UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

McCarthy, R., Gervais, W., Aczel, B., Al-Kire, R. L., Aveyard, M., Marcella Baraldo, S., Baruh, L., Basch, C., Baumert, A., Behler, A., Bettencourt, A., Bitar, A., Bouxom, H., Buck, A., Cemalcilar, Z., Chekroun, P., Chen, J. M., del Fresno- Díaz, Á., Ducham, A., ... Zogmaister, C. (2021). A Multi-Site Collaborative Study of the Hostile Priming Effect. *Collabra: Psychology*, *7*(1). https://doi.org/10.1525/collabra.18738

Social Psychology

A Multi-Site Collaborative Study of the Hostile Priming Effect

Randy McCarthy[1] [a], Will Gervais[2], Balazs Aczel[3], Rosemary L. Al-Kire[4], Mark Aveyard[5], Silvia Marcella Baraldo[6], Lemi Baruh[7], Charlotte Basch[8], Anna Baumert[9], Anna Behler[10], Ann Bettencourt[11], Adam Bitar[12], Hugo Bouxom[13], Ashley Buck[14], Zeynep Cemalcilar[7], Peggy Chekroun[13], Jacqueline M. Chen[15], Ángel del Fresno- Díaz, Alec Ducham[12], John E. Edlund[14], Amanda ElBassiouny[16], Thomas Rhys Evans[17], Patrick J. Ewell[18], Patrick S. Forscher[19], Paul T. Fuglestad[20], Lauren Hauck[1], Christopher E. Hawk[21], Anthony D. Hermann[12], Bryon Hines[22], Mukunzi Irumva[23], Lauren N. Jordan[24], Jennifer A. Joy-Gaba[10], Catherine Haley[23], Pavol Kačmár[25], Murat Kezer[7], Robert Körner[26], Muriel Kosaka[8], Marton Kovacs[3], Elicia C. Lair[24], Jean-Baptiste Légal[13], Dana C. Leighton[23], Michael W. Magee[27], Keith Markman[22], Marcel Martončik[28], Martin Müller[29], Jasmine B. Norman[15], Jerome Olsen[30], Danielle Oyler[11], Curtis E. Phills[20], Gianni Ribeiro[31], Alia Rohain[27], John Sakaluk[32], Astrid Schütz[26], Daniel Toribio-Flórez[33], Jo-Ann Tsang[4], Michela Vezzoli[6], Caitlin Williams[16], Guillermo B. Willis[34], Jason Young[8], Cristina Zogmaister[6]

[1] Department of Psychology, Northern Illinois University, DeKalb, US, [2] Centre for Culture and Evolution, Brunel University London, UK, [3] Institute of Psychology, ELTE, Eotvos Lorand University, Budapest, Hungary, [4] Department of Psychology & Neuroscience, Baylor University, Waco, US, [5] Department of Psychology, American University of Sharjah, [6] Università di Milano – Bicocca, Milan, Italy, [7] Department of Psychology, Koç University, Istanbul, Turkey, [8] Department of Psychology, Hunter College, The City University of New York, New York, US, [9] Max Planck Institute for Research on Collective Goods and School of Education, Technical University Munich, Munich, Germany, [10] Department of Psychology, Virginia Commonwealth University, Richmond, US, [11] Psychological Sciences, University of Missouri, Columbia, US, [12] Department of Psychology, Bradley University, Peoria, US, [13] University Paris Nanterre, France, [14] Department of Psychology, Rochester Institute of Technology, Rochester, US, [15] Department of Psychology, University of Utah, Salt Lake City, US, [16] Department of Psychology, California Lutheran University, Thousand Oaks, US, [17] School of Psychological, Social and Behavioural Sciences, Coventry University, UK, [18] Department of Psychology, Kenyon College, Gambier, US, [19] Université Grenoble Alpes, Grenoble, France, [20] Department of Psychology, University of North Florida, Jacksonville, US, [21] Department of Humanities and Social Sciences, DigiPen Institute of Technology, Redmond, US, [22] Department of Psychology, Ohio University, Athens, US, [23] Psychology Department, Texas A&M University—Texarkana, US, [24] University of Mississippi, Oxford, US, [25] Department of Psychology, Pavol Jozef Šafárik University in Košice, Slovakia, [26] University of Bamberg, Bamberg, Germany, [27] Department of Psychology, Saint Joseph's College-New York, New York, US, [28] University in Presov, Presov, Slovakia, [29] Department of Occupational, Economic and Social Psychology, University of Vienna, Vienna, Austria, [30] Department of Occupational, Economic and Social Psychology, University of Vienna, Vienna, Austria; Max Planck Institute for Research on Collective Goods and School of Education, Technical University Munich, Munich, Germany, [31] School of Psychology, The University of Queensland, Brisbane, AU, [32] Department of Psychology, Western University, London, Canada, [33] Max Planck Institute for Research on Collective Goods and School of Education, and School of Education, Technical University Munich, Munich, Germany, [34] Universidad de Granada, Granada, Spain

Keywords: hostile perceptions, social priming, social judgments, replication, hostile attributions, priming, crowdsourcing

https://doi.org/10.1525/collabra.18738

Collabra: Psychology

Vol. 7, Issue 1, 2021

In a now-classic study by Srull and Wyer (1979), people who were exposed to phrases with hostile content subsequently judged a man as being more hostile. And this "hostile priming effect" has had a significant influence on the field of social cognition over the subsequent decades. However, a recent multi-lab collaborative study (McCarthy et al., 2018) that closely followed the methods described by Srull and Wyer (1979) found a hostile priming effect that was nearly zero, which casts doubt on whether these methods reliably produce an effect. To address some limitations with McCarthy et al. (2018), the current multi-site collaborative study included data collected from 29 labs. Each lab conducted a close replication (total N = 2,123) and a conceptual replication (total N = 2,579) of Srull and Wyer's methods. The hostile priming effect for both the close replication (d = 0.09, 95% CI [-0.04, 0.22], z = 1.34, p = .16) and the conceptual replication (d = 0.05, 95% CI [-0.04, 0.15], z = 1.15, p = .58) were not significantly different from zero and, if the true effects are non-zero, were smaller than what most labs could feasibly and routinely detect. Despite our best efforts to produce favorable conditions for the effect to emerge, we did not detect a hostile priming effect. We suggest that researchers should not invest more resources into trying to detect a hostile priming effect using methods like those described in Srull and Wyer (1979).

In a now-classic study, Srull & Wyer (1979) demonstrated that exposing individuals to hostility-related stimuli caused them to subsequently judge a described individual as being more hostile.[1] However, whereas the original study found a "hostile priming" effect of about 3 points on a 0-10 scale, a recent Registered Replication Report (RRR;

a rmccarthy3@niu.edu

McCarthy et al., 2018) closely replicated the methods of Srull & Wyer (1979) and found hostile priming effects of only 0.07 points on the same 0-10 scale. Further, only one of the 26 labs in that study found a significant hostile priming effect. Thus, the McCarthy et al. observed effects that were much smaller than the original findings, which should decrease our confidence that the methods used by Srull & Wyer would reliably produce a hostile priming effect that is routinely and affordably detectable by researchers. However, despite the many positive aspects of their methods—e.g., having the methods vetted by an original author, pre-registered hypotheses, a large overall sample size, several independent estimates of the effect, transparent research workflow, etc.—McCarthy et al. deviated from Srull & Wyer's original methods in a few ways. And, for some, these deviations cast doubt on whether McCarthy et al. provided a critical test of the hostile priming effect, which effectively would make their results uninformative about whether methods similar to Srull & Wyer's methods reliably produce a hostile priming effect. In addition to deviating from the original methods, there are other aspects of these previously-used methods that could be improved to create the conditions that would presumably be favorable for observing a hostile priming effect.

The current study focuses exclusively on the most well-known outcome variable: Ratings of a described individual's hostility. The current study addressed several potential shortcomings of McCarthy et al. (2018) and created conditions that would be most favorable to detecting a hostile priming effect using methods similar to those originally reported by Srull & Wyer (1979). Similar to McCarthy et al., the proposed study is a multi-site collaborative study. This proposed study contained both (a) close replications of Srull & Wyer (1979) that addresses the aspects of McCarthy et al.'s methods that departed from the original study and (b) conceptual replications where each contributing researcher developed and used stimuli that were unique to their individual data collection site.

Notable Deviation Between Srull & Wyer (1979) and the RRR

The methods of McCarthy et al. (2018) deviated from those of Srull & Wyer (1979) in several ways. An exhaustive list of these known deviations is detailed in McCarthy et al. (2018), and most of these deviations are believed to be trivial. For example, the ambiguously-hostile behaviors were changed to be gender neutral, one behavior was modified from "slamming down a phone" to "abruptly hanging up a phone," and some of the stimuli had to be re-created. For the purposes of the current study, we will focus on the single deviation that has been pointed out as the most potentially consequential departure from the original methods: The setting in which the data were collected. In Srull & Wyer (1979), participants "were run in groups of four to eight" (p. 1663) in a laboratory setting. In contrast, the data collection for McCarthy et al. occurred in lecture-hall settings in groups of at least 50 participants. In working with Dr. Wyer to develop the RRR methods, this was a departure that was specifically noted by him as being potentially meaningful to detecting the effect. The decision around which setting to collect data in was ultimately made because another RRR (i.e., Verschuere et al., 2018) was run concurrently with McCarthy et al. and it was critical for this other RRR to collect data in a lecture hall.

At the heart of this criticism is that the original Srull & Wyer (1979) study was run in a relatively distraction-free laboratory environment and McCarthy et al. (2018) was run in a relatively distracting lecture hall environment. The implication is that running the study in a relatively distraction-free environment ought to provide a favorable setting in which the effect could emerge. As one reviewer noted, in addition to the "noisiness" of the environment, contexts have been considered a relevant factor in whether priming effects emerge (e.g., Cesario et al., 2010).

Other Methodological Improvements

In addition to addressing the notable departure from the original methods, we proposed addressing three other study characteristics that should result in improvements over the previously-used methods. First, in Srull & Wyer (1979) Experiment 2 (but not their Experiment 1), participants were asked if they believed any of the tasks they had completed were related. They found that participants generally did not have awareness of the relationship between these tasks. This was a crude, group-level test of participants' awareness of the study's hypotheses. Because participants' awareness of the potential influence of the prime was not a part of the original study design (i.e., Srull & Wyer, 1979, Experiment 1), this was not assessed in McCarthy et al. (2018) even though lack of awareness between the prime and the effect of the prime is theoretically crucial for priming effects to emerge (e.g., Loersch & Payne, 2014). Second, when possible, the stimuli for Srull & Wyer (1979) were used in McCarthy et al. When not possible (e.g., the original priming stimuli were not available), new stimuli were developed that were believed to be consistent with the original stimuli. These newly-developed stimuli were created by the author in a series of pretests, vetted by Dr. Wyer prior to data collection, and then these same stimuli were used at each data collection site.

Although using standard stimuli eliminates any lab-to-lab variability due to the specific stimuli that were used, this approach also has a few possible drawbacks. First, the original study was published nearly 40 years prior to McCarthy et al. (2018). Although a pretest ensured there were no obvious problems such as ceiling or floor effects in participants' perceptions of the hostility of the vignettes, the passage of time could have affected the appropriateness of these origi-

1 In addition to judgments of a described individual's hostility, Srull & Wyer (1979) also included two other outcome variables: Ratings of ambiguously-aggressive behaviors and ratings of the co-occurrence of traits. The current proposal focuses exclusively on the most well-known outcome variable: Ratings of a described individual's hostility.

nal stimuli in other unanticipated ways.

Second, and similarly, the pretesting of the stimuli was done by one researcher at one location. And the assumption was that the characteristics of the pretested stimuli would be similar across all the independent data collection sites. Arguably, it would be more informative for each researcher to customize their stimuli for their locally-available participant pool (e.g., Crandall & Sherman, 2016).

Finally, neither the original study nor McCarthy et al. (2018) included "positive controls." Because detecting a hostile priming effect rests on the assumption that several auxiliary hypotheses are true (e.g., participants paid sufficient attention to the study materials, the data were recorded properly, etc.), detecting a highly-probable effect can help to ascertain the soundness of some of these auxiliary hypotheses (e.g., Meehl, 1967). The soundness of these auxiliary hypotheses are especially important for interpreting effects that fail to replicate a previously-reported study.

Overview

The current study was a multi-site collaborative study testing whether methods similar to those of Srull & Wyer (1979) reliably produced a hostile priming effect. Each data collection site used the same stimuli as McCarthy et al. (2018) and developed their own stimuli using a standard pretesting procedure. Contributing researchers then collected a sample to obtain an estimate of the hostile priming effect using the same stimuli (i.e., close replication) and an estimate of the hostile priming effect using stimuli that were pretested specifically for their locally-available participant pool (i.e., conceptual replication). Further, the proposed methods included several characteristics to address shortcomings with previous studies. Specifically, the proposed study (a) ran participants in a laboratory setting, (b) probed participants for their awareness of the study hypotheses, (c) included both a close replication and a conceptual replication of the hostile priming effect, and (d) included several positive controls to help interpret the study results. Further, several steps ensured the data analyses were not influenced by the results obtained. Finally, because contributing researchers conducted both a close replication and a conceptual replication, it is possible to assess whether some labs are able to produce stronger effects in general (i.e., labs that produce a strong effect in the close replication also produce a strong effect in the conceptual replication). Following in-principle acceptance on August 17, 2018, the approved Stage 1 manuscript was registered at https://psyarxiv.com/gxp7u/. This registration was performed prior to data collection and analysis.

Methods

Lab Recruitment

Researchers who contributed to the current study were recruited in two ways. First, a call for the study was posted to StudySwap (https://osf.io/u6gfz/) on October 3, 2018. Second, the same call for the study was posted to the SPSP Open Forum on October 9th, 2018. After the calls were posted for four weeks, 31 labs had expressed interest in joining the current study. As per the In-Principle Acceptance, the names of the individual researchers who responded to the call were posted on the project's OSF page to demonstrate sufficient interest to continue with the project (https://osf.io/7a6ur/); note that this list does not match the final author list because some researchers [e.g., research assistants] were added after the research process began and some researchers who expressed interest to the initial call were unable to complete the study). Once we had a list of contributing labs, the lead author (RJM) provided instructions on how to proceed with stimuli pretesting, creating preregistration documents, and data collection (see the welcome email here: https://osf.io/6jyqv/).

Individual Lab Procedures

Pre-data collection activities

Contributing labs completed four steps prior to beginning data collection. For labs that used stimuli in a language other than English, the translations for each of these following steps were handled by the individual labs.

First, each contributing lab obtained ethics approval from their local IRBs or arranged a joint IRB approval with the lead author's institution.

Second, contributing labs followed a predetermined procedure for creating stimuli for their conceptual replications. This process involved generating to-be-tested stimuli, collecting data from a small sample of participants (i.e., $N >= 20$) who were drawn from the same participant pool as the participants in the main study would be drawn from, and analyzing the data to select their stimuli.

Briefly, contributing labs generated at least 50 3-word phrases that described potentially aggressive behaviors (e.g., "hit his face") and at least 50 3-word phrases that described non-aggressive behaviors (e.g., "wash the clothes"). Contributing labs also created at least two brief vignettes that described an individual who behaved in an ambiguously-aggressive manner. Participants in the pretesting sample rated the extent to which the 3-word phrases were aggressive and rated the extent to which the individual described in the vignette was hostile.

The 24 most aggressive 3-word phrases and the 30 least aggressive 3-word phrases were used for the priming stimuli in each lab's conceptual replications. These selected phrases were then used to create the "hostile priming" stimuli where 24/30 described aggressive behaviors and the "control" stimuli where 0/30 described aggressive behaviors. Contributing labs also identified the vignette in which the described individual was viewed as moderately hostile and did not have any noticeable floor or ceiling effects. Contributing labs also visually inspected a distribution of the hostility ratings to ensure the distributions did not have any floor or ceiling effects. In a small number of cases where the vignettes resulted in similar pretesting data (e.g., the mean hostility ratings were similar and the distribution of hostility ratings were similar), researchers at these labs were instructed to use their expertise/experience to select a vignette they felt would provide a "fair test" of the hostile priming effect.

Third, once labs completed their pretesting, their pretesting stimuli and data were uploaded to the projects'

Open Science Framework page and their conceptual replication stimuli were created and uploaded to the project's Open Science Framework page. A study was then created that was similar in appearance to the close replication study.

Finally, labs pre-registered their analysis plan for their conceptual replication study. That is, they specified how they would quantify the outcome variable and how they would test the hostile priming hypothesis. Researchers were instructed to choose an outcome variable they felt provided a "fair test" of the hostile priming effect.

Data collection procedures

Each contributing lab collected a sample for both the close replication and the conceptual replication. Within each of these studies, participants were randomly assigned to either a hostile priming condition or a neutral control condition.

Upon coming to the lab, each participant was greeted by a researcher and started at a computer. Participants first completed a 30-trial sentence descrambling task. In both the close replication and the conceptual replication, participants were randomly assigned to one of the two priming conditions. In the "hostile priming" condition, participants descrambled 24/30 sentences that, when descrambled, formed hostile phrases and participants who were in the "neutral" condition descrambled 0/30 sentences that, when unscrambled, form hostile phrases[2].

All participants then immediately read a brief vignette describing an individual who behaved in an ambiguously hostile manner, and then rated that individual on the traits relevant to hostility (those in the close replication condition saw the same "Ronald vignette" and provided trait ratings that were used in as used in Srull & Wyer [1979] and McCarthy et al. [2018], and those in the conceptual replication viewed a vignette and provided ratings on traits that were unique to each lab).

Thus, participants were placed into either the hostile priming condition or the control conditions for the close replication study (which was the same procedure with the same stimuli at each contributing lab) or for the conceptual replication study (which was the same procedure, but with stimuli that were unique, for each contributing lab).

All participants then viewed a screen that asked them to report the extent to which they agree with the statements "watching TV is a hobby of mine," "playing video games is a hobby of mine," and "reading books is a hobby of mine." Participants then reported how many books they have read for pleasure in the past year. At the top of this screen, participants read the following instructions:

> We are interested in whether participants actually take the time to read the directions. To demonstrate that you have read the directions, and are not mindlessly responding to items, please respond with "Completely Disagree" to the items "Watching TV is a hobby of mine" and "Playing video games is a hobby of mine." Answer honestly to the items "Reading books is a hobby of mine" and "How many books have you read for pleasure in the past year?

Finally, participants reported their age, gender, and height. After reporting their demographic information, participants then answered a few questions to probe for whether they felt the priming task influenced their judgments of the individuals described in the vignettes. These suspicion probes were modeled after the funneled debriefing example that is included in Table 2 of Bargh & Chartrand (2000). Also, prior to viewing each suspicion probe participants were reminded that "Your responses will not affect whether you get credit for completing the study or not. Please answer honestly." Participants were first asked "What do you think the purpose of the study was?" and were given the opportunity to type their responses in a text box. Participants were next asked "Do you think any of the tasks in the study were related" and provided a "yes" or "no" response. Finally, participants were asked "To what extent do you feel like the 'scrambled sentence task' influenced your ratings of the described individual?" and provided a response on a 7-point scale ranging from *1-Did not influence at all* to *7-Influenced a lot*.

All participants were then thanked and debriefed.

Sample Size Determination

To our read, the hostile priming effect is a directional prediction and does not specify the smallest effect which would be considered supportive of any particular priming hypothesis. Thus, "feasibility considerations" (p. 359, Lakens, 2017) were used to determine the smallest effects that researchers could affordably and routinely detect. Effects smaller than affordable and routinely-detectable effects would be considered too small to be of interest.

We assumed that most individual researchers could regularly obtain a sample size of 200[3]. We also assumed that individual researchers would use a Type 1 error rate of 5% (i.e., $\alpha = .05$), would interpret a one-tailed (i.e., directional) effect in the hypothesized direction as supportive of a hostile priming effect, and would desire to have 80% power to detect their effect. These parameters imply that researchers would need a minimum effect size of $d = 0.35$ to be detectable with 80% power in future studies of 200 participants. Thus, we used an effect of $d = 0.35$ to be the target (i.e., to-be-detected) effect. Notably, a review and meta-analysis of this literature (DeCoster & Claypool, 2004) found an effect reflecting the impact of priming on judg-

2 In Srull & Wyer (1979) and in the RRR, the two priming conditions were 24/30 hostile sentences or 6/30 hostile sentences. Thus, having a control condition with 0/30 hostile sentences is a slight departure from the original methods, but is being proposed to maximize the likelihood of detecting a hostile priming effect and to create a more informative comparison condition.

3 Note that in the current research that individual labs were asked to collect smaller samples because we were focused on the meta-analytic effect size estimates and not whether any individual lab could detect the effect.

ments about social targets of d = 0.35, 95% CI [0.30, 0.41] (this effect size also was used in equivalence tests reported in the Results).

We then conducted a power analysis to determine how many labs would be needed to detect an effect of d = 0.35 given each lab would be able to contribute 30 participants per cell for each of the two meta-analyses (Quintana, 2017). If 12 individual labs contributed 30 participants per cell for each meta-analysis, there would be ~97% power for the meta-analysis to correctly detect an effect of d = 0.35. If more labs than 12 labs contributed to this project, or if labs collected more than the minimum sample of 30 participants per cell, the proposed meta-analyses would nearly always detect an effect of d = 0.35 (i.e., power approached 100%). To account for some labs not completing data collection, and to account for the exclusion of individual participants, we sought for at least 15 labs to collect a sample of at least 60 participants (prior to data exclusions) in each of the close replication and conceptual replication. We surpassed this data collection goal, which means we had very high statistical power to detect an effect of d = 0.35 or we had sufficient statistical power to detect many smaller effects (d < 0.35) that would be considered theoretically meaningful.

Additionally, the lead author (RJM) collected an online sample via Mechanical Turk that was planned to be at least 200 participants for each of the direct replication and conceptual replication. This additional sample obviously was not in a laboratory setting, but it allowed us to compare whether the effects obtained in an online sample are noticeably different.

Known Deviations from the Approved Methods

There are four notable deviations from what was proposed in the IPA.

First, due to an error in creating the study templates that were provided to some labs, we did not include all the traits for the outcome variable that we described in the In-Principle Acceptance (a more detailed description of how this error came about can be found here: https://osf.io/z2g5x/). In consultation with the editor (see a copy of the email here: https://osf.io/7snj8/), we proposed using the average of the traits *hostile*, *unfriendly,* and *dislikable* (i.e., three relevant traits that were included in each lab's close replication) as the outcome variable instead. Notably, this error only affects the traits used to compute the outcome variable for the close replication because individual labs could have chosen different traits for their conceptual replication.

Second, the IPA was written with the idea that we would only include English-speaking labs to avoid the need to translate materials. However, we had several non-English-speaking labs respond to the call for the study. In consultation with the editor, these non-English-speaking labs were invited to join the project. Because the current study met the sampling goals with English-speaking labs, the addition of non-English-speaking labs is considered a bonus to the initially-proposed study.

Third, we had one lab complete the conceptual replication and not complete the close replication.

Finally, we had 5 labs who did not complete a preregistration prior to data collection. These latter two errors were due to a miscommunication and were not discovered until after data collection was complete.

Table 1: Links for study materials[a]

Material	Link
Close Replication	
Example stimuli	https://osf.io/kv7zb/
Blinded dataset	https://osf.io/h7u9j/
Unblinded dataset	https://osf.io/9fvdm/
Conceptual Replication	
Blinded dataset	https://osf.io/5dyqh/
Unblinded dataset	https://osf.io/2r4m9/
Analysis code	
Analyst 1 code	https://osf.io/wnqt4/
Analyst 2 code	https://osf.io/br53c/

[a]This table contains the links for the information on which the meta-analyses are based. The stimuli and data for each individual lab can be found on the project's Open Science Framework page: https://osf.io/j6uwa/

Results

Table 1 contains the links to the stimuli, data, and analysis code for this project. We conducted all analyses using the metafor package version 2.4-0 in R (Viechtbauer, 2010).

Data Preparation

After data collection was complete, a researcher assistant aggregated the individual datasets for the close replications and for the conceptual replications. The open-ended questions for the suspicion probe (i.e., "what do you think the purpose of the study was?") was coded for participants' suspicion (examples of responses considered "suspicious" or not can be found here: https://osf.io/nwzmt/) and then the dataset was "blinded" by the research assistant. The blinding process involved deleting the responses to the sentence descrambling task (because the content of those items would reveal which priming condition participants were in), deleting the responses to the open-ended suspicion probe (because the content of those responses may reveal which condition a participant was in), and creating a non-descriptive "condition" variable with labels condition "A" and condition "B" (which corresponded to whether a participant was in the "hostile priming" condition or the "neutral condition"). This blinded dataset was then provided to two researchers to conduct the meta-analysis (i.e., Drs. John Sakaluk and Patrick Forscher) who were not involved in collecting the data. These researchers independently analyzed the data and their results were checked against one another.

After analysis decisions were made and the data were analyzed, the dataset was "unblinded" and we revealed whether "condition A" and "condition B" referred to the "hostile priming condition" or the "neutral condition."

Effectively, this process ensured that the researchers who collected the data were not involved in analyzing the data

and vice versa. And this process meant that the analyses were done by researchers who were merely testing whether "condition A" differed from "condition B," which minimizes the possibility that any analysis decisions were made to favor or disfavor the to-be-tested hypotheses. Further, the two researchers who conducted the meta-analyses worked independently, which gives us further confidence in the analysis procedures and results are independently reproducible.

Pre-Specified Exclusions

In all, 2,123 participants completed a close replication study and 2,579 participants completed a conceptual replication study. Participants were excluded from the primary analyses if (a) they did not complete all the sentence descrambling trials, (b) they did not provide responses for all the trait ratings needed to compute the outcome variable (e.g., *hostile*, *unfriendly*, and *dislikable* for the close replications), (c) they failed one or more of the Instructional Manipulation Check items, or (d) they indicated suspicion that their trait ratings were affected by the priming task. Participants could have failed more than one of these exclusion criteria.

In the close replication, there were 1,402 participants (66.0%) who were not excluded from the primary analyses. Of these remaining participants, 413 were male and 979 were female, three provided another response, and seven were missing sex information. The average age was 22.40 years old (SD = 7.56). In the conceptual replication, there were 1,641 participants (63.6%) who were not excluded from the primary analyses. Of these, 481 were male and 1,151 were female, four provided another response, and five were missing sex information. The average age was 23.00 years old (SD = 8.79).

One potential issue was that we had different rates of exclusions across our conditions.[4] Specifically, in the close replication, 38% of participants in the hostile priming condition and 30% of participants in the neutral priming condition were excluded, X^2 (1) = 15.43, p < .001. Likewise, in the conceptual replication, 37% of participants in the hostile priming condition and 29% of participants in the neutral priming condition were excluded, X^2 (1) = 19.13, p < .001. The frequency of exclusions for each individual exclusion criterion by condition can be found here: https://osf.io/q2ngk/. For both the close replication study and the conceptual replication study, participants were more likely to be excluded for not completing the priming task, not completing the ratings, or for being flagged as suspicious if they were in the hostile priming condition. For both the close replication study and the conceptual replication study, there were no differences in rates of exclusions for failing the Instructional Manipulation Check.

Analysis of Positive Controls

Two positive controls were included to ensure the methods produced highly-probable effects.

First, a meta-analysis was conducted on the correlation between participants' self-reported agreement with the statement "reading books is a hobby of mine" and their self-reported number of books they have read for pleasure in the past year. Within the close replication samples, the meta-analytic effect size for the association between these two items was r = .59, 95% CI [.55, .62], z = 32.70, p < .001. Within the conceptual replication samples, the meta-analytic effect size for the association between these two items was r = .60, 95% CI [.56, .64], z = 32.17, p < .001.

Second, a meta-analysis was conducted of the mean difference between men's self-reported height and women's self-reported height. Within the close replication samples, males were, on average, 8.17 cm taller than women, d = -1.61, 95% CI [-1.87, -1.34], z = 11.80, p < .001. Within the conceptual replication samples, males were, on average, 8.33 cm taller than women, d = -1.45, 95% CI [-1.62, -1.28], z = 16.49, p < .001.

Thus, effects from both positive controls clearly emerged within both the close replications and conceptual replications.

Planned Meta-Analyses

Meta-analysis of close replication attempts

The effects from the close replication attempts were analyzed in a random-effects meta-analysis using the REML estimator. Among participants who were not excluded, the meta-analytic mean difference was 0.17 points on a 1 to 11 scale, which is a standardized effect size of 0.09 standard deviations, d = 0.09, 95% CI [-0.04, 0.22], z = 1.41, p = .16 (see Table 2). The heterogeneity of this effect across labs was no bigger than what would be expected due to sampling error alone, τ = 0.19; Q(df = 27) = 39.36, p = .06. As can be seen in Figure 1, this effect size is small and not significantly different from zero.

We planned the study to detect an effect of d = 0.35 because that was determined to be a feasibly-detectable effect. The planned equivalence test confirmed that our observed effect is significantly smaller than d = 0.35, z = -3.91, p < .001. This finding suggests that if the true effect is indeed non-zero, it is nevertheless too small to be routinely detected by the typical psychology lab.[5]

Meta-analysis of conceptual replication attempts

The effects from the conceptual replications were analyzed in a random-effects meta-analysis using the REML estimator. Among participants who were not excluded, the

4 We thank a reviewer for bringing this issue to our attention during the Stage 2 review.

5 One of the analysts also conducted an equivalence test using d = 0.15 as a target effect size. When using this smaller effect, the observed effect size for the close replication is not significantly smaller than d = 0.15, z = -0.86, p = .19. And the observed effect size for the conceptual replication was significantly smaller than d = 0.15, z = -1.86, p = .03.

Table 2: Descriptive statistics for close replication studies[a]

Lab	Sample size		Hostile priming condition			Neutral priming condition		
	Full sample	After exclusions	*n*	*M*	*SD*	*n*	*M*	*SD*
Aczel	58	47	19	8.12	1.84	28	7.08	1.84
Al-Kire	26	17	9	8.11	1.52	8	8.54	1.14
Aveyard	97	65	34	7.21	1.68	31	6.54	2.44
Baumert	70	42	21	9.17	1.48	21	8.22	1.42
Edlund	46	33	17	7.76	2.11	16	8.90	1.33
ElBassiouny	63	37	17	8.22	1.17	20	8.67	1.33
Evans	69	39	17	8.24	1.96	22	8.08	1.96
Ewell	-		-	-	-	-	-	-
Fuglestad	81	45	19	7.96	1.60	26	7.87	2.22
Hawk	67	48	22	8.36	1.24	26	7.78	1.88
Hermann	90	64	29	7.63	1.69	35	7.89	1.31
Hines	59	45	24	8.28	2.05	21	8.86	1.65
Joy-Gaba	64	42	18	8.57	2.17	24	8.33	2.01
Kačmár	93	69	34	7.73	1.86	35	6.91	2.07
Kezer	84	57	25	7.48	2.15	32	6.76	2.08
Lair	78	55	27	8.83	1.21	28	7.49	2.10
Légal	114	68	36	7.72	1.56	32	8.07	1.46
Leighton	37	19	10	7.70	1.98	9	8.48	1.89
Magee	66	37	17	8.71	1.50	20	7.85	1.88
McCarthy-In-person	110	62	32	8.27	1.77	30	7.77	1.92
McCarthy-Online	212	152	66	7.47	1.95	86	7.36	2.04
Norman	68	49	23	7.87	2.19	26	8.49	1.67
Olsen	76	49	22	8.36	1.97	27	8.63	1.50
Oyler	64	44	24	8.74	1.71	20	8.65	1.96
Ribeiro	62	36	17	8.53	1.07	19	8.95	1.08
Schütz	74	45	19	8.98	1.74	26	7.50	1.95
Willis	64	47	21	8.46	1.77	26	8.09	2.27
Young	66	42	18	8.13	2.39	24	8.43	1.14
Zogmaister	65	47	23	7.59	2.30	24	7.83	2.40

[a]This table contains the descriptive statistics for the close replication studies. These descriptive statistics are for the participants who were not omitted (i.e., they completed all trials of the priming task, provided ratings on all outcome variables, passed the attention checks, and did not display suspicion of the study hypotheses).

meta-analytic standardized mean difference was 0.06 standard deviations, d = 0.06, 95% CI [-0.04, 0.15], z = 1.15, p = .25 (see Table 3, individual labs could have used different rating scales, so raw mean differences are not interpretable). The heterogeneity of this effect across labs was no bigger than what would be expected due to sampling error alone, τ = 0; Q(df = 28) = 25.85, p = .58. As can be seen in Figure 1, this effect size also is small and not significantly different from zero.

An equivalence test confirmed that our observed effect is significantly smaller than d = 0.35, z = -5.88, p < .001. This finding suggests that if the true effect is indeed non-zero, it is nevertheless too small to be routinely detected by the typical psychology lab.

Comparing Close and Conceptual Replications

Labs contributed data for both a close replication and conceptual replication. This gave us an opportunity to systematically compare the close and conceptual replications while controlling for lab characteristics. We examined the following questions:

1. Did the close and conceptual replication effects differ in size?
2. Did the close and conceptual replication effects differ in how variable they were?
3. Did labs that produced large close replication effects also tend to produce large conceptual replication effects?

Table 3: Descriptive statistics for conceptual replication studies[a]

Lab	Sample size		Hostile priming condition			Neutral priming condition		
	Full sample	After Exclusions	*n*	*M*	*SD*	*n*	*M*	*SD*
Aczel	63	57	27	6.96	2.56	30	6.56	2.75
Al-Kire	179	123	61	7.31	1.66	62	7.38	1.57
Aveyard	102	70	30	7.40	1.55	40	7.83	1.55
Baumert	69	51	20	8.19	1.68	31	8.03	1.51
Edlund	48	34	14	4.79	1.72	20	4.75	1.45
ElBassiouny	57	27	11	8.88	1.91	16	8.23	1.97
Evans	71	47	23	8.02	1.44	24	8.48	1.08
Ewell	135	80	35	8.74	1.34	45	7.89	2.21
Fuglestad	82	47	21	6.34	0.74	26	6.42	1.02
Hawk	66	48	21	3.37	0.88	27	3.37	0.50
Hermann	91	59	26	8.29	1.65	33	7.91	1.56
Hines	58	44	23	8.17	1.29	21	8.25	1.23
Joy-Gaba	70	51	25	5.78	1.77	26	5.26	1.43
Kačmár	90	73	37	7.93	1.16	36	7.42	1.32
Kezer	79	48	25	8.83	1.30	23	9.01	1.16
Lair	71	40	18	10.11	2.11	22	11.32	3.01
Légal	116	77	36	6.52	1.48	41	6.84	1.61
Leighton	37	19	10	6.40	2.27	9	5.83	2.18
Magee	71	39	21	7.62	2.27	18	8.44	2.45
McCarthy-In-person	115	61	30	6.89	2.13	31	7.34	2.56
McCarthy-Online	390	208	96	6.53	2.07	112	6.28	1.97
Norman	65	45	21	8.72	1.71	24	9.03	1.45
Olsen	77	50	23	7.70	1.74	27	7.07	2.40
Oyler	60	41	18	7.11	1.97	23	7.04	2.01
Ribeiro	65	43	24	8.48	1.14	19	8.66	1.02
Schütz	65	35	17	6.82	2.24	18	5.33	2.09
Willis	57	41	18	7.19	1.76	23	6.56	2.37
Young	66	42	18	7.93	1.41	24	7.73	1.31
Zogmaister	64	41	23	4.95	1.03	18	4.47	1.04

[a]This table contains the descriptive statistics for the conceptual replication studies. These descriptive statistics are for the participants who were not omitted (i.e., they completed all trials of the priming task, provided ratings on all outcome variables, passed the attention checks, and did not display suspicion of the study hypotheses). Labs selected unique outcome variables, which means that the absolute values of the means across labs should not be directly compared in this table. These data are presented so within-lab comparisons can be seen and so the meta-analysis can be reproduced.

To answer these questions, we fit a multivariate meta-analytic model with the close and conceptual replication effects as outcome variables, an indicator variable to track whether the effect came from a close or a conceptual replication, and by-lab random intercepts to account for the non-independence of these effects. This model also included a fully unstructured between-studies variance-covariance matrix.

Close replication effects were no different in size from the conceptual replication effects, $d_{difference}$ = -0.06, 95% CI = [-0.22, 0.11], z = -0.67, p = .51. A model that constrained the heterogeneities of the close and conceptual replications to be equal had no worse fit than a model without this constraint, $\chi^2(1, k = 28) = 1.45$, p = .23, suggesting that the conceptual replication effects were no more variable than the close replication effects ($\tau_{close} = .18$, $\tau_{conceptual} = .00$). A model that constrained the covariance between the close and conceptual replication effects to zero had no worse fit than a model without this constraint, $\chi^2(1, k = 28) = 0.00$, p = .95, suggesting that the size of the close replication effects was not related to the size of the conceptual replication effects.

There is a limit to what we can infer from the comparisons between the close and conceptual replications. Only 28 labs contributed data for these analyses. While 28 is an impressive number for inferences that do not depend strongly on differences between labs, it is a small number for inferences that do depend on these differences. In par-

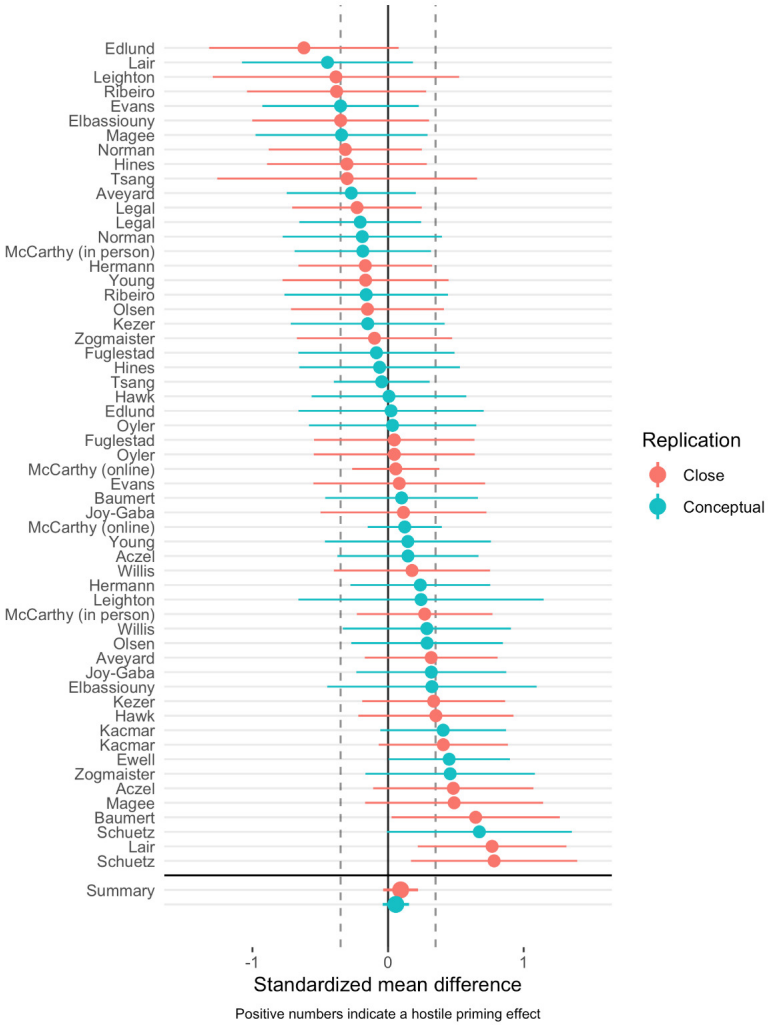

Figure 1: Combined forest plot for close replication and conceptual replication meta-analyses

This combined forest plot shows the effect sizes for both the close replications and conceptual replications and the meta-analytic effect size estimates. The vertical dashed lines represent our a priori smallest effect size of interest. The error bars represent 95% confidence intervals.

ticular, drawing firm conclusions about whether labs produce hostile priming effects when they conduct close versus conceptual replications requires observing a large number of labs. The same is true of any inference about the relationship between close and hostile priming effects. Thus, our results should not be taken as strong evidence about the relationship between close and conceptual replications.

Exploratory Analyses

Exploring the Bayesian evidence for close and conceptual replications

In addition to the preregistered meta-analyses, one of the analysts (Sakaluk) conducted a set of exploratory Bayesian meta-analytic *t*-tests (via the BayesFactor package version 0.9.2 for *R,* Morey & Rouder, 2018) of the sets of close and conceptual replication effects; the aim of these exploratory analyses was to provide greater insight into the strength of evidence for a null hostile priming effect. Given the exploratory nature of these analyses, the Bayesian synthesis was conducted using a range of scales ("medium", "wide", and "ultrawide") for the prior distribution, and the tests were effectively testing for the possibility of *any* hostile priming effect (positive or negative), and the resulting Bayes factors were interpreted using the guidelines of Lee & Wagenmakers (2013). We then calculated 95% credibility intervals by resampling from the posterior distribution of the model fit with a medium prior distribution scale.

Analysis of the close replication effects yielded moderate evidence in favor of a null hostile priming effect, BF01 = 3.02 - 5.92 (depending on prior scale), median posterior d = 0.10, 95% CR: -0.004, 0.20. Analysis of the conceptual replication effects, meanwhile, yielded moderate-to-strong evidence in favor of a null hostile priming effect, BF01 = 8.87 - 17.59 (depending on prior scale), median posterior d = 0.06, 95% CR: -0.04, 0.16.

Exploring the effects of stimuli translation

The original stimuli were created in English. Across labs, these stimuli were translated into seven other languages. The language of the stimuli was then entered as a possible moderator of the hostile priming effect. The language of the

stimuli did not significantly moderate the hostile priming effect for either the close replications, $Q(df = 7) = 10.18$, $p = .18$, or the conceptual replications, $Q(df = 7) = 9.16$, $p = .24$. Thus, the translation of materials did not seem to differentially affect the hostile priming effect.

Exploring the effects of participants' "subjective influence of primes"

Participants reported the extent to which they felt like "the 'scrambled sentence task' influenced your ratings of the described individual?" on a scale ranging from *1-Did not influence at all* to *7-Influenced a lot*. We estimated whether participants' subjective influence of the primes (measured at the individual level) interacted with the priming condition. Participants' ratings and the priming condition were both centered within lab prior to these analyses.

For the close replications, participants' ratings of the influence of the primes did not interact with the priming effect, estimate = -0.08, 95% CI [-0.20, 0.04], $z = -1.32$, $p = .19$. For the conceptual replications, participants' ratings of the influence of the primes also did not interact with the priming effect, estimate = -0.07, 95% CI [-0.17, 0.02], $z = -1.56$, $p = .12$. Although weak and not significant, the direction of both findings is consistent with the idea that increases in the subjective influence of the primes is consistent with a weaker hostile priming effect.

Robustness checks: The effect of exclusion criteria

To robustly examine the extent to which exclusion criteria affect the hostile priming effect, we explored 15 combinations of different ways of excluding participants. These results are shown in Table 4. No exclusion criterion, or combination of exclusion criteria, resulted in a hostile priming effect for either the close replication or the conceptual replication.

Discussion

The goal of the current study was to test whether Srull-and-Wyer-esque methods would reliably produce a hostile priming effect. In the end, despite our best efforts to create favorable conditions for the effect to emerge (i.e., high statistical power, customized stimuli for each local subject pool, attention checks, suspicion probes, a quiet lab environment, etc.), the observed hostile priming effects for both the close replications and conceptual replications were small in magnitude and not significantly different from zero. To put it bluntly, we did not detect a hostile priming effect in the current study.

The current results then raise the questions: Why was the hostile priming effect not detected? And what might the lack of detected effect mean?

Are There Obvious Methodological Reasons Why the Hostile Priming Effect Was Not Detected?

In addition to following the general methodology of Srull & Wyer (1979), the current study's methods addressed four factors that were considered limitations and, therefore, possible areas of contention, with the McCarthy et al. (2018) RRR.

First, the McCarthy et al. (2018) RRR collected data in a lecture hall setting, which is arguably less than ideal for subtle priming manipulations. The implication of this critique is that a quieter and more controlled setting would be a better context for the hostile priming effect to emerge. However, even though the data in the current study were all collected in a quiet laboratory setting, the hostile priming effect did not emerge.

Second, hostile priming effects presumably rely on participants' engagement with the priming stimuli and participants' lack of awareness of the influence of the primes. Although these assumptions were not tested in Srull & Wyer (1979, Experiment 1) and, hence, not included in the McCarthy et al. (2018) RRR, the current study used both awareness checks and suspicion probes to omit participants for whom the hostile priming effect would be less likely to occur. However, even though participants who failed an attention check or who expressed suspicion of the hypotheses were omitted, the hostile priming effect did not emerge.

Third, one assumption within close replications is that the stimuli operate similarly for each sample. That is, the close replications in the current study and in McCarthy et al. (2018) assumed that, for example, participants in DeKalb, IL, USA and Budapest, Hungary each interpreted the same vignette in the same way. Or, because the vignette in the close replication was from Srull & Wyer (1979), the recent close replications assume that participants today interpreted the vignette similarly as participants in the original study did about 40 years ago. Because the appropriateness of the stimuli is an assumption, and therefore debatable, we had each lab complete both a close replication and a conceptual replication. These conceptual replications involved researchers creating stimuli specifically for their local subject pools and gave (some) flexibility to the researchers to create methods they believed would produce hostile priming effects. If the stimuli or procedures in the close replications were inappropriate, for whatever reason, it should be more likely to detect a hostile priming effect within conceptual replications where the (presumably) better-suited stimuli were used. However, the hostile priming effect within both the close replications and the conceptual replications did not emerge. Further, the results of the close replications and the conceptual replications were not statistically different.

Fourth, the current study also extended Srull & Wyer (1979) and McCarthy et al. (2018) by including positive controls into the methods. Detecting a hostile priming effect relies on several factors such as properly following the procedures and selecting an outcome variable that captures the effect of the priming stimuli. Detecting a hostile priming effect also relies on fundamental, or background, factors such as the data being recorded properly, participants reading and understanding the instructions, and responding coherently. "Unsuccessful" replications can be due to one or more of these background factors being absent. In the current study, to ascertain whether some background factors were present, we included two highly-probable effects that we used as positive controls: A relationship between self-reported enjoyment for reading books and the number of books read for pleasure, and the difference in height be-

Table 4: Exploratory Analyses of Exclusion Criteria[a]

Exclusion criteria	Close replications			Conceptual replications		
	d	LL	UL	*d*	LL	UL
No exclusions	0.07	-0.03	0.16	0.07	-0.06	0.21
Primes	0.07	-0.03	0.16	0.06	-0.02	0.15
Ratings	0.07	-0.03	0.16	0.06	-0.02	0.14
IMC	0.09	-0.04	0.21	0.06	-0.03	0.15
Suspicion	0.07	-0.03	0.16	0.03	-0.05	0.12
Primes + Ratings	0.07	-0.03	0.16	0.07	-0.02	0.15
Primes + IMC	0.09	-0.04	0.21	0.07	-0.03	0.16
Primes + Suspicion	0.07	-0.02	0.16	0.04	-0.05	0.13
Ratings + IMC	0.09	-0.04	0.21	0.06	-0.03	0.15
Ratings + Suspicion	0.07	-0.03	0.16	0.04	-0.05	0.13
IMC + Suspicion	0.09	-0.04	0.22	0.04	-0.05	0.14
Primes + Ratings + IMC	0.09	-0.04	0.21	0.07	-0.02	0.16
Primes + Ratings + Suspicion	0.07	-0.02	0.16	0.05	-0.04	0.14
Primes + IMC + Suspicion	0.09	-0.04	0.22	0.05	-0.04	0.15
Primes + IMC + Suspicion	0.09	-0.04	0.22	0.05	-0.05	0.14

[a]This table explores how different exclusion criteria affect the hostile priming effect. The "No exclusions" row represents the analyses when no participants are excluded. "Primes" indicates whether a participant completed all the trials of the priming task. "Ratings" indicates whether a participant provided a rating for all the traits needed to compute the outcome variable. "IMC" indicates whether a participant passed the Instructional Manipulation Check. "Suspicion" indicates whether a participant expressed suspicion of the study hypotheses in their response to the open-ended question at the end of the study.

tween males and females. These positive controls were chosen because we were confident these effects would reliably be detected. Indeed, both positive controls were detected.

Although detecting these positive controls seems mundane, that is precisely why we included them. Further, mundane and unsurprising does not mean unimportant. These positive controls demonstrate the soundness of certain (but not all) aspects of our procedures. A simple thought experiment highlights the importance of these positive controls: Imagine the criticisms that could be raised if these effects did not emerge. Now, criticisms of the current methods must explain why the hostile priming effect did not emerge *and* why the positive controls did emerge. Effectively, the positive controls shift some of the possible blame for the lack of a detected hostile priming effect from these fundamental data collection factors onto the theory underlying the hostile priming effect.

Finally, another notable feature of the methods is that both the current study and McCarthy et al. (2018) involved vetting of the methods prior to data collection. Several individuals outside of the research teams (including one of the original authors in McCarthy et al. [2018] and experts chosen in the peer-review process of the current manuscript) approved of the methods and analysis plans prior to the data being collected. Although the current methods are not beyond critique, it is notable that the current methods were scrutinized and approved independently of the results obtained.

The current endeavor was not flawless though. We notably observed different rates of exclusion among those in the hostile priming condition and those in the neutral condition. These condition-dependent exclusions raise two possible issues. First, the analysts may have been able to infer which condition would have higher rates of exclusions, which would effectively have "unblinded" the study during data analysis. To this point, we would reiterate that we had an approved analysis plan and two analysts who independently produced identical results. Thus, even if the different rates of exclusion made the data effectively unblinded, we had other checks in place to minimize the effect of possible biases on our data analyses. Second, and perhaps more serious, is that condition-dependent exclusions might threaten the internal validity of our study (see Zhou & Fishbach, 2016). That is, even though the current study randomly assigned participants to priming condition, because participants in the hostile priming condition were omitted more often than participants in the neutral condition, the final samples may have been systematically different. To this point, we tested several ways of excluding participants and no exclusion criterion or combination of criteria dramatically affected the conclusions. We also did not find differences between the rates of exclusions for failing the Instructional Manipulation Checks, which suggests, perhaps weakly, that non-excluded participants did not differ in their attentiveness or conscientiousness. Although we do not have data to more thoroughly test why this differential exclusion occurred, this finding highlights the need to test and report condition-level rates of exclusion in future hostile priming studies.

Collectively then, we believe the current methods are at least as sound as many previously-published priming studies, we had ample statistical power to detect effects that would be considered theoretically relevant, and the current meta-analyses do not suffer from study selection biases.

When taken together, the current results and the results of McCarthy et al. (2018) strongly suggest that Srull-and-Wyer-esque methods do not reliably produce hostile priming effects that could be routinely and affordably detected by researchers.

What Does the Lack of an Effect Mean?

Does incidental exposure to hostile-related words affect subsequent perceptions of hostility? That is, does the "hostile priming effect" exist? Although this is the question we want to answer, the question is imprecise because, despite how these things are sometimes discussed, there is not *a* hostile priming effect. Methodologically, there are many priming tasks that have been used, new priming tasks that could be developed and used, endless variations of stimuli, numerous procedural details, etc. that all could be put together into studies of a "hostile priming effect." There are also several theoretical accounts of the cognitive process through which these priming effects operate. We do not claim that the current study can speak to all these methodological and theoretical variations. Indeed, no individual study can answer the more sweeping question of whether an abstractly described effect exists or not (see Yarkoni, 2019 on difficulties of generalizing results and Molden, 2014 about avoiding a monolithic view of "social priming" effects). With these caveats stated, we offer a few conclusions that are best thought of as falling on a continuum.

On one end of the continuum lies a narrow-yet-confident perspective: The current results are highly informative for studies that are methodologically similar to Srull & Wyer (1979; e.g., the construct of interest is "hostility," the priming method is the sentence descrambling task, and the outcome variable is the rated impression of a described individual). For these methodologically similar studies, the current study is obviously relevant and we conclude there is serious doubt on whether those methods readily produce a hostile priming effect. Because we believe the methods of the current study are sound, it is reasonable to doubt some of the theoretical reasons—such as how long these priming effects ought to last or the potency of these priming effects to have a measurable influence on the impression formed of a described individual—for why these methods do not produce a hostile priming effect.

As one moves towards the other end of the continuum, there is a broader-yet-less-confident perspective: The current results are still informative for studies that differ methodologically, although with decreasing confidence as those methodological differences becomes greater. For example, as the construct of interest changes from hostility to another trait, as the priming method changes from the sentence descrambling task to another priming method, as the outcome variable changes from rating a described individual to another outcome, or some combination of these changes, it requires more assumptions to generalize from the current results to others. These mounting assumptions make loose links between the results of the current study and these other studies and, thus, these mounting assumptions could make the ultimate impact of the current results more of a glance than a wallop (e.g., see Fabrigar et al., 2020).

Nevertheless, we argue that the current study is still informative, even for social priming studies whose methods do not have a great deal of overlap with the current study. For starters, the methods from Srull & Wyer (1979) have left an unavoidable impression on social-cognitive research in the decades since it was published. Indeed, literally dozens of published studies have been directly modeled after Srull & Wyer (1979), each with their own theoretical extension, but also built upon the assumption the to-be-explained empirical effect is replicable. The implication from this line of research is that any activation of cognitive representations that occurred during the priming task had relatively long-lasting and fairly potent effects (either because of residual activation or through another cognitive process that carried the effect of this initial activation). But there is mounting evidence that these to-be-explained empirical effects are not as replicable as once believed.[6] Until studies regularly produce priming effects with high methodological rigor and with enough specificity that independent labs can produce the effects (i.e., until there is a readily produced empirical phenomenon), it seems sensible to conclude that brief exposure to construct-relevant words will not have a measurable and predictable influence on a single-shot outcome that is measured several seconds after the priming manipulation. Consequently, the empirical support for theories explaining such effects might be weaker than was once thought.

To be clear, we are not declaring all possible hostile priming effects to be unreplicable. Although speculative, if a hostile priming effect exists, and if that effect is due to the residual activation of cognitive information, it is likely to be an exceedingly fleeting effect that would not be potent enough to have a detectable influence on a single-shot outcome measured immediately following the prime. That is, we believe a likely scenario for a hostile priming effect to emerge is when the priming stimuli would activate cognitive information, which would decay rapidly. And to detect this fleeting activation, it would likely take several trials where a sensitive outcome would measure the effect of the prime immediately after the presentation of the priming stimuli (see Payne et al., 2016 as an example). Thus, it is conceivable that such a study could be done, and that study could claim to be examining a "hostile priming effect," but the methods of this hypothetical study would look much different than the Srull-and-Wyer-esque methods. Such a hypothetical study also would be unlikely to have lasting implications for social judgement as Srull & Wyer sug-

6 Indeed, see the recent replications of other priming studies such as "flag priming" (Klein et al., 2014), "elderly priming" (Doyen et al., 2012), and "professor priming" (O'Donnell et al., 2018). Each of these studies replicated the methods for a longer-than-fleeting priming effect, each of these studies were not lacking for statistical power, and none detected the same effect as the original study.

gested. For example, in their discussion, Srull & Wyer (1979) state that "although these effects decrease with the time interval between their activation and the acquisition of information to be interpreted or encoded, they are sometimes detectable even after 24 hours" (p. 1670). The current results would suggest this is highly unlikely.

Conclusion

The human mind, and the methods used to study the mind, are complex. And, a criticism of some replications is that merely repeating previously-used methods and expecting the same effect is over-simplifying the matter (e.g., Cesario, 2014; Stroebe & Strack, 2014). However, we also do not want to over-complicate a matter that is relatively simple. Namely, if a method repeatedly produces an effect, then researchers should increase their confidence in those methods; if a method does not repeatedly produce an effect, then confidence ought to decrease. The current study did not detect a hostile priming effect, we now have little confidence that Srull-and-Wyer-esque methods produce a hostile priming effect, and we believe it would be unwise for researchers to place more resources into hostile priming effect studies using these methods.

Data Accessibility Statement

See Table 1.

Competing Interests

Authors do not have any competing interests to declare.

Contributions

Randy J. McCarthy and Will Gervais were the lead researchers; Patrick Forscher and John Sakaluk wrote the analysis code and analyzed the data; and all other authors collected data and reviewed and edited the manuscript.

Funding

We have no funding to declare for this project.

Submitted: September 22, 2020 PDT, Accepted: December 02, 2020 PDT

REFERENCES

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 253–285). Cambridge University Press.

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

Cesario, J., Plaks, J. E., Hagiwara, N., Navarrete, C. D., & Higgins, E. T. (2010). The ecology of automaticity: How situational contingencies shape action semantics and social behavior. *Psychological Science*, *21*(9), 1311–1317. https://doi.org/10.1177/0956797610378685

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

DeCoster, J., & Claypool, H. M. (2004). A meta-analysis of priming effects on impression formation supporting a general model of informational biases. *Personality and Social Psychology Review*, *8*(1), 2–27. https://doi.org/10.1207/s15327957pspr0801_1

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS One*, *7*(1), e29081. https://doi.org/10.1371/journal.pone.0029081

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

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B. Jr., 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: A "many labs" replication project. *Social Psychology*, *45*(3), 132–142. https://doi.org/10.1027/1864-9335/a000178

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/cbo9781139087759

Loersch, C., & Payne, B. K. (2014). Situated inferences and the what, who, and where of priming. *Social Cognition*, *32*(Supplement), 137–151. https://doi.org/10.1521/soco.2014.32.supp.137

McCarthy, R. J., Skowronski, J. J., Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., Acar, O. A., Aczel, B., Bakos, B. E., Barbosa, F., Baskin, E., Bègue, L., Ben-Shakhar, G., Birt, A. R., Blatz, L., Charman, S. D., Claesen, A., Clay, S. L., … Yıldız, E. (2018). Registered Replication Report: Srull and Wyer (1979). *Advances in Methods and Practices in Psychological Science*, *1*(3), 321–336. https://doi.org/10.1177/2515245918777487

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

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

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2*. https://CRAN.R-project.org/package=BayesFactor

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., Alshaif, N., Andringa, R., Aveyard, M., Babincak, P., Balatekin, N., Baldwin, S. A., Banik, G., Baskin, E., Bell, R., Białobrzeska, O., Birt, A. R., Boot, W. R., Braithwaite, S. R., … Zrubka, M. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*, *13*(2), 268–294. https://doi.org/10.1177/1745691618755704

Payne, B. K., Brown-Iannuzzi, J. L., & Loersch, C. (2016). Replicable effects of primes on human behavior. *Journal of Experimental Psychology: General*, *145*(10), 1269–1279. https://doi.org/10.1037/xge0000201

Quintana, D. (2017, July 29). *How to calculate statistical power for your meta-analysis*. https://towardsdatascience.com/how-to-calculate-statistical-power-for-your-meta-analysis-e108ee586ae8

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*(10), 1660–1672. https://doi.org/10.1037/0022-3514.37.10.1660

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

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., Skowronski, J. J., Acar, O. A., Aczel, B., Bakos, B. E., Barbosa, F., Baskin, E., Bègue, L., Ben-Shakhar, G., Birt, A. R., Blatz, L., Charman, S. D., Claesen, A., Clay, S. L., … Yıldız, E. (2018). Registered Replication Report: Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, *1*(3), 299–317. https://doi.org/10.1177/2515245918781032

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Yarkoni, T. (2019). *The Generalizability Crisis*. Center for Open Science. https://doi.org/10.31234/osf.io/jqw35

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493–504. https://doi.org/10.1037/pspa0000056

SUPPLEMENTARY MATERIALS

Peer Review History

Download: https://collabra.scholasticahq.com/article/18738-a-multi-site-collaborative-study-of-the-hostile-priming-effect/attachment/49515.docx?auth_token=50pW8_D_loy6l7yIBH3f