Editor's Note

In the article that follows this Editor's Note, Crüwell and colleagues report the results of an audit of the computational reproducibility of the 14 research articles published in the April 2019 issue of *Psychological Science* (Vol. 30, Issue 4). The audit was author-initiated—it was not by invitation of the journal. Crüwell and colleagues defined computational reproducibility as "the ability to recreate results using the original data and code (or at least a detailed description of the analyses)" (p. 514). They selected Volume 30, Issue 4 because it was the first in which all of the research articles were awarded the Open Data badge. Of the 14 research articles in the issue, Crüwell et al. assessed only one as meeting the requirements for the Open Data badge.

In their assessment, Crüwell and colleagues relied on the criteria provided in the Submission Guidelines of the journal at the time the 2019 authors submitted their articles (*Psychological Science*, Submission Guidelines, Open Science Badges section). The guidelines state that authors may receive an "Open Data badge for making publicly available the digitally shareable data necessary to reproduce the reported result. This includes annotated copies of the code or syntax used for all exploratory and principal analyses." In their judgments regarding Open Data badge eligibility, Crüwell and colleagues emphasized the availability of analysis code or syntax. Importantly, neither the Open Science Framework (OSF) criteria that guide badge eligibility nor the Open Practices Disclosure (OPD) form completed by the 2019 authors makes explicit reference to analysis code or syntax. The OSF criteria state that "The Open Data badge is awarded when digitally-shareable data necessary to reproduce the reported results are publicly available," and that "A data dictionary (for example, a codebook or metadata describing the data) is included with sufficient description for an independent researcher to reproduce the reported analyses and results" (https://osf.io/tvyxz/wiki/1.%20View%20the%20 Badges/). Similarly, the OPD form that the authors completed required them to "Confirm that there is sufficient information for an independent researcher to reproduce **all of the reported results**, including codebook if relevant" (emphasis in original). Neither set of criteria specifies sharing of analysis code or syntax.

The difference between the Submission Guidelines and the OPD form that authors completed is important. The Submission Guidelines provide advice, but they are not the rule of law. The rule of law is established in the OPD form, which at the time the 2019 authors completed it, made no mention of analysis code. I emphasize this point because it establishes that the 2019 authors did not openly flaunt explicit criteria when they applied for the Open Data badge, and nor did eligibility for the badge turn on provision of analysis code, as established either by *Psychological Science* or OSF.

On behalf of *Psychological Science*, I apologize for the discrepancy between the Open Data badge elements listed in the Submission Guidelines and the less explicit requirement in the OPD form. The criteria outlined in the OPD form were not sufficiently explicit regarding the elements that should be included in an open-access registry in order to ensure independent reproducibility. We have changed the wording of the OPD form such that it now provides better guidance to authors in their efforts to make their science open by making their data publicly available.

Setting aside for the moment the vagueness of the requirements of the previous version of the OPD form, it is clear that it instructed authors to provide "sufficient information for an independent researcher to reproduce all of the reported results." The OSF eligibility criteria give the same charge. By their report, for several of the articles published in the April 2019 issue, the audit team of Crüwell and colleagues was not successful in achieving the goal of independent reproduction of all of the reported results based on the information in the registry alone, with the methods they employed. Importantly, ensuring that analyses can be reproduced is only one of several possible motivations for authors to make their data openly accessible. Other possible motivations include reducing the need for duplicative data-collection efforts; facilitating collaborations; and even enabling analysis of data in different ways, thus helping to ensure findings are robust to different analytic approaches, to name a few. I venture to guess that it was goals such as these, not independent reproducibility alone, that were paramount in the minds of many of the 2019 authors as they made their data publicly available.

Critically, transparency and scientific community building are not mutually exclusive goals. In this regard, it is my pleasure to report that upon learning of the work of Crüwell and colleagues, several of the author groups with articles in Volume 30, Issue 4 of *Psychological Science* appended their registries to include elements identified in the audit as missing or insufficient. I appreciate the positive response of these author groups and their ongoing contributions to open science.

Patricia J. Bauer
Editor in Chief

What's in a Badge? A Computational Reproducibility Investigation of the Open Data Badge Policy in One Issue of *Psychological Science*

Sophia Crüwell[1,2] , Deborah Apthorp[3,4] , Bradley J. Baker[5] ,
Lincoln Colling[6] , Malte Elson[7,8] , Sandra J. Geiger[9] ,
Sebastian Lobentanzer[10], Jean Monéger[11,12] , Alex Patterson[13] ,
D. Samuel Schwarzkopf[14,15], Mirela Zaneva[16] , and
Nicholas J. L. Brown[17]

[1]Meta-Research Innovation Center Berlin (METRIC-B), QUEST Center for Transforming Biomedical Research,
Berlin Institute of Health, Charité – Universitätsmedizin Berlin; [2]Department of History and Philosophy of Science,
University of Cambridge; [3]School of Psychology, University of New England; [4]School of Computing, Australian
National University; [5]Department of Sport and Recreation Management, Temple University; [6]School of Psychology,
University of Sussex; [7]Faculty of Psychology, Ruhr University Bochum; [8]Horst Görtz Institute for IT Security, Ruhr
University Bochum; [9]Environmental Psychology, Department of Cognition, Emotion, and Methods, Faculty of
Psychology, University of Vienna; [10]Institute for Computational Biomedicine, University Hospital Heidelberg,
Germany; [11]Department of Psychology, University of Poitiers; [12]Research Center on Cognition and Learning, Centre
National de la Recherche Scientifique (CNRS) 7295; [13]Sheffield Methods Institute, The University of Sheffield;
[14]School of Optometry and Vision Science, University of Auckland; [15]Experimental Psychology, University College
London; [16]Department of Experimental Psychology, University of Oxford; and [17]Department of Psychology,
Linnaeus University

Abstract

In April 2019, *Psychological Science* published its first issue in which all Research Articles received the Open Data badge. We used that issue to investigate the effectiveness of this badge, focusing on the adherence to its aim at *Psychological Science*: sharing both data and code to ensure reproducibility of results. Twelve researchers of varying experience levels attempted to reproduce the results of the empirical articles in the target issue (at least three researchers per article). We found that all 14 articles provided at least some data and six provided analysis code, but only one article was rated to be exactly reproducible, and three were rated as essentially reproducible with minor deviations. We suggest that researchers should be encouraged to adhere to the higher standard in force at *Psychological Science*. Moreover, a check of reproducibility during peer review may be preferable to the disclosure method of awarding badges.

Open science badges are incentives for researchers to participate in open science practices such as preregistration and sharing of data and materials. Sharing data is encouraged in order to increase transparency, reuse

or reproducibility, and citations (Colavizza et al., 2020; Piwowar & Vision, 2013). *Psychological Science* adopted

Corresponding Author:
Sophia Crüwell, Meta-Research Innovation Center Berlin (METRIC-B),
QUEST Center for Transforming Biomedical Research, Berlin Institute
of Health, Charité – Universitätsmedizin Berlin
Email: sophia.cruewell@bih-charite.de

the badges in 2014 (Eich, 2014), and, in April 2019, published its first issue in which all 14 Research Articles received Open Data badges (Volume 30, Issue 4). The aim of this badge is to incentivize authors to share online the data necessary to reproduce the reported results (Blohowiak et al., 2022). *Psychological Science*'s submission guidelines state that articles may receive this badge "for making publicly available the digitally shareable data necessary to reproduce the reported result. This includes annotated copies of the code or syntax used for all exploratory and principal analyses" (Psychological Science, 2022, Open Practices Badges section; these eligibility criteria were operative in 2019).[1] The corresponding Open Practices Disclosure form uses somewhat more permissive language, requiring confirmation of "sufficient information for an independent researcher to reproduce all of the reported results." This equates to provision of analysis code or syntax for all but the simplest analyses and data sets. We understand *reproducibility* to mean computational reproducibility: the ability to recreate results using the original data and code (or at least a detailed description of the analyses). *Psychological Science* awards badges based on the *disclosure method*: Authors complete an Open Practices Disclosure form, and the journal may confirm the existence of data, materials, or a preregistration (Blohowiak et al., 2022; Psychological Science, 2022).

Kidwell et al. (2016) found that introducing badges at *Psychological Science* led to an increase in sharing, which indicates the superficial success of this policy—particularly compared with other initiatives (see Rowhani-Farid & Barnett, 2018, and Rowhani-Farid et al., 2020, who found lower and no increase in data sharing at *Biostatistics* and *BMJ Open*, respectively). Hardwicke et al. (2021) investigated the analytic reproducibility of articles that received Open Data badges at *Psychological Science* between 2014 and 2015; they were able to reproduce the results of 36% of articles without author involvement and a further 24% with author involvement. Obels et al. (2020) examined data sharing and computational reproducibility of registered reports in general psychological research; 36 of the 62 articles assessed (58%) provided both data and code, of which 21 (58%) were computationally reproducible.

Whereas Hardwicke et al. (2021) and Obels et al. (2020) were concerned with computational or analytic reproducibility per se, we focused on computational reproducibility as a measure of the effectiveness of the *Psychological Science* Open Data badge policy. If this policy was effective, the results in the April 2019 issue should be independently and precisely reproducible. If these results are wholly or partially irreproducible, then any issues we identify during reproduction attempts may inform the improvement of the policy at *Psychological Science* and other journals. Our focus on

Statement of Relevance

Open science badges are incentives for encouraging researchers to participate in open science practices such as preregistration and the sharing of data or experimental materials. These practices are thought to be desirable as a means for enhancing both transparency and reproducibility, which are important to scientific inquiry. In particular, the results of a study should be at least computationally reproducible using the same data and analyses. In the present study, we aimed specifically to investigate the effectiveness of the Open Data badge at *Psychological Science*, the stated purpose of which is to ensure the reproducibility of results. We found that the Open Data badge policy did not work as intended, and we suggest possible changes in how the badge could be awarded. We hope to contribute to improving the badge program at *Psychological Science* as well as reproducibility and transparency in psychology.

one practice in one issue of *Psychological Science* allows for in-depth examination of the effectiveness of this specific measure for incentivizing data sharing as implemented and advertised at this journal.

Open Practices Statement

The individual and summary reports, as well as the informal reproducibility ratings and code to create Tables 1 and 2, are publicly accessible at https://osf.io/xzke7/. This study was not preregistered.

Method

Sample

The scope of our investigation was all 14 Research Articles published in the April 2019 issue of *Psychological Science*, the journal's first issue in which all Research Articles were awarded the Open Data badge (Bae & Luck, 2019; Dorfman et al., 2019; Garcia & Rimé, 2019; Geniole et al., 2019; Hakim et al., 2019; Hilgard et al., 2019; Johnson & Wilson, 2019; Lindsay et al., 2019; Obaidi et al., 2019; Olsson-Collentine et al., 2019; Vardy & Atkinson, 2019; Wójcik et al., 2019; Woolley & Fishbach, 2019; Yousif & Keil, 2019). To emphasize our focus on *Psychological Science*'s Open Data badge policy and not these individual articles, we will refer to them as Articles 101 to 114, the numbers having been randomly assigned. A superficial examination of the repositories linked to the articles shows that all articles are associated

Table 1. Initial Ratings: Reproducers' Ratings of Their Initial Reproduction Attempts for Each Article

Article	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Modal rating
101	Partially	Essentially	Partially			Partially
102	Partially	Essentially	Partially	Essentially	Partially	Partially
103	Partially	Partially	Partially			Partially
104	Partially	Partially	Essentially			Partially
105	Not at all	Not at all	Not at all			Not at all
106	Not at all	Mostly not	Not at all			Not at all
107	Not at all	Not at all	Not at all	Not at all		Not at all
108	Partially	Exactly	Exactly			Exactly
109	Partially	Partially	Essentially			Partially
110	Mostly not	Essentially	Mostly not			Mostly not
111	Essentially	Essentially	Essentially			Essentially
112	Mostly not	Not at all	Mostly not			Mostly not
113	Partially	Partially	Partially			Partially
114	Partially	Essentially	Essentially	Essentially		Essentially

with at least some data. No code is provided in the linked repository for six of the articles (Articles 101, 105, 107, 111, 112, and 113).

Design

This is an observational, descriptive, one-group study. We did not compare the April 2019 issue of *Psychological Science* with any other issue or journal but rather to the ideal of the policy of the Open Data badge as implemented at *Psychological Science*.

In the present study, we were mainly concerned with this Open Data badge policy's effectiveness, not with reproducibility per se. Our informal reproducibility ratings are a proxy measure of that effectiveness. Although

we did not establish any criteria for successful reproduction in advance, for a study to count as reproducible, its results should at least be reproducible by a competent external researcher (National Academies of Sciences, Engineering, and Medicine, 2019), such as a PhD student with some experience and training in a similar field. When we say that a study was or was not reproducible, this is specific to our team of reproducers. Our informal reproducibility rating items were "exactly reproducible," which represented the ideal of the Open Data badge in which there were no deviations from the reported results; "essentially reproducible," meaning that there were minor deviations in the decimals or obvious typographical errors (e.g., 2.39 vs. 2.93); "partially reproducible," indicated that there were more than minor deviations but the

Table 2. Summary Ratings: Reproducers' Ratings of the Group's Reproduction Attempts for Each Article

Article	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Modal Rating
101	Essentially	Essentially	Partially			Essentially
102	Partially	Essentially	Partially	Partially	Partially	Partially
103	Partially	Partially	Partially			Partially
104	Partially	Partially	Partially			Partially
105	Not at all	Not at all	Not at all			Not at all
106	Mostly not	Not at all	Not at all			Not at all
107	Not at all	Mostly not	Not at all	Not at all		Not at all
108	Exactly	Exactly	Exactly			Exactly
109	Partially	Essentially	Essentially			Essentially
110	Partially	Partially	Partially			Partially
111	Essentially	Essentially	Essentially			Essentially
112	Mostly not	Not at all	Mostly not			Mostly not
113	Partially	Partially	Partially			Partially
114	Partially	Essentially	Partially	Essentially		Partially

results were mostly numerically consistent; "mostly not reproducible," meaning that there were major deviations and few numerically consistent results; and "not at all reproducible" if there was no numerical consistency between the reported results and the ones that we found, or a reproduction attempt was otherwise not possible.

Procedure

Reproducer assignment. The last author initially recruited 13 researchers of varying experience and career levels to attempt to reproduce studies from the April 2019 issue of *Psychological Science* on the basis of the data and, where available, code shared by the original authors. They were asked to indicate their ability to access and use four software packages: Excel, MATLAB, R, and SPSS. Each reproducer was asked to attempt to reproduce four of the 14 articles, the selection being determined by (a) the match between the reproducer's access to software and the format of the code or data provided by the original authors, and (b) the aim to have distinct sets of researchers working on each article, where possible. Because of an error in the assignment process, two reproducers (J.M. and S.L.) were asked to reproduce the same four articles. No two articles were reproduced by the exact same set of researchers. Two reproducers dropped out and did not complete any reproduction reports. Furthermore, reproducers were unable to complete individual reproduction attempts because of technical limitations in three cases (B. J. B., Article 106; S. C., Article 110; S. J. G., Article 112). One further reproducer joined the project at a later stage. In total, 12 reproducers completed three to five reproductions each. For each of the 14 articles, at least three researchers were assigned to, and completed, individual reproduction reports (46 individual reports in total).

Reproduction process. The reproduction process was split into two stages. In the first stage, each researcher independently attempted to reproduce their assigned studies and wrote an individual reproduction report on their experience and findings. These initial reports were unstructured; some reproducers included further information such as code, whereas others focused on the narrative report of their reproduction attempts. Results were initially not shared, and reproducers were encouraged to stay as masked as possible (i.e., not discussing results with other reproducers until their own analyses were completed). In the second stage, on the basis of the individual reports, the groups of reproducers for each article agreed on a summary report of their overall findings. After the reproduction process, they rated the reproducibility of each article they had attempted to reproduce on the basis of (a) their individual, initial experience reproducing the article and

(b) the summary findings and discussions among the group for each article.

All of our reproduction attempts were carried out independently of the articles' original authors. We then contacted the authors prior to preprinting and submission to explain the nature of the project; all our analyses and conclusions were finalized by that point. In the case of two articles, the last author of the present article had previously (i.e., before the other coauthors joined the project in May 2020) contacted the corresponding authors for reproduction advice before realizing that this was not compatible with the overarching aim of the project. Consequently, he did not write an individual report on these articles, and he did not contribute to the associated group discussions.

Results

Reproducibility

Only one of the 14 articles was rated to be exactly reproducible (Article 108), and three further articles were rated essentially reproducible with minor deviations by a majority of the researchers who reproduced them, on the basis of the summary reports (Articles 101, 109, and 111). Both the initial reproducibility ratings based on the individual reproduction attempts (Table 1) and the summary ratings based on the article group's combined reproduction attempts (Table 2) varied, and there were four changes between the modal majority-agreed initial and summary ratings (Articles 101, 109, 110, and 114).

The individual reports (46 total) and summary reports (14 total) are available on the OSF alongside further information about each reproduced article (see https://osf.io/xzke7/). The reports provide in-depth qualitative and quantitative information in the form of narrative descriptions of each reproduction attempt, often including numerical results.

Issues encountered

The following section qualitatively and nonexhaustively summarizes the issues that we encountered (for a further summary of the shared data and code, see Table 3). General issues include (a) a lack of documentation of data and/or code; (b) minor discrepancies in several results, likely due to use of random numbers without fixed seeds in bootstrapped analyses; (c) minor discrepancies in individual results, likely due to typographical or copy-paste errors; (d) unclear reporting of procedures in the article text, including the criteria for inclusion in subgroups, lack of or incorrect reporting of the variables used for regression models, and

Table 3. Summary of the Results Reported in the Summary Reproduction Reports for Each Article

Article	Results (summary rating)	Analytic code	Data	Readme file	Variable key	Other
101	Essentially reproducible	Missing	Postprocessed provided	Missing	Missing	Missing data for one experiment
102	Partially reproducible	Provided	Postprocessed provided	Missing	Missing	Inconsistencies in data from what was reported in article
103	Partially reproducible	Provided	Raw provided	Missing	Missing	Broken GitHub links, key file not linked to in repository
104	Partially reproducible	Provided	Postprocessed provided	Provided	Provided	Different reproducers had different issues running code
105	Not reproducible	Missing	Postprocessed Provided	Missing	Provided	Data for Supplemental Material were missing
106	Not reproducible	Insufficient	Raw provided	Missing	Missing	Required extra MATLAB packages
107	Not reproducible	Missing	Raw provided	Provided	Provided	Insufficient information
108	Exactly reproducible	Provided	Raw provided	Missing	Provided	Package dependency issues
109	Essentially reproducible	Provided	Postprocessed provided	Missing	Missing	Unclear whether data were raw or postprocessed
110	Partially reproducible	Insufficient	Raw provided	Missing	Missing	Corrupt data/unable to download data
111	Essentially reproducible	Missing	Postprocessed provided	Missing	Missing	Preregistration discrepancies
112	Mostly not reproducible	Insufficient	Postprocessed Provided	Provided	Provided	Required extra MATLAB packages
113	Partially reproducible	Missing	Postprocessed provided	Missing	Missing	Unclear variable identification
114	Partially reproducible	Provided	Postprocessed provided	Missing	Provided	Corrupt data/unable to download data

unreported one-sided analyses; (e) data storage issues on the OSF, including files being either corrupt or not downloadable at all (Article 110); and (f) ambiguous labeling of studies in the article's Open Practices statement (Article 109). Data-specific issues include (a) provision of cleaned data without raw data, (b) provision of raw data without cleaned data, and (c) no description of, or code for, the data-cleaning process. Code-specific issues include (a) a lack of shared analysis code or modeling code and (b) issues with package or software versions (often resolvable but sometimes only with considerable effort).

Open Data badge eligibility

Overall, we found that eight articles (Articles 101, 105, 106, 107, 110, 111, 112, and 113) did not provide, even in principle, sufficient information for independent exact reproduction of their results by our team. In these cases, reproduction would require analysis code or syntax, as the descriptions of the methodology and the shared data files did not provide enough information on their own.[2] This means that (a) these articles did not meet the standard for receiving the Open Data badge at *Psychological Science* according to the explicit requirements stated in the submission guidelines, and (b) the authors of these articles may have interpreted the less explicit requirements of the Open Practices Disclosure statement in a rather minimalist way.

Provision of both analysis code and data was a requirement for the award of an Open Data badge at *Psychological Science* at the time of submission, according to the explicit requirements stated in the submission guidelines. These requirements appear to not have been met in these cases. Articles missed these explicit requirements of the journal submission guidelines to different extents. Six

articles (Articles 101, 105, 107, 111, 112, and 113) did not provide any code in the linked repository (some modeling code was provided for Article 112 on a separate GitHub page not linked to from the article), and Article 101 additionally provided only summarized and incomplete data. Therefore, these articles do not appear to have met the requirements for receiving the Open Data badge, according to the explicit requirements in the submission guidelines that were in force at *Psychological Science* when the articles were first submitted. Arguably, given this stipulation, Articles 106 and 110 were also not eligible for the Open Data badge because they provided some code files but not the statistical analysis code. This field-leading policy was certainly introduced and implemented with the best of intentions, but there appear to have been some oversights by the journal in its execution, as the OSF guidelines recommend at least a cursory check by the journal before the badge is awarded.

On top of these clearer eligibility issues regarding the provision of sufficient information and/or analysis code for independent exact reproduction, on a strict interpretation of the badge eligibility criteria at *Psychological Science*, our reproduction results arguably imply that only one of the 14 articles met the requirements for an Open Data badge. Eight articles did not share both data and analysis code or otherwise sufficient information, and of the remaining six articles that did attempt to share sufficient information for independent reproduction in the form of analysis code, only one was exactly reproducible by our team. However, the reproducibility of the articles that shared data and analysis code likely decreased since publication (because of issues such as "software rot"; Hinsen, 2019). Therefore, it is unclear how we can make an inference from current reproducibility to past Open Data badge eligibility in the case of the articles that share both data and analysis code but were not exactly reproducible.

Discussion

The disclosure method did not ensure the required higher standard for the Open Data badge at *Psychological Science*, at least in its April 2019 issue. Of 14 articles, eight did not share both data and analysis code and so failed to meet the eligibility requirements. Of the remaining six, only one was exactly reproducible, but we do not know whether the other five were exactly reproducible at the time of submission. We make several recommendations for improving the specific badge policy at *Psychological Science* and comparable initiatives at other journals (for further general recommendations on improving data sharing and computational reproducibility, see Stodden et al., 2016; Trisovic et al.,

2022; Wilson et al., 2017). Excellent and more in-depth recommendations and tutorials for authors to ensure that their shared data and code are eligible for an Open Data badge are provided by, for example, Arslan (2019), Eberle (2022), Klein et al. (2018), Levenstein and Lyle (2018), Peikert and Brandmaier (2021), and Van Lissa et al. (2021). Moreover, the provision of further incentives, in particular by funding agencies and institutions, may help make data sharing more common and effective (Houtkoop et al., 2018).

First, authors wanting to share their data and code could take further steps to ensure eligibility for an Open Data badge. It might be argued that the average psychology researcher lacks the necessary technical skills. Any journal offering open science badges could support its authors in making their data and code reproducible and usable by providing guidance on (a) documentation of data, code, and the online repository; (b) sharing the rawest possible data (within ethical and logistical limits) alongside the cleaned data; and (c) guidance on recommendations for avoiding dependency and version issues (e.g., by using a platform such as Docker or Code Ocean; Clyburne-Sherin et al., 2019; Nüst et al., 2020; or if working in R by using, e.g., *groundhog* or *renv*; Simonsohn & Gruson, 2022; Ushey, 2022). There are many resources for making a reproducible workflow accessible, particularly concerning data and code sharing (see above). Authors can also ensure machine-actionable reusability of their data by following the findable, accessible, interoperable, and reusable (FAIR) guidelines (Wilkinson et al., 2016). It is commendable when authors attempt to share their data—data and code imperfectly shared are typically better than data and code perfectly kept to oneself. Indeed, our study would have been impossible without the introduction of the Open Data badge. The badge is a step in the right direction, but the corresponding policy needs to be improved to better support and incentivize transparent and reproducible research.

Second, there are improvements that could be made by badge-awarding journals that require both data and code for Open Data badge eligibility. If such journals rely on the disclosure method over the peer-review method, they could better describe the specific badge criteria and clarify that code, syntax, or a detailed analysis description needs to be shared alongside the data—for example, as required by the submission guidelines at *Psychological Science*. Many journals, and the baseline open science badge guidelines (Blohowiak et al., 2022), do not explicitly include the sharing of analysis code as an eligibility criterion; whether they should do so depends on the purpose of the Open Data badge. If the purpose is data reusability, not sharing code may be acceptable. If the purpose includes reproducibility,

however, code should always be included. This particularly applies to complex analyses, as verbal descriptions are unlikely to cover the information necessary for exact or essential reproduction (as demonstrated by our difficulties reproducing Article 112; see also Seibold et al., 2021). In simpler cases, not sharing code might seem acceptable (e.g., we essentially reproduced Article 111), but verbal reports can still fail, and sharing of analysis code ensures that all relevant information is available. By requiring the sharing of analysis code, *Psychological Science* is going beyond the basic requirements of the Open Data badge in order to achieve both reusability and reproducibility. Nevertheless, we still found that insufficient code was in fact shared for more than half of the examined articles. Badge-awarding journals requiring not only data but also code could more explicitly require authors to provide working code—where necessary—that enables straightforward reproducibility and produces clearly annotated output (see Bauer, 2022, for a reaffirmation of this requirement).

Third, it may be sensible to focus on other methods of awarding the open science badges. Given our results, as well as those of Hardwicke et al. (2021), a badge check may be needed as part of peer review at badge-awarding journals, including *Psychological Science.* This provides earlier verification and allows authors to upload all materials before publication and award of the badges. One way of doing this is to move to the peer-review method of awarding the Open Data badge (as opposed to the disclosure method; Blohowiak et al., 2022). The standard required by the peer-review method is open to interpretation by the specific journal: For the Open Data badge, this could range from a formal but brief review of the materials to independent reproduction of the reported results.[3] The expected standard should match up with the standard stated in the submission guidelines; in the case of *Psychological Science*, data and code are already nominally required to enable precise or exact reproducibility, at least at the time of submission (Psychological Science, 2022). This work could be done by peer reviewers, dedicated badge reviewers, editors, or dedicated editorial staff (Blohowiak et al., 2022) and should be as straightforward as running the code or scripts on the data and requiring corrections if this does not lead to an exact reproduction. A checkbox could be provided for reviewers or dedicated badge reviewers to confirm that they executed the code successfully. If the analysis methods are complex or time consuming, then it should be incumbent on the authors to provide appropriate tools and assistance to the reviewers. If this responsibility is made clear to researchers before submission, this can incentivize more straightforwardly reproducible research. Alternatively, authors could provide proof of

a successful reproduction attempt, either independently or from within the research team (which would be an improvement, as analyses are commonly carried out by single team members; Veldkamp et al., 2014).[4] This could be a condition for the award of the badge, or for an alternative Open Data+ badge, similar to the existing Preregistered+ badge (Blohowiak et al., 2022). Another approach would be to break the badge down into checkboxes of what was shared (e.g., raw and/or processed data, full or partial analysis code), thereby both lowering the threshold for participation and increasing transparency and usefulness of the badge.[5] Regardless, whether authors fill in their disclosure items appropriately should continue to be monitored—a recent study found low adherence even to mandatory data availability statements in biomedical research manuscripts (Gabelica et al., 2022).

Limitations

The focus of our study was limited to the April 2019 issue of *Psychological Science*, a nonrandom sample of all articles in *Psychological Science* that received an Open Data badge. An advantage of this approach was that we could investigate each article in more depth than would be feasible for a larger sample, resulting in 46 individual reports in total, at least three per article. In comparison, Hardwicke et al. (2021) focused only on the numerical results of a subset of substantive findings for each article, meaning that reproducibility was not as fully evaluated as in our study. Our rich qualitative and quantitative results can be a starting point for further investigation. Building on our reproduction experiences may allow us to better anticipate the roadblocks that reproducers will face.

A possible limitation of our focus is that data-sharing practices may have improved overall since the publication of the issue under investigation. However, our results show only a slight improvement over those found by Hardwicke et al. (2021), who looked at articles published between 2014 and 2015 (using their less strict definition of reproducibility, equivalent to our "essential" reproduction). The Open Data badge eligibility criteria have not substantially changed since, so there is no reason to believe that a more current issue would show substantial improvement in a shorter time frame. Specifically, the eligibility criteria for the award of an Open Data badge at *Psychological Science* have included sharing of the relevant analysis code since at least November 2017 (Psychological Science, 2017).

Where reproducers had to recreate all or part of the analyses, our reproduction attempts may not be correct. This can result from unclear reporting or a lack of code (or other issues, identified above) but also from

a reproducer's expertise and evolving abilities as a researcher. However, we believe that competent graduate students should be able to reproduce the results of an article with an Open Data badge in their field of training. For an article that was awarded the Open Data badge at *Psychological Science*, reproduction should simply be a matter of running the code on the data.

An advantage of publicly shared data—over data unshared or available "on request"—is that they are available, and ideally useful, without the original authors' involvement. Contacting authors is not always easy: Researchers change institutions or email addresses and are mortal. Sometimes authors refuse to share data, even if required by the journal. Stodden et al. (2018) assessed the effectiveness of a policy of mandatory sharing on request at the journal *Science* and found that, despite this policy, they received data for only 44% of articles. Hence, the independence of the reproduction attempts in our study is one of its strengths. Doubtless we could have exactly or essentially reproduced more articles by contacting the original authors. We did not do this, as we wanted to investigate the effectiveness of the specific Open Data badge policy at *Psychological Science*, not the analytic or computational reproducibility of individual studies. The out-of-the-box reproducibility of each article indicates that effectiveness—if a successful reproduction requires contacting the authors, the badge was unsuccessful.

Conclusion

Recent advances in open and reproducible science have been rapid, and associated journal policies are constantly improving (see *Psychological Science*'s move to Transparency and Openness Promotion [TOP] guidelines Level 2; Bauer, 2022). The stopgap, however, cannot be to award Open Data badges to articles that do not meet the minimum criteria. This study provides insight into the importance of sharing data for reproducibility and reuse as well as into the experience of reproducing studies that received the Open Data badge. We hope it can motivate improvements of the Open Data badge policy, or its implementation by the authors, at *Psychological Science* and other journals committed to promoting open science.

Transparency

ORCID iDs

Sophia Crüwell https://orcid.org/0000-0003-4178-5820
Deborah Apthorp https://orcid.org/0000-0001-5785-024X
Bradley J. Baker https://orcid.org/0000-0002-1697-4198
Lincoln Colling https://orcid.org/0000-0002-3572-7758
Malte Elson https://orcid.org/0000-0001-7806-9583
Sandra J. Geiger https://orcid.org/0000-0002-3262-5609
Jean Monéger https://orcid.org/0000-0003-1178-1896
Alex Patterson https://orcid.org/0000-0002-7780-4192
Mirela Zaneva https://orcid.org/0000-0003-3569-931X
Nicholas J. L. Brown https://orcid.org/0000-0003-1579-0730

Notes

1. We carried out our study on the basis of the clear requirements stated in the submission guidelines.
2. According to the arguably ambiguous requirements as phrased in the Open Practices Disclosure (but not according to the explicit requirements in the submission guidelines), Article 111 may be considered eligible for an Open Data badge even without code. However, although the analyses were straightforward and *mostly* described, they were not sufficiently described to enable *exact* independent reproduction by our team.

3. Note that although these standards are compatible with the Transparency and Openness Promotion (TOP) guideline levels, they are not necessarily equivalent. Thus, a journal operating on TOP Levels 0 to 2 may introduce a version of the peer-review method of awarding Open Data badges without thereby moving to TOP Level 3. Adopting TOP Level 3 is more restrictive than moving only to the peer-review badge-awarding method.

4. We thank Daniel Simons and an anonymous reviewer for this suggestion.

5. We thank a further anonymous reviewer for this suggestion.

References

Arslan, R. C. (2019). How to automatically document data with the *codebook* package to facilitate data reuse. *Advances in Methods and Practices in Psychological Science*, *2*(2), 169–187. https://doi.org/10.1177/2515245919838783

Bae, G., & Luck, S. J. (2019). Reactivation of previous experiences in a working memory task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

Bauer, P. J. (2022). Psychological science stepping up a level. *Psychological Science*, *33*(2), 179–183. https://doi.org/10.1177/09567976221078527

Blohowiak, B. B., Cohoon, J., de-Wit, L., Eich, E., Farach, F. J., Hasselman, F., Holcombe, A. O., Humphreys, M., Lewis, M., Nosek, B. A., Peirce, J., Spies, J. R., Seto, C., Bowman, S., Green, D., Nilsonne, G., Grahe, J., Wykstra, S., Hofelich Mohr, A., . . . Lowrey, O. (2022, February 4). *Badges to acknowledge open practices*. OSF. https://osf.io/tvyxz

Clyburne-Sherin, A., Fei, X., & Green, S. A. (2019). Computational reproducibility via containers in psychology. *Meta-Psychology*, *3*, Article MP.2018.892. https://doi.org/10.15626/MP.2018.892

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, *15*(4), Article e0230416. https://doi.org/10.1371/journal.pone.0230416

Dorfman, H. M., Bhui, R., Hughes, B. L., & Gershman, S. J. (2019). Causal inference about good and bad outcomes. *Psychological Science*, *30*(4), 516–525. https://doi.org/10.1177/0956797619828724

Eberle, J. W. (2022). *Improving the computational reproducibility of clinical science: Tools for open data and code*. PsyArXiv. https://doi.org/10.31234/osf.io/bf28t

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1), 3–6. https://doi.org/10.1177/0956797613512465

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological Science*, *30*(4), 617–628. https://doi.org/10.1177/0956797619831964

Geniole, S. N., Procyshyn, T. L., Marley, N., Ortiz, T. L., Bird, B. M., Marcellus, A. L., Welker, K. M., Bonin, P. L., Goldfarb, B., Watson, N. V., & Carré, J. M. (2019). Using a psychopharmacogenetic approach to identify the pathways through which—and the people for whom—testosterone promotes aggression. *Psychological Science*, *30*(4), 481–494. https://doi.org/10.1177/0956797619826970

Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. *Psychological Science*, *30*(4), 526–540. https://doi.org/10.1177/0956797619830384

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (2021). Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: An observational study. *Royal Society Open Science*, *8*(1), Article 201494. https://doi.org/10.1098/rsos.201494

Hilgard, J., Engelhardt, C. R., Rouder, J. N., Segert, I. L., & Bartholow, B. D. (2019). Null effects of game violence, game difficulty, and 2D:4D digit ratio on aggressive behavior. *Psychological Science*, *30*(4), 606–616. https://doi.org/10.1177/0956797619829688

Hinsen, K. (2019). Dealing with software collapse. *Computing in Science & Engineering*, *21*(3), 104–108. https://doi.org/10.1109/MCSE.2019.2900945

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Johnson, D. J., & Wilson, J. P. (2019). Racial bias in perceptions of size and strength: The impact of stereotypes and group differences. *Psychological Science*, *30*(4), 553–562. https://doi.org/10.1177/0956797619827529

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), Article e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1), Article 20. https://doi.org/10.1525/collabra.158

Levenstein, M. C., & Lyle, J. A. (2018). Data: Sharing is caring. *Advances in Methods and Practices in Psychological Science*, *1*(1), 95–103. https://doi.org/10.1177/2515245918758319

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science*, *30*(4), 504–515. https://doi.org/10.1177/0956797618822802

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. https://doi.org/10.17226/25303

Nüst, D., Sochat, V., Marwick, B., Eglen, S. J., Head, T., Hirst, T., & Evans, B. D. (2020). Ten simple rules for writing Dockerfiles for reproducible data science. *PLOS*

Computational Biology, *16*(11), Article e1008316. https://doi.org/10.1371/journal.pcbi.1008316

Obaidi, M., Bergh, R., Akrami, N., & Anjum, G. (2019). Group-based relative deprivation explains endorsement of extremism among Western-born Muslims. *Psychological Science*, *30*(4), 596–605. https://doi.org/10.1177/0956797619834879

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 229–237. https://doi.org/10.1177/2515245920918872

Olsson-Collentine, A., van Assen, M. A. L. M., & Hartgerink, C. H. J. (2019). The prevalence of marginally significant results in psychology over time. *Psychological Science*, *30*(4), 576–586. https://doi.org/10.1177/0956797619830326

Peikert, A., & Brandmaier, A. M. (2021). A reproducible data analysis workflow with R Markdown, Git, Make, and Docker. *Quantitative and Computational Methods in Behavioral Sciences*, *1*(1), Article e3763. https://doi.org/10.5964/qcmb.3763

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, Article e175. https://doi.org/10.7717/peerj.175

Psychological Science. (2017, November 15). *Submission guidelines*. https://web.archive.org/web/20171115110444/https://www.psychologicalscience.org/publications/psychological_science/ps-submissions#OPEN

Psychological Science. (2022, April 15). *Psychological Science submission guidelines*. https://www.psychologicalscience.org/publications/psychological_science/ps-submissions

Rowhani-Farid, A., Aldcroft, A., & Barnett, A. G. (2020). Did awarding badges increase data sharing in *BMJ Open*? A randomized controlled trial. *Royal Society Open Science*, *7*(3), Article 191818. https://doi.org/10.1098/rsos.191818

Rowhani-Farid, A., & Barnett, A. G. (2018). Badges for sharing data and code at *Biostatistics*: An observational study. *F1000Research*, *7*, Article 90. https://doi.org/10.12688/f1000research.13477.2

Seibold, H., Czerny, S., Decke, S., Dieterle, R., Eder, T., Fohr, S., Hahn, N., Hartmann, R., Heindl, C., Kopper, P., Lepke, D., Loidl, V., Mandl, M., Musiol, S., Peter, J., Piehler, A., Rojas, E., Schmid, S., Schmidt, H., . . . Nalenz, M. (2021). A computational reproducibility study of PLOS ONE articles featuring longitudinal data analyses. *PLOS ONE*, *16*(6), Article e0251194. https://doi.org/10.1371/journal.pone.0251194

Simonsohn, U., & Gruson, H. (2022). *groundhog: The simplest solution to version-control for CRAN packages*. https://cran.r-project.org/package=groundhog

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P., & Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences, USA*, *115*(11), 2584–2589. https://doi.org/10.1073/pnas.1708290115

Trisovic, A., Lau, M. K., Pasquier, T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, *9*(1), Article 60. https://doi.org/10.1038/s41597-022-01143-6

Ushey, K. (2022). *renv* (Version 0.15.5) [Computer software]. GitHub. https://rstudio.github.io/renv/

Van Lissa, C. J., Brandmaier, A. M., Brinkman, L., Lamprecht, A. L., Peikert, A., Struiksma, M. E., & Vreede, B. M. (2021). WORCS: A workflow for open reproducible code in science. *Data Science*, *4*(1), 29–49. https://doi.org/10.3233/DS-210031

Vardy, T., & Atkinson, Q. D. (2019). Property damage and exposure to other people in distress differentially predict prosocial behavior after a natural disaster. *Psychological Science*, *30*(4), 563–575. https://doi.org/10.1177/0956797619826972

Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., Van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in *Psychological Science*. *PLOS ONE*, *9*(12), Article e114876. https://doi.org/10.1371/journal.pone.0114876

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. -W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), Article 160018. https://doi.org/10.1038/sdata.2016.18

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, *13*(6), Article e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Wójcik, M. J., Nowicka, M. M., Bola, M., & Nowicka, A. (2019). Unconscious detection of one's own image. *Psychological Science*, *30*(4), 471–480. https://doi.org/10.1177/0956797618822971

Woolley, K., & Fishbach, A. (2019). Shared plates, shared minds: Consuming from a shared plate promotes cooperation. *Psychological Science*, *30*(4), 541–552. https://doi.org/10.1177/0956797619830633

Yousif, S. R., & Keil, F. C. (2019). The additive-area heuristic: An efficient but illusory means of visual area approximation. *Psychological Science*, *30*(4), 495–503. https://doi.org/10.1177/0956797619831617