Check for updates

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2018, Vol. 13(2) 268–294
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691618755704
www.psychologicalscience.org/PPS
SAGE

Registered Replication Report: Dijksterhuis and van Knippenberg (1998)

Michael O'Donnell*, Leif D. Nelson*, Evi Ackermann, Balazs Aczel, Athfah Akhtar, Silvio Aldrovandi, Nasseem Alshaif, Ronald Andringa, Mark Aveyard, Peter Babincak, Nursena Balatekin, Scott A. Baldwin, Gabriel Banik, Ernest Baskin, Raoul Bell, Olga Białobrzeska, Angie R. Birt, Walter R. Boot, Scott R. Braithwaite, Jessie C. Briggs, Axel Buchner, Desiree Budd, Kathryn Budzik, Lottie Bullens, Richard L. Bulley, Peter R. Cannon, Katarzyna Cantarero, Joseph Cesario, Stephanie Chambers, Christopher R. Chartier, Peggy Chekroun, Clara Chong, Axel Cleeremans, Sean P. Coary, Jacob Coulthard, Florien M. Cramwinckel, Thomas F. Denson, Marcos Díaz-Lago, Theresa E. DiDonato, Aaron Drummond, Julia Eberlen, Titus Ebersbach, John E. Edlund, Katherine M. Finnigan, Justin Fisher, Natalia Frankowska, Efraín García-Sánchez, Frank D. Golom, Andrew J. Graves, Kevin Greenberg, Mando Hanioti, Heather A. Hansen, Jenna A. Harder, Erin R. Harrell, Andree Hartanto, Michael Inzlicht, David J. Johnson, Andrew Karpinski, Victor N. Keller, Olivier Klein, Lina Koppel, Emiel Krahmer, Anthony Lantian, Michael J. Larson, Jean-Baptiste Légal, Richard E. Lucas, Dermot Lynott, Corey M. Magaldino, Karlijn Massar, Matthew T. McBee, Neil McLatchie, Nadhilla Melia, Michael C. Mensink, Laura Mieth, Samantha Moore-Berg, Geraldine Neeser, Ben R. Newell, Marret K. Noordewier, Asil Ali Özdoğru, Myrto Pantazi, Michał Parzuchowski, Kim Peters, Michael C. Philipp, Monique M. H. Pollmann, Panagiotis Rentzelas, Rosa Rodríguez-Bailón, Jan Philipp Röer, Ivan Ropovik, Nelson A. Roque, Carolina Rueda, Bastiaan T. Rutjens, Katey Sackett, Janos Salamon, Ángel Sánchez-Rodríguez, Blair Saunders, Juliette Schaafsma, Michael Schulte-Mecklenbeck, David R. Shanks, Martin F. Sherman, Kenneth M. Steele, Niklas K. Steffens, Jessie Sun, Kyle J. Susa, Barnabas Szaszi, Aba Szollosi, Ricardo M. Tamayo, Gustav Tinghög, Yuk-yue Tong, Carol Tweten, Miguel A. Vadillo, Deisy Valcarcel, Nicolas Van der Linden, Michiel van Elk, Frenk van Harreveld, Daniel Västfjäll, Simine Vazire, Philippe Verduyn, Matt N. Williams, Guillermo B. Willis, Sarah E. Wood, Chunliang Yang, Oulmann Zerhouni, Robert Zheng, and Mark Zrubka

*Proposing authors

Corresponding Authors:
Michael O'Donnell, 2200 Piedmont Ave., Room F501, Berkeley, CA 94720
E-mail: mo279@berkeley.edu

Leif D. Nelson, 2200 Piedmont Ave., Room F524, Berkeley, CA 94720
E-mail: leif_nelson@haas.berkeley.edu

Multilab direct replication of: A variant of Experiment 4 from 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.
Protocol vetted by: Ap Dijksterhuis
Protocol and manuscript edited by: Daniel J. Simons

Abstract
Dijksterhuis and van Knippenberg (1998) reported that participants primed with a category associated with intelligence ("professor") subsequently performed 13% better on a trivia test than participants primed with a category associated with a lack of intelligence ("soccer hooligans"). In two unpublished replications of this study designed to verify the appropriate testing procedures, Dijksterhuis, van Knippenberg, and Holland observed a smaller difference between conditions (2%–3%) as well as a gender difference: Men showed the effect (9.3% and 7.6%), but women did not (0.3% and −0.3%). The procedure used in those replications served as the basis for this multilab Registered Replication Report. A total of 40 laboratories collected data for this project, and 23 of these laboratories met all inclusion criteria. Here we report the meta-analytic results for those 23 direct replications (total N = 4,493), which tested whether performance on a 30-item general-knowledge trivia task differed between these two priming conditions (results of supplementary analyses of the data from all 40 labs, N = 6,454, are also reported). We observed no overall difference in trivia performance between participants primed with the "professor" category and those primed with the "hooligan" category (0.14%) and no moderation by gender.

Keywords
priming, replication, intelligence

Brief exposure to a category or construct can mentally activate related categories or constructs. For example, people are faster to recognize the word *doctor* after initially seeing the word *nurse* (Meyer & Schvaneveldt, 1971), presumably because the activated "nurse" construct primes a broader category that also includes "doctor," making it more accessible. Soon after this discovery, social psychologists adapted the study of lexical priming to more complex domains, such as judgments about the traits of other people. For example, people exposed to a set of negative trait words (e.g., *reckless, conceited, aloof*, and *stubborn*) judged an ambiguous person more negatively than did people exposed to positive trait words (Higgins, Rholes, & Jones, 1977; see also Srull & Wyer, 1979). More recent work explored the idea that priming a category or construct could directly affect overt behavior. In one study, participants unscrambled a set of words that were either neutral or related to stereotypes of older adults (e.g., *wrinkle, gullible, bingo*). After that task was completed, and when participants thought the study was over, the experimenters surreptitiously recorded how quickly participants walked down the hall to the elevator. Participants who had been exposed to the older-adult primes walked more slowly than those who had been exposed to the neutral primes (Bargh, Chen, & Burrows, 1996). As the authors wrote, "The same priming techniques that have been shown in prior research to influence impression formation produce similar effects when the dependent measure is switched to social behavior" (p. 239).

This finding and others like it led to an explosion of studies testing whether priming category X produced changes in behavior Y: Priming "helpfulness" increased the likelihood that participants picked up dropped items (Macrae & Johnston, 1998); priming "cheetah" increased the speed with which participants picked up a questionnaire (Aarts & Dijksterhuis, 2002); priming "politician" increased long-windedness (Dijksterhuis & Van Knippenberg, 2000); priming "superhero" increased the likelihood that participants would volunteer time with an organization (Nelson & Norton, 2005); and priming with words such as *gamble* increased the likelihood that people would bet in a simulated card game (Payne, Brown-Iannuzzi, & Loersch, 2016).

This Registered Replication Report (RRR) project examined one of the most well-cited examples, a link between priming of social categories and performance on an objective measure of knowledge (Dijksterhuis & van Knippenberg, 1998). In a set of studies, participants were first primed with either intelligence or stupidity. Some participants first imagined what their daily life would be like as a "professor" or were primed with the concept of intelligence more generally, whereas other participants imagined their life as a "soccer hooligan" or were primed with the concept of stupidity more generally. As part of the priming, all participants completed a writing task in which they either wrote a paragraph describing their life

as a professor or soccer hooligan or listed synonyms for and characteristics associated with intelligence or stupidity. They then completed an ostensibly unrelated trivia test. Participants primed with intelligence answered significantly more questions correctly than did those primed with stupidity. This study has been cited more than 800 times, and many subsequent studies have obtained findings suggesting that intelligence primes can influence intellectual performance (Dijksterhuis, van Knippenberg, & Holland, 2014). Moreover, the shorthand "professor priming" is likely to be recognized instantly by many researchers in the field of social psychology.

Over the past 6 years, a number of prominent findings of priming in social psychology, including the professor-priming effect, have come under increased scrutiny. Most notably, a series of nine studies failed to find an effect of intelligence priming (Shanks et al., 2013). Yet a more recent study using *p*-curve (Simonsohn, Nelson, & Simmons, 2014) to evaluate the "18 *p*-values from 16 studies reported in 8 articles" on professor priming indicated that "the studies contain evidential value" (Lakens, 2017, p. 8). The results of the replication attempts for professor priming, coupled with "failed" replications of other priming studies around the same time (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012, failed to replicate the effect of older-adult primes on walking speed), touched off a heated debate about the replicability of such priming effects in general (Yong, 2012, 2015). This debate led skeptics to put out a call for researchers willing to subject their own studies to direct replication according to a vetted protocol. Ap Dijksterhuis volunteered to develop a professor-priming protocol for that purpose, and this RRR presents the results of a multilab replication based on that work.

Protocol and Procedures

To verify the accuracy of the original protocol, Dijksterhuis, van Knippenberg, and Holland reran the studies using the original paradigm from Dijksterhuis and van Knippenberg (1998). In those replications, they observed the effect for men but not for women. The lead authors of this RRR (O'Donnell and Nelson), with guidance and input from Dijksterhuis, developed a protocol that included the original professor and soccer-hooligan primes, a new and normed (with two different populations) set of trivia questions, an updated procedure, and an analysis strategy.[1]

Open Science Framework project page

The plan and results for this project were uploaded to the Open Science Framework (OSF). The main OSF project page is at https://osf.io/k27hm/.

Participants

Each lab was instructed to test a minimum of 25 participants per cell in a 2 (priming condition: professor vs. hooligan) × 2 (gender: female vs. male) between-participants design, and to include approximately equal numbers of men and women within each priming condition. Labs were encouraged to recruit at least 50 participants for each cell of the design. As in the original study, participants were recruited from undergraduate psychology participant pools or from equivalent populations (e.g., behavioral-marketing students). Participants were required to be college or university students ages 18 to 24 years. Predictably, not every lab had access to large populations, so the total sample size varied from lab to lab. All sample-size targets were preregistered, and the lead researchers and Editor remained blind to the outcomes of individual studies until all data collection was completed.

Laboratories that needed a description of their study for recruiting purposes described it as involving "a series of writing tasks and general-knowledge questions."

Testing settings

Participants were tested in person either individually or in small groups (no more than 10 participants per group). They completed the study in individual cubicles or at independent workstations positioned so that they could not see each other while performing the tasks. The experimenters were required to be at least 18 years of age, and any faculty member, postdoctoral researcher, graduate student, or trained undergraduate research assistant was eligible to conduct the study. Participants were assigned to either the professor- or the hooligan-priming condition by the computerized experimental script; this ensured both that assignment to condition was random and that the experimenter was blind to this assignment.

Materials

The original study and the RRR studies were conducted entirely on the computer. For the RRR protocol, the study was programmed using PsychoPy (Peirce, 2007). The cover story used in the RRR protocol was a variant of the one used in the original study; participants were told that the priming task and the trivia task were unrelated research being conducted by students in different fields of psychology. The original study used verbal debriefing to assess suspicions about the link between the priming task and the trivia task. The RRR studies used a computer-based funnel-debriefing questionnaire as a more systematic way to test for suspicion.

Before the protocol was finalized, Andy DeSoto, at the Association for Psychological Science, gathered a large set of trivia items for use in the study and normed them using Amazon Mechanical Turk. Michael O'Donnell and Leif Nelson then normed a subset of 150 potential items in an undergraduate-student sample at the University of California, Berkeley (students participated one at a time in cubicles, in keeping with the eventual study conditions). Accuracy was similar in the two samples. O'Donnell and Nelson then selected a subset of 30 items to use in the RRR protocol, selecting items that had a mean accuracy in the range from 40% to 70% in both norming studies. Dijksterhuis reviewed that set of items, and the lead authors made some substitutions so that the items covered a broader range of topics. Three items were later changed because their translations in some languages yielded transparently obvious answers.[2]

Main study sessions

At the start of each session, the experimenter read the following to the participant or group of participants:

> This study consists of a number of unrelated tasks that will provide pilot data and help us develop materials for a variety of future studies. We will let you know the purpose of each task before you complete it, and the computer will provide the instructions for each task.

The experimenter then initiated the program and recorded each participant's gender and ID number. The remainder of the task was administered through the PsychoPy program and required no input from the experimenter.

First, participants were instructed to spend 5 min writing about themselves as if they were either a typical soccer hooligan or a typical university professor. Participants were told that the writing task was designed to generate stimuli for a social psychology student's upcoming project. Given that the term *soccer hooligan* might not be equally familiar to participants from different cultures, participants were provided with a brief description of either soccer hooligans or professors (depending on their condition assignment). Participants in the soccer-hooligan condition read:

> Imagine that you are a typical soccer hooligan. Hooligans, as a group, tend to be young men who are fanatical sports fans, generally drink a lot in public, say offensive things to passersby, and sometimes provoke fights or destroy property.

Participants in the professor condition read:

> Imagine that you are a typical university professor. Professors, as a group, tend to have completed a doctorate degree, work in colleges or universities, dedicate their time to teaching and research, and try to publish their research in academic journals.

Following the writing task, participants were told that the first task was concluded and that a second task was for a cognitive psychology student who was developing a general-knowledge scale. The experimental script further explained that the student required a pilot sample to test the differences in the difficulty of trivia items in order to develop five subscales of varying difficulty. All participants were told that they had been assigned to the most difficult set of trivia questions and then answered the 30 general-knowledge questions. The questions were presented in a fixed order, but the PsychoPy script randomized the order of the response options for each participant.

After completing the priming and trivia tasks, participants entered their age, gender,[3] native language, major, and year of study in college. Finally, they completed the funnel-debriefing questionnaire. The funnel-debriefing items were as follows:

- "In your opinion, what was the purpose of these tasks? If you have no idea, you may answer by typing 'no idea.'"
- "Do you believe that there could be a link between thinking about a [soccer hooligan/university professor] and the general-knowledge questions?" (yes or no)
  - If the answer was "yes": "What kind of link? If you have no idea, you may answer with 'no idea.'"
- Do you believe that thinking about a [university professor/soccer hooligan] affected your performance on the general-knowledge questions?" (yes or no)
  - If the answer was "yes": "How do you think that thinking about a [university professor/soccer hooligan] affected your performance on the general-knowledge questions? If you have no idea, you can answer 'no idea.'"
- "Do you have any further thoughts or comments about the tasks so far?"

At the end of the funnel debriefing, participants were asked if they had prior familiarity with the term *soccer hooligan*.

The predetermined exclusion criteria excluded participants who were aware of the other condition, but

not those who guessed the intent of the study. Participants who spontaneously mentioned a comparison condition at any point during the funnel debriefing or the optional in-person debriefing at the end of the session were flagged by the labs as being aware of the other condition.

At the end of all the tasks, the experimenter instructed the participants not to talk about the study to anyone who had yet to participate and compensated them for their time.

Stopping rules and exclusions

Each lab preregistered its rule for ending data collection, and the Editor approved those plans. The rules were designed to ensure that each lab would meet the minimum data-collection requirements for the protocol and that the decision to end data collection would not be influenced by the results obtained.

Participants' data were excluded from analyses for any of the following reasons: They were not college or university students, they were not in the required age range (18–24 years old), they failed to record their age, they did not follow instructions, they did not complete the priming and trivia tasks, they reported being aware of the other condition in the study, or the experimenter did not administer the instructions or tasks correctly. Excluded data are provided on each lab's OSF project page, and additional details are reported in the appendix.

Results

The original call for labs to participate in the study was published on August 10, 2016, on the Web site of the Association for Psychological Science and was advertised via social media. The original deadline to submit an application to participate was September 9, 2016; however, because of the extremely high level of interest in participating, the application deadline was moved up to August 28, 2016. In sum, 47 labs (including the lead lab) applied to participate. Three labs could not collect enough data, and 4 dropped out prior to data collection, so in the end, 40 labs contributed data for the project. The participating labs represent five continents and 19 countries. The breakdown was as follows: 17 labs in North America (Canada and United States), 17 labs in Europe (Belgium, France, Germany, Hungary, The Netherlands, Poland, Turkey, Slovakia, Spain, Sweden, Switzerland, and the United Kingdom), 3 labs in Oceania (Australia and New Zealand), 2 labs in Asia (United Arab Emirates and Singapore), and 1 lab in South America (Colombia).

Given that many psychology participant pools have many more women than men, a number of labs experienced difficulty recruiting enough male participants during the initial data-collection period. This problem was exacerbated somewhat by incidents of the script crashing. Although 40 labs submitted data for the project, 17 labs were unable to meet the preregistered inclusion criterion of providing data from a minimum of 25 men and 25 women in each condition. The preregistered analyses in this RRR contain data only from the 23 labs that met all inclusion criteria. However, as the 17 labs that did not meet all the criteria collected data from a large number of participants, the proposing authors and Editor made a data-blind decision to include these labs in a set of supplementary analyses that were otherwise identical to the primary analyses. The full results of these additional analyses are available through the OSF project page.

The goal of an RRR is to provide a precise estimate of the size of an effect by combining the results of multiple, independently conducted direct replications. The results of all the replications are included regardless of their outcome so that the meta-analysis of the effect will be unbiased. The analysis does not focus on null-hypothesis significance testing. Therefore, we report the meta-analytic effect size for each outcome measure, along with the confidence interval around that effect size.

Coding and analysis scripts

Each individual laboratory was provided with an R script for analyzing their data in a way that was consistent with the preregistered protocol. The output of the script reported the overall difference in trivia performance between participants who were assigned to the professor-priming condition and those who were assigned to the hooligan-priming condition (regardless of participants' gender). The script also provided an estimate of the moderation of that effect by gender, by analyzing the difference in trivia performance between the professor- and hooligan-priming conditions separately for men and for women. The individual labs were able to independently calculate means and standard deviations for trivia performance for each of the four cells of the study. Katherine Wood wrote the R scripts using simulated data, before any actual data were collected. These scripts required minor modifications after data collection to address differences in the order of output from translated scripts. These modifications did not affect the analysis functions, and the script used for each lab's analysis is available on that lab's OSF page.

A separate R script, also written before data collection, was used to conduct the meta-analysis across labs. It directly imported the raw data from all the labs and computed descriptive statistics using analysis functions similar to those used for the individual labs. This script

required minor modifications to handle data importing because of variations that were introduced during translation, as well as variations in how PsychoPy outputs .csv files from different computer platforms. The meta-analysis script included analyses of the overall effect of priming condition on trivia performance and of the moderation of that effect by gender. For each meta-analysis, we provide a forest plot showing the difference between the professor- and hooligan-priming conditions for each laboratory and the meta-analytic result across laboratories (note that the meta-analyses did not include Dijksterhuis and van Knippenberg's, 1998, original result). Tables with the summary statistics (e.g., trivia performance by priming condition, gender) for each laboratory in each forest plot are provided on the OSF project page.

Because of unforeseen inconsistencies in the operation of PsychoPy across languages and computer systems (especially in the case of text entry), some labs experienced a large number of computer crashes during testing. In many cases, those crashes occurred after participants had completed the priming and trivia tasks. During the testing process, the experiment script was updated to address some of these issues (without changing the procedures). These updates also saved a text-file backup of each participant's data as he or she moved through the program so that data from a participant could be included if a crash occurred after the primary tasks were over. Wood wrote a recovery script that converted those backup text files to the standard .csv format for data-analysis purposes. This recovery script also required minor modifications for labs testing in languages other than English. In a small number of cases, the .csv output files included additional characters that prevented the analysis scripts from running properly. In those cases, labs provided the problematic files to Wood, and she corrected the improper formatting of those individual files. Labs retained the original and corrected files, and both versions are available.

Primary analyses

In Experiment 4 of Dijksterhuis and van Knippenberg (1998), participants who were primed with intelligence scored 13% higher on the general-knowledge trivia task than did those who were primed with stupidity (i.e., they answered 2.6 more of the 20 questions correctly). The 23 labs that met all of our inclusion criteria collected data from a total of 5,146 participants. Data from 653 participants were excluded on the basis of our preregistered exclusion criteria; this left a total sample of 4,493 in our preregistered analyses. Our meta-analysis showed that, on average, participants in the professor-priming condition, compared with those in the hooligan-priming condition, answered 0.042 more of the 30 questions correctly, a difference of 0.14% (95% confidence interval, CI = [−0.71%, 1.00%]) in the expected direction (see Fig. 1). The difference in percentage correct between the professor- and hooligan-priming conditions ranged from −4.99% to 4.24% across the labs. The variability in the effect size (i.e., heterogeneity) was not significantly different from what would be expected by chance, $\tau = 0.86$, $I^2 = 17.43\%$, $H^2 = 1.21$, $Q(22) = 28.09$, $p = .17$.

Although Dijksterhuis and van Knippenberg (1998) initially found evidence for overall effects of priming condition on trivia performance, the replications conducted by Dijksterhuis, van Knippenberg, and Holland produced a smaller overall effect of priming condition (a 2%–3% difference) and showed moderation of this effect by gender: Men showed a difference (9.3% and 7.6%), but women did not (0.3% and −0.3%). Figure 2 shows that the effect of condition on trivia performance was not substantially moderated by gender in our replication study. Men showed a 0.01% difference (95% CI = [−1.38%, 1.41%]) in trivia performance between conditions, and women showed a 0.02% difference ([95% CI = −0.92%, 0.96%]).

Ancillary analyses

Analysis of all 40 participating laboratories. We repeated the main analysis including the full set of 40 laboratories that submitted data for the replication project. In this expanded set of labs, participants in the professor-priming condition, compared with those in the hooligan-priming condition, answered an average of 0.006 fewer questions correctly; this −0.02% difference (95% CI = [−0.77%, 0.73%]) was in the opposite direction of what we expected. Figure 3 summarizes the results of this analysis of all 40 labs' data. Unlike the analysis with 23 labs, this analysis did show statistically significant heterogeneity, $\tau = 1.20$, $I^2 = 26.19\%$, $H^2 = 1.35$, $Q(39) = 55.47$, $p = .04$. The analysis of all 40 labs' data also showed little difference in priming between men (−0.06%) and women (−0.20%; see Fig. 4).

Alternative operationalization of accuracy. In an exploratory analysis of data from the 23 labs that met all inclusion criteria, we treated skipped trivia answers as missing rather than incorrect (the forest plot for this analysis is available on the OSF project page). This alternative coding did not yield any meaningful difference in the output, as the meta-analytic effect size remained small, 0.13% (95% CI = [−0.74%, 0.99%]).

Restricting analysis to participants who did not think tasks were linked. Other exploratory analyses of the data from the 23 labs that met all inclusion criteria

Lab	Professor Condition Accuracy (%)	*n*	Hooligan Condition Accuracy (%)	*n*	Effect Size (%)	95% CI
Original Result	55.60		42.50		13.20	
Original Result (Replication 1)	51.40	77	48.90	101	2.50	
Original Result (Replication 2)	45.70	81	43.60	73	2.10	
Schulte-Mecklenbeck	58.47	50	54.22	60	4.24	[–0.12, 8.61]
Baskin	49.60	66	45.52	64	4.08	[–0.21, 8.36]
Braithwaite	61.01	66	57.00	70	4.01	[–0.28, 8.30]
Finnigan	52.68	137	49.79	140	2.89	[–0.39, 6.17]
Newell	54.48	64	52.82	58	1.66	[–3.05, 6.38]
O'Donnell	49.81	106	48.23	94	1.58	[–3.32, 6.49]
Tamayo	54.87	106	53.46	102	1.41	[–2.01, 4.83]
Karpinski	52.38	77	51.00	80	1.38	[–3.07, 5.83]
Klein	52.49	126	51.61	116	0.88	[–2.05, 3.81]
Keller	51.20	230	51.03	201	0.17	[–2.44, 2.79]
Shanks	53.29	72	53.17	63	0.11	[–4.90, 5.12]
Białobrzeska	52.71	64	52.82	72	–0.12	[–4.10, 3.87]
Koppel	58.04	63	58.25	76	–0.20	[–4.27, 3.87]
Philipp	52.38	70	52.81	70	–0.43	[–5.18, 4.33]
Ropovik	52.47	104	52.95	95	–0.48	[–3.62, 2.66]
Steele	50.94	267	51.52	277	–0.58	[–2.59, 1.43]
Susa	42.04	93	42.68	76	–0.63	[–4.68, 3.41]
Steffens	47.84	68	48.68	76	–0.84	[–5.63, 3.95]
Aczel	55.59	127	56.94	133	–1.35	[–4.30, 1.60]
Saunders	43.93	67	45.80	85	–1.87	[–6.31, 2.56]
McLatchie	53.21	53	55.37	54	–2.16	[–7.11, 2.78]
Aveyard	41.82	112	45.81	93	–3.99	[–7.42, –0.56]
Boot	47.92	64	52.91	86	–4.99	[–9.67, –0.31]
Meta-Analytic Average for Main Effect	51.70	2,252	51.49	2,241	0.14	[–0.71, 1.00]

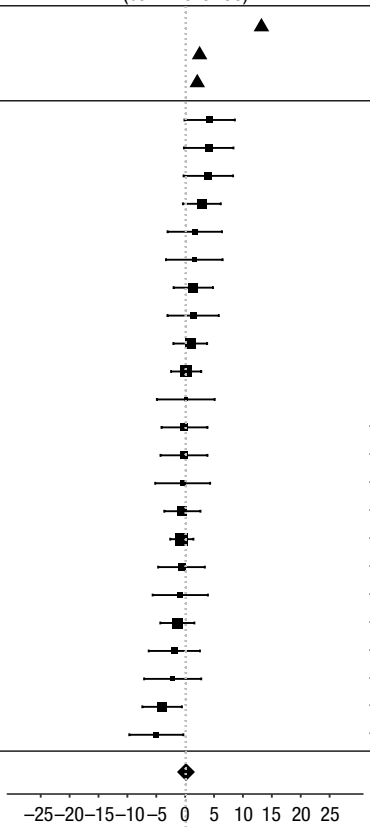


Fig 1. Results of the primary analyses: difference in trivia performance between the professor-priming and hooligan-priming conditions. For each of the 23 labs that met all the inclusion criteria, the figure shows the mean percentage correct and the sample size in each condition. The labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plot, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plot, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure are corresponding results for Dijksterhuis and van Knippenberg's (1998) original study and Dijksterhuis, van Knippenberg, and Holland's unpublished replications. The bottom row in the figure presents overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

excluded participants who, during debriefing, expressed a belief that the priming task and trivia task were related (these participants were not excluded from the primary analyses because they did not report awareness of another condition). Nearly 1 in 5 (19.9%) participants responded "yes" when asked whether they believed that thinking about a university professor or soccer hooligan affected their performance on the trivia task. The analysis excluding these participants revealed a small difference in accuracy in the expected direction, 0.17% (95% CI = [–0.68%, 1.01%]; Fig. 5). Additionally, 62.7% of participants responded "yes" when asked whether there could be a link between thinking about a university professor or soccer hooligan and the trivia task. The analysis excluding these participants revealed a difference in the expected direction; participants in the professor-priming condition performed 2.07% better on the trivia task than those in the hooligan-priming condition (95% CI = [0.57%, 3.57%]; Fig. 5). Excluding participants who responded "yes" to either or both of these questions removed 65.9% of the total sample and yielded a meta-analytic effect of 2.32% (95% CI = [0.79%, 3.86%]; Fig. 5).

a

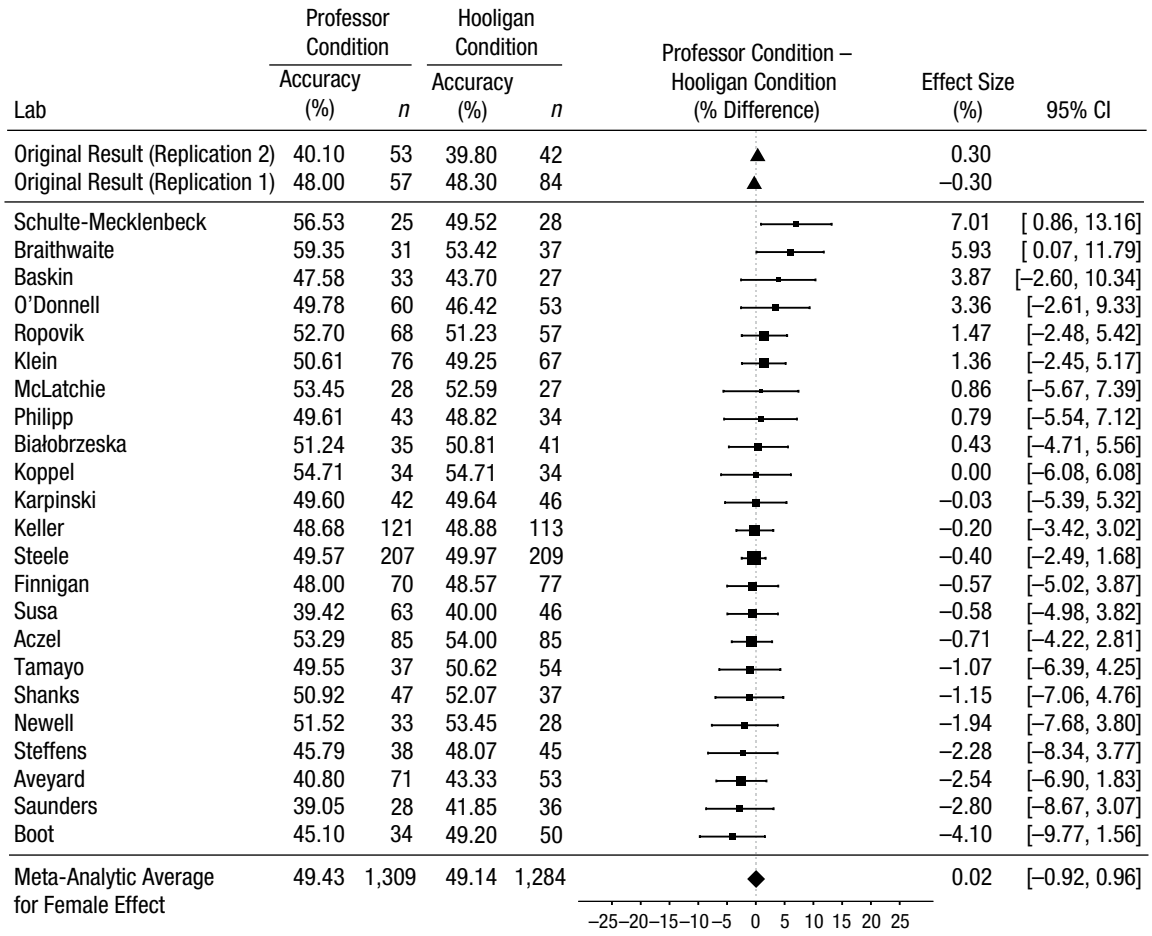


Fig. 2. Results of the primary analyses: difference in trivia performance between the professor-priming and hooligan-priming conditions, separately for (a) female participants and (b) male participants. For each of the 23 labs that met all the inclusion criteria, the figure shows the mean percentage correct and the sample size in each condition. In each panel, the labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition);

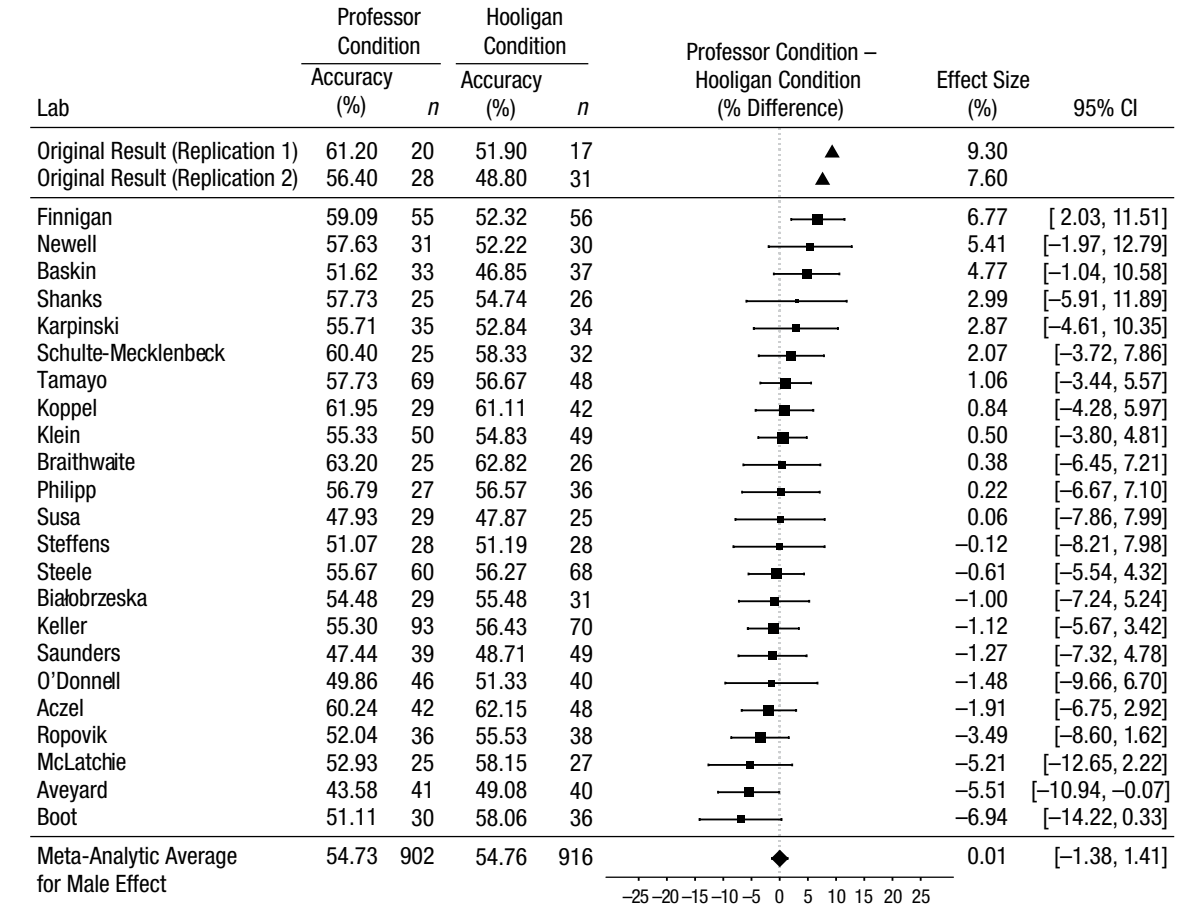
(continued)

Given that this effect was roughly consistent with the size of the overall effect in Dijksterhuis et al.'s two follow-up replication studies, and given that those studies showed moderation by gender, we conducted another exploratory analysis to examine whether gender moderated the effect we found when participants who thought the tasks were linked were removed from analysis. Contrary to the predicted pattern, this analysis revealed a smaller effect for men (1.76%, 95% CI = [−1.16%, 4.68%]) than for women (2.70%, 95% CI = [1.05%, 4.35%]). We also examined whether the effect would remain when we analyzed the data from the larger sample of 40 labs (for which we had observed some heterogeneity) and found that it was reduced to 1.24% and that the confidence interval included zero (95% CI = [−0.21%, 2.69%]).

Influence of the country where testing was conducted. We also examined whether the effect varied with the country of the participants, given that people living in different countries (N = 13) might have different familiarity with the concept of hooligans. There did not appear to be any significant variation in the professor-priming effect across countries (Fig. 6). The 95% CI for each country except the United Arab Emirates included 0, and effect sizes for the individual countries ranged from −3.99% (United Arab Emirates, 95% CI = [−7.42%, −0.56%]) to 4.24% (Switzerland, 95% CI = [−0.12%, 8.61%]).

Prior familiarity with the term *hooligan*. Finally, we looked at whether the effect varied according to whether or not participants reported having had awareness of the term *hooligan* prior to the study. Among participants

b



Lab	Professor Condition Accuracy (%)	*n*	Hooligan Condition Accuracy (%)	*n*	Effect Size (%)	95% CI
Original Result (Replication 1)	61.20	20	51.90	17	9.30	
Original Result (Replication 2)	56.40	28	48.80	31	7.60	
Finnigan	59.09	55	52.32	56	6.77	[2.03, 11.51]
Newell	57.63	31	52.22	30	5.41	[–1.97, 12.79]
Baskin	51.62	33	46.85	37	4.77	[–1.04, 10.58]
Shanks	57.73	25	54.74	26	2.99	[–5.91, 11.89]
Karpinski	55.71	35	52.84	34	2.87	[–4.61, 10.35]
Schulte-Mecklenbeck	60.40	25	58.33	32	2.07	[–3.72, 7.86]
Tamayo	57.73	69	56.67	48	1.06	[–3.44, 5.57]
Koppel	61.95	29	61.11	42	0.84	[–4.28, 5.97]
Klein	55.33	50	54.83	49	0.50	[–3.80, 4.81]
Braithwaite	63.20	25	62.82	26	0.38	[–6.45, 7.21]
Philipp	56.79	27	56.57	36	0.22	[–6.67, 7.10]
Susa	47.93	29	47.87	25	0.06	[–7.86, 7.99]
Steffens	51.07	28	51.19	28	–0.12	[–8.21, 7.98]
Steele	55.67	60	56.27	68	–0.61	[–5.54, 4.32]
Białobrzeska	54.48	29	55.48	31	–1.00	[–7.24, 5.24]
Keller	55.30	93	56.43	70	–1.12	[–5.67, 3.42]
Saunders	47.44	39	48.71	49	–1.27	[–7.32, 4.78]
O'Donnell	49.86	46	51.33	40	–1.48	[–9.66, 6.70]
Aczel	60.24	42	62.15	48	–1.91	[–6.75, 2.92]
Ropovik	52.04	36	55.53	38	–3.49	[–8.60, 1.62]
McLatchie	52.93	25	58.15	27	–5.21	[–12.65, 2.22]
Aveyard	43.58	41	49.08	40	–5.51	[–10.94, –0.07]
Boot	51.11	30	58.06	36	–6.94	[–14.22, 0.33]
Meta-Analytic Average for Male Effect	54.73	902	54.76	916	0.01	[–1.38, 1.41]

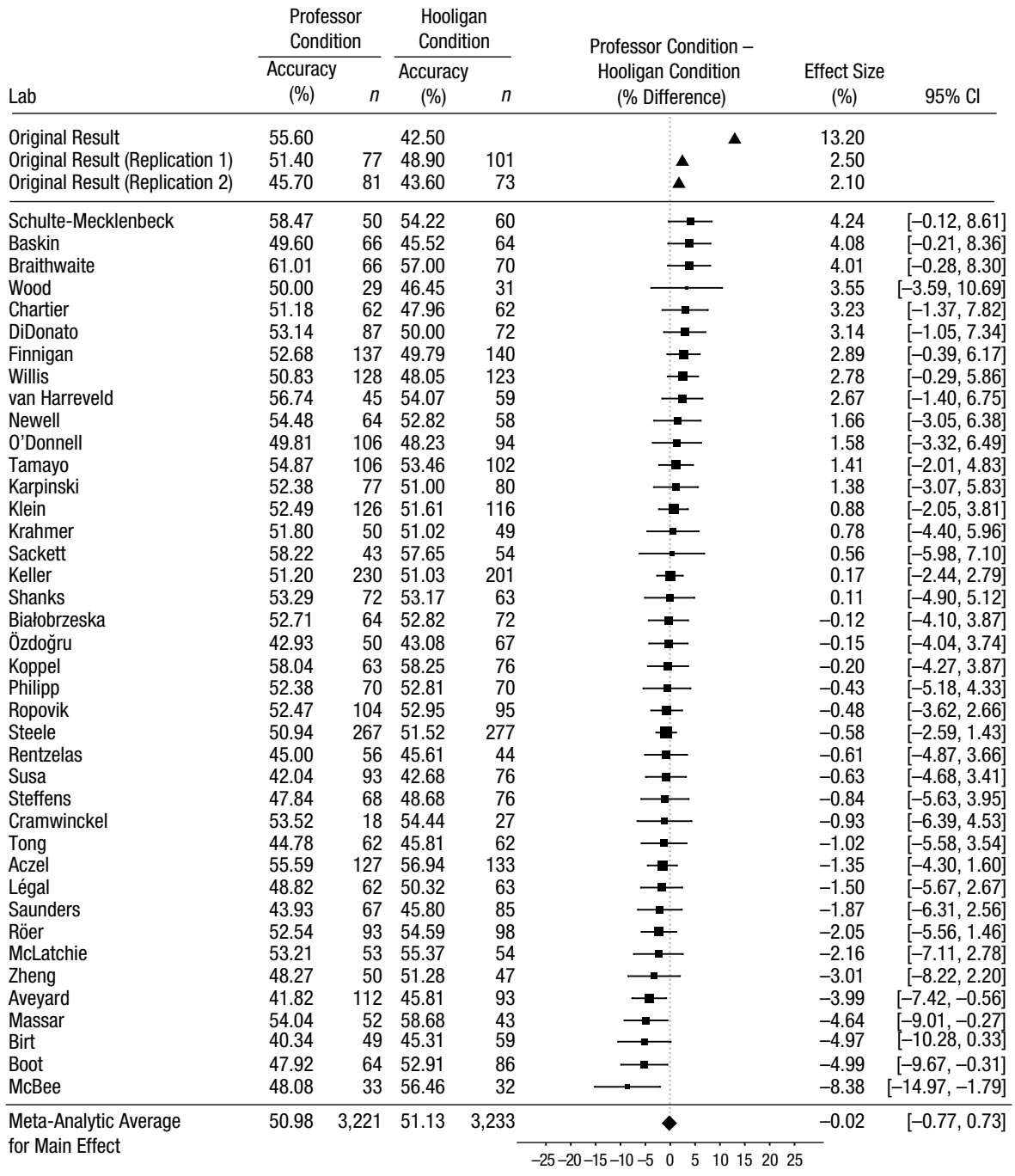
Fig. 2. *(continued)* positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plots, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plots, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of each panel are corresponding results for Dijksterhuis, van Knippenberg, and Holland's unpublished replications. The bottom row in each panel presents overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

who reported no prior exposure to the term, there was a small difference in trivia performance between the two conditions; this difference of –0.84% (95% CI = [–2.60%, 0.93%]) was in the opposite direction of we expected. In contrast, those participants who did report prior exposure to the term showed a small difference in trivia performance, 0.62% (95% CI = [–0.38%, 1.63%]), in the expected direction (Fig. 7).

General Discussion

Overall, the meta-analytic results of this multilab replication provided little empirical support for a difference in trivia performance following a writing task designed to prime high or low intelligence. We collected data from 4,493 participants across 23 labs; collectively and individually, these experiments did not find the difference in trivia performance originally observed in Dijksterhuis and van Knippenberg's (1998) Experiment 4, and they did not find the gender difference observed in the two unpublished follow-up studies that were used as the basis for the RRR protocol. In the RRR study, both the overall effect and the effect for each gender were close to zero.

It is possible that the results from this replication study differed from the original findings because of the ubiquity of the professor-priming effect in modern psychology courses. Nearly two thirds of the participants across the 23 labs expressed a belief that the writing task and the trivia task were related to each



Lab	Professor Condition Accuracy (%)	*n*	Hooligan Condition Accuracy (%)	*n*	Effect Size (%)	95% CI
Original Result	55.60		42.50		13.20	
Original Result (Replication 1)	51.40	77	48.90	101	2.50	
Original Result (Replication 2)	45.70	81	43.60	73	2.10	
Schulte-Mecklenbeck	58.47	50	54.22	60	4.24	[–0.12, 8.61]
Baskin	49.60	66	45.52	64	4.08	[–0.21, 8.36]
Braithwaite	61.01	66	57.00	70	4.01	[–0.28, 8.30]
Wood	50.00	29	46.45	31	3.55	[–3.59, 10.69]
Chartier	51.18	62	47.96	62	3.23	[–1.37, 7.82]
DiDonato	53.14	87	50.00	72	3.14	[–1.05, 7.34]
Finnigan	52.68	137	49.79	140	2.89	[–0.39, 6.17]
Willis	50.83	128	48.05	123	2.78	[–0.29, 5.86]
van Harreveld	56.74	45	54.07	59	2.67	[–1.40, 6.75]
Newell	54.48	64	52.82	58	1.66	[–3.05, 6.38]
O'Donnell	49.81	106	48.23	94	1.58	[–3.32, 6.49]
Tamayo	54.87	106	53.46	102	1.41	[–2.01, 4.83]
Karpinski	52.38	77	51.00	80	1.38	[–3.07, 5.83]
Klein	52.49	126	51.61	116	0.88	[–2.05, 3.81]
Krahmer	51.80	50	51.02	49	0.78	[–4.40, 5.96]
Sackett	58.22	43	57.65	54	0.56	[–5.98, 7.10]
Keller	51.20	230	51.03	201	0.17	[–2.44, 2.79]
Shanks	53.29	72	53.17	63	0.11	[–4.90, 5.12]
Białobrzeska	52.71	64	52.82	72	–0.12	[–4.10, 3.87]
Özdoğru	42.93	50	43.08	67	–0.15	[–4.04, 3.74]
Koppel	58.04	63	58.25	76	–0.20	[–4.27, 3.87]
Philipp	52.38	70	52.81	70	–0.43	[–5.18, 4.33]
Ropovik	52.47	104	52.95	95	–0.48	[–3.62, 2.66]
Steele	50.94	267	51.52	277	–0.58	[–2.59, 1.43]
Rentzelas	45.00	56	45.61	44	–0.61	[–4.87, 3.66]
Susa	42.04	93	42.68	76	–0.63	[–4.68, 3.41]
Steffens	47.84	68	48.68	76	–0.84	[–5.63, 3.95]
Cramwinckel	53.52	18	54.44	27	–0.93	[–6.39, 4.53]
Tong	44.78	62	45.81	62	–1.02	[–5.58, 3.54]
Aczel	55.59	127	56.94	133	–1.35	[–4.30, 1.60]
Légal	48.82	62	50.32	63	–1.50	[–5.67, 2.67]
Saunders	43.93	67	45.80	85	–1.87	[–6.31, 2.56]
Röer	52.54	93	54.59	98	–2.05	[–5.56, 1.46]
McLatchie	53.21	53	55.37	54	–2.16	[–7.11, 2.78]
Zheng	48.27	50	51.28	47	–3.01	[–8.22, 2.20]
Aveyard	41.82	112	45.81	93	–3.99	[–7.42, –0.56]
Massar	54.04	52	58.68	43	–4.64	[–9.01, –0.27]
Birt	40.34	49	45.31	59	–4.97	[–10.28, 0.33]
Boot	47.92	64	52.91	86	–4.99	[–9.67, –0.31]
McBee	48.08	33	56.46	32	–8.38	[–14.97, –1.79]
Meta-Analytic Average for Main Effect	50.98	3,221	51.13	3,233	–0.02	[–0.77, 0.73]

Fig. 3. Difference in trivia performance between the professor-priming and hooligan-priming conditions in the analysis including all 40 laboratories. For each lab, the figure shows the mean percentage correct and the sample size in each condition. The labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plot, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plot, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of the figure are corresponding results for Dijksterhuis and van Knippenberg's (1998) original study and Dijksterhuis, van Knippenberg, and Holland's unpublished replications. The bottom row in the figure presents overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

other, which suggests that there potentially was a high level of suspicion about the procedure. And when the analysis was restricted to the 34.1% who believed either that the tasks were not related or that the writing task did not affect their trivia performance, or both, there was a tendency for participants in the professor-priming condition to perform better than participants in the hooligan-priming condition (52.01% vs. 49.62%). However, even in this restricted sample, the meta-analytic effect size was substantially smaller than that reported in the original article. The effect with this more restricted sample was more similar to the overall 2% to 3% effect found in the unpublished follow-up studies that served as the basis for the RRR protocol; however, the effect in the restricted sample was substantially smaller when we analyzed the data provided by the full set of 40 labs.

Although earlier unsuccessful attempts to replicate the professor-priming effect (e.g., Shanks et al., 2013) differed from the original study in ways that Dijksterhuis et al. (2014) suggested could moderate the effect (e.g., in the original study, participants were tested individually, but group testing was used in some replications), we found little evidence that the observed effect varied with testing setting (group testing, individual testing, or a mix of the two; results reported at OSF), and results from all the settings produced similar meta-analytic results (effects close to zero).

In sum, this registered replication study found no overall effect of intelligence priming on trivia performance. The meta-analytic effect was small, and the confidence interval for the effect contained zero. Only 2 of the 23 labs that met all of the preregistered inclusion criteria found an effect with a confidence interval that did not include zero, and both of these labs found an effect in the direction opposite the anticipated direction (see Fig. 1). We also found no evidence for moderation of the effect by gender, country where testing was conducted, whether testing was conducted individually or in small groups (see OSF), or whether participants had prior familiarity with the term *hooligan*. Participants who did not express a belief that the tasks were linked showed a small effect consistent with the original, but these participants constituted a small minority of the total sample in this study, and this effect was reduced in the full sample of 40 labs. The results are somewhat surprising, as a *p*-curve analysis showed some evidential value for professor priming in the published literature (Lakens, 2017).

In considering the constraints and limitations of this replication study, we first acknowledge that the original study was conducted in the 1990s, in The Netherlands, and the social cultures of professors, hooligans, and experimental participants have changed since then. Although the protocol was designed as a test of the original hypothesis, our ability to detect the effect might have changed over time as a result of these cultural changes (e.g., hooliganism might have become less familiar as a construct, and differences in the sampled populations could also have affected our ability to observe an effect).

Although the protocol ensured that experimenters were blind to condition assignment, some participants could have intuited that the first task was meant to affect performance on the second. For example, they might have guessed that the experimenter expected poor trivia performance after they wrote about being a hooligan, and therefore not tried hard on the trivia test (demand characteristics). The analysis plan did not exclude participants who suspected a link between the tasks, so demand characteristics could have contributed to performance differences between the conditions (although we did not find differences in the primary analysis). The exploratory analysis excluding those participants who reported suspecting a link between the tasks[4] revealed a pattern more similar in magnitude to the effect in Dijksterhuis et al.'s unpublished replications. Although the effect was smaller than in the original 1998 experiment and not substantially different from zero, this self-identified naive population might have been more sensitive to the hypothesized priming effect. Our data were insufficient to test that possibility robustly, but future investigations with even larger samples could.

The professor and hooligan primes were chosen as the best possible options to reproduce the original effect, but the meaning of *professor* and *hooligan* might vary across cultures. Similarly, the trivia items were screened and normed in an online sample and at a large American public university, and we selected items with roughly similar accuracy levels (including in the subset of online participants from India). Although the absolute performance levels for individual trivia items might vary across cultures because of differences in familiarity with the topics (e.g., a question about Joan of Arc might be easier for participants in France than for those in Colombia), all the trivia items were included in both priming conditions, so such differences in absolute performance should have had relatively little impact on the effects of interest. In general, the absence of significant heterogeneity across labs is inconsistent with the possibility that differences in the materials that could have arisen during translation contributed to the size of the priming effect.

a

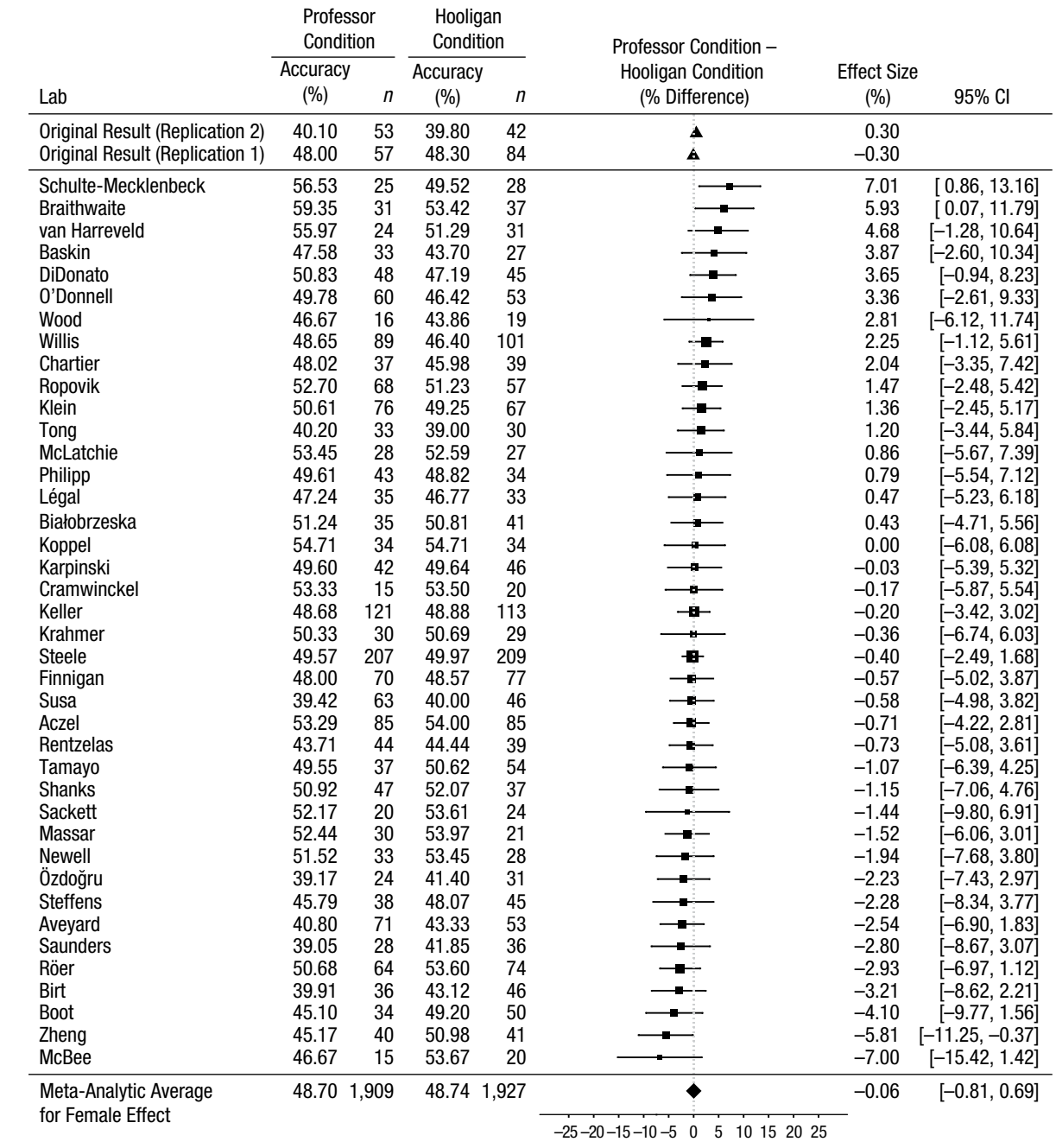


Fig. 4. Difference in trivia performance between the professor-priming and hooligan-priming conditions, separately for (a) female participants and (b) male participants, in the analysis including all 40 laboratories. Each panel shows the mean percentage correct and the sample size in each condition for each lab. The labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plots, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, and the size of each square represents the magnitude of the standard error for the lab's effect (larger squares indicate less

(continued)

b

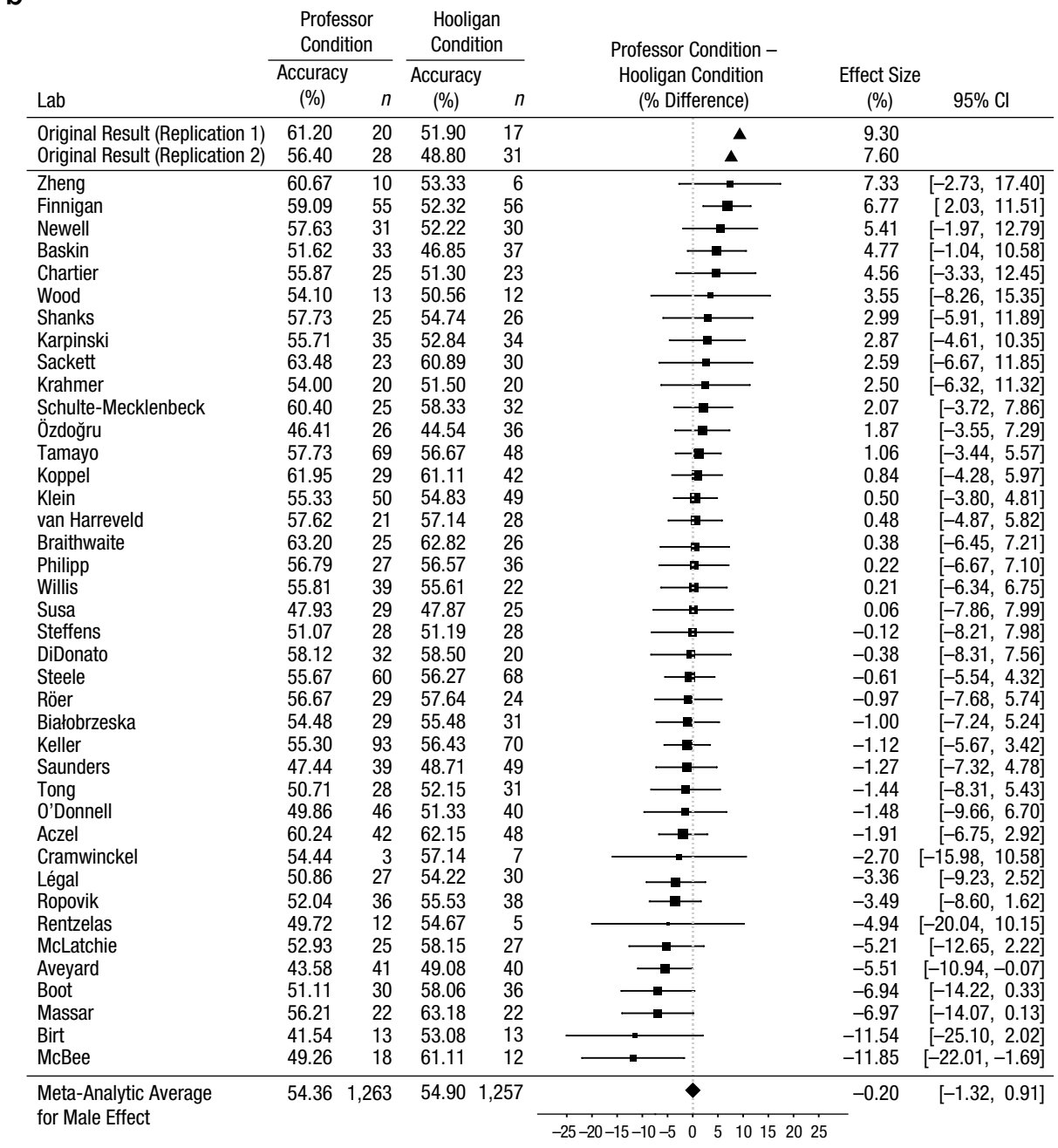


Fig. 4. *(continued)* variability in the estimate). To the right of the forest plots, the figure shows the numerical values for the effect sizes and 95% CIs. At the top of each panel are corresponding results for Dijksterhuis, van Knippenberg, and Holland's unpublished replications. The bottom row in each panel presents overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

a

Lab	Professor Condition Accuracy (%)	*n*	Hooligan Condition Accuracy (%)	*n*	Professor Condition – Hooligan Condition (% Difference)	Effect Size (%)	95% CI
Finnigan	53.64	109	49.56	113		4.08	[0.65, 7.52]
Braithwaite	61.39	48	57.52	55		3.87	[–1.04, 8.78]
Schulte-Mecklenbeck	57.46	38	53.62	47		3.84	[–1.19, 8.87]
Newell	55.35	48	52.68	51		2.67	[–2.49, 7.82]
Baskin	49.01	57	46.55	55		2.46	[–2.09, 7.01]
Tamayo	54.75	87	53.13	83		1.62	[–2.16, 5.40]
O'Donnell	53.82	83	52.67	75		1.15	[–3.48, 5.78]
Klein	52.86	105	51.86	102		0.99	[–2.15, 4.14]
Ropovik	52.98	75	52.82	71		0.16	[–3.60, 3.93]
Karpinski	51.90	63	51.83	71		0.07	[–4.70, 4.84]
Keller	52.10	173	52.04	168		0.06	[–2.83, 2.95]
Aczel	56.63	95	56.70	106		–0.07	[–3.31, 3.17]
Koppel	58.65	52	59.15	67		–0.50	[–5.03, 4.03]
Shanks	52.39	53	52.94	51		–0.55	[–6.60, 5.49]
Steffens	49.15	51	49.72	60		–0.57	[–6.12, 4.98]
Philipp	52.88	59	53.51	57		–0.63	[–5.77, 4.52]
Susa	42.04	75	42.83	60		–0.79	[–5.39, 3.81]
Steele	51.29	224	52.14	226		–0.84	[–3.05, 1.36]
Saunders	44.72	53	46.20	64		–1.48	[–6.54, 3.58]
Aveyard	43.23	95	45.71	87		–2.48	[–6.14, 1.18]
Białobrzeska	50.28	36	52.93	33		–2.65	[–8.20, 2.90]
McLatchie	53.26	44	56.38	47		–3.13	[–8.78, 2.53]
Boot	48.15	56	53.06	72		–4.90	[–9.79, –0.01]
Meta-Analytic Average After Exclusion	52.08	1,779	51.98	1,821		0.17	[–0.68, 101]

b

Lab	Professor Condition Accuracy (%)	*n*	Hooligan Condition Accuracy (%)	*n*	Professor Condition – Hooligan Condition (% Difference)	Effect Size (%)	95% CI
Finnigan	55.75	40	45.46	58		10.29	[5.16, 15.42]
O'Donnell	51.96	34	44.02	29		7.94	[0.02, 15.85]
Schulte-Mecklenbeck	58.55	23	52.38	35		6.17	[–0.37, 12.71]
Newell	57.25	17	52.19	32		5.07	[–2.86, 13.00]
Tamayo	55.36	23	51.26	37		4.10	[–2.41, 10.62]
Baskin	47.26	28	43.33	27		3.93	[–1.60, 9.46]
Steffens	49.17	20	45.40	29		3.76	[–4.03, 11.56]
Keller	52.45	68	49.12	87		3.33	[–1.04, 7.70]
Braithwaite	60.22	15	57.39	23		2.83	[–6.12, 11.78]
Klein	51.22	52	48.47	50		2.75	[–1.82, 7.33]
Aczel	57.81	32	55.44	38		2.37	[–2.96, 7.70]
Shanks	49.86	23	47.67	20		2.19	[–7.85, 12.23]
Białobrzeska	50.19	18	48.25	19		1.94	[–5.48, 9.36]
McLatchie	51.56	15	50.14	24		1.42	[–6.92, 9.75]
Philipp	54.63	18	53.21	28		1.42	[–7.68, 10.51]
Koppel	58.89	15	57.86	28		1.03	[–6.80, 8.86]
Ropovik	52.27	44	51.35	52		0.93	[–3.48, 5.34]
Karpinski	51.03	26	50.61	44		0.42	[–5.78, 6.61]
Steele	50.08	79	50.13	133		–0.04	[–3.45, 3.37]
Susa	40.75	31	41.57	34		–0.82	[–7.09, 5.46]
Saunders	41.48	18	43.41	42		–1.93	[–7.90, 4.04]
Aveyard	42.55	55	45.19	52		–2.65	[–7.12, 1.83]
Boot	46.11	24	51.14	38		–5.03	[–11.95, 1.89]
Meta-Analytic Average After Exclusion	51.58	718	49.35	959		2.07	[0.57, 3.57]

(continued)

C

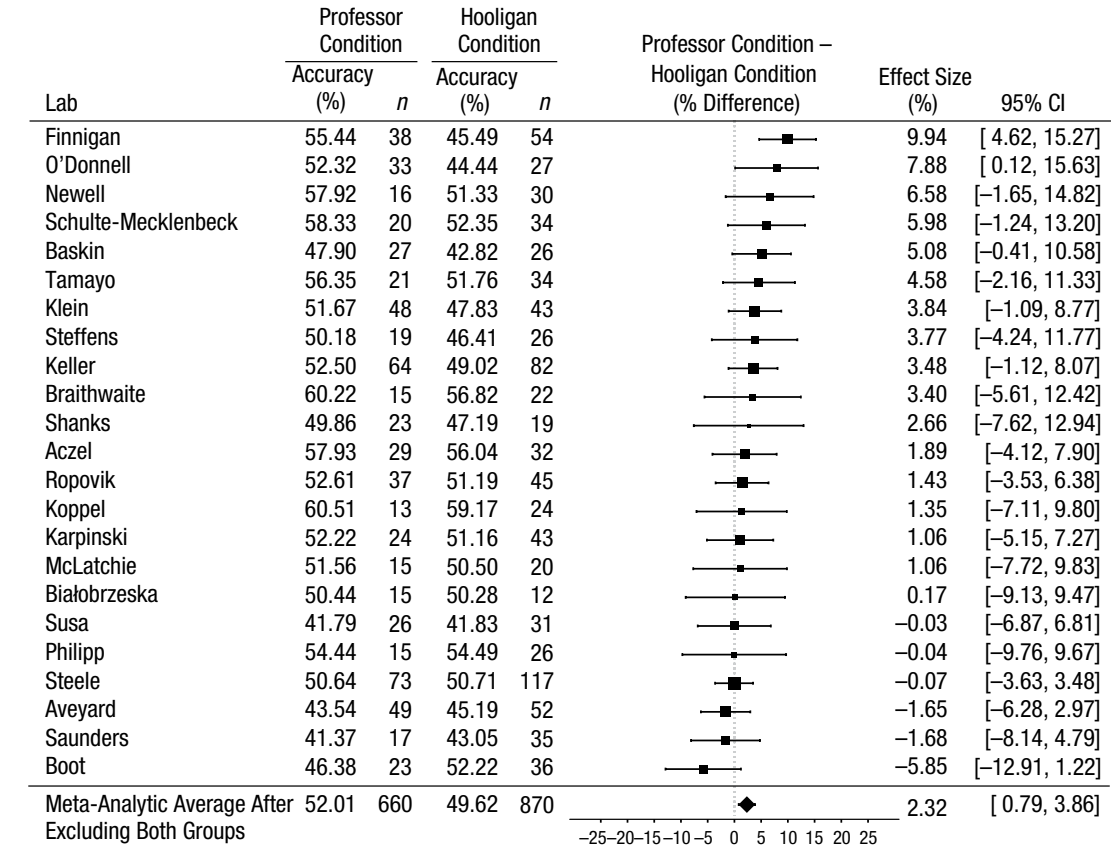


Fig. 5. Difference in trivia performance between the professor-priming and hooligan-priming conditions in analyses excluding participants who (a) thought the writing task could influence their performance, (b) thought the tasks were linked, or (c) responded "yes" to either or both of these awareness-check items. For each of the 23 labs that met all the inclusion criteria, the figure shows the mean percentage correct and the sample size in each condition. In each panel, the labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plots, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plots, the figure shows the numerical values for the effect sizes and 95% CIs. The bottom row in each panel presents overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

Lab	Professor Condition Accuracy (%)	Professor Condition *n*	Hooligan Condition Accuracy (%)	Hooligan Condition *n*	Professor Condition – Hooligan Condition (% Difference)	Effect Size (%)	95% CI
Australia							
Newell	54.48	64	52.82	58		1.66	[–3.05, 6.38]
Steffens	47.84	68	48.68	76		–0.84	[–5.63, 3.95]
Belgium							
Klein	52.49	126	51.61	116		0.88	[–2.05, 3.81]
Canada							
Saunders	43.93	67	45.80	85		–1.87	[–6.31, 2.56]
Colombia							
Tamayo	54.87	106	53.46	102		1.41	[–2.01, 4.83]
Hungary							
Aczel	55.59	127	56.94	133		–1.35	[–4.30, 1.60]
New Zealand							
Philipp	52.38	70	52.81	70		–0.43	[–5.18, 4.33]
Poland							
Białobrzeska	52.71	64	52.82	72		–0.12	[–4.10, 3.87]
Slovakia							
Ropovik	52.47	104	52.95	95		–0.48	[–3.62, 2.66]
Sweden							
Koppel	58.04	63	58.25	76		–0.20	[–4.27, 3.87]
Switzerland							
Schulte-Mecklenbeck	58.47	50	54.22	60		4.24	[–0.12, 8.61]
United Arab Emirates							
Aveyard	41.82	112	45.81	93		–3.99	[–7.42, –0.56]
United Kingdom							
Shanks	53.29	72	53.17	63		0.11	[–4.90, 5.12]
Mclatchie	53.21	53	55.37	54		–2.16	[–7.11, 2.78]
United States							
Baskin	49.60	66	45.52	64		4.08	[–0.21, 8.36]
Braithwaite	61.01	66	57.00	70		4.01	[–0.28, 8.30]
Finnigan	52.68	137	49.79	140		2.89	[–0.39, 6.17]
O'Donnell	49.81	106	48.23	94		1.58	[–3.32, 6.49]
Karpinski	52.38	77	51.00	80		1.38	[–3.07, 5.83]
Keller	51.20	230	51.03	201		0.17	[–2.44, 2.79]
Steele	50.94	267	51.52	277		–0.58	[–2.59, 1.43]
Susa	42.04	93	42.68	76		–0.63	[–4.68, 3.41]
Boot	47.92	64	52.91	86		–4.99	[–9.67, –0.31]
Meta-Analytic Average for							
Switzerland	58.47	50	54.22	60		4.24	[–0.12, 8.61]
Colombia	54.87	106	53.46	102		1.41	[–2.01, 4.83]
Belgium	52.49	126	51.61	116		0.88	[–2.05, 3.81]
United States	50.84	1,106	49.96	1,088		0.79	[–0.79, 2.37]
Australia	51.16	132	50.75	134		0.43	[–2.93, 3.79]
Poland	52.71	64	52.82	72		–0.12	[–4.10, 3.87]
Sweden	58.04	63	58.25	76		–0.20	[–4.27, 3.87]
New Zealand	52.38	70	52.81	70		–0.43	[–5.18, 4.33]
Slovakia	52.47	104	52.95	95		–0.48	[–3.62, 2.66]
United Kingdom	53.25	125	54.27	117		–1.04	[–4.56, 2.48]
Hungary	55.59	127	56.94	133		–1.35	[–4.30, 1.60]
Canada	43.93	67	45.80	85		–1.87	[–6.31, 2.56]
United Arab Emirates	41.82	112	45.81	93		–3.99	[–7.42, –0.56]

Fig. 6. Difference in trivia performance between the professor-priming and hooligan-priming conditions, by country. For each of the 23 labs that met all inclusion criteria, the figure shows the mean percentage correct and the sample size in each condition. Within each country, the labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plot, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plot, the figure shows the numerical values for the effect sizes and 95% CIs. The bottom of the figure presents overall country means, averaged across all participants in each condition without regard to lab, and the outcome of random-effects meta-analyses. Note that the meta-analytic estimates of the difference between conditions dos not necessarily equal the difference between the means.

Lab	Professor Condition Accuracy (%)	Professor Condition *n*	Hooligan Condition Accuracy (%)	Hooligan Condition *n*	Professor Condition – Hooligan Condition (% Difference)	Effect Size (%)	95% CI
No Prior Familiarity							
Braithwaite	61.62	39	54.83	40		6.79	[1.11, 12.48]
Aczel	59.00	10	53.00	10		6.00	[–2.24, 14.24]
McLatchie	57.50	8	52.12	11		5.38	[–7.13, 17.89]
Newell	53.33	25	48.64	22		4.70	[–3.60, 13.00]
Karpinski	51.42	40	47.13	43		4.28	[–1.64, 10.21]
Baskin	48.46	39	45.44	38		3.02	[–2.77, 8.81]
Klein	47.22	12	45.71	14		1.51	[–8.30, 11.31]
O'Donnell	52.41	58	51.31	51		1.11	[–4.24, 6.45]
Finnigan	49.35	67	49.02	75		0.33	[–4.36, 5.02]
Keller	50.69	131	51.00	113		–0.32	[–3.57, 2.94]
Steele	51.13	186	52.19	190		–1.06	[–3.51, 1.38]
Ropovik	47.95	13	49.72	12		–1.77	[–10.84, 7.29]
Philipp	46.08	17	48.59	26		–2.51	[–9.60, 4.58]
Saunders	42.41	36	45.28	48		–2.87	[–8.90, 3.16]
Steffens	40.69	29	43.95	38		–3.26	[–9.12, 2.61]
Susa	38.93	50	43.57	43		–4.63	[–9.74, 0.48]
Boot	46.87	33	52.33	53		–5.46	[–10.96, 0.04]
Aveyard	38.83	60	44.40	50		–5.57	[–10.43, –0.70]
Tamayo	50.48	7	57.50	12		–7.02	[–16.93, 2.88]
Shanks	46.19	21	55.50	20		–9.31	[–18.64, 0.02]
Koppel	43.33	3	65.00	2		–21.67	[–40.81, –2.53]
Prior Familiarity							
Finnigan	56.95	58	51.61	58		5.34	[0.64, 10.05]
Shanks	57.41	45	52.42	40		4.99	[–1.23, 11.21]
Susa	46.32	38	41.55	28		4.77	[–1.98, 11.52]
Schulte-Mecklenbeck	58.86	44	55.36	46		3.50	[–1.41, 8.42]
Baskin	51.06	22	47.58	22		3.48	[–3.43, 10.40]
Tamayo	55.19	99	52.93	90		2.26	[–1.39, 5.91]
Newell	56.57	33	54.90	34		1.66	[–4.18, 7.50]
Philipp	56.32	48	54.74	38		1.58	[–4.55, 7.71]
Keller	55.47	67	54.22	64		1.25	[–3.80, 6.31]
Koppel	58.99	56	58.06	74		0.93	[–3.33, 5.18]
Klein	53.18	111	52.27	100		0.92	[–2.19, 4.02]
Karpinski	55.90	26	55.06	29		0.84	[–6.94, 8.62]
Saunders	48.18	22	47.70	29		0.48	[–7.01, 7.98]
Steele	50.49	81	50.04	87		0.46	[–3.07, 3.98]
Białobrzeska	52.70	63	52.77	71		–0.07	[–4.12, 3.97]
Ropovik	53.22	90	53.50	79		–0.28	[–3.64, 3.08]
Steffens	53.78	37	55.05	35		–1.26	[–7.94, 5.41]
Braithwaite	59.80	17	61.59	23		–1.79	[–9.55, 5.97]
Aczel	55.30	117	57.26	123		–1.96	[–5.08, 1.16]
O'Donnell	56.87	33	58.93	25		–2.06	[–10.02, 5.89]
Aveyard	45.26	52	47.44	43		–2.19	[–6.81, 2.44]
McLatchie	52.28	41	55.81	39		–3.54	[–9.32, 2.25]
Boot	51.75	21	57.36	24		–5.62	[–15.39, 4.16]
Meta-Analytic Average for							
Prior Familiarity	53.99	1,221	53.40	1,201		0.62	[–0.38, 1.63]
No Prior Familiarity	48.76	884	50.30	911		–0.84	[–2.60, 0.93]

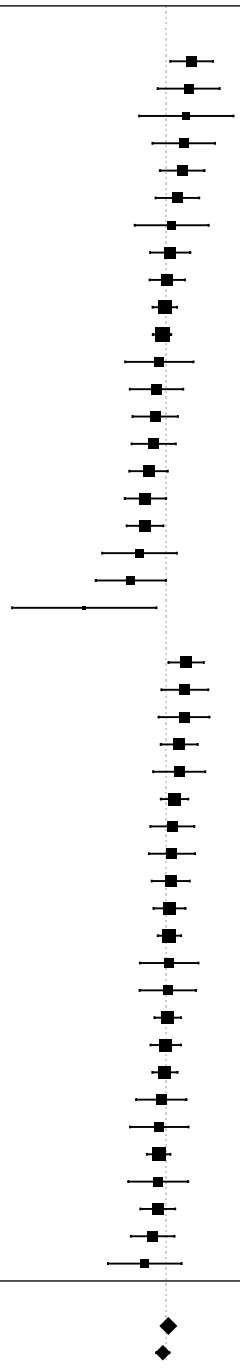


Fig. 7. Difference in trivia performance between the professor-priming and hooligan-priming conditions, separately for participants who reported prior familiarity with the term *hooligan* and those who reported being unfamiliar with the term prior to the study. For each of the 23 labs that met all the inclusion criteria, the figure shows the mean percentage correct and the sample size in each condition. Within each group of participants, the labs are listed in order of the size of the difference between the conditions (professor-priming condition minus hooligan-priming condition); positive effects correspond to the pattern observed in the original study. Each lab is identified by the last name of the corresponding author. In the forest plot, the squares show the observed effect sizes, the error bars indicate the 95% confidence intervals (CIs) around the effect sizes, 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 of the forest plot, the figure shows the numerical values for the effect sizes and 95% CIs. The bottom two rows in the figure present overall means, averaged across all participants in each condition without regard to lab, and the outcome of a random-effects meta-analysis. Note that the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means.

Appendix: Contributing Laboratories

Lead Lab

Michael O'Donnell, University of California, Berkeley
Leif D. Nelson, University of California, Berkeley
OSF page: https://osf.io/vg7ss/

A total of 234 students were recruited from the marketing participant pool at the Haas School of Business, completed the study, and received course credit for participating. Data from 200 participants remained usable after accounting for software issues and after applying the exclusion criteria (professor condition: $n = 106$; hooligan condition: $n = 94$). Participants were tested in individual cubicles. We used the provided PsychoPy scripts and were data blind until after we preregistered the preanalysis manuscript. In all other respects, we followed the official protocol.

Contributing Labs

(The labs are listed alphabetically by the last name of the first author; note that in some cases, the first author is not the corresponding author listed in the figures)[5]

Athfah Akhtar, Birmingham City University
Silvio Aldrovandi, Birmingham City University
Panagiotis Rentzelas, Birmingham City University
OSF page: https://osf.io/cwgkp/

A total of 102 students (professor condition: $n = 57$; hooligan condition: $n = 45$) were recruited from the psychology participant pool at Birmingham City University and received course credit for participating. Two participants were excluded because they did not meet the protocol's age criterion. In the data files, we changed the age entry for 3 participants who mistyped their age (e.g., as "9919 years"). We changed the occupation of 1 participant to student because after testing took place, the participant reported making an error. Participants were tested individually in separate lab rooms. We used the provided PsychoPy scripts adapted for testing in the United Kingdom. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 25 men and 25 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 56 participants (12 males) in the professor condition and 44 participants (5 males) in the hooligan condition.

Ronald Andringa, Florida State University
Nelson A. Roque, Florida State University
Walter R. Boot, Florida State University
Erin R. Harrell, Florida State University
Titus Ebersbach, University of Wuppertal
OSF page: https://osf.io/su7gr/

A total of 153 students (professor condition: $n = 89$; hooligan condition: $n = 64$) were recruited from the psychology participant pool at Florida State University and received course credit for participating. The computer program crashed before an additional 5 participants were assigned to condition and provided any data, and those participants are not included in the tallies. Because of the difficulty we experienced recruiting male participants, we posted flyers in the Psychology Building specifically seeking male participants, and also checked the participant waiting rooms in the Psychology Building for male participants waiting for other studies so we could invite them to participate in our study as well. These participants were given the same information as participants recruited through the Florida State University participant-pool Web site: They were told that we were seeking participants for a general-knowledge and writing study.

Mark Aveyard, American University of Sharjah
OSF page: https://osf.io/dz2gs/

A total of 238 students (professor condition: $n = 127$; hooligan condition: $n = 111$) were recruited from the psychology participant pool at the American University of Sharjah and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts in English. The study was listed as "Writing Task and General Knowledge" in the online registration system. In addition to receiving the scripted instructions, participants were told at the start, "When you get the last screen, it will ask you to tell us that you're done. You don't have to do that; we get a message on the main computer when you're finished. So just please wait patiently in your seat for the session to end." In all other respects, we followed the official protocol. At the end of the session, participants were informed verbally: "We will present the results of this study to you later this semester. But please do not tell other students any details about the study. You can say that there's a writing task, but don't tell them what they're writing about, and don't tell them about the specific questions they'll be asked in the study. If you tell them these details, it can ruin the study results, so please respect that." Applying the exclusion criteria, we removed 33 participants from the analysis; this left a total of 205 participants (professor condition: $n = 112$; hooligan condition: $n = 93$).

Scott A. Baldwin, Brigham Young University
Scott R. Braithwaite, Brigham Young University
Michael J. Larson, Brigham Young University
OSF page: https://osf.io/q4875/

A total of 136 students (professor condition: $n = 63$; hooligan condition: $n = 56$) were recruited from the psychology participant pool at Brigham Young University and received course credit for participating. Participants were tested in separate rooms. We used the provided PsychoPy scripts. In all respects, we followed the official protocol. Although our preregistered plan specified that we would test 35 men and 35 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 28 men in the professor condition and 29 men in the hooligan condition.

Ernest Baskin, Saint Joseph's University
Sean P. Coary, Saint Joseph's University
OSF page: https://osf.io/2bhrv/

A total of 138 students (professor condition: $n = 70$; hooligan condition: $n = 68$) were recruited from the Principles of Marketing participant pool at Saint Joseph's University and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts. In all other respects, we followed the official protocol.

Angie R. Birt, Mount Saint Vincent University
OSF page: https://osf.io/sghq6/

A total of 130 students were recruited from undergraduate courses at Mount Saint Vincent University and received course credit for participating. Because of the exclusion criteria (primarily the required age range), the data from 22 participants were omitted from analysis. Therefore, the final sample size was 108 (professor condition: $n = 49$; hooligan condition: $n = 59$). Participants were tested either individually or in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 26 men and 26 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 13 men in the professor condition and 13 men in the hooligan condition. Although Arielle Comeau, Mount Saint Vincent University, was originally listed as a contributor, she was unable to fulfill her commitments to the project.

Jessie C. Briggs, Temple University
Samantha Moore-Berg, Temple University
Andrew Karpinski, Temple University
OSF page: https://osf.io/ytjep/

A total of 224 students (professor condition: $n = 113$; hooligan condition: $n = 111$) were recruited from the psychology participant pool at Temple University and received course credit for participating. Participants were tested in one of two rooms, either individually or in groups of 2, seated facing opposite walls. We used the provided PsychoPy scripts. In all other respects, we followed the official protocol. Our preregistered plan specified that we would collect data until we had analyzable data from 30 men and 30 women in each condition. We were able to meet the minimum requirements for sample size but had greater difficulty recruiting men than women. We excluded 67 participants (either because they did not meet the a priori criteria, $n = 30$, or because their data were not recoverable following a computer failure, $n = 37$). Our final analyzable sample included 35 men and 42 women in the professor condition and 34 men and 46 women in the hooligan condition.

Desiree Budd, University of Wisconsin–Stout
Michael C. Mensink, University of Wisconsin–Stout
Sarah E. Wood, University of Wisconsin–Stout
OSF page: https://osf.io/vznyt/

A total of 68 students (30 men, 38 women) were recruited from the psychology participant pool at the University of Wisconsin–Stout and received course credit for participating. As a result of equipment failure (no data file was ever written, the script assigned participants to condition but the researchers had no knowledge of this assignment when the computer failed), 7 participants (5 men, 2 women) were excluded from analysis. The remaining participants ($n = 61$) were randomly assigned to the two conditions by the computer (professor condition: $n = 30$; hooligan condition: $n = 31$). Participants were tested either singly or as pairs in a large laboratory classroom. When participants were tested in pairs, they were seated at opposite ends of the room and faced away from each other. We used the provided PsychoPy scripts, and all materials were provided to participants in English. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 35 men and 35 women in each condition, we were unable to recruit enough participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 14 men in the professor condition and 12 men in the hooligan condition, and 16 women in the professor condition and 19 women in the hooligan condition.

Lottie Bullens, Leiden University
Florien M. Cramwinckel, Leiden University and Utrecht University
Marret K. Noordewier, Leiden University
OSF page: https://osf.io/ncyzx/

A total of 149 students (professor condition: $n = 70$; hooligan condition: $n = 79$) were recruited from the psychology participant pool at Leiden University and received course credit or a small monetary reward for participating. Participants were tested in individual cubicles. We used the provided PsychoPy scripts after translating the contents into Dutch (in accordance with the official protocol). Because of difficulties with recruitment, we extended the intended period of data collection. However, after collecting the data, we discovered that the original study had been discussed in a first-year social psychology lecture during this extension. After consulting the Editor and prior to data analysis, we decided to exclude all participants who had participated after the lecture had taken place. Therefore, although our preregistered plan specified that we would test a minimum of 25 men and 25 women in each condition, we were unable to recruit enough male participants. We ended data collection with usable data from 45 participants: 3 men in the professor condition and 6 men in the hooligan condition, 15 women in the professor condition and 20 women in the hooligan condition, and 1 participant in the hooligan condition whose gender is unknown. In all other respects, we followed the official protocol.

Christopher R. Chartier, Ashland University
Kathryn Budzik, Ashland University
OSF page: https://osf.io/3xq6b/

A total of 149 students (professor condition: $n = 75$; hooligan condition: $n = 74$) were recruited from the psychology participant pool at Ashland University and received course credit for participating. Participants were tested in a laboratory room individually. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 25 men and 25 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 25 men in the professor condition and 23 men in the hooligan condition.

Theresa E. DiDonato, Loyola University Maryland
Frank D. Golom, Loyola University Maryland
Martin F. Sherman, Loyola University Maryland
OSF page: https://osf.io/scq7z

A total of 167 students (professor condition: $n = 91$; hooligan condition: $n = 76$) were recruited from the psychology participant pool at Loyola University Maryland and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 25 men and 25 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 32 men in the professor condition and 20 men in the hooligan condition.

Julia Eberlen, Université Libre de Bruxelles
Nicolas Van der Linden, Université Libre de Bruxelles
Myrto Pantazi, Université Libre de Bruxelles
Mando Hanioti, Université Libre de Bruxelles
Olivier Klein, Université Libre de Bruxelles
Axel Cleeremans, Université Libre de Bruxelles
OSF page: https://osf.io/gbhq2/

A total of 275 students (professor condition: $n = 142$; hooligan condition: $n = 133$) participated in the study. About half of the participants were recruited from the psychology participant pool at the Université Libre de Bruxelles and received course credit for participating; the other half was recruited on campus and received payment (€5) for participating. Participants were tested in groups of no more than 8 in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts with minor (and approved) modifications; specifically, the contents were translated into French, and minor modifications were made in order to obtain a working script with French special characters. We are extremely grateful to Gillian Lucy for her evaluation of the quality of the back-translation of the script. In all other respects, we followed the official protocol. Our preregistered plan specified that we would test 50 men and 50 women in each condition. Because we felt that participants needed to be unaware of any kind of priming effect (not, as specified, simply unaware of both priming conditions), we continued data collection until we had obtained data from a sample of 275 participants before exclusion. Both the paid and the course-credit samples were balanced for gender.

Katherine M. Finnigan, University of California, Davis
Jessie Sun, University of California, Davis
Simine Vazire, University of California, Davis
OSF page: https://osf.io/hvs2z/

A total of 323 students were recruited from the psychology participant pool at the University of California, Davis, and received course credit for participating. Sixteen participants' sessions were not recorded because of a computer crash, so they were excluded. Following this

exclusion, 307 students (professor condition: n = 153; hooligan condition: n = 154) remained in the sample. Participants were tested in groups of 4 and completed the study in separate small rooms. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would aim to collect data from 50 women and 50 men in each condition, it took longer to recruit the target number of men than to recruit the target number of women, so we ended with a larger sample than intended. After implementing the study's exclusion criteria, we were left with 277 participants (professor condition: ns = 76 women and 61 men; hooligan condition: ns = 80 women and 60 men).

Natalia Frankowska, SWPS University of Social Sciences and Humanities, Warsaw
Michał Parzuchowski, SWPS University of Social Sciences and Humanities, Sopot
Katarzyna Cantarero, SWPS University of Social Sciences and Humanities, Wrocław
Olga Białobrzeska, SWPS University of Social Sciences and Humanities, Warsaw
OSF page: https://osf.io/buhpe/

A total of 269 students were recruited from the psychology participant pool at SWPS University of Social Sciences and Humanities and received course credit for participating. Of these participants, 54 were unable to complete the procedure and did not provide usable data. The final sample included 215 students (professor condition: n = 103; hooligan condition: n = 112). Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into Polish. Our preregistered target sample was 160 participants; however, we had difficulties recruiting male participants and followed the preregistered plan by continuing to recruit participants until we had usable data from 25 male participants in each condition.

Victor N. Keller, Michigan State University
Carol Tweten, Michigan State University
Jenna A. Harder, Michigan State University
David J. Johnson, Michigan State University
Richard E. Lucas, Michigan State University
Joseph Cesario, Michigan State University
OSF page: https://osf.io/ebz3j/

A total of 436 students (professor condition: n = 233; hooligan condition: n = 203) were recruited from the psychology participant pool at Michigan State University and received course credit for participating. Our preregistered plan was to test students in individual rooms. However, because of maintenance activities in some of the rooms, approximately half of the participants were tested in a shared room with tables separated by dividers. While completing the study, participants could not see other participants or computer screens other than their own. We used the provided PsychoPy scripts and followed the official protocol. Although our preregistered plan specified that we would test 100 men and 100 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 104 men in the professor condition but only 77 in the hooligan condition.

Lina Koppel, Linköping University
Gustav Tinghög, Linköping University
Daniel Västfjäll, Linköping University and Decision Research, Eugene, Oregon
OSF page: https://osf.io/5zcv6/

A total of 182 students were recruited from a participant pool at Linköping University and received 50 Swedish krone (~US$6) for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into Swedish. In all other respects, we followed the official protocol. As a result of technical issues, the data for 1 participant were not saved for analysis, and another 10 participants were excluded because the script crashed before any data could be saved. Additional exclusions were made in accordance with the official protocol. Our final sample included usable data from 29 men and 34 women in the professor condition and 42 men and 34 women in the hooligan condition.

Jean-Baptiste Légal, Université Paris Nanterre
Anthony Lantian, Université Paris Nanterre
Peggy Chekroun, Université Paris Nanterre
Oulmann Zerhouni, Université Paris Nanterre
OSF page: https://osf.io/bsngu/

A total of 137 students (professor condition: n = 67; hooligan condition: n = 70) were recruited from the psychology participant pool at Université Paris Nanterre and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into French. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 30 men and 30 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 26 men in the professor condition.

Karlijn Massar, Maastricht University
Philippe Verduyn, Maastricht University
OSF page: https://osf.io/w4hzw/

A total of 106 students (professor condition: n = 55; hooligan condition: n = 51) were recruited from the psychology participant pool at Maastricht University, The Netherlands, and received partial course credit or a €5 voucher for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into German and Dutch. In all other respects, we followed the official protocol. We ended data collection with data from 25 men and 30 women in the professor condition and 25 men and 25 women in the hooligan condition. However, after inspection of the data, we had to exclude participants because they were younger than 18 years or older than 25 years. The final data set consisted of usable data from 22 men and 30 women in the professor condition and 22 men and 21 women in the hooligan condition.

Matthew T. McBee, East Tennessee State University
Stephanie Chambers, East Tennessee State University
Jacob Coulthard, East Tennessee State University
OSF page: https://osf.io/hva2k/

A total of 77 students (professor condition: n = 33; hooligan condition: n = 32; condition unknown: n = 12) were recruited from the psychology participant pool at East Tennessee State University and received course credit for participating. (Condition was unknown for some students because of computer crashes that resulted in no usable saved data). Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts for data collection and followed the official protocol. Although our preregistered plan specified that we would test 25 men and 25 women in each condition, we were unable to recruit enough participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 33 students in the professor condition (16 male, 15 female, 2 whose gender is unknown) and 32 students in the hooligan condition (12 male, 20 female).

Neil McLatchie, Lancaster University
Dermot Lynott, Lancaster University
OSF page: https://osf.io/rykbu/

A total of 113 students (professor condition: ns = 29 women and 28 men; hooligan condition: ns = 27 women and 29 men) were recruited from the psychology participant pool at Lancaster University and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts. Although we followed the official protocol in all respects, we did deviate from our initial preregistration plan, in which we had stated that we would recruit only participants who spoke English as a first language; we also recruited participants who spoke English fluently as a second language. This modification was approved by the Editor in the middle of recruitment and prior to data analysis. We had initially intended to stop data collection after obtaining usable data from 25 male and 25 female participants in each condition, but following exclusions, we discovered that we had data from only 24 male participants in the hooligan condition; however, this number increased to 25 once we included the usable data from participants whose sessions had been cut short because the script crashed.

Ben R. Newell, University of New South Wales
Aba Szollosi, University of New South Wales
Thomas F. Denson, University of New South Wales
OSF page: https://osf.io/6dhx4/

A total of 142 students were recruited from the psychology participant pool at the University of New South Wales and received either a flat fee of $7.50 Australian or course credit for participating. The data of 4 participants were unrecoverable because the experimental program crashed, and the data of another 8 participants were deleted because they were under 18 years old. Our final sample consisted of 69 participants in the professor condition and 61 participants in the hooligan condition. Participants were tested individually in separate cubicles; a maximum of 4 participants were tested simultaneously per session. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol.

Asil Ali Özdoğru, Üsküdar University
Nursena Balatekin, Üsküdar University
OSF page: https://osf.io/ctkup/

A total of 121 students (professor condition: n = 52; hooligan condition: n = 69) were recruited from the undergraduate programs at Üsküdar University and received course credit for participation. Participants were tested one at a time in a small room. We used the provided PsychoPy scripts after translating the contents into Turkish. After completing the computer tasks, participants responded to two brief paper-and-pencil self-report measures. In all other respects, we followed the official protocol. This lab was excluded from the primary analyses because, after the exclusion criteria were applied, there were fewer than 25 men in each condition.

Michael C. Philipp, Massey University
Matt N. Williams, Massey University
Peter R. Cannon, Massey University
Aaron Drummond, Massey University
OSF page: https://osf.io/7tny4/

A total of 170 participants were recruited via online advertisements, presentations in classes, and in-person recruitment on campus at Massey University. Eighty-six participants were assigned to the professor condition, 82 were assigned to the hooligan condition, and 2 participants' condition assignments were not recorded because of script crashes. Participants were provided with shopping vouchers as compensation for their time. They were tested at both the Palmerston North and the Auckland campuses (in a room with dividers at Palmerston North, and in separate sound-proofed, adjoining booths at Auckland). The provided PsychoPy scripts were used. Our protocol deviated from the official protocol in two respects. First, as noted in our lab-specific preregistration, we excluded participants who had taken an Introduction to Psychological Research course, in which a version of the professor-priming experiment is used as a class research project. Second, one of our research assistants had no previous experience conducting a laboratory-based human-participants study (although she did have experience with other human-participants research), so she received extra supervision, including several practice runs, to ensure compliance with the experimental protocols. Although our preregistered plan specified that we would collect usable data from a minimum of 30 men and 30 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable (postexclusion) data from 27 men in the professor condition and 36 men in the hooligan condition. We obtained usable data from 43 women in the professor condition and 34 women in the hooligan condition.

Monique M. H. Pollmann, Tilburg University
Emiel Krahmer, Tilburg University
Juliette Schaafsma, Tilburg University
OSF page: https://osf.io/43fw8/

A total of 121 students (professor condition: n = 61; hooligan condition: n = 54; condition not recorded: n = 6) were recruited from the Communication and Information Sciences participant pool at Tilburg University and received course credit for participating. Participants were tested in separate cubicles. We used the provided PsychoPy scripts after translating the contents into Dutch. In all other respects, we followed the official protocol. Following our preregistered plan specifying that we would test 25 men and 25 women in each condition, we ended data collection after recruiting 25 male participants in each condition. Screening of the data revealed that 15 participants were younger than 18, were 25 or older, or did not provide their age; that 6 participants did not provide their gender; and that the data from several participants were not recorded. We ended up with usable data from 19 men and 29 women in the professor condition and 18 men and 27 women in the hooligan condition.

Jan Philipp Röer, Witten/Herdecke University
Raoul Bell, Heinrich Heine University Düsseldorf
Laura Mieth, Heinrich Heine University Düsseldorf
Axel Buchner, Heinrich Heine University Düsseldorf
OSF page: https://osf.io/2xw7n/

A total of 220 participants (professor condition: n = 111; hooligan condition: n = 109) were recruited at Heinrich Heine University and received course credit or a small honorarium for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into German. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would recruit a target sample of 200 participants, we decided to continue with data collection after consulting the Editor, because in several instances the script had crashed, and it was unclear at that time whether the data would be recoverable or not. In addition, although the preregistered plan specified that we would recruit only male participants after we had reached the desired number of female participants, we decided not to do this because it would have introduced a systematic difference between the male and female participants in our sample. This decision was made in consultation with the Editor. We ended data collection with usable data from 96 participants in the professor condition and 95 participants in the hooligan condition. This lab was excluded from the primary analyses because, after the exclusion criteria were applied, there were fewer than 25 men in each condition.

Ivan Ropovik, University of Presov
Gabriel Banik, University of Presov
Peter Babincak, University of Presov
OSF page: https://osf.io/mj7yn/

A total of 210 students (professor condition: n = 110; hooligan condition: n = 100) were recruited from the social sciences participant pool at the University of Presov and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts

after translating the contents into Slovak. In all other respects, we followed the official protocol.

Katey Sackett, Rochester Institute of Technology
John E. Edlund, Rochester Institute of Technology
OSF page: https://osf.io/n2974/

A total of 104 students (professor condition: $n = 45$; hooligan condition: $n = 53$; condition unknown because of data failure: $n = 6$) were recruited through the Sona participant pool at Rochester Institute of Technology and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts in the original format and did not deviate from the original protocol in any way. Although our preregistered plan specified that we would test 50 men and 50 women in each condition, significant computer malfunctions prevented us from meeting these quotas. We ended data collection with usable data from 49 men (professor condition: $n = 22$; hooligan condition: $n = 27$) and 38 women (professor condition: $n = 14$; hooligan condition: $n = 24$). Gender imbalances were due to random assignment by the software.

Blair Saunders, University of Dundee
Michael Inzlicht, University of Toronto
OSF page: https://osf.io/25p8z/

A total of 152 students (professor condition: $n = 67$; hooligan condition: $n = 85$) were recruited from the psychology participant pool at the University of Toronto Scarborough and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test at least 30 men and 30 women in each condition, we ended with a larger proportion of men than women. This meant that we collected usable data from only 28 female participants in the professor condition. Although this number was below our preregistered target, the minimum sample sizes ($n = 25$) were exceeded for both conditions when data collection ended.

Michael Schulte-Mecklenbeck, University of Bern and Max Planck Institute for Human Development, Berlin, Germany
Evi Ackermann, University of Bern
Geraldine Neeser, University of Bern
OSF page: https://osf.io/t47wp/

A total of 111 students (professor condition: $n = 50$; hooligan condition: $n = 61$) were recruited from the psychology participant pool at the University of Bern, Switzerland, and received a voucher for a lottery in return for their time. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into German. In all other respects, we followed the official protocol.

David R. Shanks, University College London
Miguel A. Vadillo, Universidad Autónoma de Madrid
Marcos Díaz-Lago, Universidad de Deusto
Chunliang Yang, University College London
OSF page: https://osf.io/ah9w2/

We followed the official protocol and our preregistered plan.

Kenneth M. Steele, Appalachian State University
Corey M. Magaldino, Appalachian State University
Andrew J. Graves, Appalachian State University
Justin Fisher, Appalachian State University
OSF page: https://osf.io/dhgkv/

A total of 634 students (professor condition: $n = 316$; hooligan condition: $n = 318$) were recruited from the psychology participant pool at Appalachian State University and received course credit for participation. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts, without modification. We followed the official protocol in all other respects. Our preregistered plan specified that we would test 300 students, including 100 men. However, because we changed the location of our experiment prior to beginning data collection, we exceeded these goals. The new location allowed us to run twice as many participants per session (maximum = 6). An unanticipated event was that one version of the script produced no records (not even records of condition assignment) for 53 participants. These participants are not included in the total count.

Niklas K. Steffens, University of Queensland
Kim Peters, University of Queensland
Richard L. Bulley, University of Queensland
OSF page: https://osf.io/264p5/

A total of 158 students (professor condition: $n = 72$; hooligan condition: $n = 86$) were recruited from the psychology participant pool at the University of Queensland and received course credit or a monetary incentive for participating. Testing took place in a room with dividers that separated participants from one another, so that they could not see each other's screens. For all testing, we used the most up-to-date PsychoPy script provided

by the Editor. In all respects, we followed the official protocol.

Kyle J. Susa, California State University, Bakersfield
Nasseem Alshaif, California State University, Bakersfield
Heather A. Hansen, California State University, Bakersfield
OSF page: https://osf.io/pwcsh/

A total of 241 students (professor condition: n = 132; hooligan condition: n = 109) were recruited from either the psychology participant pool or classes at California State University, Bakersfield. Participants received either course credit or $5 cash for their time. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts. In all other respects, we followed the official protocol. We ended data collection with usable data from 29 men and 63 women in the professor condition and 26 men and 48 women in the hooligan condition.

Barnabas Szaszi, Institute of Psychology and Doctoral School of Psychology, Eötvös Loránd University
Mark Zrubka, Institute of Psychology, Eötvös Loránd University
Janos Salamon, Institute of Psychology and Doctoral School of Psychology, Eötvös Loránd University
Balazs Aczel, Institute of Psychology, Eötvös Loránd University
OSF page: https://osf.io/ps6fz/

A total of 269 students (professor condition: n = 130; hooligan condition: n = 139) were recruited from the psychology participant pool at Eötvös Loránd University, Budapest, Hungary, and received course credit for participating. Testing took place in a room with dividers that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into Hungarian. In all other respects, we followed the official protocol.

Ricardo M. Tamayo, Universidad Nacional de Colombia
Carolina Rueda, Universidad Nacional de Colombia
Deisy Valcarcel, Universidad Nacional de Colombia
OSF page: https://osf.io/7vahw/

A total of 292 students were recruited from the psychology participant pool at Universidad Nacional de Colombia and received course credit for participating. The data from 72 participants were lost because of software crashes. Of the remaining 220 participants, 113 were assigned to the professor condition, and 107 were assigned to the hooligan condition. Testing took place in a room with dividers that prevented participants from seeing each other's screens. A maximum of 6 participants were tested simultaneously per session. We used the provided PsychoPy scripts after translating the contents into Spanish in collaboration (for initial and blind back-translation) with the laboratory at the University of Granada. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test only 40 men and 40 women in each condition, after consulting the Editor and prior to data analysis, we recruited participants until the proposed deadline. We ended data collection with usable data from 66 men in the professor condition, 48 men in the hooligan condition, 47 women in the professor condition, and 59 women in the hooligan condition.

Yuk-yue Tong, Singapore Management University
Andree Hartanto, Singapore Management University
Nadhilla Melia, Singapore Management University
Clara Chong, Singapore Management University
OSF page: https://osf.io/92ujp/

A total of 149 students were recruited from the psychology participant pool at Singapore Management University and received $6 Singaporean (~US$4.30) for participating. Testing took place in a room with dividers separating participants from each other. We used the provided PsychoPy scripts in English. Our preregistered plan specified that we would test at least 30 men and 30 women in each condition, and we aimed to recruit 35 in each cell as a buffer against potential exclusion. However, there were not enough male participants after exclusions due to software crashes and exceeding the age limit. Hence, we conducted a second phase of data collection 2 weeks after the initial data collection, recruiting only male participants. We ended data collection with usable data from 29 men and 34 women in the professor condition and 32 men and 30 women in the hooligan condition. In all other respects, we followed the official protocol.

Frenk van Harreveld, University of Amsterdam
Michiel van Elk, University of Amsterdam
Bastiaan T. Rutjens, University of Amsterdam
OSF page: https://osf.io/zx928/

A total of 140 students (professor condition: n = 70; hooligan condition: n = 70) were recruited from the psychology participant pool at the University of Amsterdam and received course credit or money (€5) for participating. Testing took place in a room with individual cubicles that separated participants from each other. We used the provided PsychoPy scripts after translating the contents into Dutch, in collaboration with the other Dutch research teams. In all other respects, we followed the official protocol. Two participants' data were incomplete because of a computer crash and were not included in the analyses.

The final sample consisted of 69 participants in each condition. This lab was excluded from the primary analyses because, after the exclusion criteria were applied, there were fewer than 25 men in each condition.

Guillermo B. Willis, University of Granada
Efraín García-Sánchez, University of Granada
Ángel Sánchez-Rodríguez, University of Granada
Rosa Rodríguez-Bailón, University of Granada
OSF page: https://osf.io/tx7dg/

A total of 278 students (professor condition: n = 144; hooligan condition: n = 134) were recruited from the psychology, human-resources, and occupational-therapy participant pools at the University of Granada and received course credit for participating. They were tested in separate and isolated rooms. We used the provided PsychoPy scripts after translating the contents into Spanish in collaboration with the laboratory at the Universidad Nacional de Colombia. In all other respects, we followed our preregistered official protocol. Although we originally recruited enough participants, after exclusions and computer crashes, we did not meet the target for male participants in the hooligan condition. We ended with usable data from 190 women (professor condition: n = 89; hooligan condition: n = 101) and 61 men (professor condition: n = 39; hooligan condition: n = 22).

Robert Zheng, University of Utah
Kevin Greenberg, University of Utah
OSF page: https://osf.io/wujkd/

A total of 122 students (professor condition: n = 61; hooligan condition: n = 61; mean age = 20.4 years) were recruited from the educational psychology and psychology participant pools at the University of Utah and received course credit for participating. Testing took place in a room with dividers separating participants from each other. We used the provided PsychoPy scripts. In all other respects, we followed the official protocol. Although our preregistered plan specified that we would test 35 men and 35 women in each condition, we were unable to recruit enough male participants. After consulting the Editor and prior to data analysis, we ended data collection with usable data from 12 men in the professor condition and 8 men in the hooligan condition.

Acknowledgments

We thank Ap Dijksterhuis for providing materials and helping to ensure the accuracy of the protocol, Katherine Wood for coding the experiment and analysis scripts and for helping to adapt the code for translated versions of the scripts, and Andy DeSoto for gathering trivia items and conducting the Mechanical Turk norming studies for them.

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

We thank the Association for Psychological Science and the Arnold Foundation, who provided funding to participating laboratories to defray the costs of running the study.

Notes

1. In the original publication, professor primes were not compared directly with soccer-hooligan primes in a single two-cell design. One experiment compared participants who had received the professor prime with both participants who had received a secretary prime and participants who had not been primed at all. Another experiment compared participants who had received the soccer-hooligan prime with those who had not been primed. Experiment 4 did include both professor and hooligan primes, along with two other prime types in a 2 × 2 design. However, the results were combined across the different prime types, and the authors did not report a direct comparison of the professor- and hooligan-priming conditions. In our replication study, the two primary between-participants conditions (professor and hooligan primes) were combined into a single experiment at each lab, to allow for a direct comparison.
2. An example of such an item is, "Where do arboreal animals live?" In Latin-based languages, the question gives away the answer, but it was answered correctly by only 65% of undergraduates at the University of California, Berkeley.
3. Although the experimenters recorded each participant's gender at the beginning of the survey, we used participant-reported gender in our analyses, as this was standardized across labs, whereas the experimenter-entered value was not.
4. The wording of this particular debriefing item was quite broad, and it was used to identify participants who believed the studies were linked in *any* way, not just those who intuited that the priming task was meant to affect performance on the trivia task.
5. In some cases, the final sample sizes reported by the individual labs differ from the sample sizes in the reported analyses because the labs made manual exclusions in accordance with the preregistered protocol and additional exclusions were made when the analysis script was applied.

References

Aarts, H., & Dijksterhuis, A. (2002). Category activation effects in judgment and behaviour: The moderating role of perceived comparability. *British Journal of Social Psychology*, *41*, 123–138. doi:10.1348/014466602165090

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.

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.

Dijksterhuis, A., & Van Knippenberg, A. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition*, *18*, 55–74.

Dijksterhuis, A., van Knippenberg, A., & Holland, R. W. (2014). Evaluating behavior priming research: Three observations and a recommendation. *Social Cognition, 32*(Suppl.), 196–208.

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

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.

Lakens, D. (2017). *Professors are not elderly: Evaluating the evidential value of two social priming effects through p-curve analyses*. Retrieved from osf.io/preprints/psyarxiv/3m5y9

Macrae, C. N., & Johnston, L. (1998). Help, I need somebody: Automatic action and inaction. *Social Cognition*, *16*, 400–417.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Nelson, L. D., & Norton, M. I. (2005). From student to superhero: Situational primes shape future helping. *Journal of Experimental Social Psychology*, *41*, 423–430.

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.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., . . . Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLOS ONE*, *8*(4), Article e56515. doi:10.1371/journal.pone.0056515

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534–547.

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.

Yong, E. (2012, October 3). *Nobel laureate challenges psychologists to clean up their act*. Retrieved from https://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535

Yong, E. (2015, August 27). *How reliable are psychology studies?* Retrieved from https://www.theatlantic.com/science/archive/2015/08/psychology-studies-reliability-reproducability-nosek/402466/