

Citation Patterns Following a Strongly Contradictory Replication Result: Four Case Studies From Psychology



Advances in Methods and Practices in Psychological Science
July-September 2021, Vol. 4, No. 3,
pp. 1–14
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/25152459211040837
www.psychologicalscience.org/AMPPS
 SAGE

**Tom E. Hardwicke^{1,2}, Dénes Szűcs³, Robert T. Thibault^{4,5},
Sophia Crüwell^{2,6}, Olmo R. van den Akker⁷, Michèle B. Nuijten⁷,
and John P. A. Ioannidis^{2,8,9}**

¹Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; ²Meta-Research Innovation Center Berlin (METRIC-B), QUEST Center for Transforming Biomedical Research, Charité – Universitätsmedizin Berlin, Berlin, Germany; ³Department of Psychology, University of Cambridge, Cambridge, England; ⁴School of Psychological Science, University of Bristol, Bristol, England; ⁵MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, England; ⁶Department of History and Philosophy of Science, University of Cambridge, Cambridge, England; ⁷Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, the Netherlands; ⁸Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, USA; and ⁹Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, Stanford University, Stanford, California, USA

Abstract

Replication studies that contradict prior findings may facilitate scientific self-correction by triggering a reappraisal of the original studies; however, the research community's response to replication results has not been studied systematically. One approach for gauging responses to replication results is to examine how they affect citations to original studies. In this study, we explored postreplication citation patterns in the context of four prominent multilaboratory replication attempts published in the field of psychology that strongly contradicted and outweighed prior findings. Generally, we observed a small postreplication decline in the number of favorable citations and a small increase in unfavorable citations. This indicates only modest corrective effects and implies considerable perpetuation of belief in the original findings. Replication results that strongly contradict an original finding do not necessarily nullify its credibility; however, one might at least expect the replication results to be acknowledged and explicitly debated in subsequent literature. By contrast, we found substantial citation bias: The majority of articles citing the original studies neglected to cite relevant replication results. Of those articles that did cite the replication but continued to cite the original study favorably, approximately half offered an explicit defense of the original study. Our findings suggest that even replication results that strongly contradict original findings do not necessarily prompt a corrective response from the research community.

Keywords

replication, citations, meta-research, self-correction, citation bias, open data, open materials, preregistered

Received 2/9/21; Revision accepted 7/23/21

It is often assumed that science is a self-correcting enterprise: The veracity of scientific knowledge should progressively improve as inaccurate claims are abandoned and accurate claims are reinforced (Vazire & Holcombe, 2020). Replication studies are considered to be a key driver of this process because they may indicate that prior results are exaggerated or erroneous (Ioannidis, 2012; Zwaan et al., 2018). Although interpreting the outcome

of replication studies is not necessarily straightforward (Collins, 1985; Earp & Trafimow, 2015; Maxwell et al., 2015), one might expect a replication result that strongly

Corresponding Author:

Tom E. Hardwicke, Department of Psychology, University of Amsterdam
Email: tom.hardwicke@uva.nl



Table 1. Progressive or Regressive Responses to Strongly Contradictory Replication Results and Their Expected Impact on Citation Patterns for Original Studies

Progressive responses	Regressive responses
<p>Belief correction A decrease in the number of favorable citations may reflect a decline in belief in the credibility of the original finding (a <i>belief correction pattern</i>). This may be accompanied by a relative increase in the number of unfavorable citations (an <i>active belief correction pattern</i>) or relatively fewer/no unfavorable citations (a <i>passive belief correction pattern</i>).</p>	<p>Belief perpetuation A maintenance of or increase in the number of favorable citations may reflect a maintenance of or increase in belief in the credibility of the original finding (a <i>belief perpetuation pattern</i>). This may be accompanied by a relative increase in unfavorable citations (a <i>challenged belief perpetuation pattern</i>) or relatively fewer/no unfavorable citations (an <i>unchallenged belief perpetuation pattern</i>).</p>
<p>Citation balance Articles that cite the original study may also cite the contradictory replication, reflecting that relevant evidence has been considered (a <i>balanced citation pattern</i>).</p>	<p>Citation bias Articles that cite the original study may neglect to also cite the contradictory replication, reflecting that relevant evidence has been neglected either through lack of awareness or deliberate omission (a <i>citation bias pattern</i>).</p>
<p>Explicit defense Articles that favorably cite the original study and unfavorably cite the contradictory replication may offer concrete counterarguments that state why the credibility of the original finding has not been undermined (an <i>explicit defense pattern</i>).</p>	<p>Absent defense Articles that favorably cite the original study and unfavorably cite the contradictory replication may offer no concrete counterarguments that state why the credibility of the original finding has not been undermined (an <i>absent defense pattern</i>).</p>

contradicts¹ and outweighs the results of a prior (“original”) study to affect how that study is cited in subsequent academic literature. For example, if a replication undermines belief in the credibility of an original finding, one might expect to see a change in the frequency and valence (favorability) of citations to the original study, that is, a decrease in favorable citations accompanied by an increase in unfavorable citations. However, as discussed below, a variety of interesting patterns could emerge depending on how the research community responds to a replication result. The goal of the present study was to empirically explore and describe postreplication citation patterns in the context of four prominent multilaboratory replication attempts published in the field of psychology that strongly contradicted and outweighed the findings of prior studies.

Table 1 outlines several citation patterns that might follow a contradictory replication result, each reflecting different types of response by the research community. We have tentatively categorized these response patterns as being “progressive” or “regressive,” depending on their expected impact on the accumulation of scientific knowledge.² The first set of patterns, *belief correction/perpetuation*, refers to what is often considered a primary functional role of (contradictory) replication studies—to change belief in the credibility of exaggerated or erroneous original findings (Ioannidis, 2012; Vazire & Holcombe, 2020; Zwaan et al., 2018). In the absence of an explicit defense of an original study that convincingly explains a strongly contradictory replication result (see explicit

defense explanation below), a progressive response might involve a decrease in favorable citations and an increase in unfavorable citations, reflecting updated beliefs about the credibility of the original finding. Conversely, a regressive response might involve maintenance of (or even increase in) favorable citations and relatively few unfavorable citations, which suggests a perpetuation of belief in the credibility of the original finding despite the contradictory replication result. Prior research has documented how favorable citations to observational epidemiology studies can persist despite the claims of those studies being strongly contradicted in subsequent randomized trials (Tatsioni et al., 2007). Likewise, it has been reported that even when articles are retracted, they can continue to receive favorable citations (Budd et al., 1998; Fernández & Vadillo, 2020). Thus, there is evidence that belief in the credibility of original findings can perpetuate even when subsequent events cast doubt on their credibility; however, we are unaware of similar evidence in the context of studies that were explicitly designed to test the replicability of prior findings.

The second set of patterns in Table 1, *citation balance/bias*, generally refers to whether positive (supportive) evidence is preferentially cited relative to negative (nonsupportive) evidence (Bastiaansen et al., 2015; Greenberg, 2009). This pattern has previously been observed in the context of research on inclusion body myositis; citation content analysis showed that the accumulating literature heavily cited the theory that beta amyloid is involved, ignoring multiple studies that

Table 2. Sample Sizes and Effect Sizes for Replication Studies and Original Studies

Original study	Replication study	Effect	Total citations (original) ^a	Original sample size	Replication sample size	Original effect size [95% CI]	Replication effect size [95% CI]
Baumeister et al. (May, 1998)	Hagger et al. (July, 2016) ^b	Ego depletion	1,974	$k = 1$ $N = 67$	$k = 23$ $N = 2,141$	$d = 2.05$ [1.31, 2.79]	$d = 0.04$ [-0.07, 0.15]
Sripada et al. (April, 2014)	Hagger et al. (July, 2016) ^b	Ego depletion	36	$k = 1$ $N = 26$	$k = 23$ $N = 2,141$	$d = 0.68$ [0.09, 1.27]	$d = 0.04$ [-0.07, 0.15]
Strack et al. (May, 1988)	Wagenmakers et al. (October, 2016)	Facial feedback	708	$k = 1$ $N = 92$	$k = 17$ $N = 2,124$	MD = 0.82 [-0.05, 1.69]	MD = 0.03 [-0.11, 0.16]
Caruso et al. (July, 2012)	Klein et al. (January, 2014)	Money priming	57	$k = 1$ $N = 30$	$k = 36$ $N = 6,333$	$d = 0.8$ [0.05, 1.54]	$d = 0.01$ [-0.06, 0.09]
T. J. Carter et al. (July, 2011)	Klein et al. (January, 2014)	Flag priming	54	$k = 1$ $N = 70$	$k = 36$ $N = 4,896$	$d = 0.50$ [0.01, 0.99]	$d = 0.01$ [-0.07, 0.08]

Note: Publication dates are earliest available (i.e., “online first” if relevant). d = Cohen’s d ; MD = mean difference; k = number of data-collection sites; N = total number of participants; CI = confidence interval.

^aTotal citations to the original study between the publication date and December 31, 2019.

^bFor methodological reasons (see Hagger et al., 2016), the ego-depletion replication was aimed at a classic study in the field (Baumeister et al., 1998) but actually employed a modified computer-based version of the original paradigm (Sripada et al., 2014). We examined postreplication citation patterns for both studies.

contradicted this theory (Greenberg, 2009). In the present study, these patterns specifically refer to whether articles citing an original study also cite the subsequent contradictory replication study. A progressive response would be to cite both studies (citation balance) because this involves considering and reporting highly relevant evidence (even if the implications of the replication are disputed; see explicit defense explanation below). By contrast, citation bias could occur if articles citing an original study neglect to cite a relevant replication study. Regardless of whether this occurs through lack of awareness or deliberate omission, it can be considered a regressive response pattern because highly relevant evidence is not being reported or considered.

The third set of patterns in Table 1, *explicit/absent defense*, refers to whether researchers who continue to favorably cite the original finding despite the strongly contradictory replication result offer a concrete defense of the original study. As implied above, even when the results of a replication study strongly contradict the results of an original study, this does not necessarily nullify the credibility of the original findings; the same criticisms that one might apply to an original study to infer that its findings are erroneous or exaggerated may also be applied to replication studies (Collins, 1985; Earp & Trafimow, 2015; Maxwell et al., 2015). Thus, if proponents of the original claim mount an explicit defense that counters the implications of the replication results, this might still be considered a progressive response (although obviously one could disagree with the arguments that are presented). By contrast, if favorable citations to the

original study are not accompanied by explicit argumentation about the replication result (an absent defense), this might be considered a regressive response because the replication result is apparently discounted without providing any rationale.

In the present study, we explored the postreplication citation patterns described above in the context of four case studies in the field of psychology in which the findings of a replication study strongly contradicted and outweighed the findings of an original study (Table 2). In two of the cases, the replication studies were part of a single Many Labs project (Klein et al., 2014) and addressed the “flag priming effect” (T. J. Carter et al., 2011) and “money priming effect” (Caruso et al., 2013), respectively. The other two cases involved Registered Replication Reports (Simons et al., 2014) that examined influential demonstrations of the “facial feedback effect” (cf. Strack et al., 1988; Wagenmakers et al., 2016) and the “ego-depletion effect” (cf. Baumeister et al., 1998; Hagger et al., 2016; Sripada et al., 2014), respectively. For methodological reasons (see Hagger et al., 2016), the ego-depletion replication was aimed at a classic study in the field (Baumeister et al., 1998) but actually employed a modified computer-based version of the original paradigm (Sripada et al., 2014). For this particular case study, we examined citation patterns to both of these original studies.

We adopted a case-study approach to develop a “narrow and deep” understanding of the topic, as opposed to a “broad and shallow” approach, which would have required a deliberate representative sampling strategy. We chose these particular case studies because they involved

prominent preregistered multilaboratory replication attempts with sample sizes 23 times to 211 times larger than the original studies, thus providing highly visible and highly credible evidence that strongly contradicted and outweighed earlier findings.³ This facilitates additional interpretative clarity about the citation patterns one might expect to observe.

The present study was exploratory in nature and intended to provide descriptive observations rather than test hypotheses. The three sets of expected citation patterns outlined in Table 1 were used to guide our study design and interpretation, but we do not claim that this is a comprehensive typology of the postreplication patterns that may occur. Such patterns may be more complex and idiosyncratic in other topic domains. To examine belief correction/perpetuation patterns, we downloaded citation histories (a list of citing articles) for each original study and classified the valence (favorable, equivocal, or unfavorable) of a set of prereplication and postreplication citations. To examine citation balance/bias patterns, we manually checked whether postreplication citations of the original study were accompanied by citations to the replication study. Finally, to examine explicit/absent defense patterns, we extracted and categorized any counterarguments offered in articles that cited the replication.

Method

The study protocol (rationale, methods, and analysis plan) was preregistered on April 7, 2018 (<https://osf.io/eh5qd/>). An amended protocol was registered partway through data collection on May 1, 2019, primarily because we extended the sampling frame to cover additional months (<https://osf.io/pdvb5/>). All deviations from these protocols are explicitly acknowledged in Supplementary Information A in the Supplemental Material available online. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Design

This was a retrospective observational study consisting of four case studies. Primary outcome variables were annual citation counts for original studies, citation valence (favorable, equivocal, unfavorable), co-citation of original and replication studies, and frequency/type of counterarguments.

Sample

We examined four case studies in which a prominent preregistered and multilaboratory replication study strongly contradicted and outweighed the findings of an

original study (Table 2). As shown, all the original studies found modestly large to very large effects, and all of them were relatively small studies (thus they would have been underpowered to detect small effects). Conversely, all the replication efforts comprised very large sample sizes, and they would be very well powered to detect even small effects; however, they all obtained null results.

Procedure

Annual citation counts. Citation histories (i.e., bibliographic records for all articles that cite the original study) from the publication date of each original study through December 31, 2019, were downloaded from Clarivate Analytics Web of Science Core Collection, accessed via the Charité – Universitätsmedizin Berlin on August 12, 2020. We also obtained citation histories for a reference class—all articles published in the same journal and the same year as each original study—from the same source. For example, for Baumeister et al. (1998), the reference class was all articles published in 1998 in the *Journal of Personality and Social Psychology*. Citation counts were standardized in each case study by setting the citation count in the replication year to the standardized value of 100 and then adjusting the counts in other years according to the same transformation ratio. For example, if the raw citation count in the replication year was 1,000, citation counts in each year would be standardized by dividing by 10. This computation was performed separately for citations to the reference class and citations to the original article.

Qualitative assessment. Qualitative assessment of citation patterns was limited to a time period starting 1 year before the year of publication of the replication study until December 31, 2019, excluding the year in which the replication was published. We excluded the replication year because it may be unreasonable to expect citing articles already in the publication pipeline to cite the replication study. For the Baumeister case, the qualitative analysis was based on a random sample of 40% of citing articles from the prereplication period and postreplication period because of the large number of citations to the original study ($n = 1,974$; for details, see Supplementary Information B in the Supplemental Material).

For each citing article undergoing qualitative assessment, we attempted to retrieve the full text via several methods in the following order: (a) search of at least two of the institutional libraries we are affiliated with; (b) general Internet search for the article title, including the Google and Google Scholar search engines and Research Gate; (c) email requests to the corresponding author; and (d) interlibrary loan request. Articles that remained inaccessible after all of these methods were

exhausted were excluded. Articles written in a non-English language were translated by one of the authors or by using Google Translate (see Supplementary Information D in the Supplemental Material). For articles for which we could obtain the full text, we classified the research design according to the categories in Table 3 and recorded whether the replication study was cited after manual inspection of the reference section (see Table 1: citation balance/bias).

To examine the belief correction/perpetuation pattern (Table 1), one of six primary coders (T. E. Hardwicke, D. Szűcs, R. T. Thibault, S. Crüwell, O. R. van den Akker, and M. B. Nuijten) manually extracted the “citation context” of the original study and the replication study (i.e., all relevant verbatim text surrounding each in-text citation). The primary coder then classified the citation valence as “favorable,” “equivocal,” “unfavorable,” or “unclassifiable.” Favorable citations were those used to support a positive claim about the phenomenon of interest, whereas unfavorable citations were used to support a negative claim about the phenomenon of interest. Citations were considered equivocal if the authors did not take a predominantly favorable or unfavorable position. Citations that did not endorse or oppose the phenomenon of interest (e.g., simply referring to the procedures of the original study) were designated as unclassifiable. Because this process was inherently subjective, the citation contexts and classifications were also examined by one of six secondary coders (T. E. Hardwicke, D. Szűcs, R. T. Thibault, S. Crüwell, O. R. van den Akker, and M. B. Nuijten). Disagreements were resolved through discussion, and a third coder arbitrated when necessary. Valence classifications by the primary coder were modified after discussion with the secondary coder in 31 (5%) cases.

To examine the explicit/absent defense pattern (Table 1), the primary coder flagged articles that co-cited the original and replication studies and also contained any explicit defense of the original study. Subsequently, two team members (O. R. van den Akker and S. Crüwell) reexamined all of the flagged cases, extracted verbatim counterarguments, and developed a post hoc categorization scheme that summarized them as concisely and informatively as possible. Coding disagreements were resolved through discussion, and a third coder (T. E. Hardwicke) arbitrated when necessary.

In additional exploratory (not preregistered) analyses, we examined overlap of authorship for articles that provided counterarguments with (a) any of the authors of the original studies and (b) any prior collaborators of the first authors of the original studies. These analyses are complicated by the fact that author names in bibliographic records do not always adhere to the same grammatical standards—for example, whether forenames are initialized or middle names are included—so

Table 3. Counts and Percentages for Article Type Classifications of Articles Included in Qualitative Analyses

Article type	Count (%)
No empirical data (e.g., editorials, commentaries [without reanalysis], simulations, news, and reviews)	197 (33)
Data synthesis - meta-analysis	11 (2)
Empirical data - commentary including analysis	4 (1)
Empirical data - case study	1 (< 1)
Empirical data - survey	79 (13)
Empirical data - field study	40 (7)
Empirical data - laboratory study	248 (41)
Empirical data - multiple study types are reported	24 (4)

it is not straightforward to isolate individual authors within bibliographic databases. To identify prior collaborators of the first authors of the original studies, we downloaded bibliographic records (on February 2, 2021) for all articles published by each of the original study first authors according to their author record in the Web of Science Core Collection. These author records are automatically generated by an algorithm that attempts to identify all documents likely published by an individual author using several variations of their name (e.g., “Hardwicke, Tom E.,” “Hardwicke, Tom,” “Hardwicke, T. E.”), but errors can still occur, and incomplete database coverage means that this method likely misses some of the authors’ prior publications and consequently some of their collaborators. Nevertheless, the method supports a reasonable lower bound estimate of authorship overlap with articles providing counterarguments. To identify authorship overlap, we used string manipulation tools in R to extract only author surnames from bibliographic records and then used string matching to automatically detect the presence of original author or collaborator surnames among the surnames of authors of articles that provided counterarguments. When a match was detected, it was verified by manual examination of the authors’ full names.

Results

In total, 2,829 articles cited one of the original studies, of which 632 articles (after taking a 40% random sample in the Baumeister case) fell within the time period designated for qualitative assessment. Of these 632 articles, we excluded 28 from the qualitative analysis because (a) we could not access the full text ($n = 22$), (b) they included a citation to the original study in the reference section but not in the main text ($n = 5$), or (c) manual inspection indicated that they did not actually appear to cite the original study at all ($n = 1$). Article type classifications for the remaining 604 articles included in the qualitative analysis are shown in Table 3.

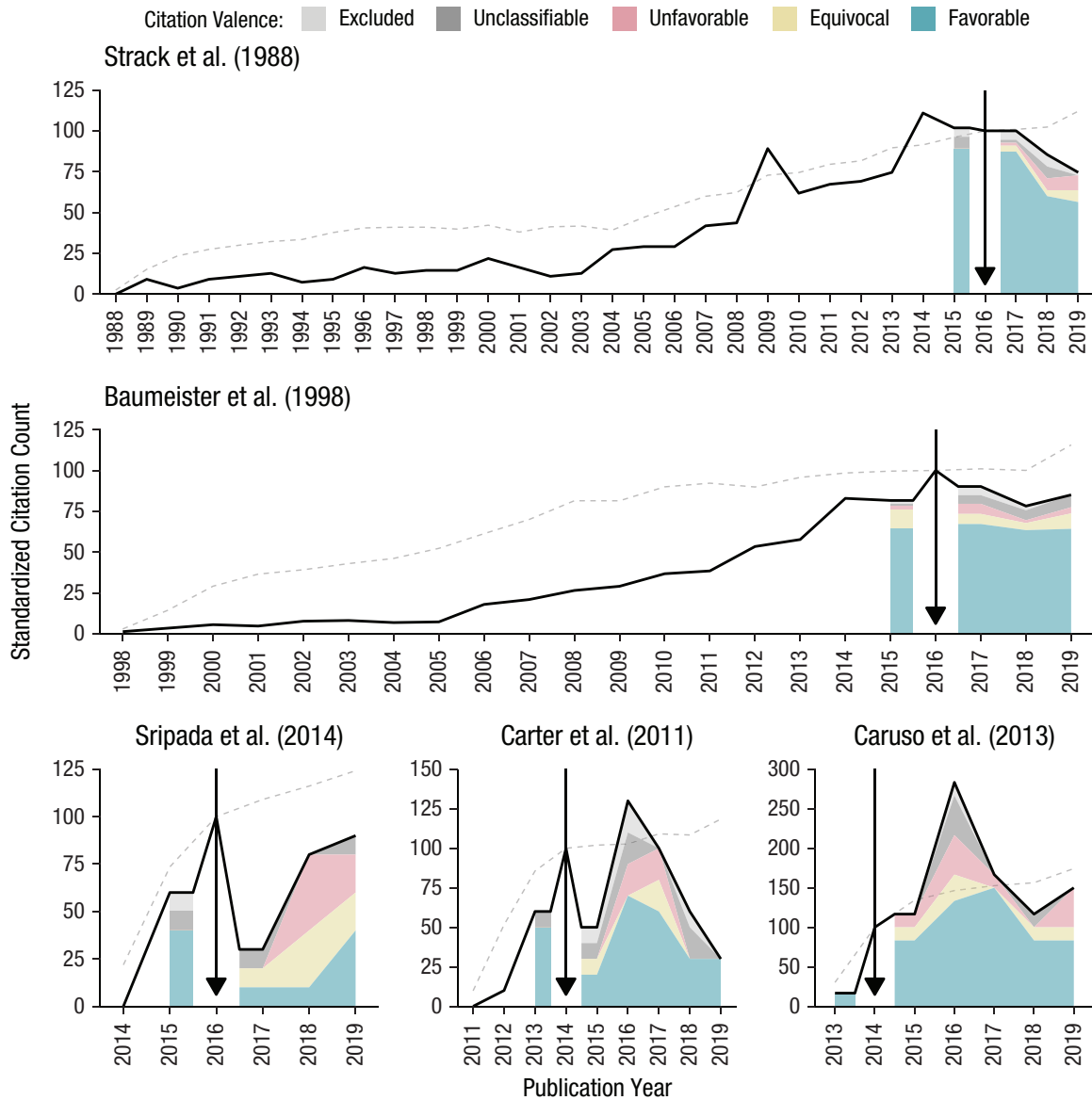


Fig. 1. Standardized annual citation counts (solid line) for the five original studies with citation valence (favorable, equivocal, unfavorable, unclassifiable) illustrated by colored areas in prereplication and postreplication assessment periods. The dashed line depicts citations to the reference class (all articles published in the same journal and same year as the target article). Annual citation counts are standardized against the year in which the replication was published (citation counts in the replication year, indicated by a black arrow, are set at the standardized value of 100). Citation valence classifications for the Baumeister case are extrapolated to all articles in the assessment period according to a 40% random sample.

Annual citation counts and citation valence

Figure 1 shows standardized annual citation counts for each original study and the respective reference class (citations to all articles published in the same year and same journal as the original study) and classifications of citation valence (favorable, equivocal, unfavorable, unclassifiable, or excluded). The data can also be viewed in tabular format in Supplementary Table C1 in the

Supplemental Material. All counts (n) reported in the text and table are raw counts (i.e., not standardized).

After the replication was published, citations to the reference classes were continuing their trend to plateau (Baumeister case) or increase (other cases). By contrast, citations to the original study appeared to undergo a modest decline in the Strack case (decreasing from 56 to 41 between 2015 and 2019) and a small decline followed by a small increase in the Baumeister case (increasing from 191 to 199 between 2015 and 2019). In the other

cases (Sripada, Carter, Caruso), the total citation counts were much lower, and there was considerable variability in the postreplication citation patterns; nevertheless, there was no substantial change in annual citations from prereplication to postreplication in these three cases (the maximum difference was +8 citations).

Before the replication, the vast majority of citations were favorable for all five articles (range = 67%–100%). In most cases (Strack, Sripada, Carter, and Caruso), there was a small postreplication increase in unfavorable citations and a small decrease in favorable citations, indicating a modest active correction pattern. However, the overall number of unfavorable citations was very low, and there was still a substantial majority of favorable citations. For example, in the Strack case, unfavorable citations increased from 0% in the prereplication period (2015) to 7% in the postreplication period, whereas favorable citations decreased from 88% to 78%. In the Baumeister case, the proportion of favorable citations remained stable from prereplication (79%) to postreplication (77%), a pattern consistent with belief perpetuation. The very small number of unfavorable citations (2017: $n = 7$, 7%; 2018: $n = 2$, 2%; 2019: $n = 2$, 4%) suggests that this is largely an unchallenged belief perpetuation pattern (see Table 1).

Citation balance and citation bias

Figure 2 shows the proportion of citing articles that also cited or did not cite the replication study after it was published (excluding the publication year itself). The data can also be viewed in tabular format in Supplementary Table C1 in the Supplemental Material. In the Strack and Baumeister cases, a considerable majority of articles citing the original study did not cite the replication study, which indicates substantial citation bias. In the Baumeister case, the proportion of articles citing the replication study remained stable (20% in 2017, 18% in 2019). In the Strack case, the proportion increased from 13% to 41%. In the Carter and Caruso cases, the proportion never exceeded 50%, also consistent with substantial citation bias. In the Sripada case, it was much more common for the replication study to be cited (> 88%), which reflects a balanced citation pattern.

Explicit defense and absent defense

Table 4 shows whether articles that cited the original study and replication study (“co-citing articles”) and the subset of co-citing articles that cited the original study favorably provided any explicit counterarguments to defend the credibility of the original finding (an explicit defense) or not (an absent defense). Overall, fewer than half of the 127 co-citing articles provided any

counterarguments. Of the 60 co-citing articles that cited the original study favorably, around half provided counterarguments. We identified 58 discrete counterarguments in 51 citing articles (45 of which were unique articles because six of them were cited in two of the case studies) and allocated them to one of three categories (Table 5).

In additional exploratory analyses (not preregistered), we examined other characteristics of the 45 unique articles that contained counterarguments. The articles were published in 34 individual journals; *Frontiers in Psychology* published seven of the articles, *Social Psychology* published four of the articles, and all other journals published only one or two of the articles. Seventeen of the articles did not involve empirical data, three involved reanalysis or meta-analysis of existing data, and 25 involved collection of novel data. The articles had 112 individual authors, of whom all contributed to a single article except for nine individuals who had authored or coauthored two articles. Three articles were authored or coauthored by one of the original authors, and nine articles were authored or coauthored by at least one prior collaborator of one of the first authors of the original articles. Seven of these articles did not involve empirical data, and five of them involved novel data collection.

Citations to replication studies and co-citation of original studies

A reviewer requested that we examine citation counts for replication studies and check whether citing articles also co-cited the relevant original study. To obtain the data, we downloaded bibliographic records from the Clarivate Analytics Web of Science Core Collection, accessed via the University of Amsterdam on April 16, 2021, for articles that cited each replication study (up to the study endpoint—December 2019) and cross-checked them with our sample of articles that cited the original study. As shown in Table 6, the replication studies also have a life of their own, and they are often cited independently of the specific original study. Often this is easy to explain. For example, Klein et al. replicated 13 original studies, not just the two that were of interest in our analysis. These studies are also likely to have accrued citations by virtue of being among the first highly prominent examples of preregistered multilaboratory replications in psychology.

Discussion

It has been proposed that replication studies can facilitate scientific self-correction by modifying scientists’ belief in the credibility of published findings (Ioannidis, 2012; Vazire & Holcombe, 2020; Zwaan et al., 2018); however,

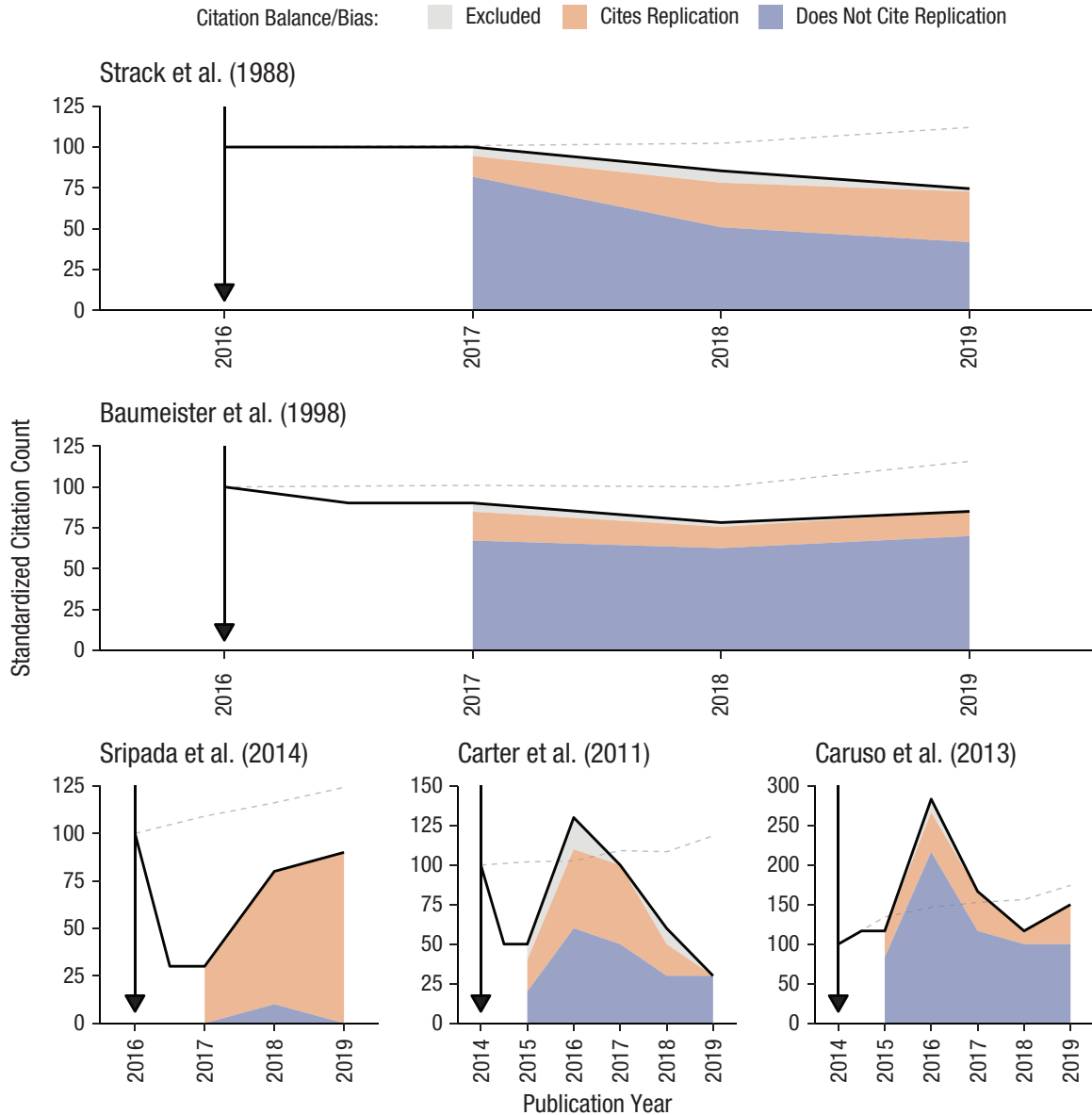


Fig. 2. Standardized annual citation counts (solid line) for the five original studies with citation balance/bias (i.e., whether the replication is cited) illustrated by colored areas in the postreplication assessment period. The dashed line depicts citations to the reference class (all articles published in the same journal and same year as the target article). Annual citation counts are standardized against the year in which the replication was published (citation counts in the replication year, indicated by a black arrow, are set at the standardized value of 100). Replication citation proportions for the Baumeister case are extrapolated to all articles in the assessment period according to a 40% random sample.

the extent to which this occurs in practice is unclear. In this study, we investigated how the research community responded to four strongly contradictory replications in the field of psychology by examining postreplication citation patterns for original studies. We observed some progressive response patterns in the form of modest active correction (a small decline in favorable citations and a small increase in unfavorable citations) and even more prominent regressive response patterns in the form of unchallenged belief perpetuation (sustained levels of favorable citations and few unfavorable citations,

particularly in the Baumeister case) and considerable citation bias (neglecting to cite the replication study; in all cases aside from Sripada). When authors cited the original study favorably despite the replication result, only half of the articles provided any explicit counterarguments in defense of the original study (an explicit defense). Overall, these findings are consistent with prior observations that favorable citation patterns appear relatively unperturbed by subsequent publication of contradictory results in studies with lower risk of bias (Tatsioni et al., 2007) or even by full retraction (Budd et al., 1998;

Fernández & Vadillo, 2020) and that positive (supportive) evidence is preferentially cited relative to negative (non-supportive) evidence once a theory gets entrenched despite overwhelming evidence against it (Bastiaansen et al., 2015; Greenberg, 2009).

It is reasonable to question whether the replication results in these case studies should (rationally) have instigated belief change in the research community and thus triggered a more sizable decline in favorable citations than we observed. Note that the case studies we selected were deliberately chosen because the replication results were superior to the original results in terms of both credibility and evidential value. The replications were preregistered, multilaboratory studies with large sample sizes. By contrast, the original studies had much smaller sample sizes and arguably had a much higher risk of bias given that they were not preregistered, were performed by single teams, and arose in domains affected by publication bias and other questionable research practices (E. C. Carter et al., 2015; Coles et al., 2019; Vadillo, 2019; Vadillo et al., 2016). Thus, it seems reasonable to suggest that the compelling replication results should have reduced belief in the credibility of the original results. However, one could also contend that if there were some flaw in the replication study that undermined its validity (Brandt et al., 2014; Fabrigar et al., 2020; Vazire et al., 2021), then this may provide justification to continue favorably citing the original study. Indeed, in these particular case studies, the validity of the replication results has been challenged by proponents of the original findings (Table 5; Baumeister & Vohs, 2016; Strack, 2016). Because this debate remains far from settled (Coles et al.,

Table 4. Counts and Percentages for Whether Articles That Cited Both the Original Study and Replication Study Provided Any Explicit Argumentation to Defend the Original Study

Case	Did co-citing articles provide argumentation to defend the original study?			
	All citation valences		Favorable citation valence	
	No	Yes	No	Yes
Baumeister	24 (56%)	19 (44%)	11 (46%)	13 (54%)
Carter	11 (79%)	3 (21%)	3 (75%)	1 (25%)
Caruso	8 (67%)	4 (33%)	1 (50%)	1 (50%)
Sripada	10 (53%)	9 (47%)	2 (40%)	3 (60%)
Strack	23 (59%)	16 (41%)	12 (48%)	13 (52%)
All cases	76 (60%)	51 (40%)	29 (48%)	31 (52%)

Note: Data are displayed for co-citing articles with any citation valence classification and the subset of co-citing articles with favorable citation valence classifications.

2019; Vadillo, 2019), ideally any favorable citation of the original studies should at a minimum be accompanied by co-citation of the replication results and some discussion of the discrepant findings.

The clear evidence of citation bias that our study documents may have two main contributory factors: (a) a lack of awareness about the replication results and/or (b) a decision to ignore the replication results. The practical issue of awareness is not necessarily straightforward to address. Individual scientists can find it difficult to keep up to date with the voluminous literature that is

Table 5. Categorization of Counterarguments Provided to Defend the Original Study in Light of the Contradictory Replication Result

Category	Definition and examples	Count
Methodological differences and moderators	Methodological features of the replication study or other moderating factors may explain the absence of an effect. Example: “Ego depletion exists but its occurrence seems to depend on moderating conditions. Therefore, we think that the search of moderators concerning ego depletion (or ego depletion as a moderator, respectively) is justified” (Kühl & Bertrams, 2019, p. 9).	45
Additional evidence	Evidence from other studies supports the existence of the effect. Example: “Although there has been a challenge to this original finding, there have been many replications of the principle and a meta-analysis shows a robust facial-feedback effect” (Lewis, 2018, p. 2).	11
Expertise	Inadequate expertise of the replicating authors explains why they could not replicate the effect. Example: “In other words, it is easier to be successful at non-replications while it takes expertise and diligence to generate a new result in a reliable fashion” (Strack, 2017, p. 3).	2

Note: Fifty-eight discrete counterarguments were identified in 51 articles (45 unique articles across cases).

Table 6. Citations Counts for Replication Studies and Co-Citation Counts to Relevant Original Studies

Replication study	Citation count	Co-citations of original study
Hagger et al. (2016)	258	136 (Baumeister et al., 1998) 22 (Sripada et al., 2014)
Klein et al. (2014)	316	15 (T. J. Carter et al., 2011) 12 (Caruso et al., 2013)
Wagenmakers et al. (2016)	80	36 (Strack et al., 1988)

relevant to their research. Recently, the reference manager Zotero introduced a feature that alerts users when an article in its database has been retracted (Zotero, 2019). One could imagine a similar feature being introduced for replication studies, perhaps based on databases that explicitly identify replication studies. However, it is much less straightforward to define and identify a relevant replication study (Neuliep & Crandall, 1993), and users would need to be alerted that this requires some scientific judgment rather than simple article metadata. Another solution could be to encourage researchers to focus less on individual studies and more on up-to-date evidence summaries (i.e., reviews and meta-analyses) in which relevant evidence is systematically identified and collated. This would require that high-quality and contemporary evidence summaries are available; however, in psychology, systematic reviews and meta-analyses can be of low quality, and their results may still be inflated and nonreproducible (Kvarven et al., 2020; Maassen et al., 2020; Polanin et al., 2020). Moreover, empirical studies are often not included in any form of evidence synthesis (Hardwicke et al., 2021).

The second issue of authors ignoring highly relevant replications seems undesirable and implies a biased appraisal or presentation of the evidence. Contradictory replication results do not necessarily nullify the credibility of an original study (Collins, 1985; Earp & Trafimow, 2015; Maxwell et al., 2015), but we would still expect highly relevant replication results to be cited and explicitly debated. In fact, we found that when authors continued to favorably cite original studies despite the replication results, around half did not provide explicit counterarguments in defense of the original study. Although our study did not examine researchers' individual beliefs, one recent study reported that when research psychologists were confronted with replication evidence, they often did update their (self-reported) beliefs, albeit modestly (McDiarmid et al., 2021). However, there are several reasons to be uncertain about whether the results from this artificial setting might generalize to real-world settings, including potential participant reactivity (i.e., participants behaving

differently because they are under observation and/or responding to the perceived expectations of the research team) and the possibility that individuals may behave differently in settings in which they have substantial personal investment and may be publicly scrutinized. In addition, cognitive psychology studies have obtained some evidence of a "continued influence effect" wherein an individual's beliefs and behavior can continue to be influenced by false or misleading information despite subsequent efforts to reject it (Lewandowsky et al., 2012). It is plausible that various cognitive biases, such as confirmation bias (preferentially seeking out and processing evidence that supports preexisting beliefs) or motivated reasoning (constructing and evaluating arguments according to what is desirable rather than what is rationally justifiable), may partly explain researchers' tendency to ignore or dismiss the replication evidence (Bishop, 2019; Kunda, 1990; Nickerson, 1998).

We observed that when counterarguments were raised, most of them tried to dismiss the contradictory replication by claiming that the original and the replication studies differed in important ways that moderated the absence/presence of the effect under investigation. In some cases, authors pointed to evidence from other studies as a rationale for their continued belief in the effect. In a minority of cases, the authors challenged the competence of replicators. We found that articles presenting counterarguments were published in a variety of journals (rather than clustered in a few journals) and involved collection of new data in around half of the cases. They were also published by a sizable group of investigators, only a minority of whom were one of the original authors or had previously collaborated with one of the original first authors. This suggests that the explicit defense of the original study came from a variety of sources rather than being confined to a small number of investigators. However, note that this analysis may underestimate authorship overlap because of the difficulties isolating individual researcher identities (see Method section) and articles published in the same year as the replication not being included.

The findings presented here are inherently limited by the observational nature of the study design, which complicates straightforward conclusions about the causal impact of the replications. Although the use of a reference class enables us to detect the influence of exogenous factors to some extent, we cannot rule out their contribution. For example, the modest decline in favorable citations observed in most cases could be attributable, at least in part, to a more general awareness in the research community about methodological issues (e.g., that the sample sizes of the original studies may not have provided adequate statistical power). We have also focused only on the replication study and the original study in each case study without considering the impact

of other potentially relevant events. Note that meta-research studies have detected signatures of publication bias and other questionable research practices in the fields to which these case studies belong (E. C. Carter et al., 2015; Coles et al., 2019; Vadillo, 2019; Vadillo et al., 2016), and other relevant replication studies contesting prior findings have been published (e.g., Rohrer et al., 2015).

We have been able to gauge reactions to replications only to the extent that they are reflected in relatively short-term citation patterns. A potential explanation of the apparently cursory treatment of the replication studies could be that researchers became aware of them only after their own research projects had begun and/or even had been completed. If one of the original studies had been a key motivator for one's own study, then it may be difficult to accommodate the strongly contradictory replication results. It may even be tempting to ignore them or give them only superficial treatment. Moreover, researchers who are convinced that the replication study has squarely refuted the original may no longer be interested in doing research on a topic for which they see no future potential. In addition, examination of citation patterns would not detect whether there had been a correction effect among individuals who would not typically cite the original study, such as students, members of the public, or researchers working in other fields. A contested study may continue to be cited favorably by its proponents who remain working in the field. This will suffice to create belief perpetuation in the published literature even though other scientists may simply no longer be interested in getting involved with such a strongly contested research topic. Relatedly, we did not examine citation patterns beyond 3 to 5 years after replication. Some perspectives envision scientific self-correction unfolding over a much longer time scale (Lauden, 1981; Peterson & Panofsky, 2020) and suggest that the process is characterized less by the impact of individual study results and more by the gradual accumulation of converging evidence, gradual revision of theoretical understanding, and/or informal sociological processes (e.g., researchers choosing alternative topics to study). Thus, although the current findings may contradict the expectations of a more direct and expedited view of the corrective impact of replication studies (Ioannidis, 2012; Vazire & Holcombe, 2020), they are not necessarily inconsistent with a slower and more indirect process of self-correction. Future research could employ longitudinal designs or older historical case studies to evaluate citation patterns unfolding over a longer time scale. Finally, because we examined only citation patterns, this study could not capture other potential responses to replication results, such as changes in research practices.

Generalization beyond these four case studies requires caution. Note that the replication studies examined here

were some of the first large-scale, multilaboratory replication attempts conducted in the field of psychology. This gave them particular prominence and initiated considerable debate, which resulted in broader ramifications beyond the research community that typically studies the topics under scrutiny (Nelson et al., 2018). Also note that we deliberately selected case studies in which the replication studies were high-profile and had yielded high credibility evidence that strongly contradicted and outweighed the original findings. A correction effect may be less expected in cases in which replication results are more ambiguous, less consequential, or less well known. For example, in a situation in which two high-credibility studies with similar evidential value yield contradictory results, it would be premature to lose confidence in one of the studies before further investigation has probed the cause of the discrepancy. Pursuit of potential moderating factors may be entirely rational in such circumstances (Gershman, 2019).

We also note that particular aspects of our study were inherently subjective, specifically, the identification of citation context, the classification of citation valence, and the identification, extraction, and categorization of counterarguments. To minimize subjectivity, a team of six investigators performed coding in duplicate, with a third investigator arbitrating if necessary. Because disagreements between primary and secondary coders were infrequent, we are confident that the classifications are meaningful, but there may be some edge cases when an independent observer might reasonably disagree with our classifications.

In conclusion, postreplication citation patterns in four case studies indicated that the anticipated corrective impact of strongly contradictory replication results did not materialize to any substantive degree. A lack of awareness of replications and/or a decision to discount or omit them appears to have played a significant role. This highlights potential practical problems with the discoverability of replication studies and psychological or sociological issues related to belief change. The findings also indicate that scientific self-correction may not be as expedient or straightforward as one might hope (Ioannidis, 2012), which adds further impetus toward efforts to improve the quality of the academic literature (Hardwicke et al., 2020; Nelson et al., 2018).

Transparency

Action Editor: Julia Strand

Editor: Daniel J. Simons

Author Contributions

T. E. Hardwicke, D. Szűcs, and J. P. A. Ioannidis designed the study. T. E. Hardwicke, D. Szűcs, R. T. Thibault, S. Crüwell, O. R. van den Akker, and M. B. Nuijten performed the data extraction and coding. T. E. Hardwicke and S. Crüwell performed the data analysis. T. E. Hardwicke

wrote the manuscript. All of the authors provided feedback and approved the final manuscript for submission.

Declaration of Conflicting Interests

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

Funding

The authors received no specific funding for this work. The Meta-Research Innovation Center at Stanford (METRICS) is supported by a grant from the Laura and John Arnold Foundation. The Meta-Research Innovation Center Berlin (METRIC-B) is supported by a grant from the Einstein Foundation and Stiftung Charité. The work of J. P. A. Ioannidis is supported by an unrestricted grant from Sue and Bob O'Donnell. R. T. Thibault is supported by a postdoctoral fellowship from the Fonds de la recherche en santé du Québec. T. E. Hardwicke receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement 841188. D. Szűcs receives funding from the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition (Grant 220020370). O. R. van den Akker is supported by a Consolidator Grant (IMPROVE) from the European Research Council (Grant 726361).

Open Practices

Open Data: <https://osf.io/w8h2q/>

Open Materials: <https://osf.io/w8h2q/>


Preregistration: <https://osf.io/eh5qd/>


All data and materials have been made publicly available via OSF and can be accessed at <https://osf.io/gyzbm/files/>. The protocol and analysis plans were preregistered via OSF and can be accessed at <https://osf.io/eh5qd/>. To facilitate reproducibility, this manuscript was written by interleaving regular prose and analysis code using knitr (Xie, 2017) and papaja (Aust & Barth, 2020) and is available in a Code Ocean container (<https://doi.org/10.24433/CO.4225975.v3>) that recreates the software environment in which the original analyses were performed. 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 iDs

Tom E. Hardwicke  <https://orcid.org/0000-0001-9485-4952>

Robert T. Thibault  <https://orcid.org/0000-0002-6561-3962>

Sophia Crüwell  <https://orcid.org/0000-0003-4178-5820>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/25152459211040837>

Notes

1. Determining whether a replication finding contradicts an original finding is a complex issue that we largely sidestep here by

examining only cases that are *strongly contradictory*—that is, according to several reasonable quantitative criteria (e.g., effect size magnitude, *p* values, Bayes's factors), the results of the original study and replication study lead to opposing inferences (e.g., absence vs. presence of an effect). Also note that one can accept that results are strongly contradictory and remain agnostic about the explanation for the contradiction.

2. This terminology was inspired by but does not directly mirror terminology proposed by Imre Lakatos in his work on the rationality of the research community's response when a scientific theory is contracted by empirical evidence (Lakatos, 1970).

3. We deliberately focused on a select group of case studies rather than other potentially larger samples to aid interpretative clarity; for example, the extent to which the results of the large-scale Reproducibility Project in Psychology (RPP; Open Science Collaboration, 2015) actually contradicted the original studies has been contested (Etz & Vandekerckhove, 2016). In addition, original studies were not actually cited in the RPP research report, which may have diluted awareness about relevant replication studies.

References

- Aust, F., & Barth, M. (2020). *Papaja: Create APA manuscripts with Rmarkdown*. <https://github.com/crsh/papaja>
- Bastiaansen, J. A., de Vries, Y. A., & Munafò, M. R. (2015). Citation distortions in the literature on the serotonin-transporter-linked polymorphic region and amygdala activation. *Biological Psychiatry*, *78*(8), E35–E36. <https://doi.org/10.1016/j.biopsych.2014.12.007>
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–1265. <https://doi.org/10.1037/0022-3514.74.5.1252>
- Baumeister, R. F., & Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science*, *11*(4), 574–575. <https://doi.org/10.1177%2F1745691616652878>
- Bishop, D. V. (2019). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *73*(1), 1–19. <https://doi.org/10.1177/1747021819886519>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Budd, J. M., Sievert, M., & Schultz, T. R. (1998). Phenomena of retraction: Reasons for retraction and citations to the publications. *JAMA*, *280*(3), 296–297. <https://doi.org/10.1001/jama.280.3.296>
- Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General*, *144*(4), 796–815. <https://doi.org/10.1037/xge0000083>
- Carter, T. J., Ferguson, M. J., & Hassin, R. R. (2011). A single exposure to the American flag shifts support toward

- republicanism up to 8 months later. *Psychological Science*, 22(8), 1011–1018. <https://doi.org/10.1177/0956797611414726>
- Caruso, E. M., Vohs, K. D., Baxter, B., & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. *Journal of Experimental Psychology: General*, 142(2), 301–306. <https://doi.org/10.1037/a0029288>
- Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. *Psychological Bulletin*, 145(6), 610–651. <https://doi.org/10.1037/bul0000194>
- Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. SAGE.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, Article 621. <https://doi.org/10.3389/fpsyg.2015.00621>
- Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the Reproducibility Project: Psychology. *PLOS ONE*, 11(2), Article e0149794. <https://doi.org/10.1371/journal.pone.0149794>
- Fabrigar, L. R., Wegener, D. T., & Petty, R. E. (2020). A validity-based framework for understanding replication in psychology. *Personality and Social Psychology Review*, 24(4), 316–344. <https://doi.org/10.1177/1088868320931366>
- Fernández, L. M., & Vadillo, M. A. (2020). *Retracted papers die hard: Diederik Stapel and the enduring influence of flawed science*. PsyArXiv. <https://doi.org/10.31234/osf.io/cszpzy>
- Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, 26(1), 13–28. <https://doi.org/10.3758/s13423-018-1488-8>
- Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *BMJ*, 339, Article b2680. <https://doi.org/10.1136/bmj.b2680>
- Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., . . . Zwienerberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573. <https://doi.org/10.1177/1745691616652873>
- Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. A. (2020). Calibrating the scientific ecosystem through meta-research. *Annual Review of Statistics and Its Application*, 7(1), 11–37. <https://doi.org/10.1146/annurev-statistics-031219-041104>
- Hardwicke, T. E., Thibault, R. T., Kosie, J., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. (2021). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science*. Advance online publication. <https://doi.org/10.1177%2F1745691620979806>
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6), 645–654. <https://doi.org/10.1177/1745691612464056>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Kühl, T., & Bertrams, A. (2019). Is learning with elaborative interrogation less desirable when learners are depleted? *Frontiers in Psychology*, 10, Article 707. <https://doi.org/10.3389/fpsyg.2019.00707>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kvarven, A., Strömmland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423–434. <https://doi.org/10.1038/s41562-019-0787-z>
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In A. Musgrave & I. Lakatos (Eds.), *Criticism and the growth of knowledge* (Vol. 4, pp. 91–196). Cambridge University Press. <https://doi.org/10.1017/CBO9781139171434.009>
- Lauden, L. (1981). Peirce and the trivialization of the self-corrective thesis. In *Science and hypothesis* (pp. 226–251). Springer.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewis, M. B. (2018). The interactions between botulinum-toxin-based facial treatments and embodied emotions. *Scientific Reports*, 8, 14720. <https://doi.org/10.1038/s41598-018-33119-1>
- Maassen, E., van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLOS ONE*, 15(5), Article e0233107. <https://doi.org/10.1371/journal.pone.0233107>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487–498. <https://doi.org/10.1037/a0039400>
- McDiarmid, A., Tullett, A., Whitt, C. M., Vazire, S., Smaldino, P. E., & Stephens, E. E. (2021). *Self-correction in psychological science: How do psychologists update their beliefs in response to replications?* PsyArXiv. <https://doi.org/10.31234/osf.io/hjcm4>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology’s Renaissance. *Annual Review of Psychology*, 69(1), 511–534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Neuliep, J. W., & Crandall, R. (1993). Everyone was wrong: There are lots of replications out there. *Journal of Social Behavior and Personality*, 8(6), 1–8.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), Article aac4716. <https://doi.org/10.1126/science.aac4716>

- Peterson, D., & Panofsky, A. (2020). *Self-correction in science: The diagnostic and integrative motives for replication*. SocArXiv. <https://doi.org/10.31235/osf.io/96q xv>
- Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in psychology: A meta-review. *Perspectives on Psychological Science*, *15*(4), 1026–1041. <https://doi.org/10.1177/1745691620906416>
- Rohrer, D., Pashler, H., & Harris, C. R. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General*, *144*(4), e73–e85. <https://doi.org/10.1037/xge0000058>
- Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to Registered Replication Reports at *Perspectives on Psychological Science*. *Perspectives on Psychological Science*, *9*(5), 552–555. <https://doi.org/10.1177/1745691614543974>
- Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science*, *25*(6), 1227–1234. <https://doi.org/10.1177/0956797614526415>
- Strack, F. (2016). Reflection on the smiling registered replication report. *Perspectives on Psychological Science*, *11*(6), 929–930. <https://doi.org/10.1177/1745691616674460>
- Strack, F. (2017). From data to truth in psychological science. A personal perspective. *Frontiers in Psychology*, *8*, Article 702. <https://doi.org/10.3389/fpsyg.2017.00702>
- Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the Facial Feedback Hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768–777. <https://doi.org/10.1037/0022-3514.54.5.768>
- Tatsioni, A., Bonitsis, N. G., & Ioannidis, J. P. A. (2007). Persistence of contradicted claims in the literature. *JAMA*, *298*(21), 2517–2526. <https://doi.org/10.1001/jama.298.21.2517>
- Vadillo, M. A. (2019). Ego depletion may disappear by 2020. *Social Psychology*, *50*(5–6), 282–291. <https://doi.org/10.1027/1864-9335/a000375>
- Vadillo, M. A., Hardwicke, T. E., & Shanks, D. R. (2016). Selection bias, vote counting, and money-priming effects: A comment on Rohrer, Pashler, and Harris (2015) and Vohs (2015). *Journal of Experimental Psychology: General*, *145*(5), 655–663. <https://doi.org/10.1037/xge0000157>
- Vazire, S., & Holcombe, A. O. (2020). *Where are the self-correcting mechanisms in science?* PsyArXiv. <https://doi.org/10.31234/osf.io/kgqzt>
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2020). Credibility beyond replicability: Improving the four validities in psychological science. *PsyArXiv*. <https://doi.org/10.31234/osf.io/bu4d3>
- Wagenmakers, E.-J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin-Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin-Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L., DeGicco, J. M., . . . Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*(6), 917–928. <https://doi.org/10.1177/1745691616674458>
- Xie, Y. (2017). *Dynamic documents with R and knitr*. CRC Press.
- Zotero. (2019, June 14). *Retracted item notifications with retraction watch integration*. <https://www.zotero.org/blog/retracted-item-notifications>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, Article e120. <https://doi.org/10.1017/S0140525X17001972>