Meta-Psychology, 2024, vol 8, MP.2022.3308
https://doi.org/10.15626/MP.2022.3308
Article type: Original Article

Open data: Yes
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: Yes

Edited by: Malte Elson
Reviewed by: Hans Ijzerman, Ian Hussey, Ivan Ropovik
Analysis reproduced by: Federico D'Atri
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/ZYNG2

Evaluating the Replicability of Social Priming Studies

Erik Mac Giolla[1 2], Simon Karlsson[2 3], David Neequaye[2 4], and Magnus Bergquist[2]
[1]Department of Nursing, Health Sciences and Integrated Care, Atlantic Technological University, Ireland
[2]Department of Psychology, University of Gothenburg, Gothenburg, Sweden
[3]Institute for Futures Studies, Stockholm, Sweden
[4]Department of Psychology, Lancaster University, Lancaster, United Kingdom

To assess the replicability of social priming findings we reviewed the extant close replication attempts in the field. In total, we found 70 close replications, that replicated 49 unique findings. Ninety-four percent of the replications had effect sizes smaller than the effect they replicated and only 17% of the replications reported a significant *p*-value in the original direction. The strongest predictor of replication success was whether or not the replication team included at least one of the authors of the original paper. Twelve of the 18 replications with at least one original author produced a significant effect in the original direction and the meta-analytic average of these studies suggest a significant priming effect (d = 0.40, 95% CI[0.23; 0.58]). In stark contrast, none of the 52 replications by independent research teams produced a significant effect in the original direction and the meta-analytic average was virtually zero (d = 0.002, 95% CI[-0.03; 0.03]). We argue that these results have shifted the burden of proof back onto advocates of social priming. Successful replications from independent research teams will likely be required to convince sceptics that social priming exists at all.

Keywords: Priming, Replication, Preregistration, Meta-analysis

Introduction

Theorists have proposed that incidental exposure to environmental cues can influence subsequent cognitions, attitudes, and behavior. This phenomenon—dubbed social priming—has produced striking findings in contemporary social psychology. As an example, consider the finding that participants primed with the word *professor* performed 13 percent better on a subsequent trivia test compared to participants who were primed with the words *soccer hooligans* (Dijksterhuis and van Knippenberg, 1998; for reviews see Bargh, 2006; Molden, 2014). One can see why social priming findings have attracted so much attention: If mere exposure to subtle primes can have vast behavioral consequences, social priming could provide an unprecedented tool to influence human behavior. However, recently, researchers have found it difficult to replicate several key social priming findings, despite using large samples and highly similar experimental setups (e.g., Cheung et al., 2016; Pecher et al., 2015 ; Rohrer et al., 2019). These failed replications have sparked extensive debate on the credibility of the field (Newell and Shanks, 2014; Shanks, 2017). We examine the scale of this credibility issue by reviewing the extant close replication attempts of social priming findings.

Social priming is perhaps best understood when contrasted with *semantic priming* (Molden, 2014). Semantic priming builds on cognitive models of spreading activation, which hold that closely related words and concepts are wired more strongly than unrelated ones (Collins & Loftus, 1975; Doyen et al., 2014). Semantic priming thus concerns construct activation and accessibility. Accordingly, a person primed with the word *nurse* will generally be faster at recognizing the subsequent word *doctor* compared to a person primed with the unrelated word *cat* (Neely, 1977). The activation of the mental representation nurse will spread and thus lower the amount of stimulation necessary for reaching the threshold to recognize and process the related concept *doctor*, more so than the word *cat*.

Social priming also holds that primes increase the accessibility of related constructs (Molden, 2014). For example, a person primed with words related to old age should show an increased activation and accessibility of mental concepts related to old age. However, social priming goes a step further and contends that this increased accessibility can spill over into behavioral outcomes through a mechanism called the perception-behavior link (Dijksterhuis & Bargh, 2001). The perception-behavior link led researchers to predict that priming people with concepts related to old age would not only increase the accessibility of related con-

cepts, but would lead them to walk more slowly (Bargh et al., 1996). Our review focuses on this type of social priming that goes beyond mere construct activation, and claims that subtle primes can elicit cognitions, attitudes, and behaviors. [1]

Replication studies also have their nomenclature. One important distinction is between *close* (sometimes called *direct*) and *conceptual* replications. While close replications stay as close as possible to the original study's methodology, conceptual replications depart from the original methodology in some meaningful way (Brandt et al., 2014). Close and conceptual replications perform different functions. Conceptual replications are central to advancing theory by examining whether a phenomenon will hold in analogous contexts. In contrast, close replications help assess the credibility of a single finding. If researchers cannot closely replicate a single finding, other explanations for the effect, such as measurement error, cannot be ruled out (Simons, 2014). Here, we review close replication attempts of social priming findings, reasoning that if a single close replication assesses the credibility of a single finding, then close replications of many findings assess the robustness of a phenomenon.

A more typical way to assess the robustness of a phenomenon is to synthesize all research on a topic through a traditional meta-analysis. Indeed, a recent and comprehensive meta-analysis on social priming concluded that "the priming effect is a *real* psychological phenomenon that remains robust when using the most advanced bias detection methods" (Dai et al., 2023, p.86, emphasis in original). However, if the studies included in the meta-analysis are severely biased—for instance, because of publication bias or a widespread use of questionable research practices—then the meta-analysis will be severely biased too. This limitation will lead to inflated estimates of effect size (Bakker et al., 2012). Meta-analytic statistical approaches to correct for suspected bias can lead to more accurate measures of effect size, and Dai et al., did include several such bias correction techniques. However, no bias correction method is perfect, and their performance will depend on unknown contextual factors such as the prevalence of questionable research practices in the field (Carter et al., 2019). To rectify the issue of a biased literature, others have called for a greater focus on large-scale replications, at the expense of traditional meta-analyses (e.g., Kvarven et al., 2019; van Elk et al., 2015). The argument being that large-scale replication attempts are the best source of unbiased estimates of effect size (Kvarven et al., 2019). By synthesizing the extant close replication attempts of social priming studies, we hope to utilize the respective strengths of both meta-analyses and replications.

Just as there is no single correct way to replicate a study, neither is there a single measure to assess if a replication attempt is successful. The Open Science Collaboration (2015)—who replicated 100 findings in psychology—used several measures of replication success. They focused on the statistical significance, direction, and size of the effect observed in the replication study compared to the effect observed in the original study. Specifically, they examined whether the effect observed in the replication study was in the same direction as the original effect, and, if so, whether the p-value of the replication effect was significant ($p < .05$). In addition, they examined whether the effect size in the replication study was smaller than the original effect size and whether the 50% confidence interval of the replication effect size included the original effect. We adopt a similar analysis plan (see below for details).

In addition to this primary analysis plan, we also quantified the evidence for priming effects in replication studies using a meta-analytic approach. Replication studies can fail for many reasons other than the original study being a false positive (e.g., if the replication study has insufficient power to detect the effect in question). A meta-analytic approach allows us to estimate the average size of priming effects reported in replication studies and to examine if any moderators influence these effects. Analysts note that preregistered replication studies may present less biased estimates of effect sizes (Kvarven et al., 2019; van Elk et al., 2015). Thus, a meta-analysis of preregistered replication studies should provide more accurate estimates of priming effects and ultimately address whether social priming exists at all.

We were also interested in potential moderators of replication success and priming effects. These moderators can be categorized into attributes of the original study and attributes of the replication study (for a similar strategy see Open Science Collaboration, 2015). For attributes relating to the original study, we focused on the effect size, the sample size, and the *p*-value of the original study. All things being equal, larger sample sizes and larger effect sizes increase the precision of effect size estimates, reducing the risk of false-positives (Lakens & Evers, 2014). Assuming the replication study is appropriately conducted, more precise estimates of effect size in the original study should increase the chances of a successful replication. Similarly, smaller p-values in the original study should increase

[1] Theorists have proposed other mechanisms of priming besides the perception-behavior link (see Molden, 2014). Nonetheless, all those theories essentially claim that priming can influence behavior via construct accessibility.

the chances of successful replication, since smaller p-values reduce the risk of the original finding being a false-positive (Benjamin et al., 2018).

For attributes relating to the replication study, we focused on the sample size of the replication, whether the replication attempt was preregistered, and whether the replication team was independent from the researchers of the original study. The replication studies' sample size follows a similar logic to the previously mentioned moderators. Namely, that all things being equal, larger sample sizes increase the chances of reliably detecting a real effect. The two remaining moderators—preregistration and researcher independence—concern meta-scientific discussions on replicability.

One suggested explanation for the replication crisis relates to the incentive structure of academia (Bakker et al., 2012; Grimes et al., 2018). The argument goes that in academia publications are hard currency, integral to career advancement, resulting in many researchers facing considerable pressure to publish. Furthermore, scientific journals have a tendency to favor the publication of significant (p <.05) findings (e.g., Grimes et al., 2018; Francis, 2012; Kühberger et al., 2014). Hence, researchers may engage in questionable research practices (QRPs) to meet the implicit criteria for publication. Surveys on the prevalence of QRPs lends credence to the argument just described (John et al., 2012; but see Fiedler and Schwarz, 2016). Preregistering a study should reduce the prevalence of QRPs by constraining researchers' degrees of freedom, thereby inhibiting their ability to engage in QRPs. Insofar as researchers may be incentivized to obtain successful replications, preregistered replications may provide less biased rates of replication success and less biased estimates of effect size.

Our final moderator—the independence of the replication team—is arguably of most theoretical interest. One explanation put forward to account for recent replication failures within social priming is that priming effects are highly sensitive to contextual variations (Cesario, 2014). Since the current state of knowledge does not encompass a complete understanding of these conditions, some analysts claim that failures to directly replicate other researchers' findings will be uninformative in regard to what inferences one can make (e.g., Cesario, 2014; Stroebe & Strack, 2014). That is, we cannot know whether a failed replication is because the original finding was a false positive or whether it was due to some unknown moderator present during the replication attempt.

Contrasting that view, Simons (2014) argues that if an effect can only be replicated by the same research team operating under some special but unknown moderators—that cannot be specified and tested empirically—it is questionable what value the finding brings to our current state of knowledge. Instead, Simons suggests that self-replication is a suitable method for providing initial evidence for an effect, but the verification of an effect can only come from independent replications. Following Simons' argument, we maintain that if independent replications are substantially less likely to replicate than same-lab replications, the credibility and generalizability of the original effect remains in doubt.

We contribute to the ongoing discussion about the credibility of social priming by reviewing the replicability of social priming findings. Our main objective was to evaluate the success of close replications by examining their statistical significance and effect size. We hoped to overcome the limitations with evidence provided by isolated replication attempts by reviewing close replications of many social priming findings. Additionally, to uncover why some studies may be more replicable than others, we investigated the moderating role of study characteristics related to both the original and replication studies.

Method

Transparency and Openness

The research question, search strategy, inclusion criteria, and original analysis plan were preregistered https://osf.io/dc94y. Additionally, we preregistered an extended version of this original analysis plan, and the exact code for the statistical tests https://osf.io/3btse. The extended analysis plan and the code were written and preregistered during data extraction, but before any statistical analyses of the full data set were conducted. In addition to this original analysis plan, we conducted analogous mixed-effects analyses, as well as a meta-analysis. The mixed-effects analyses and the meta-analysis were not pre-registered. Data were analyzed using R, version 4.0.0 (Team, 2019), the package ggplot2 (Wickham, 2016), and the package metafor (Viechtbauer, 2010). All data and code to perform all analyses are available at https://osf.io/ndxf4/.

Inclusion Criteria

The studies included in the analysis had to fit four criteria. First, the study should be what we refer to as an "explicit replication", meaning that the authors of the replication attempt should explicitly state the original experiment they are trying to replicate (e.g., "Bargh et al., 1996, Experiment 1"). Second, the study should be a 'close' or 'exact' replication (Brandt et al., 2014). Put differently, the authors of the replication attempt should have the goal of following—as exactly as possible—the

methods of the original study in terms of stimuli, measures, and procedures. Third, the original experiment (and by extension the replication) should examine what we have defined as *social priming*. That is, the studies should examine priming effects on cognitions, attitudes, or behaviors, not merely increased construct accessibility. Finally, the original study, but not necessarily the replication attempt, should be published in a peer-reviewed journal. We included this final condition so that we could ensure that the original study was accessible.

Search Strategy

To find relevant literature, during the spring 2020 we searched for empirical studies in the database PsycINFO (default setting) using the following search string: *replicat* AND (prime OR priming OR primed OR automatic OR automatically OR nonconscious* OR incidental* OR "embodied cognition")*. The search terms designating social priming were inspired by those used by Weingarten et al. (2016). Additionally, when scanning the articles identified in our original database search, there were instances where researchers referred to previous replication attempts within the field of social priming. These articles were noted and later exposed to the same scrutiny as the articles identified in the original database search. To search for grey literature, we assessed oral- and poster presentations accepted for the conference Association of Personality and Social Psychology from 2003 to 2021. Specifically, we searched the term "replicat*" in the event scheme. For records lacking sufficient statistical information for inclusion, we contacted the first authors asking for a) means, standard deviations, and n / cell for both the control and priming conditions, or b) raw data. The grey literature search generated thirty replications of social priming studies, of which two fulfilled the criteria for inclusion.

An overview of the study selection process is provided in Figure 1. The original search yielded 2,342 hits after duplicates were removed. In addition, we added 14 records from the reference lists of included studies and 14 from searching in grey. The exclusion criteria were then applied in a consecutive manner by four independent coders. First, based on title and abstract screening, articles were removed that did not satisfy our definition of social priming. Second, based on full-text screenings, articles were removed that did not fulfill our criteria of being a replication. After removing articles that did not satisfy the preregistered requirements for inclusion, 67 studies were deemed eligible for data extraction. However, during data extraction additional articles had to be removed due to issues not foreseen before preregistration. Such issues included studies that did not provide sufficient information to calculate the statistics necessary for the analyses and studies using within-subject designs. As noted by Pashler et al. (2012), most within-subject priming studies do not provide information about the variability of outcome measures across participants, which is a necessary component for calculating a within-subject's variant of Cohen's d. A further 29 studies were excluded during data extraction, yielding a final sample of 36 studies (+ 2 from grey literature search), containing 65 replication attempts. After posting the paper as a pre-print (as required by *Meta-Psychology*), we were contacted by four researchers suggesting, in total, an additional 4 studies. Two of these met the inclusion criteria and contained a further 5 replication attempts. This brought our final sample to 38 studies containing 70 replication attempts (see the Data and Analysis folder on OSF for additional information https://osf.io/ndxf4/).

Analysis plan

Measures of Replication Success

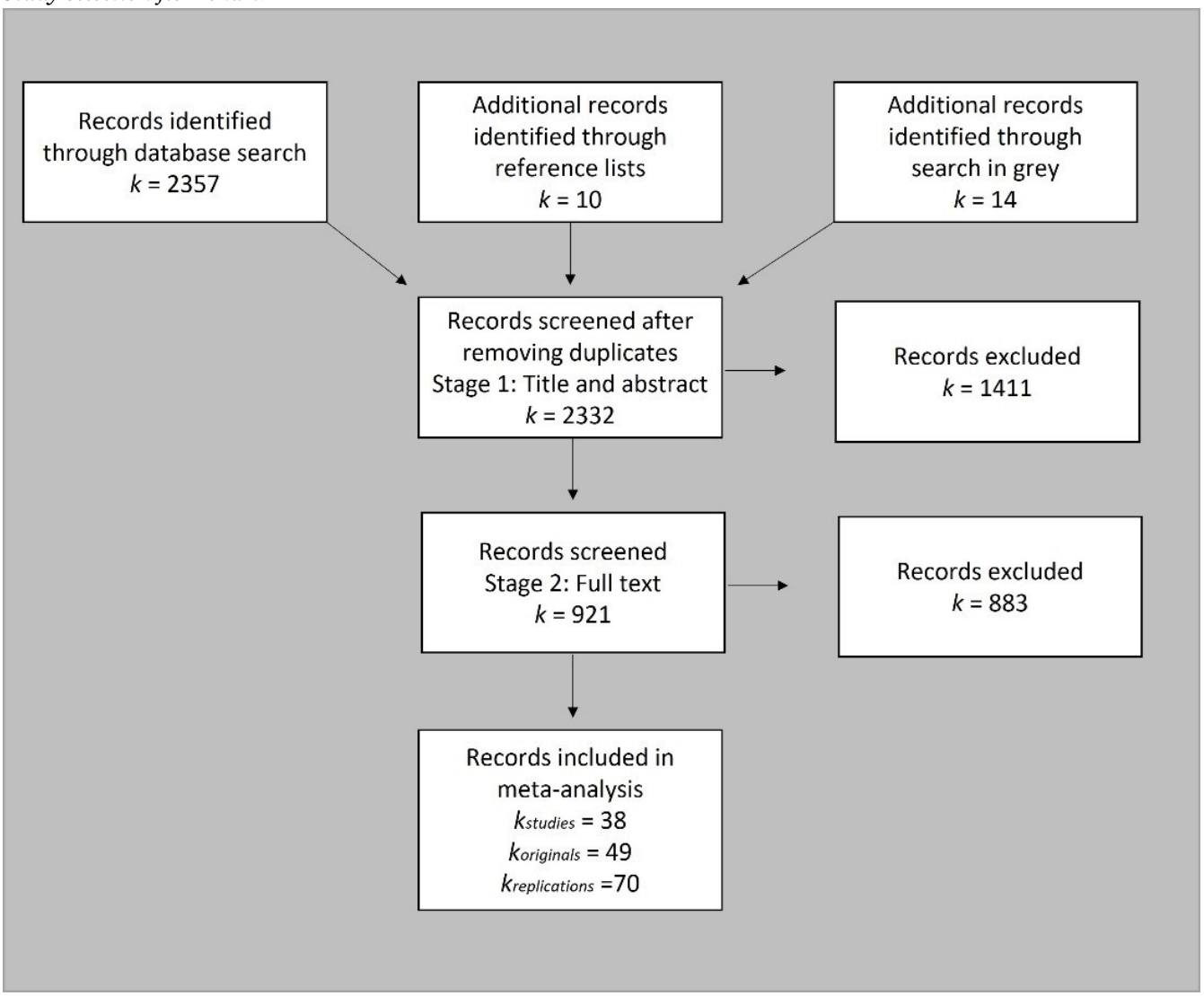
Our analyses of replicability were broadly similar to the approach taken by the Open Science Collaboration (2015). To assess replicability, we followed a preregistered analysis plan that focused on: (i) an analysis of *p*-values; (ii) an analysis of effect sizes; and (iii) an analysis of moderating variables. These are discussed in turn.

To examine the consistency between *p*-values obtained in the original studies and replication studies, three primary analyses were conducted. First, we performed a McNemar's test to determine whether the proportion of significant effects in the replication study (replication studies with $p < .05$ / total number of replication studies) significantly differed from the proportion of significant effects in the original studies (original studies with $p < .05$) / total number of original studies). For this test, studies were categorized as successful if the effect was in the hypothesized direction and p (two-tailed) $\leq .05$. The analysis was identical to that of the Open Science Collaboration (2015), with one exception. On occasion, our recalculated *p*-values of the original study were greater than .05, but were interpreted by the original authors as statistically significant. We treated these as non-significant original effects in our analyses. However, when this occurred in the Open Science Collaboration study, they were treated as significant effects.

Second, we performed a Wilcoxon signed-rank test and a paired *t*-test to examine whether the median and average *p*-value, respectively, differed between original studies and replication studies. Regarding the tests in-

Figure 1

Study selection flow chart.



volving raw *p*-values, we had not anticipated that so many of the replication effects would be in the opposite direction to the original effect. Since *p*-values do not take an effect's direction into account, the planned comparison between *p*-values of original studies and replication studies became problematic. To deal with this issue, we recoded replication *p*-values with an effect in the opposite direction as 1.

We conducted four analyses to examine the consistency between effect sizes obtained in original studies and replication studies. Specifically, we conducted a Wilcoxon signed-rank test and a paired *t*-test to examine whether the median and average effect size (Cohen's *d*) differed between original studies and replication studies. We then conducted a binomial test to explore whether the proportion of pairs in which the effect size (Cohen's *d*) was stronger in original studies compared to replication studies was larger than would be expected by chance (i.e., 50% probability). We also examined the proportion of study pairs in which the original effect was within the 95% CI of the replication effect size. Though it should be noted that this measure of replication success is not without criticism (Morey & Lakens, 2016; Patil et al., 2016). Specifically, we used a χ^2 goodness of fit test to examine whether the proportion of original studies within the 95% CI of the replication effect size differed from the expected proportion. The expected proportion was calculated using

the R function developed by Open Science Collaboration (2015). In this analysis, the expected proportion is the sum of expected probabilities across study-pairs. The test assumes that the original study and its replication attempt have the same population effect size. For details see the analysis code on https://osf.io/ndxf4/.

Finally, we analyzed the influence of moderating variables by including them as independent variables in correlations predicting replication success (for a similar approach see Open Science Collaboration, 2015). In these analyses, replication success was defined by two different measures: (i) $p \leq .05$ in the original direction; and (ii) the difference in effect size between the original study and the replication attempt. The following six moderators were examined: (1) the effect size obtained in the original study (Cohen's d); (2) the sample size in the original study; (3) the p-value in the original study; (4) the sample size in the replication study; (5) whether or not the replication study was preregistered; and (6) whether or not the replicating lab was independent to the originating lab. Regarding moderator number 6, replication attempts with at least one author included in the original study were coded as same-lab replications, whereas studies with no authorship overlap were coded as independent replications. We deviated from our pre-registered analysis plan in one aspect. In the pre-registration we only specified Spearman's rank ordered correlation (ρ) for these analyses. However, this analysis is not appropriate when we have one dichotomous variable and one continuous variable (e.g., when predicting if $p \leq .05$ was in the original direction from the original study sample size) or two dichotomous variables (e.g., when predicting if $p \leq .05$ was in the original direction based on whether the replication was preregistered or not). In these situations we used Glass' rank biserial correlation (r_{rb} or the Phi Coefficent (ϕ), respectively.

To facilitate interpretation, all original social priming effects were coded as positive. Consequently, if a replication attempt found an effect in the opposite direction to the original finding, the effect was coded as negative.

Mixed-Effects model

When preregistering our analysis plan, we did not anticipate that some original studies would be replicated more than once. This created a dependency in the data that violated the assumption of independence relevant for our planned analyses. To account for this dependence, we ran, when possible, mixed-effects models analogous to our planned analyses, where we included the original study as a random effect (specifically a random intercept). We used the R package lme4 (Bates et al., 2015) to calculate all mixed effects models and the R package lmerTest (Kuznetsova et al., 2017) to calculate associated p-values.

Meta-analytic Approach

We meta-analyzed the effect sizes of the replication studies using the Metafor package (Viechtbauer, 2010) for R (Team, 2019). We opted for a random effects model as we expected substantial between-study variance due to differences in populations and experimental procedures (Cooper, 2010). To analyze potential moderating effects, we used the same six moderators discussed above. Furthermore, to account for the dependency in the data set (replication studies nested in original studies), we initially ran a multi-level meta-analysis with two levels, where we included the original study as the higher order level, and the replication study as the lower order level. However, the sigma squared for the higher order level was zero; hence, it did not account for any additional variance over the lower order level. For this reason, we report results from the simpler single level meta-analysis at the level of replication study. This analysis was proposed by an anonymous reviewer, and hence was not preregistered.

Results

Characteristics of the Literature

In total, 70 replication attempts were included in the analyses, replicating 49 unique original studies. For an overview of all replication attempts and original studies see supplemental materials S1. Sixty-eight of the replication studies came from articles, published between 1996 and 2019. The remaining two studies came from unpublished conference presentations. In total, 30 372 participants were included in the analyses of the replication attempts, compared to 3087 in the original studies. The average sample size of the replication studies was 434 (Mdn = 126, SD = 1098), compared to 63 (Mdn = 57, SD = 33) in the original studies.

Analyses of p-values

All 49 original social priming effects were interpreted as statistically significant (i.e., $p \leq .05$) by the original authors. However, the recalculated (two-tailed) p-values indicated that only 41, 85.71%, of the original findings showed significant priming effects. Worth noting is that every recalculated non-significant p-value among original studies balanced on the .05 threshold, ranging from .052 to .079. For replication studies, 12 experiments, 17.14%, showed statistically significant effects in the same direction, a significant reduction compared to original studies (McNemar's test, $\chi^2(1) = 46.08, p = < .001$).

For the analogous mixed-effect analysis, we used a mixed-effects logistic regression. Specifically, the dichotomous variable *p*-value (significant vs. non-significant) was predicted by the dichotomous variable study type (original vs. replication), with original study as a random effect. The results were comparable to the planned analysis: study type was a significant predictor of *p*-value, $b = 14.15$, OR < .001, $p < .001$, suggesting that replication attempts produced significantly fewer significant p-values compared to the original studies (full models are available in the supplemental material S2).

One possibility is that replication attempts tended to end up on the "wrong" side of the significance threshold by a close margin. To account for the limitations associated with classifying data based on an arbitrary cut-off point, additional analyses were conducted using raw *p*-values. Both the paired *t*-test and Wilcoxon signed-rank test indicated that original studies ($M = .023$; $Mdn = .017$; $SD = .018$) obtained significantly lower *p*-values than their associated direct replication attempts ($M = .593$; $Mdn = .630$; $SD = .400$), $t(69) = -11.90$, $p < .001$ and $W = 625$, $p < .001$.

For the analogous mixed-effects analysis, we used a mixed-effects linear regression where the continuous variable *p*-value was predicted by the dichotomous variable study type (original vs. replication), with original study as a random effect. The results were comparable to the planned analyses: study type was a significant predictor of *p*-value, $b = .567$, p < .001, suggesting that replication attempts produced significantly larger *p*-values compared to the original studies (full models are available in the supplemental material S2).

There are limitations in what can be said about the evidence for social priming effects based on comparisons between original studies and replication attempts. Perhaps, true social priming effects are substantially smaller than indicated by original studies, resulting in power issues to detect the true effect among replication attempts. One method for estimating the evidence for the effect without using the original findings as a benchmark is to examine the distribution of non-significant p-values among replication attempts. If the null hypothesis is false, and there is a true social priming effect to detect, we would expect the distribution of non-significant p-values to deviate from uniformity and be weighted towards zero. Contrary to this prediction, the distribution of non-significant *p*-values among replication attempts did not show signs of deviating from uniformity, $\chi^2(100) = 109.90$, $p = .538$. See Figure 2A for a graph depicting the distribution of *p*-values.

Analyses of Effect Sizes

Effect sizes were measured in Cohen's *d*. Overall, effect sizes of original studies ($M = 0.72$; $Mdn = 0.70$; $SD = 0.19$) were substantially larger than those of replication studies ($M = 0.14$; $Mdn = 0.06$; $SD = 0.32$), $t(69) = 15.33$, $p < .001$ and $W = 4489$, p < .001. For the analogous mixed-effects analysis, we used a mixed-effects linear regression where effect sizes were predicted by the dichotomous variable study type (original vs. replication) with original study as a random effect. The results were comparable to the planned analysis: study type was a significant predictor, $b = -0.556$, $p < .001$, suggesting that replication attempts produced significantly smaller effect sizes compared to the original studies (full models are available in the supplemental material S2). Out of the 70 study pairs included in the analysis, 66 study pairs showed an original effect size larger than that of the replication. This proportion was significantly different from what would be expected by chance (94.29%, $p < .001$, binomial test). We had no analogous mixed-effects analysis for the binomial test.

The 95% CI of the replication effect sizes overlapped with the original study in only 17 of 70 study pairs, or 24.29%. This finding was significantly lower than the expected proportion of 65.32%, $p < .001$. Due to the calculations that were required to estimate the expected proportion, we had no analogous mixed-effects version of this analysis.

Finally, the original effect sizes were moderately related to their corresponding replication effect sizes, Pearson's $r = 0.31$, $p = .009$, but only weakly so, when using a non-parametric equivalent, Spearman's $\rho = 0.10$, $p = .426$.

Moderation Analyses

Moderators relating to the characteristics of the original studies

An overview of our moderation analyses is presented in Table 1. Considering the characteristics of original studies, larger original effect sizes were related to greater effect size differences between original and replication studies ($\rho = .42$). Since the effect estimates among replications were generally low, this result could simply be a consequence of high original effect estimates increasing the distance to the replication effect. Moreover, larger original sample sizes were associated with smaller effect size differences between original and replication studies ($\rho = -.35$). This relationship could result from effect estimates generally being higher among original studies with small sample sizes. Last, studies with lower original p-values were

Table 1

Relationship between study characteristics and replication success. For the correlations, we used Spearman's rank order correlation (ρ), Glass' rank biserial correlation (r_{rb}), or the Phi coefficient (ϕ).

	Replication p-value		Effect size difference		
	≤ .05	> .05	Correlation		Correlation
	Mdn	*Mdn*			
Original study characteristics					
Effect size	0.70	0.66	.01		.42***
Sample size	60	45.5	.34		–.35**
p-value	.012	.021	-.33		–.14
Replication study characteristics					
Sample size	60	132	-.48**		.15
Independence of replication team	Same: k = 12	Same: k = 6		*Mdn* same = 0.24	
	Indep.: k = 0	Indep.: k = 52	–.76 ∗ **[a]	*Mdn* indep. = 0.68	.64***
Preregistration	Yes: k = 1	Yes: k = 26		*Mdn* Yes = 0.57	
	No: k = 11	No: k = 32	-.26*†	*Mdn* No = 0.66	.18

Note. $P \leq .05$ *, $P \leq .01$ **, $P \leq .001$ ***. [a]*The analogous mixed effects model for this analysis would not converge because there was a zero in the contingency table, as there was no independent replication with a significant p-value in the correct direction.* †*This effect was not significant in the analogous mixed effects model.*

somewhat more likely to be significant in the replication study (r_{rb} = –.33), though this result was not statistically significant. This suggests that lower original *p*-values were somewhat more likely to replicate.

Moderators relating to the characteristics of the replication studies

We turn now to the characteristics of replication studies. Replications with larger sample sizes were less likely to produce a significant *p*-value (r_{rb} = –.48), suggesting that increased power and precision in the replication attempt did not favor replication success. The single strongest predictor of replication success was whether there was an authorship overlap between the replication study and the original study. Replication attempts conducted by researchers unaffiliated to the original finding were substantially less likely to obtain significant p-values in the original direction (ϕ = -.76) and deviated more from original studies in their effect sizes (r_{rb} = .64). Finally, replication attempts that were preregistered were less likely to obtain significant effects (ϕ = -.26), indicating that preregistration lowered the chance of successfully replicating the original study.

Summary of moderation analyses

Taken together, these analyses suggest that the independence of the replicating lab is the single most important factor to consider when predicting replication success of social priming findings. To illustrate the moderating role of study characteristics related to the replicating lab, Figure 2 provides an overview of *p*-values and effect sizes highlighting researcher independence, preregistration, and sample size. As illustrated in Figure 2A, for same-lab replications the distribution of *p*-values is weighted towards zero, and 12 out of 18 replication attempts show a significant effect in the original direction. In stark contrast, of the 52 replication attempts conducted by independent labs, not a single replication attempt yielded a statistically significant effect in the original direction. Furthermore, the *p*-values among independent replications were widely distributed and did not show any clear signs of deviating from a uniform distribution.

Figure 2B shows how the effect size patterns are markedly different between original studies, same-lab replications, and independent replications. Original studies and same-lab replications showed moderate to substantial social priming effects. This pattern was not reflected among replications conducted by independent labs, where the distribution of effect sizes is approximately normally distributed around zero.

Analogous mixed-effects analysis of moderators

For the analogous mixed-effects analyses, we used the moderators as predictor variables in mixed-effects logistic regressions (when predicting whether *p* was significant or not) and mixed-effects linear regressions (when predicting the difference in effect size between original and replication studies). For all of these analyses, original study was included as a random effect.

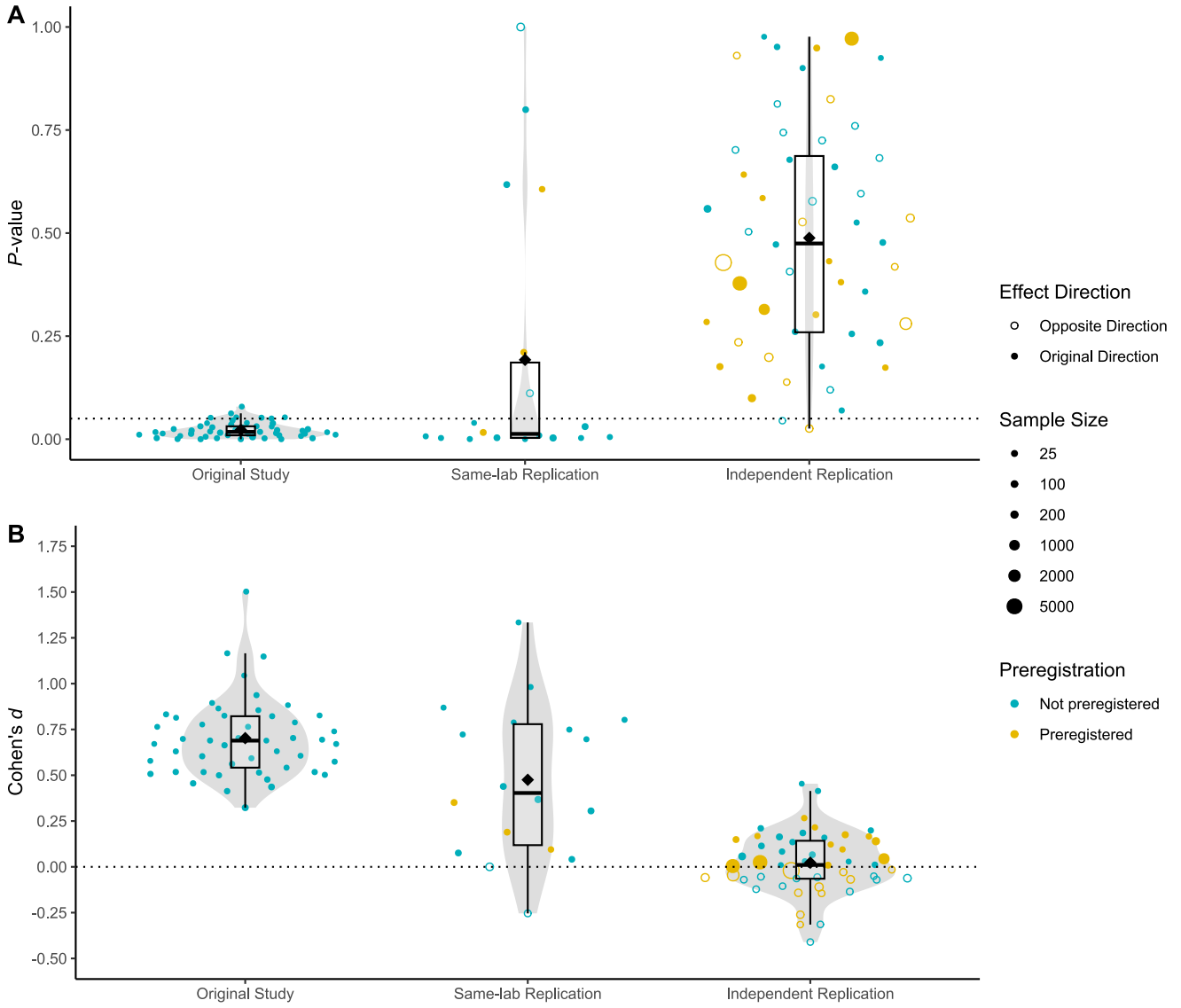


Figure 2

Beeswarm boxplots illustrating the distribution of p-values (A) and effect sizes (B). The horizontal dotted line in (A) equals 0.05. The horizontal dotted line in (B) equals zero. Diamonds equal unweighted means and the horizontal line in the boxplots equal unweighted medians.

The results were broadly comparable to the planned analyses. Again, larger original effect sizes were related to greater effect size difference between original and replication studies (b = 0.45, p = .039); larger sample sizes were associated with smaller effect size differences between original and replication studies (b = -0.003, p = .023); replication studies with larger samples were less likely to produce significant effects in the original direction (b = -83.41, OR < 0.001, p < .001); and replications from independent labs showed significantly smaller effect sizes (b = 0.43 , p < .001).

There was only one substantive difference between the planned analyses and their mixed-effects equivalents. In the planned analyses, a significant relationship between preregistration and the p-value of the replication study was found, showing that preregistered studies were less likely to produce a significant effect in the original direction. This effect, though in the same direction, was not significant in the analogous mixed effects analysis (b = -2.13, OR = 0.119, p = .662.

Furthermore, the mixed-effects logistic regression, where lab-independence was used to predict whether the replication *p*-value was significant, did not converge. The reason the model did not converge was that there was a zero in the contingency table: none of the independent replications had a significant *p*-value in the correct direction (full models are available in the supplemental material S2).

Meta Analytic Approach

Results for the random effects meta-analysis showed a small but statistically significant priming effect in the replication studies (see Table 2). However, there was considerable heterogeneity in effect sizes, as indicated by the large I^2 and wide prediction intervals. For this reason, we conducted a series of moderation analyses, using the same six moderators discussed above. For a forest plot depicting the effect sizes of the original studies and their respective replications see Supplemental Materials S3 and for funnel plots, centered on zero, depicting the original studies and the replication studies see Supplemental Materials S4.

The effect size, sample size, and *p*-value of the original study had no notable moderating effects: the moderators were not significant and accounted for less than 1% of the heterogeneity in priming effects. Nor was the sample size of the replication study a significant moderator, accounting for only 2.5% of heterogeneity in priming effects. The two remaining moderators—preregistration and lab independence—were significant (see Table 2).

Whether a replication was pre-registered or not accounted for approximately 26% of heterogeneity in effect sizes. Replication studies that were not preregistered displayed a small meta-analytic priming effect, $d = 0.12$, 95% CI[0.06; 0.19]. Pre-registered replication studies produced a notably smaller effect, with the 95% CI including zero, $d = 0.02$, 95% CI[-0.13; 0.17]. The strongest moderator by far was whether the replication team contained an author from the original study (same lab) or not (independent lab). This moderator accounted for over 98% of heterogeneity in priming effects. Results show that replication studies conducted by the same lab produced small to moderate priming effects, $d = 0.28$, 95% CI[0.20; 0.37]. In contrast, replication studies conducted by independent labs showed no indication of priming effects, $d = 0.003$, 95% CI[-0.17; 0.17].

Figure 2 shows signs of heteroscedasticity between effect sizes from same labs and independent labs, with the variance of effect sizes being notably smaller for replications conducted by independent labs. For this reason, we computed meta-analytic estimates for same and independent labs using separate τ^2 for each subgroup, rather than a pooled τ^2. This approach is preferred when heteroscedasticity between subgroups is large (Rubio-Aparicio et al., 2019). In essence, this analysis equates to calculating separate meta-analytic models for each subgroup. The results are in line with those reported above: replication studies conducted by the same lab found small to moderate priming effects, whereas those conducted by independent labs showed no indication of priming effects. Of note, are the narrow prediction intervals for the independent lab analysis and the extremely low τ and I^2 (see Table 2). These findings indicate that virtually all of the variance in effect size estimates from replications by independent labs can be accounted for by sampling variation alone.

Discussion

Our aim was to assess the replicability, and by extension the credibility, of social priming findings. We conducted this assessment by reviewing extant close replication attempts of social priming studies. The results could hardly be clearer: although some researchers successfully replicated their own findings, we did not find a single successful replication by an independent team of researchers. None of the independent replications produced a significant effect in the predicted direction, despite having considerably larger samples sizes than the original studies. The *p*-values of the independent replications were uniformly distributed, consistent with a world in which the effects replicated were null. Furthermore, all independent replication attempts produced effect sizes smaller than their corresponding original study, and, more importantly, the average effect was approximately zero and virtually all of the variation in effect sizes could be accounted for by sampling variation alone. Our results paint a sorry picture of the state of social priming for anyone who believes that independent replications are central in establishing the authenticity of a finding (Simons, 2014). Below we discuss previous criticisms that have been levied against failed replications in social psychology and their relevance to the current findings.

Critics may suggest that the independent labs lacked the flair, intuition, and skill, of the original researchers (cf. Baumeister, 2016). Considering the sheer number of researchers who were involved in these independent replications—over 160 researchers—we find such a position unconvincing. If true, then the ability to find an effect of social priming must be a very rare skill indeed.[2]

[2]We would also like to draw the reader's attention to the recent correspondence between Dijksterhuis and Schimmack suggesting an alternative to the skill and flair argument.

Table 2

Meta-analytic estimates of replication studies

	k	d	95% CI	95% PI	τ	I^2
Basic Model	70	0.07	0.03; 0.12	-0.17; 0.32	0.12	59.46%

Moderation Analyses		Qm(df=1)	p	Heterogeneity accounted for
Original Study characteristics				
	Effect size	2.68	.102	<1%
	Sample size	0.38	.540	<1%
	p-value	.02	.895	<1%
Replication Study characteristics				
	Sample size	2.17	.141	<2.54%
	Lab (same vs. independent)	40.59	<0.001	98.54%
	Preregistered (no vs. yes)	5.26	.022	26.31%

Subgroups analysis using separate τ^2

	k	d	95% CI	95% PI	τ	I^2
Same Lab	18	0.40	0.23; 0.58	-0.24; 1.05	0.32	75.39%
Independent Lab	52	0.002	-0.02; 0.03	-0.02; 0.03	<0.01	<0.01

Note. k = number of studies included, d = effect size Cohen's d, 95% CI = the 95% confidence interval of the mean effect size, 95% PI = dispersion of effect size in 95% of all comparable populations, τ = estimated of the standard deviation of the distribution of the true effect sizes, I^2 = ratio of true heterogeneity to total observed variation, Q_m = test statistic for the omnibus test of moderators.

A related criticism is that priming effects are subtle and context dependent (Cesario, 2014; Stroebe & Strack, 2014). As such, failed replications do not cast doubt on the original finding, but rather uncover boundary conditions of the effect (Asendorpf et al., 2013). We sympathize with this perspective. However, in the current context, it would mean that all 52 independent replication attempts happened to strike upon a hidden moderator that negates the original effect. If the research community is to be convinced by that argument, original researchers must better specify the circumstances under which a finding can reasonably be expected to replicate. Otherwise, the infinite number of potential hidden moderators will render it virtually impossible to identify genuine false positive findings in the literature. In its extreme, ascribing all failed replication attempts to reasons beyond the robustness of the effect itself will result in unfalsifiable scientific claims.

Stroebe (2019) proposed yet another set of criticisms: 1) experimental procedures in original studies are typically not complete and 2) following the exact same experimental procedures can become a weakness if manipulations are sensitive to culture and context. Again, we sympathize with these points. Nevertheless, for these criticisms to explain the current results, it would once again mean that all 52 independent replication attempts failed to induce social priming to the extent that an originally strong effect was nullified.

A broader criticism is that a failed replication of an auxiliary hypothesis does not say anything about the theory (Stroebe, 2019). This is a valid critique of isolated replication attempts where only one auxiliary hypothesis is tested. However, our review assess replication attempts of at least 49 auxiliary hypotheses, 35 of which were tested by research teams independent from the original authors. Considering that none of these independent replications were successful, this finding would appear to provide some of the strongest evidence to date against social priming theory.

One could still argue that the body of evidence supporting social priming theory is much more than the

Namely, that many social priming effects may not replicate because of the original researchers' liberal use of questionable research practices, such as failing to report results when predicted effects did not emerge (see Dijksterhuis' reply to Schimmack available at, Schimmack, 2021).

35 auxiliary hypotheses addressed by the independent replications we reviewed (Dai et al., 2023). These 35 findings were not randomly sampled from the universe of social priming studies and may have been chosen by the replication teams because they were skeptical of the original results. Such a claim suggests that the failed replications are unrepresentative of social priming studies and, in fact, may have represented particularly fragile priming effects. Hence, the replications by independent research teams may only say something about these 35 findings, and little about the grand theory of social priming. We maintain that our results say more than this. If nothing else, the results have shifted the burden of proof back onto advocates of social priming. To convince skeptics that social priming exists at all, at least *some* successful independent replications will surely be required.

Limitations

We did not explicitly assess preregistration deviations when sampling studies for this review. Recent research suggests that authors sometimes deviate from preregistration plans without reporting those deviations (e.g., Claesen et al., 2021). Hence, our analysis of the influence of preregistration on replication might suffer limitations we cannot identify at this time. We encourage future investigations to conduct their preregistration quality checks in accordance with the scope of their objectives.

One might also question the objectivity of meta-analyzing social priming effects by citing the broader conceptual question: what counts as social priming? That question is an issue stakeholders continue to grapple with (Molden, 2014). Determining a consensus definition of social priming that all would agree upon is beyond the scope of this study. Thus, an inevitable limitation of our review—and any meta-analysis of social priming (e.g., Dai et al., 2023)—is that a critic can highlight any number of studies included in our analysis as unfit. Conversely, the critic can cite an unpredictable range of reasons to support a claim that we should have included their preferred list of studies. Such an objection is yet another species of concern, besides our findings, pointing to the elusiveness of social priming as a robust phenomenon.

Molden (2014) notes that there is much diversity in the social priming effects researchers have examined. "[. . .][A]ny classification of such effects with the common label social priming can only broadly characterize this area of research rather than enumerate necessary and sufficient criteria that precisely define any related phenomenon (p. 6)." Therefore, our sampling strategy was pragmatic, consisting of studies wherein authors *explicitly* claim to examine the effect of incidental stimuli exposure on cognitions, attitudes, or behavior. Even with such a broad sampling strategy, a single successful replication by an independent team failed to emerge.

Limitations can also be raised regarding our search strategy. Examples include using one database (PsychInfo), not issuing public calls in relevant channels for grey literature, and limited search terms (e.g., not searching for synonyms of replicability, such as 'reproducibility'). These limitations mean that our review may have missed some applicable studies. While such limitations are worth mentioning, it seems unlikely that we missed a sufficient number of studies to change the clear pattern observed in our data.

A final concern of note is our classification for a successful replication. How to define a successful replication remains an area of debate (Armbruster, 2021; Nosek & Errington, 2020; Peels & Bouter, 2023). However, given the clear trends observed in the data, see Figure 2, we believe our findings would speak against the replicability of social priming for any reasonable measure of replication success.

Conclusion

The current meta-analysis reports an unambiguous pattern: independent replications of social priming effects fail. A reasonable question is where do we go from here? Based on our findings, it may be warranted to suggest that the limited resources available to conduct psychological research are better spent elsewhere. If researchers nonetheless wish to continue investigating social priming, we recommend further efforts to conduct preregistered, and ideally independent, replications of key studies. Such replication efforts will be required to create the credible evidence-base that social priming theory sorely needs.

Author note and Contact

We would like to thank Marcel Van Assen and Timothy Luke for their input on statistical analyses, and Madison Tunon for helping with data collection. Parts of this article are based on the unpublished master's thesis of the second author. Correspondence concerning this article should be addressed to Erik Mac Giolla, Department of Nursing, Health Sciences and Integrated Care, Atlantic Technological University, F23 X853, Mayo, Ireland. E-mail: erik.macgiolla@atu.ie

Conflict of Interest and Funding

The authors declare no conflict of interest. There was no funding for this project.

Author Contributions

EMG, DN, and MB conceived of the idea for the study. SK was primarily responsible for the data collection. SK wrote the first draft. EMG and SK had primary responsibility for the analyses. All authors contributed substantially to further revisions.

Open Science Practices

This article earned the Preregistration+, Open Data, Open Materials, and Open Code badge for preregistering the hypothesis and analysis before data collection, and for making the data, materials, and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Armbruster, S. (2021). *What makes a replication successful? an investigation of frequentist and bayesian criteria to assess replication success* [Doctoral dissertation, Ludwig Maximilian University of Munich]. https://doi.org/10.5282/ubm/epub.77434

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., & Wicherts, J. M. (2013). Replication is more than hitting the lottery twice. *European Journal of Personality*, *27*, 138–138.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Bargh, J. A. (2006). What have we been priming all these years? on the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, *36*(2), 147–168. https://doi.org/10.1002/ejsp.336

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244. https://doi.org/10.1037/0022-3514.71.2.230

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, 153–158.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Johnson, V. E., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6–10.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*, 115–144.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*(1), 40–48. https://doi.org/10.1177/1745691613513470

Cheung, I., Campbell, L., LeBel, E. P., Ackerman, R. A., et al. (2016). Registered replication report. *Perspectives on Psychological Science*, *11*(5), 750–764. https://doi.org/10.1177/1745691616664694

Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, *8*(10), 211037. https://doi.org/10.1098/rsos.211037

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295x.82.6.407

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). SAGE Publications, Inc.

Dai, W., Yang, T., White, B. X., Palmer, R., Sanders, E. K., McDonald, J. A., Leung, M., & Albarracín, D. (2023). Priming behavior: A meta-analysis of the effects of behavioral and nonbehavioral primes on overt behavioral outcomes. *Psychological Bulletin*, *149*(1-2), 67–98. https://doi.org/10.1037/bul0000374

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, *33*, 1–40. https://doi.org/10.1016/s0065-2601(01)80003-4

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology*, *74*(4), 865–877. https://doi.org/10.1037/0022-3514.74.4.865

Doyen, S., Klein, O., Simons, D. J., & Cleeremans, A. (2014). On the other side of the mirror: Priming in cognitive and social psychology. *Social Cognition*, *32*, 12–32. https://doi.org/10.1521/soco.2014.32.supp.12

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, *7*, 45–52.

Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, *19*(2), 151–156. https://doi.org/10.3758/s13423-012-0227-9

Grimes, D. R., Bauch, C. T., & Ioannidis, J. P. A. (2018). Modelling science trustworthiness under publish or perish pressure. *Royal Society Open Science*, *5*(171511), 171511. https://doi.org/10.1098/rsos.171511

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PloS one*, *9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kvarven, A., Strømland, E., & Johannesson, M. (2019). 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

Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science*, *9*(3), 278–292. https://doi.org/10.1177/1745691614528520

Molden, D. C. (2014). Understanding priming effects in social psychology: An overview and integration. *Social Cognition*, *32*, 243–249. https://doi.org/10.1521/soco.2014.32.supp.243

Morey, R. D., & Lakens, D. (2016). Why most of psychology is statistically unfalsifiable. *Unpublished manuscript*.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254. https://doi.org/10.1037/0096-3445.106.3.226

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*, 1–19.

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, 1–8. https://doi.org/10.1126/science.aac4716

Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? failure to replicate effects on social and food judgments. *PLoS ONE*, *7*(8), e42510. https://doi.org/10.1371/journal.pone.0042510

Patil, P., Peng, R. D., & Leek, J. T. (2016). What should researchers expect when they replicate studies? a statistical view of replicability in psychological science. *Perspectives on Psychological Science*, *11*(4), 539–544.

Pecher, D., van Mierlo, H., Cañal-Bruland, R., & Zeelenberg, R. (2015). The burden of secrecy? no effect on hill slant estimation and beanbag throwing. *Journal of Experimental Psychology: General*, *144*(4), e65–e72. https://doi.org/10.1037/xge0000090

Peels, R., & Bouter, L. (2023). Replication and trustworthiness. *Accountability in Research*, *30*(2), 77–87.

Rohrer, D., Pashler, H., & Harris, C. R. (2019). Discrepant data and improbable results: An examination of vohs, mead, and goode (2006). *Basic and Applied Social Psychology*, *41*(4), 263–271. https://doi.org/10.1080/01973533.2019.1624965

Schimmack, U. (2021). Replicability audit of ap dijksterhuis [Replicability-Index, April 23]. https://replicationindex.com/2021/04/23/r-audit-ap-dijksterhuis/

Shanks, D. R. (2017). Misunderstanding the behavior priming controversy: Comment on payne, brown-iannuzzi, and loersch (2016). *Journal of Experimental Psychology: General*, *146*, 1216–1222. https://doi.org/10.1037/xge0000307

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Stroebe, W. (2019). What can we learn from many labs replications? *Basic and Applied Social Psychology*, *41*, 91–103.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59–71. https://doi.org/10.1177/1745691613514450

Team, R. C. (2019). R: A language and environment for statistical computing.

van Elk, M., Matzke, D., Gronau, Q. F., Guan, M., Vandekerckhove, J., & Wagenmakers, E.-J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1–7. https://doi.org/10.3389/fpsyg.2015.01365

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, *36*, 1–48.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, *142*(5), 472–497. https://doi.org/10.1037/bul0000030

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org