 99.87% power to detect the effect observed in Study 4 (d = 0.70, α = .05) and a 95% power to demonstrate in an equivalence test that an observed effect is considerably smaller than the effect observed in Study 4 (preregistered equivalence bound of d < 0.5 compared with d = 0.70 in Study 4).

Procedure. Participants were presented with five different research findings (for an overview of research findings used as stimuli, see Table 2). The findings were described as being published either as a peer-reviewed journal article or as a preprint, depending on condition. For each research finding, participants indicated their perceived credibility ("How credible is this study result?") on a 7-point scale (1 = *not at all credible*, 7 = *very credible*). Participants received no further information (e.g., an explanation of the peer-review process). In fact, all five findings (Gervais & Norenzayan, 2012; Hauser et al., 2014; Nishi et al., 2015; Shah et al., 2012; Wilson et al., 2014) were published in the peer-reviewed journals *Nature* or *Science*. Descriptions of these findings were adapted from prior work and were proved to be comprehensible to nonscientists (Hoogeveen et al., 2020). Findings covered various psychological and economic behavioral science topics, and participants judged the credibility of these five research findings. An average credibility score across all five ratings was computed and served as the dependent variable.

Results

In line with our preregistration, we computed an average credibility score across all five credibility ratings. As predicted, without a brief explanation, participants considered research findings published as preprints (M = 4.09, SD = 0.80) to be equally credible compared with findings published as peer-reviewed journal articles (M = 4.24, SD = 0.88), $t(202)$ = 1.25, p = .211, d = 0.18, 95% confidence interval [CI] = [–0.10, 0.46]. This finding is presented in Figure 2. A preregistered equivalence test, a test that provides support for the absence of a meaningful effect, showed that the observed effect size, which is conventionally considered very small, was equivalent with an interval containing only small to medium effects (d < 0.5), $t(202)$ = 2.29, p = .012. Descriptive statistics for the perceived credibility across studies and conditions throughout this article are presented in Table 3.

Study 2

In Study 2, we aimed to replicate the findings from Study 1 in a different population using an even larger sample

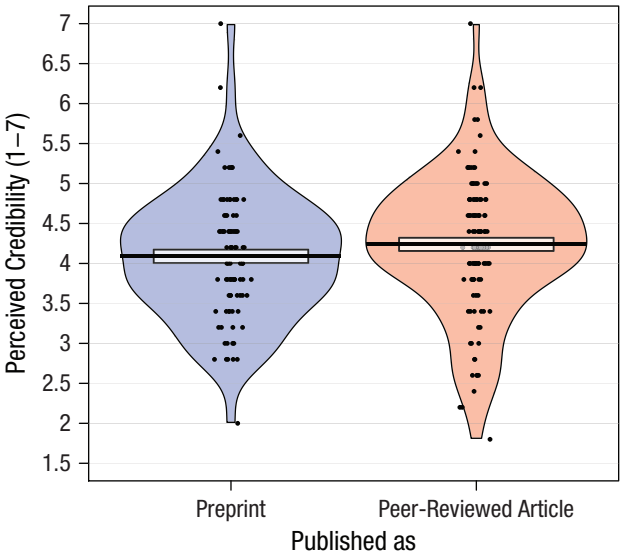

Fig. 2. Pirate plot showing perceived credibility as a function of publication mode in Study 1 (German participants), in which participants received no explanation of preprints. The black dots represent the jittered raw data, which are shown with smoothed densities indicating the distributions in each condition. The central tendency is the mean, and the intervals represent two standard errors around the mean.

Table 3. Perceived Credibility of Research Findings Depending on Source and Explanation Across All Studies Presented in This Article

Study	*M*	*SD*	*t* value[a]	*df*	*p* value[b]	Cohen's *d* [95% confidence interval]
Study 1 (without explanation)						
Peer-reviewed article	4.24	0.88				
Preprint	4.09	0.80	1.25	202	.211	0.18 [–0.10, 0.46]
Study 2 (without explanation)						
Peer-reviewed article	4.73	1.11				
Preprint	4.58	1.00	1.50	464	.136	0.14 [–0.04, 0.32]
Study 3 (with explanation)						
Peer-reviewed article	5.63	1.34				
Preprint	4.00	0.93	10.06	51	< .001 (one-sided)	*dz* = 1.39
Study 4 (with explanation)						
Peer-reviewed article	4.15	0.65				
Preprint	3.67	0.72	3.74	111	< .001 (one-sided)	0.70 [0.32, 1.09]
Study 5 (with explanation)						
Peer-reviewed article	4.65	1.00				
Limited information	4.42	1.15	2.02	379	.044	0.21 [0.01, 0.41]
Authors' explanation	4.39	1.07	2.31	359	.010 (one-sided)	0.25 [0.04, 0.45]
External explanation	4.31	1.02	3.25	379	< .001 (one-sided)	0.33 [0.13, 0.54]

[a]The *t*-tests results refer to the comparison of the respective condition with the peer-review condition in each study. These are *t* tests for dependent samples in Study 3 and for independent samples in the other studies.
[b]One-sided *p* values are reported for directional hypotheses.

size (N = 466; U.S. sample) and a stricter preregistered criterion of what constitutes a negligible difference (d < 0.3). The design was identical to Study 1 except that Study 2 also included a basic text-comprehension check that had to be answered correctly to ensure that participants were aware that the five research findings were published as preprints or peer-reviewed journal articles, respectively.

Method

Participants and design. Participants were U.S.-based individuals recruited on the Amazon Mechanical Turk (MTurk) platform in exchange for $0.50. The target sample size was set to 578, which allowed us to detect group differences of d = 0.30 ($1 - \beta = 0.95$, $\alpha = .05$) and, moreover, provided sufficient power for an equivalence tests ($1 - \beta$ = 0.95, equivalence bounds of d = 0.3). To increase data quality, we opted to exclude participants who failed a basic text-comprehension check (see below). This decision was based on previous research raising concerns about MTurk workers not reading study materials or even being bots (Chmielewski & Kucker, 2020). To compensate for potential exclusions, we recruited 753 participants, of which 476 passed the preregistered comprehension check. Finally, 466 participants answered all credibility items and were therefore included in the main analysis (42.15% female; age: M = 37.00 years, SD = 11.97). Despite this reduced sample size, a sensitivity analysis revealed that the final sample size had an 80% power (with $\alpha = .05$) to detect an effect of d = 0.26 and a 95% power to detect d = 0.33.

Procedure. For the text-comprehension check, participants had to answer how the research findings were published and were presented with eight options (e.g., "as textbooks," "as preprints"). If participants answered the text-understanding question incorrectly, they were asked to carefully read the text again. If they failed the text-understanding question again, they were excluded from our analyses. We also added a few exploratory questions about whether participants perceived the research findings as strictly quality controlled, whether they believed that the researchers adhered to the standard publication procedure, and participants' education and familiarity with the publication process (to ensure comparability with Study 5). Apart from this, the procedure and design were identical to Study 1.

Results

We computed an average credibility score across all five credibility ratings. As predicted and in line with Study 1, participants rated research findings from preprints (M = 4.58, SD = 1.00) as equally credible as research

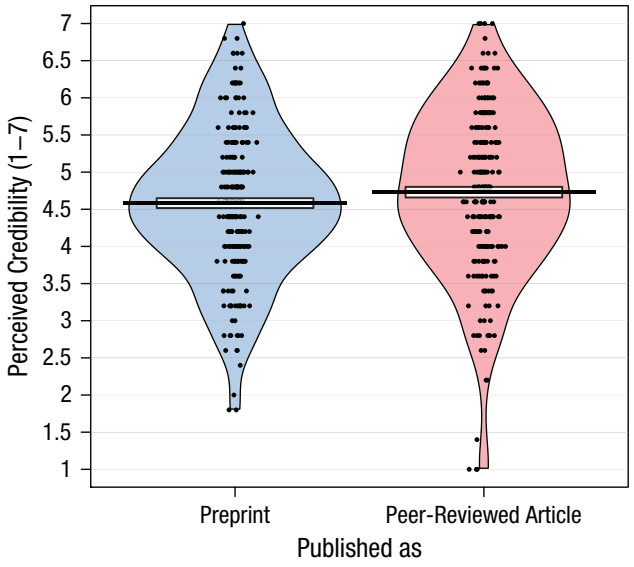

Fig. 3. Pirate plot showing perceived credibility as a function of publication mode in Study 2 (U.S. participants), in which participants received no explanation of preprints.

results from peer-reviewed journal articles (M = 4.73, SD = 1.11), $t(464) = 1.50$, $p = .136$, $d = 0.14$, 95% CI = [–0.04, 0.32]. This finding is presented in Figure 3. An equivalence test showed that this observed effect size, which is conventionally considered very small, was equivalent with an interval containing only small effects ($d < 0.3$), $t(464) = 1.741$, $p = .041$.

Study 3

Studies 1 and 2 found that without an explanation, nonscientists rated research findings from preprints as equally credible as research findings from peer-reviewed journal articles. Study 3 tested whether nonscientists truly believe that the two types are equally credible or whether they start to differentiate once they get an explanation of preprints and the peer-review process and can directly compare these two options. Study 3 straightforwardly tested this by employing a within-subjects design in which participants rated research findings in general.

Method

Participants and design. Participants were recruited through postings in public German social media groups for voluntary research participation. The targeted sample size was set to 45, based on an a priori power analysis for 95% power (one-sided α of .05) to detect a moderate effect of $dz = 0.5$ that would be typical for similar social-psychological research. The recruited sample was slightly larger, as is often the case in online studies, and consisted of 65 participants. Of these participants, 52 responded to all credibility items and were therefore included in the main analysis (73.08% female; age: M = 30.83 years, SD = 9.71).

Procedure. This study employed a within-subjects design. Participants read a short, jargon-free description of the peer-review process, which highlighted that peer review serves as a quality-control process and that peer review currently is the standard procedure for scientific publication. They were also informed that some research findings are initially published before the peer-review process as preprints to achieve rapid dissemination of results. The full description reads as follows (translation by authors):

> Usually, scientific articles are subject to an extensive peer-review process. This means that other scientists anonymously review articles submitted to a scientific journal. They then speak out for or against a publication and provide important suggestions for article improvement. This procedure is considered the gold standard of scientific journals. Only articles that receive positive reviews have a chance of being published. This procedure is intended to ensure that the articles are of particularly high quality. However, some articles are now published online as preprints without having been peer reviewed. This allows scientists to make their results available to the public very rapidly, whereas the time-consuming peer-review process can take several months. Normally, peer review is then carried out after the article has been submitted to a scientific journal.

Afterward, participants reported the perceived credibility of research findings published as peer-reviewed articles ("How credible are research findings that are published as journal articles [with peer review]?") and as preprints ("How credible are research findings that are published as preprints [without peer review]?") on a 7-point rating scale (1 = *not at all credible*, 7 = *very credible*). Finally, participants indicated whether they had heard about preprints and peer-reviewed articles before the study, completed demographics, and were debriefed.

Results

As predicted, participants rated research findings from preprints (M = 4.00, SD = 0.93) as less credible than research results from peer-reviewed journal articles (M = 5.63, SD = 1.34), $t(51) = 10.06$, one-sided $p < .001$, dz = 1.39 (see Fig. 4).

Study 4

Study 4 tested whether the finding of Study 3 generalizes to a more realistic situation in which participants do not

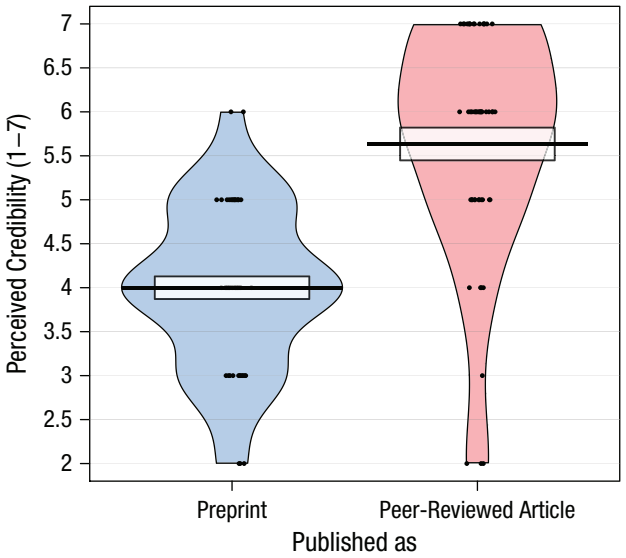

Fig. 4. Pirate plot showing perceived credibility as a function of publication mode in Study 3 (German participants), in which participants received an explanation and directly compared the two options.

directly compare preprints and peer-reviewed articles with each other. Instead, participants judged specific research findings, and the design was largely identical (and thus directly comparable) to Studies 1 and 2.

Method

Participants and design. Participants were German university students recruited online in exchange for course credits. The study employed a between-subjects experimental design. We randomly assigned participants to one of two between-subjects conditions (preprint condition, peer-review condition). The target sample size was set to 102, based on an a priori power analysis for 80% power (one-sided α of .05) to detect a moderate effect of $d = 0.5$. The recruited sample consisted of 140 participants, of which 113 responded to all credibility items and were therefore included in the main analysis (76.11% female; age: $M = 23.75$ years, $SD = 5.10$).

Procedure. Participants read the same short descriptions of the peer-review process and preprints as in Study 3 and answered two exploratory text-comprehension questions. Participants judged the credibility of five research findings (the same research findings used in Studies 1 and 2). The findings were described as being published either as peer-reviewed journal articles or as preprints. Ratings were made on a 7-point scale (1 = *not at all credible*, 7 = *very credible*). Participants also indicated whether they had heard about preprints and peer-reviewed articles before the study, received an exploratory open-entry question on how they made their credibility judgments, completed demographics, and were debriefed.

Results

In line with our preregistration, we computed an average credibility score across all five credibility ratings. As predicted, participants rated research findings from preprints ($M = 3.67$, $SD = 0.72$) as less credible than research findings from peer-reviewed journal articles ($M = 4.15$, $SD = 0.65$), $t(111) = 3.74$, one-sided $p < .001$, $d = 0.70$, 95% CI = [0.32, 1.09]. This pattern is depicted in Figure 5.

Study 5

In Study 5, we developed a shortened version of the explanation used in Studies 3 and 4, which could be easily added to preprints. We tested whether this explanation allows nonscientists to differentiate between preprints and the peer-reviewed literature. We further tested whether it matters if this brief explanation is provided by the authors or by an external source but expected the explanation to be effective in both cases. Because most preprints are published in English, we tested this in an English-speaking population ($N = 727$; U.S. sample). We also aimed to explore the underlying mechanism of our explanation and tested preregistered mediators (perceived quality control and perceived adherence to publication standards) and moderators (education and familiarity with the publication process).

Method

Participants and design. Participants were U.S.-based individuals recruited on the Amazon MTurk platform in exchange for $0.50. We randomly assigned participants to one of four between-subjects conditions (peer-review condition, preprint: limited-information condition, preprint: authors'-explanation condition, preprint: external-explanation condition). The target sample size was set to 1,000, which allowed us to detect group differences of $d = 0.29$ ($1 - \beta = 0.95$, one-sided α of .05) and, moreover, provided sufficient power for an equivalence tests ($1 - \beta = 0.91$, equivalence bounds of $d = 0.3$). We recruited 1,051 participants, of which 739 passed the preregistered text-comprehension check. For the text-comprehension check, participants had to answer how the research findings were published (see Study 2). If an additional explanation of peer review and preprints was given, they also indicated for three additional text-comprehension questions whether they were true or false ("Scientific articles are usually peer reviewed"; "As part of the peer-review process, independent researchers evaluate the quality of the work"; and "Preprints have been peer reviewed"). If participants answered any of the questions incorrectly, they were asked to read the text carefully again. If they again failed any of the text-comprehension questions, they were excluded from our analyses. Finally, 727 participants responded to

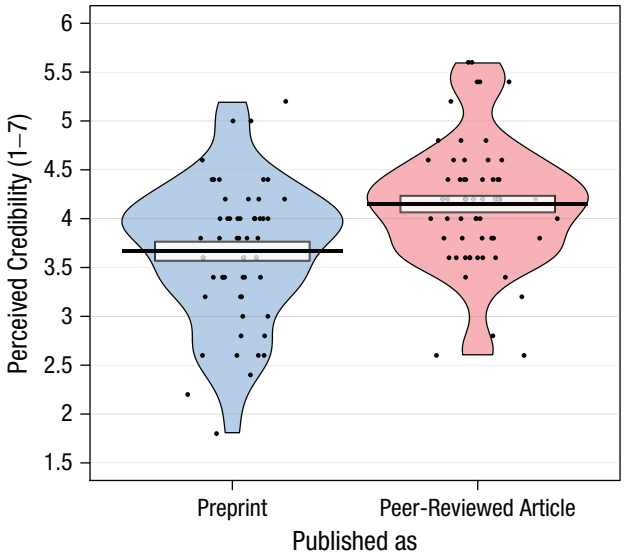

Fig. 5. Pirate plot showing perceived credibility as a function of publication mode in Study 4 (German participants). This study provided participants with an explanation of preprints and the peer-review process.

all credibility items and were thus included in the main analyses (43.39% female; age: M = 39.24 years, SD = 12.68). One participant did not respond to the remaining items, which reduced the sample size for secondary analyses to 726. Despite this reduced sample size, a sensitivity analysis revealed that for all possible group comparisons, our sample had at least an 80% power (with α = .05) to detect an effect of d = 0.30 and a 95% power to detect d = 0.39.

Procedure. Participants learned that they would judge the credibility of five research findings and were randomly assigned to one of four conditions. In the peer-review condition, participants were informed that the findings went through a peer-review process and were published in a scientific journal. The preprint: limited-information condition stated that the findings were preprints, but in contrast to Studies 1 and 2, it was also added that preprints are not peer reviewed (without further information, however, what is meant by peer review). In the other two conditions, the research findings were presented as nonpeer-reviewed preprints, but participants received an additional explanation of preprints and the peer-review process. This additional explanation was allegedly either provided by the authors of the preprint (preprint: authors'-explanation condition) or without any reference to the source in the introduction of the study (preprint: external-explanation condition).

The explanation was drafted building on the information provided in Studies 3 and 4 but incorporated further feedback from colleagues from various disciplines (anthropology, biology, psychology, and sociology) and from nonscientists to ensure an interdisciplinary perspective and comprehensibility. The explanation highlighted two important aspects: that peer review serves as a quality-control process and that peer review currently is the standard procedure for scientific publication. Compared with Studies 3 and 4, we aimed to keep this explanation as comprehensive as possible. This explanation read:

> Scientific articles usually go through a peer-review process. This means that independent researchers evaluate the quality of the work, provide suggestions, and speak for or against the publication. Please note that the present article has not (yet) undergone this standard procedure for scientific publications.

After judging the credibility of the research findings, participants were also asked about the perceived quality control of the research findings, the perceived adherence to scientific publication standards, their education, and their familiarity with the publication process. Credibility ratings were given on a 7-point rating scale (1 = *not at all credible*, 7 = *very credible*). Familiarity with the publication process ("I am familiar with the scientific publication process"), perceived quality control of the research findings ("The quality of the research findings has been strictly controlled"), and perceived adherence to scientific publication standards ("When publishing their findings, the researchers followed the standard procedure of the research community") were measured on a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*).

Results

Main analyses. We again computed an average credibility score across all five credibility ratings. As predicted, across both preprint-explanation conditions, participants reported lower credibility of research findings compared with the peer-review condition (M = 4.65, SD = 1.00). This was the case when participants received the explanation by the authors (M = 4.39, SD = 1.07), $t(359) = 2.32$, one-sided p = .010, d = 0.25, 95% CI = [0.04, 0.45], and by an external source (M = 4.31, SD = 1.02), $t(379) = 3.25$, one-sided $p < .001$, d = 0.33, 95% CI = [0.13, 0.54] (see Figure 6). Unexpectedly, this was also the case when participants received only very limited information (M = 4.42, SD = 1.15), $t(379) = 2.02$, p = .044, d = 0.21, 95% CI = [0.01, 0.41]. The three preprint conditions did not significantly differ from each other (all $ps > .317$, all $ds < .10$), and the observed differences between these conditions were all equivalent with an interval containing only small effects ($d < .3$), all $ps < .031$ (see OSF analyses for details).

Quality control and adherence to scientific publication standards. However, the three preprint explanations differed regarding the perceived quality control of the research findings and the perceived adherence to

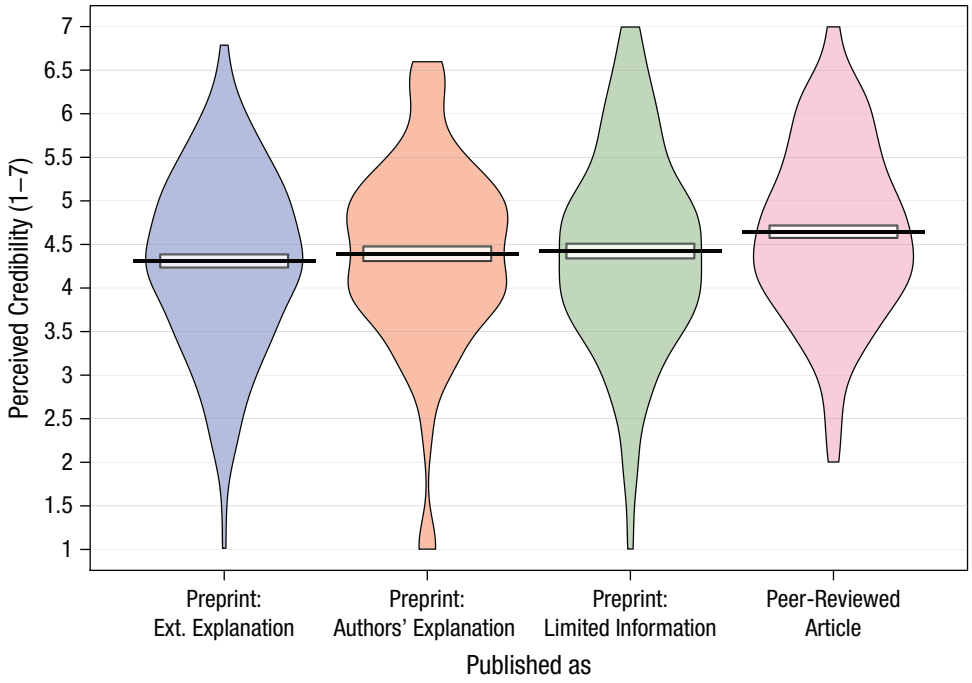

Fig. 6. Pirate plot showing perceived credibility as a function of publication mode and explanation in Study 5 (U.S. participants). Raw data are not visualized in the figure because of the large number of data points.

scientific publication standards. Participants who received an explanation reported lower perceived quality control of preprints compared with the limited information condition (M = 4.27, SD = 1.79). This was the case no matter whether participants received this explanation by the authors (M = 3.72, SD = 1.63), $t(344) = 2.97$, one-sided p = .002, d = 0.32, 95% CI = [0.11,0.53], or by an external source (M = 3. 81, SD = 1.71), $t(363) = 2.51$, one-sided p = .006, d = 0.26, 95% CI = [0.06, 0.47]. Likewise, after receiving an explanation, participants reported lower perceived adherence to scientific publication standards compared with the limited-information condition. This was again the case no matter whether participants received this explanation by the authors (M = 3.83, SD = 1.72), $t(344) = 3.15$, one-sided $p < .001$, d = 0.34, 95% CI = [0.13,0.55], or by an external source (M = 3.95, SD = 1.79), $t(363) = 2.50$, one-sided p = .007, d = 0.26, 95% CI = [0.05, 0.47].

Moderation analyses. In line with our preregistration, we also tested whether education or familiarity with the scientific publication process moderated the effect of our explanation on the perceived credibility of research findings (compared with the peer-review condition). For these analyses, we merged the preprint: authors'-explanation condition and the preprint: external-explanation condition because they did not differ on any of the relevant variables. We conducted multiple linear regression analyses to test whether any of our potential moderator variables moderated the relationship between explanation (detailed explanation vs. peer review) and credibility. Indeed, whereas centered education did not significantly interact with our explanation (b = –0.22, SE = 0.14), $t(539) = 1.58$, p = .115, centered familiarity with the publication process was a significant moderator, which indicates that our explanation was more effective for people who indicated a higher familiarity with the scientific publication process (b = –0.12, SE = 0.05), $t(539) = 2.49$, p = .013 (see Table 4).

Mediation analyses. Finally, we explored preregistered mediators of the effect of our explanation on credibility (compared with the peer-review condition). For these cross-sectional analyses, we again merged the authors'-explanation condition and the external-explanation condition because they did not differ on any of the relevant variables. We investigated whether perceived quality control or perceived adherence to publication standards mediated the negative effect of explaining preprints on perceived credibility. To test this, we ran a parallel mediation model (see Fig. 7) with 10,000 bootstrap resamples using the R package *lavaan* (Rosseel, 2012). This model revealed that both perceived quality control (b = –0.23, 95% CI = [–0.14, –0.33]) and perceived adherence to publication standards (b = –0.25, 95% CI = [–0.15, –0.35]) simultaneously mediated the effect.

Because this mediation model relied on cross-sectional data, these results should be considered with caution because the mediating variables were not experimentally manipulated and may be biased (Bullock et al.,

Table 4. Multiple Linear Regression Predicting Perceived Credibility From Condition, Centered Familiarity With the Publication Process, and Their Interaction Term

Predictor	*B*	*SE B*	*t*(539)	*p*
Condition (0 = peer review, 1 = explanation)	–0.22	0.09	2.46	.014
Familiarity (centered)	0.23	0.04	5.85	< .001
Condition × Familiarity (Centered)	–0.12	0.05	2.49	.013

2010). An observed statistical mediation cannot conclusively prove actual mediation (Fiedler et al., 2011) and should, rather, be seen as a tentative hint for mediation.

General Discussion

A central argument against preprints is that nonscientists might fail to differentiate them from the peer-reviewed literature (Fox, 2018; Vazire, 2020). Indeed, nonscientists from Germany (Study 1) and the United States (Study 2) perceived research findings published as preprints as equally credible as research findings published in peer-reviewed journals. However, a brief explanation of the peer-review process combined with the information that preprints are not peer reviewed led nonscientists to perceive identical research findings published as preprints as less credible than the peer-reviewed literature. This effect was observed for research findings in general (Study 3) and specific psychological research findings (Studies 4 and 5). Study 5 further suggested that even a very brief explanation, which could be added to all preprints, allowed nonscientists to differentiate them from the peer-reviewed literature. Note that this effect emerged independently of whether this explanation was allegedly provided by the preprints' authors or by an external source, albeit the effect was descriptively smaller in the former situation. The explanation seemed to be especially effective for individuals who are rather familiar with the scientific publication system, and it seems to work by influencing whether nonscientists see preprints as quality controlled and as adhering to publication standards. In other words, when nonscientists are well informed about the source of information, they can adjust their credibility ratings accordingly.

In practice, however, most psychological preprints do not contain such an explanation. The pilot study, in which we coded recent preprints from two popular psychological preprint servers, revealed that less than 30% of preprints contained information that they are a preprint. Even fewer mentioned that they are not peer reviewed, and virtually none provided an explanation similar to the one used in our studies. Taking this current status quo into account, our findings suggest that nonscientists might

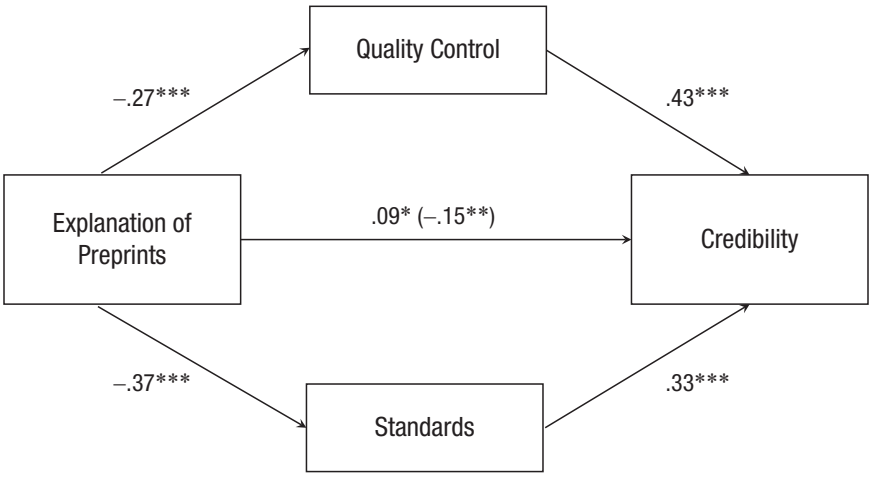

Fig. 7. Parallel mediation analyses involving perceived quality control and adherence to publication standards as dual, simultaneous mediators for the link between explanations (0 = peer-review condition, 1 = merged-explanation conditions) and perceived credibility. Values represent standardized path coefficients. The total effect is presented in parentheses. Asterisks indicate significance at the $p < .05$ level (*), at the $p < .01$ level (**), and at the $p < .001$ level (***).

currently be unable to differentiate between preprints and the peer-reviewed literature.

Some scholars (e.g., Elmore, 2018) have pointed to the fact that the term "preprint" is a misnomer because there may never be a future print version in a scientific journal (e.g., if the preprint does not pass the peer-review process). Nonscientists might, however, believe that preprints are in fact earlier versions of already published and peer-reviewed articles. This discrepancy could explain why Study 5 found that simply stating that preprints have not yet passed peer review—something that many individuals are probably not aware of—already reduced perceived credibility. The same study, however, also demonstrated that a more detailed explanation led to a stronger differentiation between preprints and peer-reviewed literature regarding their perceived quality control and their perceived adherence to publication standards, which were relevant mediators. We thus recommend that future authors of preprints, but also preprint servers or science journalists covering preprints, should briefly explain the peer-review process and highlight that preprints are not peer reviewed. Our research suggests that such an explanation might be especially effective if it includes elements that indicate that peer review serves as a quality-control process and that it is the standard procedure for scientific publication.

One important discussion point, however, is whether it is desirable that nonscientists differentiate between preprints and peer-reviewed literature in terms of credibility. Although the peer-review system leads to improvements of a manuscript (Carneiro et al., 2019; Godlee et al., 1998; Goodman et al., 1994; Schroter et al., 2004), it also has serious drawbacks (Heesen & Bright, 2021; Huisman & Smits, 2017; Jukola, 2017), and one might argue that preprints are not necessarily less credible than peer-reviewed articles. Regardless of whether peer-reviewed articles are objectively more credible, we find that if provided with information about the differences between preprints and peer-reviewed articles, participants used this information to inform their credibility judgments. We, therefore, argue that this information should not be withheld. In contrast to more patronizing statements, such as the statement by BioRxiv (preprints "should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information"), our approach leaves it up to the reader to decide whether a preprint is less credible.

Even if one agrees that preprints are on average less credible than peer-reviewed articles, it could be argued that it is not desirable to reduce the perceived credibility of all preprints because some preprints may in fact be highly credible. However, because psychological research findings are often nonreplicable (Open Science Collaboration, 2015) and context sensitive (Van Bavel et al., 2016), we argue that it is better to err on the side of caution by increasing nonscientists' vigilance toward preprints even if this may not always be necessary. This does not imply, of course, that nonscientists should rely solely on whether a manuscript is a preprint when evaluating its content. In fact, recent work suggests that nonscientists are also sensitive to other important aspects of scientific research, such as the strength of evidence (Hoogeveen et al., 2020) or successful replications of the presented work (Hendriks et al., 2020).

It is also important to discuss the generality of our findings (Simons et al., 2017). First, because we replicated our findings in rather different samples (U.S. MTurk users and German students), we expect our findings to replicate also in more representative samples for these and other Western countries. Note, however, we found that our explanation was more effective for participants who reported a high familiarity with the publication process. This might explain why we observed substantially larger effects in Germany: Because the German samples mostly consisted of undergraduate students, they might be more familiar with the publication process compared with the U.S. samples of Amazon MTurk users. Thus, familiarity with the publication process might constrain the generality of our findings. From an applied perspective, it seems likely that nonscientists seeking out preprints might be rather familiar with the publication process (e.g., journalists), which means that our explanation would be rather effective in such a situation. However, it is also possible that this is a methodological artifact: Participants who read our materials more closely might consequently report a higher familiarity with the publication process and being more strongly affected by the manipulation.

Moreover, it seems likely that the effectiveness of our explanation depends on participants' general trust in science because our explanation highlights that preprints do not follow the established scientific publication procedure. If, however, participants' trust in the established scientific knowledge is generally low, a deviation from established standards might not reduce trust but could even increase it. This could, for example, be the case for politically highly conservative participants, who are contemporarily characterized by relatively low trust in science (Gauchat, 2012).

Finally, it would also be vital to test whether our findings generalize to other forms of non-peer-reviewed science communication, such as blogs, podcasts, or popular science magazines. For example, during the COVID-19 crisis, some scientists shared their findings through non-peer-reviewed podcasts and even press conferences (Kupferschmidt, 2020). In such a situation, it might also be desirable to inform the public that the presented research findings have not been peer reviewed to avoid public overconfidence in the presented research. It, however, remains possible that the public already perceives such publication formats as rather uncommon and

thus less credible, which would leave no room for such an explanation to have an additional effect. This remains an interesting question for future research.

In sum, our work suggests that concerns about non-scientists not differentiating between preprints and peer-reviewed psychological literature are legitimate. However, we also suggest and test a solution: Preprint authors, preprint servers, and other relevant institutions can likely mitigate this problem by briefly explaining the concept of preprints and their lack of peer review. This would allow harvesting the benefits of preprints, such as faster and more accessible science communication, while reducing concerns about public overconfidence in the presented findings.

Transparency

Action Editor: Alexa Tullett
Editor: Daniel J. Simons
Author Contributions

T. Wingen generated the idea for the research project, with feedback from J. B. Berkessel and S. Dohle. T. Wingen and J. B. Berkessel jointly programmed the study and collected the data. T. Wingen wrote the analysis code and analyzed the data, and J. B. Berkessel verified the accuracy of those analyses. T. Wingen wrote the first draft of the manuscript, and all authors critically edited it. All of the authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The research was partly funded by a DFG grant (Deutsche Forschungsgemeinschaft, DO 1900/3-1) awarded to S. Dohle.

Open Practices

Open Data: https://osf.io/r3p6t/
Open Materials: https://osf.io/r3p6t/
Preregistration: https://osf.io/r3p6t/

All data have been made publicly available via OSF and can be accessed at https://osf.io/r3p6t/. All materials have been made publicly available via OSF and can be accessed at https://osf.io/r3p6t/. The analysis plan was preregistered at OSF prior to data collection and can be accessed at https://osf.io/r3p6t/. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Tobias Wingen https://orcid.org/0000-0002-1559-859X

Acknowledgments

We thank Nicolas Alef and Antonia Dörnemann for their valuable practical support. We further thank Paul Davies and Luzie U. Wingen for their extensive feedback on materials. We finally thank Angela Dorrough and Jan Landwehr for valuable comments on an earlier version of this manuscript. The submitted manuscript was previously posted on a preprint archive, doi:10.31219/osf.io/7n3mj.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/25152459211070559

References

Adam, D. (2019, March 12). Does a 'Dark Triad' of personality traits make you more successful? *Science|AAAS*. https://www.sciencemag.org/news/2019/03/does-dark-triad-personality-traits-make-you-more-successful

Adena, M., Alizade, J., Bohner, F., Harke, J., & Mesters, F. (2019). Quality certification for nonprofits, charitable giving, and donor's trust: Experimental evidence. *Journal of Economic Behavior & Organization*, *159*, 75–100. https://doi.org/10.1016/j.jebo.2019.01.007

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/ghzsdc

Bachmann, R., & Inkpen, A. C. (2011). Understanding institutional-based trust building processes in inter-organizational relationships. *Organization Studies*, *32*(2), 281–301. https://doi.org/10/bp6q87

Blanchard, A. L., Welbourne, J. L., & Boughton, M. D. (2011). A model of online trust: The mediating role of norms and sense of virtual community. *Information, Communication & Society*, *14*(1), 76–106. https://doi.org/10/db32d9

Boiral, O. (2012). ISO 9000 and organizational effectiveness: A systematic review. *Quality Management Journal*, *19*(3), 16–37. https://doi.org/10/gk9pn7

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*(4), 550–558. https://doi.org/10.1037/a0018933

Carneiro, C. F., Queiroz, V. G., Moulin, T. C., Carvalho, C. A., Haas, C. B., Rayêe, D., Henshall, D. E., De-Souza, E. A., Espinelli, F., & Boos, F. Z. (2019). *Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature*. bioRxiv. https://doi.org/10.1101/581892

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2018). *Package 'pwr'* (R package version 1.3-0) [Computer software]. https://cran.r-project.org/web/packages/pwr/pwr.pdf

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using nonnaive participants can reduce effect sizes. *Psychological Science*, *26*(7), 1131–1139. https://doi.org/10/f7ktxj

Chivers, T. (2020). Geek tip: If it doesn't say "Registered Report," don't trust it - The Post. *UnHerd*. https://unherd.com/thepost/geek-tip-if-it-doesnt-say-registered-report-dont-trust-it/

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, *11*(4), 464–473. https://doi.org/10/gf92b6

Chow, D. (2021, September 14). More Europeans are taking climate change seriously. In the U.S., not so much. *NBC News*. https://www.nbcnews.com/science/environment/europeans-are-taking-climate-change-seriously-us-not-much-rcna1990

Cobo, E., Cortés, J., Ribera, J. M., Cardellach, F., Selva-O'Callaghan, A., Kostov, B., García, L., Cirugeda, L., Altman, D. G., González, J. A., Sànchez, J. A., Miras, F., Urrutia, A., Fonollosa, V., Rey-Joly, C., & Vilardell, M. (2011). Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: Masked randomised trial. *BMJ*, *343*, Article d6783. https://doi.org/10.1136/bmj.d6783

Cooke, S. J., Nguyen, V. M., Wilson, A. D. M., Donaldson, M. R., Gallagher, A. J., Hammerschlag, N., & Haddaway, N. R. (2016). The need for speed in a crisis discipline: Perspectives on peer-review duration and implications for conservation science. *Endangered Species Research*, *30*, 11–18. https://doi.org/10/f8vtbf

Dohle, S., Wingen, T., & Schreiber, M. (2020). Acceptance and adoption of protective measures during the COVID-19 pandemic: The role of trust in politics and trust in science. *Social Psychological Bulletin*, *15*(4), 1–23. https://doi.org/10.32872/spb.4315

Eisenhart, M. (2002). The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education*, *32*(2), 241–255. https://doi.org/10/fm3v6x

Elmore, S. A. (2018). *Preprints: What role do these have in communicating scientific results?* SAGE.

Elson, M., Huff, M., & Utz, S. (2020). Metascience on peer review: Testing the effects of a study's originality and statistical significance in a field experiment. *Advances in Methods and Practices in Psychological Science*, *3*(1), 53–65. https://doi.org/10.1177/2515245919895419

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*(6), 1231–1236. https://doi.org/10.1016/j.jesp.2011.05.007

Forster, V. (2020, February 2). No, the Coronavirus was not genetically engineered to put pieces of HIV in it. *Forbes*. https://www.forbes.com/sites/victoriaforster/2020/02/02/no-coronavirus-was-not-bioengineered-to-put-pieces-of-hiv-in-it/

Fox, F. (2018). *The preprint dilemma: Good for science, bad for the public?* A discussion paper for the scientific community. Science Media Centre. https://www.sciencemediacentre.org/the-preprint-dilemma-good-for-science-bad-for-the-public-a-discussion-paper-for-the-scientific-community/

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, *19*(4), Article e3000959. https://doi.org/10.1371/journal.pbio.3000959

French, C. (2012, March 15). Precognition studies and the curse of the failed replications. *The Guardian*. https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications

Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review*, *77*(2), 167–187. https://doi.org/10/gc3d6j

German Psychological Society. (2018). *Ethisches Handeln in der Psychologischen Forschung. Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* [Ethical Guidelines for Psychological Research]. Hogrefe Verlag.

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, *336*(6080), 493–496. https://doi.org/10/r2x

Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *JAMA*, *280*(3), 237–240. https://doi.org/10/b5vnc9

Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Annals of Internal Medicine*, *121*(1), 11–21. https://doi.org/10/gk9pph

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, *511*(7508), 220–223. https://doi.org/10/f59dm3

Heesen, R., & Bright, L. K. (2021). Is peer review a good idea? *British Journal for the Philosophy of Science*, *72*(3), 635–663. https://doi.org/10/ggkwvx

Heimstädt, M. (2020). Between fast science and fake news: Preprint servers are political. *Impact of Social Sciences*. https://blogs.lse.ac.uk/impactofsocialsciences/2020/04/03/between-fast-science-and-fake-news-preprint-servers-are-political/

Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. *Elife*, *6*, Article e21718. https://doi.org/10.7554/eLife.21718

Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research. *Public Understanding of Science*, *29*(3), 270–288. https://doi.org/10.1177/0963662520902383

Hoogeveen, S., Sarafoglou, A., & Wagenmakers, E-J. (2020). Laypeople can predict which social-science studies will be replicated successfully. *Advances in Methods and Practices in Psychological Science*, *3*(3), 267–285. https://doi.org/10/gg877f

Huisman, J., & Smits, J. (2017). Duration and quality of the peer review process: The author's perspective. *Scientometrics*, *113*(1), 633–650. https://doi.org/10/gb3b8s

Imhoff, R., & Lamberty, P. (2020). A bioweapon or a hoax? The link between distinct conspiracy beliefs about the Coronavirus disease (COVID-19) outbreak and pandemic behavior. *Social Psychological and Personality Science*, *11*(8), 1110–1118. https://doi.org/10/gg4cq5

Jukola, S. (2017). A social epistemological inquiry into biases in journal peer review. *Perspectives on Science*, *25*(1), 124–148. https://doi.org/10/gk9ppm

Koch, T., Frischlich, L., & Lermer, E. (2021). *The effects of warning labels and social endorsement cues on credibility perceptions of and engagement intentions with fake news*. PsyArXiv. https://doi.org/10.31234/osf.io/fw3zq

Koerber, A. (2021). Is it fake news or is it open science? Science communication in the COVID-19 pandemic. *Journal of Business and Technical Communication*, *35*(1), 22–27. https://doi.org/10/ghcwmd

Kupferschmidt, K. (2020, April 28). How the pandemic made this virologist an unlikely cult figure. *Science|AAAS*. https://www.sciencemag.org/news/2020/04/how-pandemic-made-virologist-unlikely-cult-figure

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nature*, *581*(7807), 130–131. https://doi.org/10/ggvwt6

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10/gbf8nt

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. https://doi.org/10/gdj7s9

Lindner, J. (2020, May 28). Warum über die Drosten-Studie gestritten wird? [Why is the Drosten-study debated?]. *ZDF.de*. https://www.zdf.de/uri/606f8734-2eca-42fe-b811-73b5a6860710

Niggemeier, S. (2020, May 26). Die Machtprobe: Worum es beim Kampf von "Bild" gegen Drosten geht [A test of power: What is the fight between "Bild" and Drosten all about]. *Übermedien*. https://uebermedien.de/49542/die-machtprobe-worum-es-beim-kampf-von-bild-gegen-drosten-geht/

Nishi, A., Shirado, H., Rand, D. G., & Christakis, N. A. (2015). Inequality and visibility of wealth in experimental social networks. *Nature*, *526*(7573), 426–429. https://doi.org/10.1038/nature15392

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*(3), 217–243. https://doi.org/10/gcsk27

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *JAMA*, *316*(12), 1315–1316. https://doi.org/10/gh7f7z

Olson, C. M., Rennie, D., Cook, D., Dickersin, K., Flanagin, A., Hogan, J. W., Zhu, Q., Reiling, J., & Pace, B. (2002). Publication bias in editorial decision making. *JAMA*, *287*(21), 2825–2828. https://doi.org/10/b26cf3

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

Phillips, N. (2017). *Yarrr: A companion to the e-book "yarrr!: The pirate's guide to r"* (R package version 0.1.5) [Computer software]. https://CRAN.R-project.org/package=yarrr

Polka, J. K., Dey, G., Pálfy, M., Nanni, F., Brierley, L., Fraser, N., & Coates, J. A. (2021). *Preprints in motion: Tracking changes between posting and journal publication*. bioRxiv. https://doi.org/10/gh5mhm

Rahal, R-M., & Heycke, T. (2020). *Hoarding in science, no thanks. Openness and transparency in crisis mode and beyond*. OSF Preprints. https://doi.org/10.31222/osf.io/akd6c

R Development Core Team. (2021). *R: A language and environment for statistical computing*. https://www.R-project.org/

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, *9*(2), 79–101. https://doi.org/10/c3qhkb

Revelle, W. R. (2021). *psych: Procedures for personality and psychological research* (R package version 2.1.6) [Computer software]. https://CRAN.R-project.org/package=psych

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. https://doi.org/10.18637/jss.v048.i02

Schroter, S., Black, N., Evans, S., Carpenter, J., Godlee, F., & Smith, R. (2004). Effects of training on quality of peer review: Randomised controlled trial. *BMJ*, *328*(7441), Article 673. https://doi.org/10/d52tpr

Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, *338*(6107), 682–685. https://doi.org/10/f4ccsq

Sheldon, T. (2018). Preprints could promote confusion and distortion. *Nature*, *559*(7714), 445–446. https://doi.org/10.1038/d41586-018-05789-4

Silva, R. R., & Topolinski, S. (2018). My username is IN! The influence of inward vs. outward wandering usernames on judgments of online seller trustworthiness. *Psychology & Marketing*, *35*(4), 307–319. https://doi.org/10/gk9pp3

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10/gcmvgf

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, *99*(4), 178–182.

Soderberg, C. K., Errington, T. M., & Nosek, B. A. (2020). Credibility of preprints: An interdisciplinary survey of researchers. *Royal Society Open Science*, *7*(10), Article 201520. https://doi.org/10.1098/rsos.201520

Tenopir, C., Levine, K., Allard, S., Christian, L., Volentine, R., Boehm, R., Nichols, F., Nicholas, D., Jamali, H. R., & Herman, E. (2016). Trustworthiness and authority of scholarly information in a digital age: Results of an international questionnaire. *Journal of the Association for Information Science and Technology*, *67*(10), 2344–2361. https://doi.org/10/gfs4nn

Torchiano, M. (2020). *effsize: Efficient effect size computation* (R package version 0.8.1) [Computer software]. https://doi.org/10.5281/zenodo.1480624

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, USA*, *113*(23), 6454–6459. https://doi.org/10/f8qqzh

Vazire, S. (2020, June 25). Peer-reviewed scientific journals don't really do their job. *Wired*. https://www.wired.com/story/peer-reviewed-scientific-journals-dont-really-do-their-job/

Wenegrat, B., Castillo-Yee, E., & Abrams, L. (1996). Social norm compliance as a signaling system. II. Studies of fitness-related attributions consequent on a group norm violation. *Ethology and Sociobiology*, *17*(6), 417–429.

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. L., & Shaked, A. (2014). Just think: The challenges of the disengaged mind. *Science*, *345*(6192), 75–77. https://doi.org/10/th9

Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science*, *11*(4), 454–463. https://doi.org/10.1177/1948550619877412