Check for updates

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(3) 321–336
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245918777487
www.psychologicalscience.org/AMPPS

Registered Replication Report on Srull and Wyer (1979)

Randy J. McCarthy*, John J. Skowronski*, Bruno Verschuere*, Ewout H. Meijer*, Ariane Jim*, Katherine Hoogesteyn*, Robin Orthey*, Oguz A. Acar, Balazs Aczel, Bence E. Bakos, Fernando Barbosa, Ernest Baskin, Laurent Bègue, Gershon Ben-Shakhar, Angie R. Birt, Lisa Blatz, Steve D. Charman, Aline Claesen, Samuel L. Clay, Sean P. Coary, Jan Crusius, Jacqueline R. Evans, Noa Feldman, Fernando Ferreira-Santos, Matthias Gamer, Coby Gerlsma, Sara Gomes, Marta González-Iraizoz, Felix Holzmeister, Juergen Huber, Rafaele J. C. Huntjens, Andrea Isoni, Ryan K. Jessup, Michael Kirchler, Nathalie klein Selle, Lina Koppel, Marton Kovacs, Tei Laine, Frank Lentz, David D. Loschelder, Elliot A. Ludvig, Monty L. Lynn, Scott D. Martin, Neil M. McLatchie, Mario Mechtel, Galit Nahari, Asil Ali Özdoğru, Rita Pasion, Charlotte R. Pennington, Arne Roets, Nir Rozmann, Irene Scopelliti, Eli Spiegelman, Kristina Suchotzki, Angela Sutan, Peter Szecsi, Gustav Tinghög, Jean-Christian Tisserand, Ulrich S. Tran, Alain Van Hiel, Wolf Vanpaemel, Daniel Västfjäll, Thomas Verliefde, Kévin Vezirian, Martin Voracek, Lara Warmelink, Katherine Wick, Bradford J. Wiggins, Keith Wylie, and Ezgi Yıldız
*Lead authors

Multilab direct replication of: Experiment 1 from 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*, 1660–1672. doi:10.1037/0022-3514.37.10.1660

Protocol vetted by: Robert S. Wyer

Abstract

Srull and Wyer (1979) demonstrated that exposing participants to more hostility-related stimuli caused them subsequently to interpret ambiguous behaviors as more hostile. In their Experiment 1, participants descrambled sets of words to form sentences. In one condition, 80% of the descrambled sentences described hostile behaviors, and in another condition, 20% described hostile behaviors. Following the descrambling task, all participants read a vignette about a man named Donald who behaved in an ambiguously hostile manner and then rated him on a set of personality traits. Next, participants rated the hostility of various ambiguously hostile behaviors (all ratings on scales from 0 to 10). Participants who descrambled mostly hostile sentences rated Donald and the ambiguous behaviors as approximately 3 scale points more hostile than did those who descrambled mostly neutral sentences. This Registered Replication Report describes the results of 26 independent replications (N = 7,373 in the total sample; k = 22 labs and N = 5,610 in the

Corresponding Authors:
Randy J. McCarthy, Center for the Study of Family Violence and Sexual Assault, Northern Illinois University, DeKalb, IL 60115
E-mail: rmccarthy3@niu.edu
John J. Skowronski, Department of Psychology, Northern Illinois University, DeKalb, IL 60115
E-mail: jskowron@niu.edu

primary analyses) of Srull and Wyer's Experiment 1, each of which followed a preregistered and vetted protocol. A random-effects meta-analysis showed that the protagonist was seen as 0.08 scale points more hostile when participants were primed with 80% hostile sentences than when they were primed with 20% hostile sentences (95% confidence interval, CI = [0.004, 0.16]). The ambiguously hostile behaviors were seen as 0.08 points less hostile when participants were primed with 80% hostile sentences than when they were primed with 20% hostile sentences (95% CI = [−0.18, 0.01]). Although the confidence interval for one outcome excluded zero and the observed effect was in the predicted direction, these results suggest that the currently used methods do not produce an assimilative priming effect that is practically and routinely detectable.

Keywords
hostility, priming, impression formation, replication, Many Labs, open data, open materials, preregistered

In a now-classic study, Srull and Wyer (1979) demonstrated that exposure to hostility-related stimuli affected how people subsequently interpreted the actions of a person (Donald) described in a brief vignette and how they rated ambiguously hostile behaviors. Srull and Wyer's report has had considerable influence on the field of social cognition: It is heavily cited, the Donald vignette has been used in several subsequent studies (e.g., Bartholow & Heinz, 2006; Devine, 1989; Philippot, Schwarz, Carrera, De Vries, & Van Yperen, 1991), the original findings have inspired many conceptual replications and extensions (e.g., Bargh & Pietromonaco, 1982; Herr, 1986; Mussweiler & Damisch, 2008), and the report is considered foundational both in the hostility-priming literature and for studies that have extended priming effects beyond the domain of social judgments (e.g., Bargh, Chen, & Burrows, 1996; Dijksterhuis & van Knippenberg, 1998). A review and meta-analysis of the literature on priming effects in impression-formation tasks (DeCoster & Claypool, 2004) found a moderately sized effect of priming on judgments about social targets (d = 0.35, 95% confidence interval, CI = [0.30, 0.41]).

However, in recent years, the robustness and replicability of some prominent social priming findings have been questioned (e.g., Cesario, 2014; Molden, 2014). Given its foundational role and continued citation as evidence of how priming can influence social judgments (e.g., Bargh, 2006, 2014; Higgins & Eitam, 2014; Strack & Schwarz, 2016), Srull and Wyer's study meets the Registered Replication Report (RRR) criterion of having high "replication value." In the current RRR project, we sought to estimate the magnitude and reliability of the hostility-priming effects reported by Srull and Wyer through a series of independently conducted direct replications.

Original Hostility-Priming Methods and Effects

The primary effect of interest in the current RRR is a phenomenon known as *assimilative priming*: an effect in which exposure to priming stimuli causes subsequent judgments to incorporate more of the qualities of the primed construct.[1] Srull and Wyer tested two predictions regarding social assimilative priming. First, the amount of "activation" of a primed mental representation (manipulated by exposing people to more or fewer of the priming stimuli) should be associated with the extent to which social judgments are affected. Second, the activation of primed mental representations should decay with the passage of time, thereby reducing the influence of the primes on subsequent social judgments.[2]

In Srull and Wyer's Experiment 1 (the focus of this RRR), participants first completed a sentence-descrambling task in which they underlined three of four words that could then be used to create a grammatically correct three-word sentence (e.g., "hand break his nose" can form the sentence "break his nose" or "break his hand"). Different groups of participants completed sets of scrambled sentences that, when unscrambled, referred to different proportions of hostile behaviors. After the sentence-descrambling task, participants were directed to a second researcher, who was ostensibly conducting a different study. The "other study" consisted of three tasks. In the first task, participants read a vignette about a day in the life of a man named Donald who displayed a number of behaviors that were ambiguously hostile (e.g., "Donald insisted that the waitress replace all the silverware because it was dirty"). They then rated Donald on 12 traits using a scale from 0 (*not at all*) to 10 (*extremely*). Ratings for 6 of these traits (i.e., hostile, unfriendly, dislikeable, kind, considerate, and thoughtful) were averaged (after the latter 3 were reverse-scored) to form an index of the extent to which Donald was perceived as hostile. In the second task, participants rated the hostility of 15 individual behaviors (e.g., "Refusing to let a salesperson enter their house") using a scale from 0 (*not at all hostile*) to 10 (*extremely hostile*). Five behaviors were clearly hostile, 5 behaviors were clearly not hostile, and 5 behaviors were ambiguous with respect to hostility. Responses to the 5 ambiguously hostile behaviors were averaged to form an index of the extent to which the ambiguous behaviors were

perceived to be hostile. Finally, participants estimated the co-occurrence of hostility with 11 other traits. However, Srull and Wyer did not report the results from these co-occurrence ratings, so they were not included in the current replication project.

The design of Srull and Wyer's Experiment 1 included a number of between-participants variables:

- Participants descrambled a total of either 30 sentences or 60 sentences;
- Either 80% or 20% of the descrambled sentences referred to hostile behaviors;
- The three rating tasks were completed immediately after the descrambling task, after a 1-hr delay, or after a 24-hr delay; and
- Participants read one of two different versions of the Donald vignette.

Experiment 1 was completed by a total of 96 participants, 4 in each cell of the 2 × 2 × 3 × 2 between-participants factorial design.[3] Srull and Wyer hypothesized that participants who descrambled a greater proportion of hostile sentences would view both Donald and the ambiguously hostile behaviors as more hostile.

The priming effect Srull and Wyer reported was large. For the ratings of Donald, the mean difference between the two cells most comparable to the conditions tested in this replication project (the 30-trials/no-delay conditions; see the Method section for details) was approximately 3 scale points on the 11-point scale. For the ratings of the ambiguously hostile behaviors, the mean difference between these two cells also was approximately 3 scale points on the 11-point scale. However, there may have been an error in the statistics reported in the original article (R. S. Wyer, personal communication to D. J. Simons, August 22, 2016). The possibility of an erroneously reported statistic is consistent with the fact that for a similar study (Srull & Wyer, 1980), the standard deviations reported were approximately 6 times as large and the effect size was substantially smaller (see DeCoster & Claypool, 2004, for a detailed discussion). The uncertainty about the size and credibility of the original effect underscores the need for precise estimates of social assimilative priming effects.[4]

Disclosures

Preregistration

The approved protocol for the RRR was posted on the Open Science Framework project page at https://osf.io/3bwx5/. Each laboratory preregistered their Editor-approved implementation of the official protocol on their individual project page, and those preregistrations are available by visiting the labs' project pages (linked from the Contributing Labs section at https://osf.io/hrju6/wiki/home/). Each laboratory team reported (on their project page) how they determined their sample size and documented all data exclusions. Any departures from the official protocol or the lab's preregistered implementation are documented in the Lab Implementation Appendix at https://osf.io/uskr8/ (also at http://journals.sagepub.com/doi/suppl/10.1177/2515245918777487). Drafts of the meta-analysis scripts were written in a data-blind manner, using simulated data. Those preregistered versions are posted at https://osf.io/jp45u/. The final scripts were updated to address minor formatting inconsistencies across labs, to improve the appearance of figures, and to add exploratory analyses. All changes from the data-blind scripts are noted in the final scripts posted at https://osf.io/mcvt7/.

Data, materials, and online resources

All materials are available at https://osf.io/rbejp/. All data and analyses are available at https://osf.io/mcvt7/wiki/home/. Supplementary online materials include the Lab Implementation Appendix, which documents the individual labs' contributions to the project (https://osf.io/uskr8/ and http://journals.sagepub.com/doi/suppl/10.1177/2515245918777487).

Reporting

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

Ethical approval

Each laboratory obtained any necessary institutional-review-board or ethical approval from their home institution to accommodate differences in the requirements at different universities and in different countries.

Method

Contributing labs

The current replication project involved a total of 26 labs (see the appendix following the Discussion section for a list of the authors participating at each lab). Data were collected between November 2016 and November 2017. The study materials, which were originally created in English, were translated into eight different languages (13 labs used materials in English, 5 labs used German, 4 used Dutch, 1 used French, 1 used Hebrew, 1 used Hungarian, 1 used Portuguese, 1 used Swedish, and 1 used Turkish; note that 2 labs used materials in two languages).

Study participants

Total sample sizes for the individual contributing labs ranged from 207 to 377 participants (total *N* before exclusions = 7,373; 2,147 men, 5,175 women, and 51 participants with missing gender information; mean age = 20.77 years, *SD* = 2.90). Table 1 summarizes the demographics of each individual sample. Each lab preregistered its data-collection stopping rules prior to beginning data collection.

Procedure

Participants completed the study as part of a packet that included other tasks (see Table 2). After providing consent and then demographic information, participants completed the tasks for this study. These tasks always came before the tasks for the companion replication project (see the next section).

Participants first completed the sentence-descrambling task. In this task, they viewed 30 groups of four words (e.g., "him yell swear at") and were instructed to underline three words that would create a grammatically correct sentence (e.g., "yell at him" or "swear at him").[5] Some of these 30 items could be completed only as sentences describing hostile behaviors, and others could be completed only as sentences describing nonhostile behaviors. Participants were randomly assigned to one of two conditions: mostly hostile sentences (24 of the 30, or 80%, described hostile behaviors) or mostly neutral sentences (6 of the 30, or 20%, described hostile behaviors). Participants then read the vignette and rated the protagonist of the vignette on the same traits and using the same response scale (0 = *not at all*, 10 = *extremely*) as in Srull and Wyer's Experiment 1. Next, participants viewed and rated the hostility of the same set of behaviors as in Srull and Wyer's Experiment 1 (with minor modifications described in the next section), again using the same response scale (0 = *not at all hostile*, 10 = *extremely hostile)* as in that experiment.

Thus, the experimental design had one between-participants variable (i.e., 80% hostile primes vs. 20% hostile primes) and two separate dependent variables (average hostility ratings of the vignette's protagonist and average hostility ratings of the ambiguously hostile behaviors).

Known differences between this RRR study and Srull and Wyer's Experiment 1

This replication project was developed in parallel with a replication project (Verschuere et al., 2018, this issue) focusing on Mazar, Amir, and Ariely's (2008) Experiment 1. The two projects were developed to be combined into one data-collection effort, which allowed them to be framed as a series of unrelated tasks. Whenever possible, the current project used Srull and Wyer's original materials, including the Donald vignette and the materials for rating Donald, and the ambiguously hostile behaviors. However, we had to either re-create or modify some of the study materials, and we had to modify some aspects of the procedure to accommodate the constraints of this RRR project. Our decisions concerning these modifications were driven by goals to minimize the differences between our methods and Srull and Wyer's original methods and to maintain the theoretically necessary conditions for an assimilative priming effect to emerge. These modifications were made in consultation with Wyer.

The original sentence-descrambling stimuli were unavailable, so the first author generated and pretested new stimuli that were consistent with the description of the original stimuli (see https://osf.io/32pkz/ for details on the pretesting). Further, in consultation with Wyer, we modified the pronouns in the original list of behaviors to make them gender neutral and to fix minor wording errors. Given that young adults may be unfamiliar with the action of slamming a handset onto a receiver to hang up a phone, we also changed the listed behavior of "slamming down a phone" to "abruptly hanging up a phone." Finally, because the name Donald might have activated unwanted associations with Donald Trump following the 2016 election in the United States, we changed the name of the protagonist of the vignette from Donald to Ronald.

The purpose of the current project was to attempt to replicate the assimilative priming effect originally reported by Srull and Wyer. To do so, rather than including all of the factors in the original 2 × 2 × 3 × 2 design, we focused on a comparison of two conditions that showed a clear effect in Srull and Wyer's experiment. Given that all variables in the original study were manipulated between groups, excluding some of the variables should not have affected the primary outcome measure. Thus, for both practical reasons (to avoid the need for participants to return later) and because it showed strong priming effects in the original study, we chose to focus on the immediate-testing condition. Specifically, the sentence-descrambling task in the current replication project always included 30 trials; for half of the participants, 80% of the descrambled sentences (i.e., 24 out of 30) described hostile behaviors, and for the other half, 20% of the descrambled sentences (i.e., 6 out of 30) described hostile behaviors. All participants completed the ratings of Ronald and of the ambiguously hostile behaviors immediately after the priming task. Though this design did not permit an assessment of all

Table 1. Demographic Information on Each Lab's Sample

	Full sample			Included in primary analyses?[a]	Sample after exclusions[b]		
Lab	*N*	Gender	Mean age		*N*	Gender	Mean age
Acar	237	82 males, 153 females, 2 unrecorded	21.15 (2.03)	Yes	214	76 males, 138 females	20.96 (1.58)
Aczel	245	53 males, 191 females, 1 unrecorded	20.82 (1.73)	Yes	225	47 males, 178 females	20.76 (1.63)
Baskin	207	105 males, 102 females	19.63 (0.90)	No	198	99 males, 99 females	19.60 (0.79)
Birt	234	46 males, 188 females	21.50 (4.52)	Yes	205	37 males, 168 females	20.37 (2.09)
Blatz	320	48 males, 264 females, 8 unrecorded	22.05 (3.58)	No	212	24 males, 188 females	20.66 (2.19)
Evans	332	97 males, 234 females, 1 unrecorded	21.68 (3.20)	Yes	243	69 males, 174 females	20.94 (1.68)
Ferreira-Santos	291	76 males, 214 females, 1 unrecorded	19.99 (4.34)	Yes	234	59 males, 175 females	19.35 (1.60)
González-Iraizoz	235	39 males, 196 females	18.65 (0.88)	Yes	229	38 males, 191 females	18.64 (0.87)
Holzmeister	274	130 males, 143 females, 1 unrecorded	21.89 (2.13)	Yes	253	118 males, 135 females	21.62 (1.61)
Huntjens	216	62 males, 152 females, 2 unrecorded	20.85 (2.06)	No	190	54 males, 136 females	20.64 (1.77)
klein Selle and Rozmann	337	76 males, 258 females, 3 unrecorded	22.29 (1.72)	Yes	299	65 males, 234 females	22.21 (1.52)
Koppel	263	119 males, 143 females, 1 unrecorded	22.03 (2.20)	Yes	242	108 males, 134 females	21.76 (1.73)
Laine	313	41 males, 269 females, 3 unrecorded	19.39 (2.14)	Yes	253	32 males, 221 females	19.24 (1.31)
Loschelder	248	83 males, 156 females, 9 unrecorded	21.30 (2.00)	Yes	226	79 males, 147 females	21.13 (1.63)
McCarthy	318	123 males, 193 females, 2 unrecorded	21.41 (2.95)	Yes	279	106 males, 173 females	20.88 (1.66)
Meijer	377	97 males, 279 females, 1 unrecorded	20.31 (1.90)	Yes	348	86 males, 262 females	20.20 (1.59)
Özdoğru	365	42 males, 323 females	20.27 (2.63)	Yes	332	36 males, 296 females	19.96 (1.32)
Pennington	255	51 males, 196 females, 8 unrecorded	20.29 (4.44)	Yes	217	45 males, 172 females	19.31 (1.40)
Roets	253	28 males, 224 females, 1 unrecorded	18.44 (2.02)	Yes	204	23 males, 181 females	18.47 (0.96)
Suchotzki	256	46 males, 207 females, 3 unrecorded	20.35 (1.68)	Yes	246	44 males, 202 females	20.30 (1.65)
Sutan	304	154 males, 148 females, 2 unrecorded	20.64 (0.91)	Yes	252	129 males, 123 females	20.62 (0.93)
Tran	277	77 males, 200 females	24.59 (3.55)	No	194	38 males, 156 females	22.95 (1.36)
Vanpaemel	288	64 males, 224 females	20.27 (3.16)	Yes	237	48 males, 189 females	20.25 (1.76)
Verschuere	302	88 males, 213 females, 1 unrecorded	19.76 (2.20)	Yes	285	83 males, 202 females	19.60 (1.62)
Wick	367	219 males, 148 females	19.30 (1.91)	Yes	343	205 males, 138 females	19.15 (1.26)
Wiggins	259	101 males, 157 females, 1 unrecorded	20.85 (2.04)	Yes	244	93 males, 151 females	20.80 (1.93)
Total	7,373	2,147 males, 5,175 females, 51 unrecorded	20.77 (2.90)		6,404	1,841 males, 4,563 females	20.38 (1.85)

Note: Numbers in parentheses are standard deviations.
[a]Labs were not included in the primary analyses if they had fewer than 100 participants in each condition in the final sample. [b]Individual participants were not included in analyses if they (a) did not complete all of the items in the sentence-descrambling task, (b) were not currently students, (c) did not complete all the ratings of Ronald, (d) did not complete the ratings of all the behaviors, (e) were less than 18 years old or older than 25 years old, or (f) did not provide gender information, or if (g) the experimenters recorded any other information that warranted exclusion (e.g., they did not follow instructions).

Table 2. List of Tasks in the Combined Procedure for the Two Registered Replication Reports (RRRs)

Task	Description	RRR
Demographics and informed consent	Participants provided their age, sex, and major and gave written informed consent.	Both
Sentence descrambling (hostility priming) (Srull & Wyer, 1979, Experiment 1)	For each of 30 groups of four words, participants marked the three words that would make a complete sentence (e.g., "child the question watch"). Either 80% or 20% of the descrambled sentences described hostile behaviors.	Current
Vignette (Srull & Wyer, 1979, Experiment 1)	Participants read a short story about a man named Ronald who behaved in a manner that could be seen as hostile (e.g., he told a beggar to find a job).	Current
Judgments of the vignette's protagonist (Srull & Wyer, 1979, Experiment 1)	Participants rated Ronald on 12 characteristics (e.g., unfriendly).	Current
Judgments of behaviors (Srull & Wyer, 1979, Experiment 1)	Participants judged the hostility of 15 behaviors (e.g., refusing to let a salesperson into one's house).	Current
Abstract reasoning (materials provided by C. Chabris)	Participants solved a 10-item nonverbal-intelligence task.	Filler
Priming (moral reminder)	Participants wrote as many of the Ten Commandments as they could remember or the names of 10 books they had read in high school.	Verschuere et al.
Matrix (cheating opportunity) (Mazar et al., 2008, Experiment 1)	Participants tried to find the numbers that added up exactly to 10 (e.g., 3.18 and 6.82) in as many of 20 matrices as time allowed. They then tore either a blank page or the matrix page out of the task booklet.	Verschuere et al.
Collection slip (Mazar et al., 2008, Experiment 1)	Participants reported how many matrices they had solved.	Verschuere et al.
Alternative Uses Test (Guilford, 1967)	Participants listed as many possible uses of a paper clip as they could think of.	Filler
Religiousness[a]	Participants used a scale from 1 (*not at all*) to 5 (*completely*) to answer three questions: "How religious are you?"; "To what extent do you believe in a God?"; and "To what extent do you believe in a punishing God?"	Verschuere et al.
Fatigue[a] (Profile of Mood States; McNair, Lorr, & Droppleman, 1971) and sleep	Participants rated their fatigue, by using a scale from 1 (*not at all*) to 5 (*extremely*) to indicate how much they felt worn out, fatigued, exhausted, sluggish, weary, and bushed; participants also reported how many hours they had slept the previous night.	Filler
Time estimation[a]	Participants estimated how much time they had taken in the timed tasks of this battery.	Verschuere et al.
HEXACO[a] (Ashton & Lee, 2009)	Participants completed this 60-item personality scale.	Filler

Note: This table lists the order of all of the tasks included in the combined procedure for the current RRR, on Srull and Wyer's (1979) Experiment 1, and for Verschuere et al.'s (2018, this issue) RRR, on Mazar, Amir, and Ariely's (2008) Experiment 1. All between-participants conditions were counterbalanced.

[a]These tasks were included to allow exploratory analyses of possible moderators of cheating. The religiousness task was included in the preregistered plan.

the variables (i.e., delay, number of priming sentences) manipulated by Srull and Wyer, the pair of conditions that we chose to include provides a test of the replicability of the assimilative hostility-priming effect they reported.

We also used only one of the two vignettes from the original study. One vignette was reported in the text of Srull and Wyer's article, and the other was provided by Wyer in preparation for this project. Given the possibility that cultural norms for hostility have changed since 1979, the first author conducted a norming study (details available at https://osf.io/32pkz/) to assess how hostile Donald was viewed in the two vignettes in the absence of priming. The vignette we ultimately used elicited somewhat lower and slightly more variable ratings of Donald's hostility than the Srull and Wyer reported. Given the results of this norming study, and in consultation with Wyer, we elected to use the vignette that was not included in the text of the original article.

Finally, one consequence of the need to include this project's tasks as part of a larger packet of tasks was that a modification to the cover story was required. Srull

and Wyer's participants were asked to complete the sentence-descrambling task ahead of another study that was described as unrelated. In the current project, the sentence-descrambling task and ratings tasks were completed as part of a single administration in a large classroom setting. Further, although the tasks for this project always came first, the anticipation of additional and presumably unrelated tasks could have induced a different task-completion mind-set (e.g., "I need to move along fast to get this done") than might have been present in Srull and Wyer's study. As the RRR project was being developed, Wyer noted that these features were potentially meaningful departures from the conditions of the original study. However, we believe that the spirit of the original cover story was maintained: The packet was described as a collection of separate writing, memory, imagination, judgment, and problem-solving tasks, and the priming and social judgment tasks were distinct enough that participants likely viewed them as unrelated. Finally, other studies have successfully used sentence-descrambling tasks to examine hostile attributions without using the procedures Srull and Wyer described (e.g., Bargh et al., 1996; Crouch, Skowronski, Milner, & Harris, 2008; DeWall & Bushman, 2009; Srull & Wyer, 1980; Wann & Branscombe, 1990).

Prespecified exclusions

Given that this study was conducted in conjunction with another replication project, inclusion criteria that were specific to that study applied to the current one as well. Participants were not included if they did not complete the critical items or if they did not follow the study's instructions. Also, participants who were less than 18 years old or more than 25 years old (an exclusion criterion for the other replication project) or who did not provide gender information were not included. Labs were not included if they did not collect data from a minimum of 100 participants in each condition (see https://osf.io/9afwn/ for details of the exclusion criteria).

In total, four labs did not collect data from the minimum of 100 participants in each condition. Although these labs were omitted from the primary analyses, they were included in the ancillary analyses. Among the 22 labs that were included in the primary analyses, sample sizes after exclusions ranged from 204 to 348 participants (1,626 men, 3,984 women; mean age = 20.30 years, $SD = 1.82$; see Table 1 for information about each individual lab).

Results

The meta-analyses we report used a random-effects model and the restricted maximum likelihood estimator for estimating the amount of heterogeneity. They were conducted using the metafor package in R (e.g., Viechtbauer, 2010).

Primary analyses

Judgments of Ronald's hostility. As in Srull and Wyer's Experiment 1, ratings of the vignette's protagonist on the six traits—hostile, unfriendly, dislikeable, kind, considerate, and thoughtful—were averaged (after reverse-coding the last three traits) to yield a hostility index score for each participant. We then obtained an average hostility rating for each priming condition for each lab. Using these average ratings, we conducted a random-effects meta-analysis on the difference between conditions to obtain an overall estimate of the size of the hostility-priming effect.

Our results are summarized in Figure 1 (see Supplemental Tables, in the Supplemental Material, for the individual labs' results). Srull and Wyer's Figure 1 showed that participants in the 80%-hostile priming condition rated Donald as approximately 3 scale units more hostile (on a scale from 0 to 10) than did those in the 20%-hostile priming condition. The meta-analysis of the 22 studies that met our inclusion criteria of having at least 100 participants in each condition revealed an overall difference of 0.08 points (95% CI = [0.004, 0.16]). The heterogeneity of this effect across labs was no bigger than what would be expected as a result of sampling error alone, $\tau = 0.08$, $Q(21) = 25.31$, $p = .23$, and the I^2 statistic indicated that about 17.73% of the observed variance of the effect sizes was caused by systematic differences between studies.

Judgments of ambiguously hostile behaviors. As did Srull and Wyer, we averaged each participant's hostility ratings for the five ambiguously hostile behaviors separately for each condition for each lab. These five behaviors were as follows:

- "Telling a garage mechanic that they will have to go somewhere else if the mechanic cannot fix their car that same day"
- "Refusing to let a salesperson enter their house"
- "When asked to donate blood to the Red Cross, lying by saying they had diabetes and therefore could not do so"
- "Demanding their money back from a sales clerk"
- "Refusing to pay their rent until the landlord paints their apartment"

Using these average ratings, we conducted a random-effects meta-analysis on the difference between conditions to obtain an overall estimate of the size of the hostility-priming effect.

Lab	20%-Hostile Condition Mean Rating	20%-Hostile Condition *n*	80%-Hostile Condition Mean Rating	80%-Hostile Condition *n*	Effect Size	95% CI
Srull & Wyer (1979)	5.56	8	8.57	8	3.01	
Birt	7.06	102	7.61	103	0.55	[0.15, 0.94]
Aczel	6.19	116	6.50	109	0.32	[–0.03, 0.67]
Suchotzki	7.37	125	7.65	121	0.29	[–0.01, 0.58]
Ferreira-Santos	6.97	130	7.25	104	0.28	[–0.09, 0.66]
Wick	6.90	170	7.16	173	0.26	[–0.05, 0.58]
Holzmeister	7.21	128	7.42	125	0.22	[–0.16, 0.59]
Loschelder	7.26	112	7.43	114	0.17	[–0.17, 0.51]
McCarthy	7.35	125	7.51	154	0.16	[–0.19, 0.51]
klein Selle & Rozmann	7.13	142	7.28	157	0.14	[–0.16, 0.44]
Koppel	7.20	120	7.33	122	0.13	[–0.23, 0.48]
Acar	6.66	108	6.72	106	0.07	[–0.37, 0.50]
Meijer	6.92	170	6.99	178	0.07	[–0.20, 0.33]
Özdoğru	6.58	171	6.61	161	0.03	[–0.30, 0.36]
Roets	7.06	100	7.08	104	0.02	[–0.30, 0.34]
González-Iraizoz	7.31	116	7.33	113	0.02	[–0.31, 0.34]
Vanpaemel	7.03	116	7.04	121	0.01	[–0.28, 0.31]
Wiggins	6.96	124	6.95	120	–0.01	[–0.41, 0.39]
Laine	6.19	125	6.18	128	–0.01	[–0.34, 0.31]
Sutan	6.02	129	5.99	123	–0.03	[–0.36, 0.30]
Evans	7.29	124	7.18	119	–0.11	[–0.49, 0.27]
Verschuere	7.14	146	6.89	139	–0.26	[–0.54, 0.03]
Pennington	7.60	114	7.31	103	–0.29	[–0.63, 0.05]
Meta-Analytic Average	6.97	2,813	7.06	2,797	0.08	[0.00, 0.16]



Fig. 1. Results of the primary analyses: forest plot of the difference in ratings of Ronald's hostility between the 80%-hostile and 20%-hostile priming conditions. For each of the 22 labs that met all the inclusion criteria, the figure shows the mean rating and sample size in each condition. The labs are listed in order of the size of the difference between the conditions (80%-hostile priming condition minus 20%-hostile priming condition). The squares show the observed effect sizes, the error bars represent 95% confidence intervals (CIs), and the size of each square represents the magnitude of the standard error for the lab's effect (larger squares indicate less variability in the estimate). To the right, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure, the estimated effect from Srull and Wyer's (1979) Experiment 1 is shown (the data are no longer available, and we could not compute confidence intervals from the available information). The bottom row in the figure presents the unweighted means of the individual sample means and the outcome of a random-effects meta-analysis.

Our results are summarized in Figure 2. Srull and Wyer's Figure 2 showed that participants in the 80%-hostile priming condition rated the ambiguous behaviors as approximately 3 scale units more hostile (on a scale from 0 to 10) than did those in the 20%-hostile priming condition. The meta-analysis of the 22 studies that met our inclusion criteria of having at least 100 participants in each condition revealed a difference of −0.08 points (95% CI = [−0.18, 0.01]). The heterogeneity of this effect across labs was no bigger than what would be expected as a result of sampling error alone, $\tau = 0.10$, $Q(21) = 24.39$, $p = .27$, and the I^2 statistic indicated that about 18.03% of the observed variance of the effect sizes was caused by systematic differences between studies.

Ancillary analyses

We conducted two sets of ancillary analyses. The first set examined the pattern of results when we included all laboratories and participants regardless of the size of the final sample. The second set examined whether the language of the stimuli moderated the hostility-priming effects.

The impact of the exclusion criteria. The primary analyses excluded data from laboratories that collected data on fewer than 100 participants in each priming condition. The first ancillary analysis included data from all laboratories even if they did not meet that criterion. All the exclusion criteria for individual participants (e.g.,

Lab	20%-Hostile Condition Mean Rating	20%-Hostile Condition *n*	80%-Hostile Condition Mean Rating	80%-Hostile Condition *n*	Effect Size	95% CI
Srull & Wyer (1979)	4.50	8	7.49	8	2.99	
Laine	4.32	125	4.59	128	0.27	[–0.13, 0.67]
Özdoğru	3.63	171	3.86	161	0.22	[–0.13, 0.57]
Acar	4.48	108	4.65	106	0.16	[–0.26, 0.58]
Aczel	5.30	116	5.42	109	0.12	[–0.27, 0.51]
Loschelder	3.94	112	4.01	114	0.07	[–0.36, 0.50]
Wick	4.87	170	4.93	173	0.07	[–0.32, 0.45]
Ferreira-Santos	4.55	130	4.56	104	0.01	[–0.47, 0.49]
Wiggins	4.88	124	4.89	120	0.01	[–0.43, 0.44]
Roets	5.58	100	5.58	104	–0.01	[–0.40, 0.38]
Meijer	4.17	170	4.14	178	–0.03	[–0.38, 0.32]
McCarthy	4.88	125	4.81	154	–0.07	[–0.51, 0.37]
Suchotzki	4.50	125	4.40	121	–0.10	[–0.48, 0.28]
Holzmeister	4.51	128	4.40	125	–0.11	[–0.53, 0.31]
Birt	5.18	102	5.07	103	–0.11	[–0.64, 0.41]
Verschuere	4.69	146	4.55	139	–0.14	[–0.52, 0.23]
Koppel	4.00	120	3.86	122	–0.14	[–0.57, 0.28]
Vanpaemel	5.34	116	5.16	121	–0.17	[–0.53, 0.18]
González-Iraizoz	5.12	116	4.92	113	–0.21	[–0.63, 0.22]
Evans	4.38	124	4.16	119	–0.22	[–0.68, 0.24]
Sutan	5.27	129	4.78	123	–0.49	[–0.84, -0.13]
klein Selle & Rozmann	3.95	142	3.44	157	–0.51	[–0.90, -0.13]
Pennington	5.19	114	4.67	103	–0.52	[–0.97, -0.07]
Meta-Analytic Average	4.67	2,813	4.58	2,797	–0.08	[–0.18, 0.01]

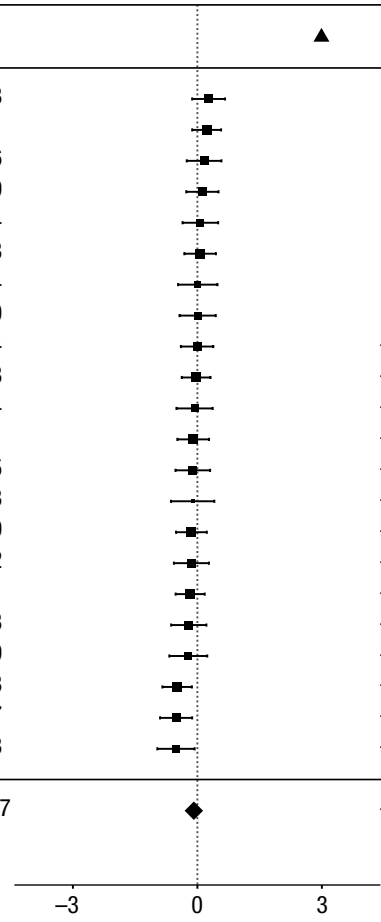

Fig. 2. Results of the primary analyses: forest plot of the difference between the 80%-hostile and 20%-hostile priming conditions in ratings of hostility for the five ambiguously aggressive behaviors. For each of the 22 labs that met all the inclusion criteria, the figure shows the mean rating and sample size in each condition. The labs are listed in order of the size of the difference between the conditions (80%-hostile priming condition minus 20%-hostile priming condition). The squares show the observed effect sizes, the error bars represent 95% confidence intervals (CIs), and the size of each square represents the magnitude of the standard error for the lab's effect (larger squares indicate less variability in the estimate). To the right, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure, the estimated effect from Srull and Wyer's (1979) Experiment 1 is shown (the data are no longer available, and we could not compute confidence intervals from the available information). The bottom row in the figure presents the unweighted means of the individual sample means and the outcome of a random-effects meta-analysis.

failure to complete all priming trials or to follow instructions) were still applied in this analysis.

In this full sample, which included 26 labs with 6,404 total participants, we observed a between-conditions difference of 0.07 (95% CI = [0.003, 0.14]) for the trait ratings of Ronald (see Fig. 3) and a between-conditions difference of −0.10 (95% CI = [−0.19, −0.001]) for the behavior ratings (see Fig. 4). For the trait ratings of Ronald, the heterogeneity of this effect across labs was no bigger than what would be expected as a result of sampling error alone, $\tau = 0.05$, $Q(25) = 25.89$, $p = .41$, $I^2 = 7.10\%$. For the behavior ratings, the heterogeneity of this effect across labs was also no bigger than what would be expected as a result of sampling error alone, $\tau = 0.13$, $Q(25) = 35.03$, $p = .09$, $I^2 = 28.86\%$.

Overall, the results with the full sample were nearly identical to the results based on labs with at least 100 participants per condition.

Moderation by language. The original stimuli were created in English. We examined whether the language of the materials moderated the hostility-priming effect. Two labs administered the tasks using both a nontranslated version and a translated version of the materials. This allowed us to compute an effect for each version in the case of these labs. Thus, to test for moderation by

Lab	20%-Hostile Condition Mean Rating	20%-Hostile Condition *n*	80%-Hostile Condition Mean Rating	80%-Hostile Condition *n*	Effect Size	95% CI
Srull & Wyer (1979)	5.56	8	8.57	8	3.01	
Birt	7.06	102	7.61	103	0.55	[0.15, 0.94]
Aczel	6.19	116	6.50	109	0.32	[–0.03, 0.67]
Suchotzki	7.37	125	7.65	121	0.29	[–0.01, 0.58]
Ferreira-Santos	6.97	130	7.25	104	0.28	[–0.09, 0.66]
Wick	6.90	170	7.16	173	0.26	[–0.05, 0.58]
Holzmeister	7.21	128	7.42	125	0.22	[–0.16, 0.59]
Loschelder	7.26	112	7.43	114	0.17	[–0.17, 0.51]
McCarthy	7.35	125	7.51	154	0.16	[–0.19, 0.51]
klein Selle & Rozmann	7.13	142	7.28	157	0.14	[–0.16, 0.44]
Koppel	7.20	120	7.33	122	0.13	[–0.23, 0.48]
Acar	6.66	108	6.72	106	0.07	[–0.37, 0.50]
Meijer	6.92	170	6.99	178	0.07	[–0.20, 0.33]
Tran	7.23	95	7.27	99	0.04	[–0.35, 0.43]
Özdoğru	6.58	171	6.61	161	0.03	[–0.30, 0.36]
Roets	7.06	100	7.08	104	0.02	[–0.30, 0.34]
Baskin	7.21	104	7.23	94	0.02	[–0.39, 0.43]
González-Iraizoz	7.31	116	7.33	113	0.02	[–0.31, 0.34]
Vanpaemel	7.03	116	7.04	121	0.01	[–0.28, 0.31]
Wiggins	6.96	124	6.95	120	–0.01	[–0.41, 0.39]
Huntjens	6.89	92	6.88	98	–0.01	[–0.34, 0.32]
Laine	6.19	125	6.18	128	–0.01	[–0.34, 0.31]
Blatz-Crusius	7.43	122	7.42	90	–0.01	[–0.40, 0.37]
Sutan	6.02	129	5.99	123	–0.03	[–0.36, 0.30]
Evans	7.29	124	7.18	119	–0.11	[–0.49, 0.27]
Verschuere	7.14	146	6.89	139	–0.26	[–0.54, 0.03]
Pennington	7.60	114	7.31	103	–0.29	[–0.63, 0.05]
Meta-Analytic Average	7.01	3,226	7.09	3,178	0.07	[0.00, 0.14]

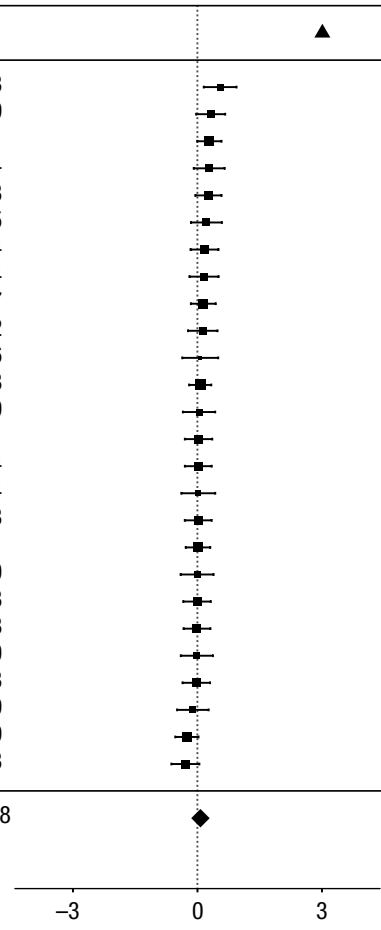

Fig. 3. Results of the ancillary analyses: forest plot of the difference in ratings of Ronald's hostility between the 80%-hostile and 20%-hostile priming conditions. For each of the 26 labs in the full sample, the figure shows the mean rating and sample size in each condition. The labs are listed in order of the size of the difference between the conditions (80%-hostile priming condition minus 20%-hostile priming condition). The squares show the observed effect sizes, the error bars represent 95% confidence intervals (CIs), and the size of each square represents the magnitude of the standard error for the lab's effect (larger squares indicate less variability in the estimate). To the right, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure, the estimated effect from Srull and Wyer's (1979) Experiment 1 is shown (the data are no longer available, and we could not compute confidence intervals from the available information). The bottom row in the figure presents the unweighted means of the individual sample means and the outcome of a random-effects meta-analysis.

language, we ran analyses that included 28 effects (i.e., effects for 26 labs, 2 of which provided 2 effects each). The original English version of the materials was used with 13 samples, and these stimuli were translated into eight languages (German: $k = 5$; Dutch: $k = 4$; French: $k = 1$; Hebrew: $k = 1$; Hungarian: $k = 1$; Portuguese: $k = 1$; Swedish: $k = 1$; and Turkish: $k = 1$). For purposes of the moderation analysis, we tested whether the effects obtained using the translated versions (regardless of the language) differed from the effects obtained using the nontranslated (i.e., English) version. Thus, the comparison had 1 degree of freedom.

For the trait ratings of Ronald, the translated versions of the stimuli yielded hostility-priming effects that were not significantly different from those obtained with the nontranslated, English version, $QM(1) = 0.12$, $p = .73$. For the ratings of the ambiguous behaviors as well, the translated versions of the stimuli yielded hostility-priming effects that were not significantly different from those obtained with the nontranslated, English version, $QM(1) = 1.36$, $p = .24$.

Discussion

In recent years, the replicability of assimilative priming effects has come into question. Other RRRs (e.g., Cheung et al., 2016; O'Donnell et al., 2018), Many Labs studies (e.g., Klein et al., 2014), and individual studies

Lab	20%-Hostile Condition Mean Rating	20%-Hostile Condition *n*	80%-Hostile Condition Mean Rating	80%-Hostile Condition *n*	Effect Size	95% CI
Srull & Wyer (1979)	4.50	8	7.49	8	2.99	
Huntjens	4.36	92	4.79	98	0.43	[–0.03, 0.88]
Laine	4.32	125	4.59	128	0.27	[–0.13, 0.67]
Özdoğru	3.63	171	3.86	161	0.22	[–0.13, 0.57]
Acar	4.48	108	4.65	106	0.16	[–0.26, 0.58]
Aczel	5.30	116	5.42	109	0.12	[–0.27, 0.51]
Loschelder	3.94	112	4.01	114	0.07	[–0.36, 0.50]
Wick	4.87	170	4.93	173	0.07	[–0.32, 0.45]
Ferreira-Santos	4.55	130	4.56	104	0.01	[–0.47, 0.49]
Wiggins	4.88	124	4.89	120	0.01	[–0.43, 0.44]
Roets	5.58	100	5.58	104	–0.01	[–0.40, 0.38]
Meijer	4.17	170	4.14	178	–0.03	[–0.38, 0.32]
McCarthy	4.88	125	4.81	154	–0.07	[–0.51, 0.37]
Suchotzki	4.50	125	4.40	121	–0.10	[–0.48, 0.28]
Holzmeister	4.51	128	4.40	125	–0.11	[–0.53, 0.31]
Birt	5.18	102	5.07	103	–0.11	[–0.64, 0.41]
Verschuere	4.69	146	4.55	139	–0.14	[–0.52, 0.23]
Koppel	4.00	120	3.86	122	–0.14	[–0.57, 0.28]
Vanpaemel	5.34	116	5.16	121	–0.17	[–0.53, 0.18]
González-Iraizoz	5.12	116	4.92	113	–0.21	[–0.63, 0.22]
Blatz-Crusius	5.01	122	4.80	90	–0.22	[–0.68, 0.24]
Evans	4.38	124	4.16	119	–0.22	[–0.68, 0.24]
Baskin	5.16	104	4.68	94	–0.48	[–0.96, –0.00]
Sutan	5.27	129	4.78	123	–0.49	[–0.84, –0.13]
Tran	4.37	95	3.88	99	–0.50	[–0.98, –0.02]
klein Selle & Rozmann	3.95	142	3.44	157	–0.51	[–0.90, –0.13]
Pennington	5.19	114	4.67	103	–0.52	[–0.97, –0.07]
Meta-Analytic Average	4.68	3,226	4.58	3,178	–0.10	[–0.19, -0.00]

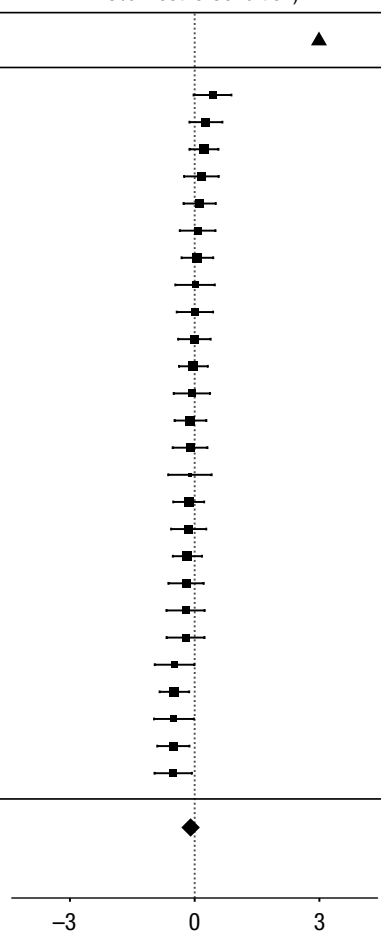

Fig. 4. Results of the ancillary analyses: forest plot of the difference between the 80%-hostile and 20%-hostile priming conditions in ratings of hostility for the five ambiguously aggressive behaviors. For each of the 26 labs in the full sample, the figure shows the mean rating and sample size in each condition. The labs are listed in order of the size of the difference between the conditions (80%-hostile priming condition minus 20%-hostile priming condition). The squares show the observed effect sizes, the error bars represent 95% confidence intervals (CIs), and the size of each square represents the magnitude of the standard error for the lab's effect (larger squares indicate less variability in the estimate). To the right, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure, the estimated effect from Srull and Wyer's (1979) Experiment 1 is shown (the data are no longer available, and we could not compute confidence intervals from the available information). The bottom row in the figure presents the unweighted means of the individual sample means and the outcome of a random-effects meta-analysis.

(e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; McCarthy, 2014; Pashler, Coburn, & Harris, 2012) have not found evidence of such priming effects. This context of doubt provided a reason to explore the replicability of one of the most influential assimilative priming effects in the field of social cognition: the hostility-priming effect reported by Srull and Wyer in 1979.

The current replication project had two outcome variables. The first was the average hostility rating of the vignette's protagonist. Participants who completed the version of the sentence-descrambling task that had 80% hostile primes—the group theorized to be more primed by hostility—rated the protagonist to be 0.08 points more hostile (on an 11-point scale) than did participants who completed the version of the task that had 20% hostile primes. The 95% CI around this estimate excluded zero (i.e., the meta-analytic assimilative priming effect was significantly different from zero), and the effect observed at 18 of the 26 labs was numerically in the predicted direction. However, the overall effect was much smaller than both the original effect reported by Srull and Wyer and the expected effect size derived from reviews of the published literature (e.g., DeCoster & Claypool's, 2004, meta-analysis).

The second outcome was the average hostility rating of five ambiguously hostile behaviors. Participants in the 80%-hostile priming condition rated these behaviors as 0.08 points less hostile (on an 11-point scale) than did participants in the 20%-hostile priming condition. Not only is this effect smaller than the original effect reported by Srull and Wyer, but it is numerically in the opposite direction. An effect in the predicted direction was observed at only 9 of the 26 labs. In short, the meta-analytic effects of assimilative priming for both outcome measures were close to 0 scale units—much smaller differences than the approximately 3-scale-unit differences reported by Srull and Wyer.

One possible explanation for the discrepancies between our results and the previously reported effects is that the published literature exhibits publication bias that leads to an inflated view of the magnitude and replicability of the hostility-priming effect. Indeed, in DeCoster and Claypool's (2004) meta-analysis, the magnitude of the published effects was negatively related to the precision of those effects, a pattern that is consistent with (but not definitive proof of) the presence of publication bias. In the presence of publication bias, the literature might paint a misleading picture of the replicability and magnitude of assimilative priming effects. Unsurprisingly, then, when publication bias is eliminated from the data, as in the current replication project, the obtained effect size is much smaller than a simple synthesis of the published literature would suggest.

Method differences between the original study and our project also might have contributed to the discrepant results. In comparison with Srull and Wyer's study, ours used different sentence-descrambling primes, only one of the two original vignettes, and a different name for the protagonist (Ronald rather than Donald). Although such procedural details, either individually or in combination, could change the outcome of a study, it is hard to construct a cogent explanation for how they could do so. Moreover, we pretested the priming stimuli and the vignette to ensure that they activated the relevant constructs, and there is no obvious reason to believe that the protagonist's name or other procedural differences should matter for obtaining an assimilative priming effect.

However, other differences in methods might more plausibly have contributed to the differences in outcomes. In Srull and Wyer's Experiment 1, participants were exposed to an unexpected task (the sentence-descrambling task) before completing the task for which they had signed up (which was supposedly unrelated to the sentence-descrambling task). In our study, the priming task and the person judgment tasks were framed as unrelated, but both appeared in the same lengthy booklet. This difference in the cover story could have led to different results. For example, the booklet's length could have induced a task-completion mind-set (e.g., "I have to move along fast to get this done") that might not have been present in Srull and Wyer's study, leading to shallower stimulus processing than in the original. The group context also might have led our participants to be less attentive to the study materials, and assimilative priming effects might be weakened as a result. During the planning phase of the project, Wyer noted this change in the cover story as a possible reason to expect a different outcome. However, in a study subsequent to the one we focused on in this replication project, Srull and Wyer (1980) replicated their original assimilative priming effects using a procedure that involved only one researcher who gave participants a study packet containing "a wide array of experiments, contributed by various members of the psychology faculty, [to be completed] over the course of 2 hours" (p. 845). Srull and Wyer justified this procedural choice by stating that "these instructions, along with the fact that the tasks were highly dissimilar, were intended to make subjects think there was no relationship between any two tasks in the sequence" (p. 845). Given this precedent, it seems that neither using a single experimenter nor a lengthy packet of "unrelated" tasks has historically been considered a barrier to creating the conditions necessary to produce an assimilative priming effect.

We can exclude one difference as a plausible explanation for the different outcomes. Several labs contributing to this RRR translated their priming-task materials into non-English languages, and priming effects might have been reduced because of subtle differences in meaning despite quality controls for these translations. However, our ancillary analyses showed that the effects observed in the current project were generally homogeneous across labs, so language differences do not appear to explain the difference between the effect sizes we observed and those reported by Srull and Wyer.

In sum, we observed a small assimilative priming effect in the predicted direction for ratings of Ronald (i.e., the confidence interval for ratings of Ronald excluded zero) and a similarly small effect in the opposite direction for judgments about behaviors. Both effect-size estimates were close to zero and were substantially smaller than those previously reported in published research. Our results suggest that the procedures we used in this replication study are unlikely to produce an assimilative priming effect that researchers could practically and routinely detect. Indeed, to detect priming effects as small as the 0.08-scale-unit difference we observed (which works out to approximately d = 0.06, 95% CI = [0.01, 0.12]), a study would need 4,362 participants in each priming condition to have 80% power with an alpha set to .05. Although the current

procedures were unfavorable for producing assimilative priming effects, other procedures, such as within-participants repeated measures designs with a brief delay between the priming stimuli and the outcome measure, might provide a more promising approach for future assimilative priming research (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Payne, Brown-Iannuzzi, & Loersch, 2016; Payne, Cheng, Govorun, & Stewart, 2005).

Appendix: Author Affiliations

(The Supplemental Material includes an additional appendix with a one-paragraph summary for each lab that specifies any departures from the protocol or from their own preregistered plan, as well as which analyses included the data from that lab. This Lab Implementation Appendix is also available at https://osf.io/vxz7q/).

Lead Labs

Randy J. McCarthy, Northern Illinois University
John J. Skowronski, Northern Illinois University

Bruno Verschuere, University of Amsterdam
Ariane Jim, University of Amsterdam, now at Ghent University

Ewout H. Meijer, Maastricht University
Katherine Hoogesteyn, Maastricht University
Robin Orthey, Maastricht University and University of Portsmouth

Contributing Labs
(Alphabetical by last name of first author)

Oguz A. Acar, City, University of London
Irene Scopelliti, City, University of London

Balazs Aczel, Institute of Psychology, ELTE Eötvös Loránd University
Bence E. Bakos, Institute of Psychology, ELTE Eötvös Loránd University
Marton Kovacs, Institute of Psychology, ELTE Eötvös Loránd University
Peter Szecsi, Institute of Psychology, ELTE Eötvös Loránd University

Ernest Baskin, Haub School of Business, Saint Joseph's University
Sean P. Coary, Haub School of Business, Saint Joseph's University

Angie R. Birt, Mount Saint Vincent University

Lisa Blatz, University of Cologne
Jan Crusius, University of Cologne

Jacqueline R. Evans, Florida International University
Keith Wylie, Florida International University
Steve D. Charman, Florida International University

Fernando Ferreira-Santos, University of Porto
Fernando Barbosa, University of Porto
Rita Pasion, University of Porto

Marta González-Iraizoz, University of Warwick
Andrea Isoni, University of Warwick
Elliot A. Ludvig, University of Warwick

Felix Holzmeister, University of Innsbruck
Juergen Huber, University of Innsbruck
Michael Kirchler, University of Innsbruck

Rafaele J. C. Huntjens, University of Groningen
Coby Gerlsma, University of Groningen

Nathalie klein Selle, Hebrew University of Jerusalem
Noa Feldman, Hebrew University of Jerusalem
Gershon Ben-Shakhar, Hebrew University of Jerusalem
Nir Rozmann, Bar-Ilan University
Galit Nahari, Bar-Ilan University

Lina Koppel, Linköping University
Gustav Tinghög, Linköping University
Daniel Västfjäll, Linköping University and Decision Research, Eugene, Oregon

Tei Laine, Université Grenoble Alpes
Kévin Vezirian, Université Grenoble Alpes
Laurent Bègue, Université Grenoble Alpes

David D. Loschelder, Leuphana University of Lueneburg
Mario Mechtel, Leuphana University of Lueneburg

Asil Ali Özdoğru, Üsküdar University
Ezgi Yıldız, Üsküdar University

Charlotte R. Pennington, University of the West of England
Neil M. McLatchie, Lancaster University
Lara Warmelink, Lancaster University

Arne Roets, Ghent University
Alain Van Hiel, Ghent University

Kristina Suchotzki, University of Würzburg
Matthias Gamer, University of Würzburg

Angela Sutan, Université Bourgogne Franche-Comté, Burgundy School of Business - CEREN
Frank Lentz, Université Bourgogne Franche-Comté, Burgundy School of Business - CEREN
Jean-Christian Tisserand, Université Bourgogne Franche-Comté, Burgundy School of Business - CEREN
Eli Spiegelman, Université Bourgogne Franche-Comté, Burgundy School of Business - CEREN

Ulrich S. Tran, University of Vienna
Martin Voracek, University of Vienna

Wolf Vanpaemel, University of Leuven
Aline Claesen, University of Leuven
Sara Gomes, University of Leuven
Thomas Verliefde, University of Leuven

Katherine Wick, Abilene Christian University
Ryan K. Jessup, Abilene Christian University
Monty L. Lynn, Abilene Christian University

Bradford J. Wiggins, Brigham Young University-Idaho
Scott D. Martin, Brigham Young University-Idaho
Samuel L. Clay, Brigham Young University-Idaho

Action Editor

Daniel J. Simons served as action editor for this article.

Author Contributions

R. J. McCarthy proposed the replication project reported in this article. R. J. McCarthy and J. J. Skowronski were responsible for developing and gathering the materials necessary for the project, as well as for writing the manuscript, and R. J. McCarthy wrote the analysis code. All the lead authors were involved with designing the overall procedure for the combined project that included the study reported by Verscheure et al. (2018, this issue). Each author contributed by conducting the study in his or her respective lab and providing valuable input on the manuscript.

Acknowledgments

We thank Robert S. Wyer for providing materials for the study and guidance about necessary changes to the protocol, Chris Chabris for providing the abstract-reasoning task included as part of the battery, and Katherine Wood for assisting in creating the forest plots.

Declaration of Conflicting Interests

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

Funding

This project was partially supported by an NWO (Netherlands Organisation for Scientific Research) Replication Grant (No. 401.16.001). The Association for Psychological Science and the Arnold Foundation provided funding to participating laboratories to defray the costs of running the study.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/2515245918777487

Open Practices



All data, analysis scripts, and materials have been made publicly available via the Open Science Framework. The data and scripts can be accessed at https://osf.io/mcvt7/wiki/home/, and the materials can be accessed at https://osf.io/rbejp/wiki/home/. The design and analysis plans were preregistered at the Open Science Framework and can be accessed at https://osf.io/3bwx5 and https://osf.io/hrju6/wiki/home/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/2515245918777487. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Notes

1. There also are *contrastive priming effects*, wherein increasing exposure to priming stimuli causes judgments that social targets have *less* of the quality of the primed construct (e.g., Bless & Schwarz, 2010; Martin, 1986). An example of a contrastive hostility-priming effect is Herr's (1986) demonstration that participants exposed to more extreme exemplars of hostility subsequently judge a social target as less hostile.
2. It is not a given that the influence of priming stimuli will weaken over time. For example, some researchers have primed goals, which theoretically involve auxiliary cognitive processes that can maintain or even increase the effect of the priming stimuli on outcome variables with the passage of time (e.g., Bargh, Lee-Chai, Barndollar, Gollwitzer, & Trötschel, 2001).
3. The logistics of the current replication project precluded us from manipulating the delay between the priming task and the social judgment tasks. Thus, we did not include any of the delay conditions that Srull and Wyer did.
4. Notably, Srull and Wyer conceptually replicated their hostility-priming findings (with somewhat weaker effects) by assessing the impact of "kindness" priming on social judgments of kindness in their Experiment 2. However, the current project focused only on their hostility-priming result.
5. Some labs reported difficulty when literally translating each word of the sentence-descrambling task from English into other languages (e.g., the labs encountered issues with gendered words or the way articles are used). In some cases, to allow for successful translations, the words were changed slightly, or the instructions were changed so that participants were told to unscramble "4 words or phrases." The individual labs' translations are available at https://osf.io/rbejp/.

References

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*, 340–345. doi:10.1080/00223890902935878

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*, 147–168. doi:10.1002/ejsp.336

Bargh, J. A. (2014). The historical origins of priming as the preparation of behavioral responses: Unconscious carryover and contextual influences of real-world importance. *Social Cognition*, *32*, 209–224. doi:10.1521/soco.2014.32.supp.209

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*, 230–244. doi:10.1037/0022-3514.71.2.230

Bargh, J. A., Lee-Chai, A., Barndollar, K., Gollwitzer, P. M., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*, 1014–1027. doi:10.1037/0022-3514.81.6.1014

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437–449. doi:10.1037/0022-3514.43.3.437

Bartholow, B. D., & Heinz, A. (2006). Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias. *Psychological Science*, *17*, 30–37. doi:10.1111/j.1467-9280.2005.01661x

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 42, pp. 319–373). San Diego, CA: Academic Press.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*, 40–48. doi:10.1177/1745691613513471

Cheung, I., Campbell, L., LeBel, E. P., Ackerman, R. A., Aykutoğlu, B., Bahník, Š., . . . Yong, J. C. (2016). Registered Replication Report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*, 750–764. doi:10.1177/1745691616664694

Crouch, J. L., Skowronski, J. J., Milner, J. S., & Harris, B. (2008). Parental responses to infant crying: The influence of child physical abuse risk and hostile priming. *Child Abuse & Neglect*, *32*, 702–710. doi:10.1016/j.chiabu.2007.11.002

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*, 2–27. doi:10.1207/S15327957PSPR0801_1

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5

DeWall, C. N., & Bushman, B. J. (2009). Hot under the collar in a lukewarm environment: Words associated with hot temperature increase aggressive thoughts and hostile perceptions. *Journal of Experimental Social Psychology*, *45*, 1045–1047. doi:10.1016/j.jesp.2009.05.003

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*, 865–877. doi:10.1037/0022-3514.74.4.865

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), Article e29081. doi:10.1371/journal.pone.0029081

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027. doi:10.1037/0022-3514.69.6.1013

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, *51*, 1106–1115. doi:10.1037/0022-3514.51.6.1106

Higgins, E. T., & Eitam, B. (2014). Priming…shmiming: It's about knowing *when* and *why* stimulated memory representations become active. *Social Cognition*, *32*, 225–242. doi:10.1521/soco.2014.32.supp.225

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*, 132–142. doi:10.1027/1864-9335/a000178

Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, *51*, 493–504. doi:10.1037/0022-3514.51.3.493

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*, 633–644. doi:10.1509/jmkr.45.6.633

McCarthy, R. J. (2014). Close replication attempts of the heat priming-hostile perception effect. *Journal of Experimental Social Psychology*, *54*, 165–169. doi:10.1016/j.jesp.2014.04.014

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Profile of Mood States manual*. San Diego, CA: Educational and Industrial Testing Service.

Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition*, *32*, 1–11. doi:10.1521/soco.2014.32.supp.1

Mussweiler, T., & Damisch, L. (2008). Going back to Donald: How comparisons shape judgmental priming effects. *Journal of Personality and Social Psychology*, *95*, 1295–1315. doi:10.1037/a0013261

O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., . . . Zrubka, M. (2018). Registered Replication Report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*, *13*, 268–294. doi:10.1177/1745691618755704

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

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

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293. doi:10.1037/0022-3514.89.3.277

Philippot, P., Schwarz, N., Carrera, P., De Vries, N., & Van Yperen, N. W. (1991). Differential effects of priming at the encoding and judgment stage. *European Journal of Social Psychology*, *21*, 293–302. doi:10.1002/ejsp.2420210403

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*, 1660–1672. doi:10.1037/0022-3514.37.10.1660

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, *38*, 841–856. doi:10.1037/0022-3514.38.6.841

Strack, F., & Schwarz, N. (2016). Social priming: Information accessibility and its consequences. *Current Opinion in Psychology*, *12*, iv–vii. doi:10.1016/j.copsyc.2016.11.001

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., . . . Yıldız, E. (2018). Registered Replication Report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, *1*, 299–317.

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

Wann, D. L., & Branscombe, N. R. (1990). Person perception when aggressive or nonaggressive sports are primed. *Aggressive Behavior*, *16*, 27–32. doi:10.1002/1098-2337(1990)16