

Registered Replication Report: Dijksterhuis & van Knippenberg (1998)

Multilab direct replication of: A variant of study 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(4), 865-877.

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

Proposing/Lead researchers: Michael O'Donnell & Leif Nelson

Protocol vetted by: Ap Dijksterhuis

Protocol and manuscript edited by: Daniel J. Simons

Address correspondence to: Michael O'Donnell, mo279@berkeley.edu or Leif Nelson, leif_nelson@haas.berkeley.edu

Acknowledgments: Thanks to the Association for Psychological Science (APS) and the Arnold Foundation who provided funding to participating laboratories to defray the costs of running the study. 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 help in adapting the code for translated versions of the scripts, and Andy DeSoto at APS for gathering trivia items and conducting the MTurk norming studies for them.

Keywords: priming, replication, intelligence

Abstract

Dijksterhuis and van Knippenberg (1998) reported that participants primed with an intelligent category (“professor”) subsequently performed 13.1% better on a trivia test than participants primed with an unintelligent category (“soccer hooligans”). Two unpublished replications of this study by the original authors, designed to verify the appropriate testing procedures, 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 multi-lab Registered Replication Report (RRR). A total of 40 laboratories collected data for this project, with 23 laboratories meeting all inclusion criteria. Here we report the meta-analytic result of those 23 direct replications (total $N = 4,493$) of the updated version of the original study, examining the difference between priming with professor and hooligan on a 30-item general knowledge trivia task (a supplementary analysis reports results with all 40 labs, $N = 6,454$). We observed no overall difference in trivia performance between participants primed with professor and those primed with hooligan (0.14%) and no moderation by gender.

Brief exposure to one category or construct can 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. Social psychologists soon adapted the study of lexical priming to more complex domains like 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 was 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 (Bargh, Chen, & Burrows, 1996). As the original 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 a participant picks up dropped items (Macrae & Johnston, 1998); priming “cheetah” increases the speed with which a participant picks up a questionnaire (Aarts & Dijksterhuis, 2002); priming “politician” increases long-windedness (Dijksterhuis & van Knippenberg, 2000); priming “superhero” increases the likelihood of volunteering time with an organization (Nelson

& Norton, 2005); or priming with words such as “gamble” increased the likelihood that people would bet in a simulated card game (Payne et al., 2016).

This Registered Replication Report (RRR) project examines 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). Across 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, while other participants imagined their life as a “soccer hooligan,” or were primed with the concept of stupidity more generally. All participants completed a writing task as part of the prime, in which they wrote a paragraph describing their life as either of these types of people, or they listed synonyms for and characteristics associated with intelligence and stupidity. They then completed an ostensibly unrelated trivia test. Participants primed with intelligence answered significantly more questions correctly. This study has been cited over 800 times, and many studies have reported 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 in the field of social psychology.

Over the past 6 years, a number of prominent findings of priming in social psychology have come under greater scrutiny, the professor priming study among them. Most notably, a series of 9 studies failed to find an effect of intelligence priming (Shanks et al., 2013). Yet, a more recent evaluation of the 18 significant *p*-values in 16 published findings of professor priming using *p*-curve (Simonsohn, Nelson, & Simmons, 2014), suggested that the studies contain evidential value (Lakens, 2017). The replication attempts for “professor priming,”

coupled with “failed” replications of other priming studies around the same time (e.g., Doyen et al., 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 represents the results of a multi-lab replication based on that work.

OSF Project page

From the main OSF project page for this RRR (<https://osf.io/k27hm/>), readers can access the experimental protocol (<https://osf.io/gxpm5/>), all materials and experiment scripts (<https://osf.io/6whj3/>), data and analysis scripts along with additional analyses (<https://osf.io/fyptm/>), and a list of participating labs with links to their pre-registration information and descriptions of their testing setting (<https://osf.io/e32su/>). The project page also includes the draft of the manuscript compiled before data analysis began. This final manuscript included a few modifications from that pre-analysis draft, including the addition of an abstract and the discussion section as well as some minor editing for clarity (e.g., expanded figure captions).

Protocol and Procedures

To verify the accuracy of his original protocol, Dijksterhuis re-ran his studies using the original paradigm from Dijksterhuis and van Knippenberg (1998). In those replications, he observed the effect for men but not for women. The lead authors (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 set of trivia questions (with two different populations), an updated procedure, and an analysis strategy.¹

Participants

Each lab was instructed to test a minimum of 25 participants per cell in a 2 (prime: professor vs. hooligan) × 2 (gender: female vs. male) between participant design, with approximately the same proportion of men and women in 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 an equivalent population (e.g., behavioral marketing) in similar ways. Participants were required to be college or university students aged 18-24 years old, with an average age within each lab of approximately 18-20 years. Predictably, not every lab had access to large populations, so the total sample size collected in each lab varied. All sample-size targets were pre-registered and the lead researchers and editor remained blind to the outcomes of individual studies until all data collection was completed.

Testing setting

Participants were tested in person either individually or in small groups (no more than 10). All participants were required to complete the study in individual cubicles or at independent workstations positioned so that participants 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

¹ In the original paper, professor primes were not compared directly to soccer hooligan primes in a single, 2-cell design. One study compared professor-primed participants to both secretary-primed and non-primed participants. Another study compared soccer hooligan-primed participants to non-primed participants. Experiment 4 in the original paper did include both professor and hooligan as primes in a single study, along with two other prime types in a 2x2 design. However, the results were combined across the different prime types and the paper did not report a direct comparison of the professor and hooligan priming conditions. The RRR combined the two primary between-participant conditions (professors and hooligans) into a single experiment to allow for a direct comparison.

to conduct the study. Participants were assigned to either the professor- or hooligan-priming condition by the computerized experimental script, ensuring both randomization and that the experimenter was blind to condition assignment.

Materials

In the original study and in the RRR protocol, the entire study was conducted on the computer. For the RRR protocol, the study was programmed using PsychoPy (Peirce, 2007). The cover story used in the RRR is a variant of the one used in the original study, in which 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 prime and the trivia task. The RRR study used a computer-based funnel-debriefing questionnaire as a more systematic way to test for suspicion.

Generating Trivia Items

Prior to finalizing the protocol, Andy DeSoto at the Association for Psychological Science (APS) gathered a large set of trivia items for use in the study and normed them using Amazon Mechanical Turk (MTurk). 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 (collected one at a time in cubicles, in keeping with the eventual study conditions). The two samples showed similar accuracy. O'Donnell and Nelson then selected a subset of 30 items to use in the RRR protocol, with a goal to select items that had a mean accuracy in the 40-70% range in both norming studies. That set of items was reviewed by Ap Dijksterhuis, with some substitutions made in the original set to ensure that the items covered a broader range of topics.

Three items were changed because their translations in some languages yielded transparently obvious answers.²

Main Study Session

Laboratories that needed a study description for recruiting purposes described the study as: “Complete a series of writing tasks and general knowledge questions.” Prior to the study, the experimenter read the following to 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 initiated the program and recorded the participant’s sex and ID number for each session. The remainder of the task was administered through the PsychoPy program and required no input from the experimenter.

First, participants were instructed to spend five minutes 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 an upcoming social psychology student project. Given that the term “soccer hooligan” might not be equally familiar to participants across cultures, participants were provided with a brief description of either soccer hooligans or of professors (depending on their condition assignment). For the soccer hooligan condition, participants 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.”

² For example, “Where do arboreal animals live?” In Latin-based languages, the question gives away the answer, but it was answered correctly by only 65% of UC Berkeley undergraduates.

For the professor condition, participants 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 between trivia questions of varying levels of difficulty. All participants were told that they had been assigned to the most difficult set of trivia questions, and then answered 30 general knowledge questions. The questions were presented in a fixed order, but the PsychoPy script randomized the order of the answer options for each participant.

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

- “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 | no]
 - If 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 | no]

- If 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?”

The pre-determined exclusion criteria excluded participants who were aware of the other condition, but not those who guessed the intent of the study.

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

Stopping rules and exclusions

Each lab pre-registered its stopping rule to end 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 of the study.

Data from participants 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 from each lab are provided on their OSF project page, and additional details are reported in the appendix.

Results

The original call for labs to participate in the RRR was published on August 10, 2016 on the APS website and advertised via social media. The original deadline to submit an application to participate was September 9, 2016, however due to 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 in the RRR. Seven labs were unable to participate (3 could not collect enough data; 4 dropped out prior to data collection) leaving 40 labs contributing data for the project. The participating labs represent 5 continents and 19 countries. The breakdown of participation was 17 labs from North America (countries represented: Canada & USA), 17 labs from Europe (countries represented: Belgium, France, Germany, Hungary, the Netherlands, Poland, Turkey, Slovakia, Spain, Sweden, Switzerland, and the United Kingdom), 3 labs from Oceania (countries represented: Australia and New Zealand), 2 labs from Asia (countries represented: United Arab Emirates and Singapore) and 1 lab from South America (country represented: 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 issues with the script crashing. Although 40 labs submitted data for the project, 17 labs were unable to meet the pre-registered inclusion criteria of providing data from 25 men and 25 women in each condition. Therefore, the pre-registered analyses in this RRR contain data from the 23 labs that met all inclusion criteria. However, as the remaining 17 labs that did not meet the full inclusion criteria collected data from a large number of participants, the lead lab and editor made a data-blind decision to include these labs in a set of supplementary analyses that were otherwise identical in output to the primary analyses. The full set of these additional analyses is available on the OSF project page, and we provide the results of the primary analysis below.

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 studies are

included regardless of their outcome, providing an unbiased meta-analysis of the effect. The analysis does not focus on null-hypothesis significance testing. Instead, 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 to analyze their data in a way that is consistent with the pre-registered protocol. The output of the script reports the overall difference in trivia performance between participants who were assigned to the professor and hooligan primes (ignoring participant gender). It also reports an estimate of the moderation of that effect by providing separate analyses for the difference in trivia performance between the professor and hooligan primes for men and 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 imports the raw data from all labs and uses similar analysis functions to compute descriptive statistics. Note that this script also required minor modifications to handle data importing across variations introduced during translation and due to differences in how PsychoPy outputs csv files across computer platforms. The meta-analysis script includes analyses of the overall effect of priming condition on trivia performance and of the moderation

of that effect by gender. For each meta-analytic result, we provide a forest plot showing the overall difference between professor and hooligan primes for each laboratory. At the top of each forest plot we show the original result from Dijksterhuis and van Knippenberg (1998), and below the forest plot we provide the results of a random-effects meta-analysis across laboratories for that measure (the meta-analysis does not include the original Dijksterhuis and van Knippenberg result). Tables with the details for each laboratory that went into each forest plot are provided on the OSF project page.

Due to unforeseen inconsistencies in the operation of PsychoPy across languages and computer systems (especially with text entry), some labs experienced a large number of computer crashes during testing. In many cases, those crashes occurred after the priming and trivia tasks were complete. The experiment script was updated to address some of these issues during the testing process (without changing the procedures). These updates also saved a text file backup of each participant's data as they moved through the program so that data from a participant could be included provided that the crash occurred after the primary tasks. Katherine 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 Katherine Wood, and she corrected the improper formatting of those individual files. Labs retained the original and corrected files and both versions are available.

Some additional analyses described below were suggested during the writing and editing of the pre-analysis manuscript. These analyses were coded using simulated data and the analysis

scripts were uploaded and pre-registered prior to conducting the analyses using the actual data. These analyses include examining the effect of moderation by the country in which the study was conducted and familiarity with the concept of hooligan. These additional analyses are flagged as exploratory below.

Primary Analyses

Insert Figure 1 about here

In Study 4 of Dijksterhuis and van Knippenberg (1998), participants who were primed with intelligence scored 13% higher (2.6 more questions answered correctly out of 20) on the general knowledge trivia task than those primed with stupidity. The 23 labs that met all of our inclusion criteria collected data from a total of 5,146 participants. Data from 653 subjects were excluded based on our preregistered exclusion criteria, leaving a total sample of 4,493 included in our preregistered analyses. Our meta-analysis showed an average difference of 0.042 more questions answered correctly out of 30, or 0.14% (95% confidence interval: -0.71% to 1%) between the two priming conditions in the expected direction (see Figure 1). The difference in percentage correct between the professor and hooligan prime conditions ranged from -4.99% to 4.24% across the included labs. The variability in the effect size among the labs (i.e., heterogeneity) was not significantly different from what would be expected by chance ($\tau = 0.86$, $I^2 = 17.43\%$, $H^2 = 1.21$, $Q_{22} = 28.09$, $p = .17$).

Insert Figure 2 about here

While Dijksterhuis and van Knippenberg (1998) initially predicted overall effects of priming condition on trivia performance, based on two follow-up studies his lab conducted to verify the procedures for this RRR, Dijksterhuis expected the difference to be larger for men and possibly absent for women. The follow-up experiments conducted by Dijksterhuis produced a smaller overall effect of the priming condition (a 2-3% difference), with men showing a difference (9.3% and 7.6%) and women not showing a difference (0.3% and -0.3%). Figure 2 shows that the difference between conditions in trivia performance was not substantially moderated by gender in the RRR. Men showed a 0.01% difference (95% CI: -1.38% to 1.41%) in trivia performance between the professor and hooligan primes and women showed a 0.02% difference (95% CI: -0.92% to 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 RRR project. The main result for this expanded set of labs showed an average difference of -0.006 questions answered correctly, or a -0.02% difference (95% confidence interval: -0.77% to 0.73%) between the two priming conditions in the opposite direction of what we expected. Figure 3 shows the forest plot analysis with all 40 laboratories included). Unlike the analysis with 23 labs, this analysis did show statistically significant heterogeneity ($\tau = 1.20$, $I^2 = 26.19\%$, $H^2 = 1.35$, $Q_{39} = 55.47$, $p = .04$). This analysis of all 40 labs showed little difference in priming for men (-.06 points) and women (-.20 points; see Figure 4).

Insert Figures 3 & 4 about here

Alternative Operationalization of Accuracy

An additional exploratory analysis with the 23 labs meeting all inclusion criteria repeated these analyses while treating skipped trivia answers as missing rather than incorrect (forest plot 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% to 0.99%).

Restricting Analysis to Participants Who Did Not Think Tasks Were Linked

Another exploratory analysis repeated the same analyses while excluding participants who, during debriefing, expressed a belief that the priming task and trivia task were related (but who were not excluded from the primary analyses because they did not report awareness of another condition; Figure 5). Nearly 1 in 5 (19.9%) participants responded “yes” when asked, “Do you believe that thinking about a [university professor | soccer hooligan] affected your performance on the general knowledge task.” The analysis excluding these participants reveals a small difference in the expected direction, 0.17% (95% CI: -0.68% to 1.01%). Additionally, 62.7% of participants responded “yes” when asked, “Do you believe that there could be a link between thinking about a [university professor | soccer hooligan] and the general knowledge questions?” The analysis excluding these participants revealed a difference in the expected direction, with participants primed with professor performing 2.07% better on the trivia task than those primed with hooligan (95% CI: 0.57% to 3.57%). Excluding participants who responded “yes” to either or both questions removed 65.9% of the total sample and yielded a meta-analytic effect of 2.32% (95% CI: 0.79% to 3.86%).

Given that this effect is roughly consistent with the size of the overall effect reported by Dijksterhuis for his two follow-up studies and that those studies showed gender moderation, we

conducted further exploratory analysis to examine whether gender moderated that 2.32% effect. Contrary to the predicted pattern, men showed a smaller effect (1.76%, 95% CI: -1.16% to 4.68%) than women (2.70%, 95% CI: 1.05% to 4.35%). We also examined whether the 2.32% effect was robust with the larger sample of 40 labs (where we had observed some heterogeneity) and found that it was reduced and that the confidence interval included zero (1.24%, 95% CI: -0.21% to 2.69%).

Insert Figure 5 about here

Influence of Country Where Testing Was Conducted

We also examined whether the effect varied with the country of the participants (Figure 6) given that different countries ($N_{countries} = 13$) 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, as the 95% CI for each country except the United Arab Emirates included 0, with effect sizes for the individual countries ranging from -3.99% (UAE, 95% CI: -7.42% to -0.56%) to 4.24% (Switzerland, 95% CI: -0.12% to 8.61%).

Prior Familiarity with “Hooligan”

Finally, we looked at whether the effect varied based on whether or not the participant reported having had awareness of the term hooligan prior to the study (Figure 7). Among participants who reported no prior exposure to the term hooligan, there was a small difference in trivia performance, -0.84% (95% CI: -2.60% to 0.93%), in the opposite direction than we expected, while those participants who did report prior exposure to the term hooligan showed a small difference in trivia performance, 0.62% (95% CI -0.38% to 1.63%) in the expected direction.

Insert Figures 6 & 7 about here

General Discussion

Overall, the meta-analytic results of this multi-lab replication observed 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 studies did not observe the difference in trivia performance originally reported in Study 4 of Dijksterhuis and van Knippenberg (1998), and they did not find the gender difference reported in the two unpublished follow-up studies that were used as the basis for the RRR protocol. In the RRR, both the overall effect and the effect for each gender were close to zero.

It is possible that the results from this RRR differed from the original findings due to 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 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 answered “no” when asked either if the tasks were related or if the writing task affected their trivia performance, or both, there was a tendency for professor-primed participants to perform better than hooligan-primed participants (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 paper. The effect with this more restricted sample is more similar to the overall 2-3% effect reported in the unpublished follow-up studies that served as the basis for this protocol, however this difference was substantially smaller when we considered 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 in ways that Dijksterhuis et al. (2014) suggested could moderate the effect (e.g., the original effect tested participants individually, yet some replications used group testing settings) we found little evidence for a difference in testing setting on the observed effect, and results from all settings produced similar meta-analytic results (effects close to zero).

In sum, the findings of this RRR show 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 pre-registered inclusion criteria found an effect with a confidence interval that did not include zero, yet both of these labs found an effect in the opposite direction of the anticipated finding. 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, or whether participants had prior familiarity with the term “hooligan.” Participants who failed to express a belief that the tasks are linked showed a small effect consistent with the original, but these participants constitute a small minority of the total number collected in this RRR and that effect was reduced in the full sample of 40 labs. The results of the RRR are somewhat surprising, as a p-curve analysis showed some evidential value for professor priming in the published literature (Lakens, 2017).

Constraints and Limitations

The original study was conducted in the 1990s, in the Netherlands, and the social culture of professors, hooligans, and experimental participants have changed since then. While the protocol was designed as a test of the original hypothesis, the original effect might have changed

over time (e.g., hooliganism might be less familiar as a construct) and differences in the sampled populations could also affect the ability to observe an effect.

Although the protocol ensured that experimenters were blind to condition assignment, some participants might have intuited the link between the tasks. For example, they might guess that the experimenter expected worse trivia performance after writing about being a hooligan, leading them to try less hard on a trivia test (demand characteristics). The analysis plan did not exclude participants who suspected a link between tasks, meaning that demand characteristics could contribute to differences between conditions (although we did not find differences in the primary analysis). The exploratory analysis excluding those participants who reported a link observed a pattern directionally more similar to the original effect. Although the effect was smaller than in the original and not substantially different from zero, this self-identified naïve population might be more sensitive to the hypothesized priming effects. The data here are 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 a subset of the online participants from India). Although the absolute performance levels for individual trivia items might vary across cultures due to differences in familiarity with the topic (e.g., a question about Joan of Arc might be easier for participants in France than in Colombia), all items were tested in both priming conditions, so such differences in absolute performance should have 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 contributed to the size of the priming effect.

Unedited Manuscript

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*(1), 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*(2), 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*(4), 865-877.
- Dijksterhuis, A., & Van Knippenberg, A. D. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition, 18*(1), 55-74.
- Dijksterhuis, A., Van Knippenberg, A., & Holland, R. W. (2014). Evaluating behavior priming research: Three observations and a recommendation. *Social Cognition, 32*(Supplement), 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), e29081.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141-154.
- Lakens, D. (2017, January 15). 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*(4), 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*(2), 227-234.
- Nelson, L. D., & Norton, M. I. (2005). From student to superhero: Situational primes shape future helping. *Journal of Experimental Social Psychology, 41*(4), 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*(10), 1269-1279.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 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), e56515.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 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*(10), 1660-1672.

Yong, E. (2012, October 3). Nobel laureate challenges psychologists to clean up their act. Retrieved June 04, 2017, 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 June 04, 2017, from <https://www.theatlantic.com/science/archive/2015/08/psychology-studies-reliability-reproducibility-nosek/402466/>

Figure Captions

Figure 1. Difference in percentage correct on trivia performance after priming with professor or priming with hooligan. Results in the forest plot are ordered by the size of the difference between the professor priming condition and the hooligan priming condition, with positive effects corresponding to the pattern in the original study. Laboratories are identified by the last name of the corresponding author. The figure also shows the mean percentage correct for each condition for each lab as well as the sample size contributing to that mean. The difference in percentages between conditions and the confidence interval around that are depicted in the forest plot and are reported to the right of the forest plot. Note that the overall means appearing in the same row as the meta-analytic result in this and all other forest plots in this manuscript are an average across all participants in that condition, without regard to lab. In contrast, the meta-analytic result is the outcome of a random effects meta-analysis of the difference scores and variability from each lab. Consequently, the meta-analytic estimate of the difference between conditions does not necessarily equal the difference between the means in the corresponding row of the figure. (Figure 3 shows the same forest plot analysis with all 40 laboratories included.)

Figure 2. Difference between priming with professor and priming with hooligan on trivia performance, separated by the gender of the participants.

Figure 3. Difference between priming with professor and priming with hooligan on trivia performance for all 40 laboratories.

Figure 4. Difference between priming with professor and priming with hooligan on trivia performance for all 40 laboratories, separated by the gender of the participants.

Figure 5. Difference between priming with professor and priming with hooligan on trivia performance, excluding participants who (a) thought the writing task could influence their trivia performance, (b) thought the tasks were linked, or (c) responded “yes” to either or both of these awareness check items.

Figure 6. Difference between priming with professor and priming with hooligan on trivia performance, separated by the country where testing took place.

Figure 7. Difference between priming with professor and priming with hooligan on trivia performance, separated into those participants who reported prior understanding of the term “hooligan” and those who reported being unfamiliar with the term prior to the study.

Appendix – Lab information

Athfah Akhtar, Birmingham City University
Silvio Aldrovaldi, Birmingham City University
Panagiotis Rentzelas, Birmingham City University
OSF Project: <https://osf.io/cwgkp/>

A total of 102 students (Professor n=57; Hooligan n=45) were recruited from the psychology subject pool at Birmingham City University and received course credit for participating. Two participants were excluded because they did not meet the age study protocol criterion. In the data files, we changed the age entry for three participants as they mistyped their age (e.g., as 9919 years). For one participant, we changed the occupation to student, because the participant reported an error after testing took place. 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 pre-registered 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 (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 Project: <https://osf.io/su7gr/>

A total of 153 students (Professor n=89; Hooligan n=64) were recruited from the psychology subject pool at Florida State University and received course credit for participating. An additional 5 participants experienced a program crash before providing any data (including the condition assignment), and those participants are not included in the tallies above. Due to 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 to invite them to participate in our study as well. These participants were given the same information as participants recruited through the FSU subject pool website: we were seeking participants for a general knowledge and writing study.

Mark Aveyard, American University of Sharjah
OSF Project: <https://osf.io/dz2gs/>

A total of 239 students (Professor n=129; Hooligan n=112) were recruited from the psychology subject pool at the American University of Sharjah and received course credit for participating. Participants tested in a room with dividers separating 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 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."

Scott Baldwin, Brigham Young University
Scott Braithwaite, Brigham Young University
Michael J. Larson, Brigham Young University
OSF Project: <https://osf.io/q4875/>

A total of 136 students (Professor $n = 63$; Hooligan $n = 56$) were recruited from the psychology subject pool at Brigham Young University and received course credit for participating. Participants tested in separate rooms. We used the provided PsychoPy scripts. In all respects, we followed the official protocol. Although our pre-registered 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 in the Hooligan condition.

Ernest Baskin, Haub School of Business, Saint Joseph's University
Sean P. Coary, Haub School of Business, Saint Joseph's University
OSF Project: <https://osf.io/2bhrv/>

A total of 138 students (Professor $n=70$; Hooligan $n=68$) were recruited from the Principles of Marketing subject pool at Saint Joseph's University and received course credit for participating. Participants were tested 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.

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

A total of $N=130$ students were recruited from undergraduate courses at Mount Saint Vincent University and received course credit for participating. Due to exclusion criteria (primarily age range), the data from 22 participants were omitted from analysis, resulting in a final sample size of 108 (Professor $n=49$; Hooligan $n=59$). Participants were tested either individually or in a room with dividers separating participants from each other. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our pre-registered 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 in the Hooligan condition. Although Arielle Comeau, Mount Saint Vincent University, was originally listed as a contributor, she was unable to fulfil her commitments to the project.

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

A total of 227 students (Professor $n = 113$; Hooligan $n = 111$) were recruited from the psychology subject pool at Temple University and received course credit for participating. Participants were tested in one of two rooms, either individually or in groups of two seated facing opposite walls. We used the provided PsychoPy scripts. In all other respects, we followed the official protocol. Our pre-registered plan specified that we would collect data until we had an analyzable sample of 30 men and 30 women in each condition. We were able to fulfill our minimum required sample, but had greater difficulty recruiting men than women (male Professor $n = 35$; female Professor $n = 42$; male Hooligan $n = 34$; female Hooligan $n = 46$).

Desiree Budd, University of Wisconsin-Stout
Michael C. Mensink, University of Wisconsin-Stout
Sarah E. Wood, University of Wisconsin-Stout
OSF Project: <https://osf.io/vznyt/>

A total of 68 students (Professor $n = 32$; Hooligan $n = 34$) were recruited from the psychology subject pool at the University of Wisconsin-Stout and received course credit for participating. Participants were tested either singly or as a pair 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 pre-registered 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 13 men in the professor condition and 12 in the hooligan condition, and 17 women in the professor condition and 19 in the hooligan condition.

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

A total of 149 students (Professor $n=70$; Hooligan $n=79$) were recruited from the psychology subject 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 having translated the contents into Dutch (in accordance with the official protocol). Due to difficulties with recruitment, we extended the intended period of data collection. However, after collecting the data, we discovered that the original study was discussed in a first-year social psychology lecture in this extended period. After consulting the editor and prior to data analysis, we decided to exclude all participants who participated after the lecture had taken place. As such, although our pre-registered 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; 1 participant in the hooligan condition whose gender is unknown. In all other respects, we followed the official protocol.

Christopher R. Chartier, Ashland University

Kate Budzik, Ashland University

OSF Project: <https://osf.io/3xq6b/>

A total of 149 students (Professor n=75; Hooligan n=74) were recruited from the psychology subject 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 pre-registered 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 in the Hooligan condition.

Theresa E. DiDonato, Loyola University Maryland

Frank D. Golom, Loyola University Maryland

Martin F. Sherman, Loyola University Maryland

OSF Project: <https://osf.io/scq7z>

A total of 167 students (Professor n=91; Hooligan n=76) were recruited from the psychology subject pool at Loyola University Maryland and received course credit for participating. Participants tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts, English version. In all other respects, we followed the official protocol. Although our pre-registered 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 in the Hooligan condition.

Julia Eberlen, Université Libre de Bruxelles (ULB)

Nicolas Van der Linden, Université Libre de Bruxelles (ULB)

Myrto Pantazi, Université Libre de Bruxelles (ULB)

Mando Hanioti, Université Libre de Bruxelles (ULB)

Olivier Klein, Université Libre de Bruxelles (ULB)

Axel Cleeremans, Université Libre de Bruxelles (ULB)

OSF Project: <https://osf.io/gbhq2/>

A total of 275 students (Professor n=142; Hooligan n=133) participated in the study. About half of the participants were recruited from the psychology subject pool at the Université Libre de Bruxelles (ULB) and received course credit for participating, while the other half was recruited on campus and received payment (5€) for participating. Participants tested in a room with dividers separating participants from each other, in groups of max. 8. We used the provided PsychoPy scripts with minor (and approved) modifications, i.e., the contents were translated into French, and minor modifications were done 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 pre-registered plan specified that we would test 50 men and 50 women in each condition. Due to our perception that participants needed to be unaware of any kind of priming effect (not, as specified, unaware of both priming conditions), we continued data collection until we had obtained a sample of 275 participants before exclusion. Participants in both the paid and the course credit sample are balanced for gender.

Katherine M. Finnigan, University of California, Davis

Jessie Sun, University of California, Davis

Simine Vazire, University of California, Davis

OSF Project: <https://osf.io/hvs2z/>

A total of 323 students were recruited from the psychology subject pool at the University of California, Davis, and received course credit for participating. Sixteen participants experienced a computer crash that left no record of their session, so they were excluded. Following this exclusion, $N = 307$ students (Professor $n = 153$; Hooligan $n = 154$) 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 pre-registered plan specified that we would aim for 50 women and 50 men in each condition, it took longer to collect 50 men in each condition than 50 women, leaving us with a larger sample than intended. After implementing the study exclusion criteria, we were left with $N = 277$ participants (Professor: $n = 76$ women, $n = 61$ men; Hooligan: $n = 80$ women, $n = 60$ 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 Project: <https://osf.io/buhpe/>

A total of 269 students (Professor $n=103$; Hooligan $n=112$) were recruited from the psychology subject pool at SWPS University of Social Sciences and Humanities and received course credit for participating. Participants were tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts after translating the contents into Polish. Our pre-registered target sample was 160 participants, however we were unable to recruit 25 male participants per condition, therefore, as preregistered, we kept recruiting participants until we had usable data from 25 male participants per condition.

Frenk van Harreveld, University of Amsterdam

Michiel van Elk, University of Amsterdam

Bastiaan Rutjens, University of Amsterdam

OSF Project: <https://osf.io/zx928/>

A total of 140 students (Professor $n=70$; Hooligan $n=70$) were recruited from the psychology subject pool at University of Amsterdam and received course credit or money (5 euros) for participating. Participants tested in a room with individual cubicles separating participants from each other. We used the provided PsychoPy scripts with after translating the contents into Dutch, in collaboration with the other Dutch research teams. In all other respects, we followed the

official protocol. Data of two participants were incomplete due to a computer crash and are not included in the analyses, leaving 69 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 Project: <https://osf.io/ebz3j/>

A total of 436 students (Professor $n=233$; Hooligan $n=203$) were recruited from the psychology subject pool at Michigan State University and received course credit for participating. Our pre-registered plan was to test students in individual rooms. However, due to maintenance 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 pre-registered 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, OR
OSF Project: <https://osf.io/5zcv6/>

A total of 182 students were recruited from a subject pool at Linköping University and received 50 SEK (approx. 6 USD) for participating. Participants tested in a room with dividers separating 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. Data for one participant was lost due to technical issues during saving 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 includes usable data from 29 men and 34 women in the professor condition 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 Project: <https://osf.io/bsngu/>

A total of 137 students (Professor $n=67$; Hooligan $n=70$) were recruited from the psychology subject pool at Univ. Paris Nanterre and received course credit for participating. Participants tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts with after translating the contents into French. In all other respects, we

followed the official protocol. Although our pre-registered 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 Project: <https://osf.io/w4hzw/>

A total of 106 students (Professor n=55; Hooligan n=51) were recruited from the psychology subject pool at Maastricht University, The Netherlands, and received partial course credit or a 5€ voucher for participating. Participants were tested in a room with dividers separating 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 25 men and 30 females in the professor condition, and 25 men and 25 females 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 dataset consisted of usable data from 22 men and 21 women in the hooligan condition, and 22 men and 30 women in the professor condition.

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

A total of 78 students (Professor n=33; Hooligan n=32; Condition unknown=12) were recruited from the psychology subject pool at East Tennessee State University and received course credit for participating. (Condition was unknown for some subjects due to computer crashes resulting in no usable saved data). Participants tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts for data collection and followed the official protocol. Although our pre-registered 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 unknown) and 32 in the Hooligan condition (12 male, 20 female).

Neil McLatchie, Lancaster University
Dermot Lynott, Lancaster University
OSF Project: <https://osf.io/rykbu/>

A total of 113 students (Female Professor n=29; Female Hooligan n=27; Male Professor n=28, Male Hooligan n=29) were recruited from the psychology subject pool at Lancaster University and received course credit for participating. Participants were tested in a room with dividers separating 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 pre-registration plan in which we had stated 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 was approved mid-recruitment by the Editor and prior to data analysis. Also, although we had

initially intended to stop data collection after recruiting usable data from 25 male and 25 female participants in both the professor and soccer hooligan conditions, following exclusions we only recruited 24 male participants in the hooligan condition, although this increased to 25 once we had included the usable data from participants where the script had crashed prior to the completion of the experiment.

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

A total of 142 students were recruited from the psychology subject pool at the University of New South Wales and either received a flat fee of 7.50 AUD or course credit for participating. From these participants, 69 were assigned to the Professor condition and 61 to the Hooligan condition; the data of 4 participants were unrecoverable due to a crash in the experimental program, and the data of further 8 participants were deleted because they were under 18 years old. Participants were tested individually in separate cubicles with a maximum of 4 participants tested simultaneously per session. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol.

Michael O'Donnell, Haas School of Business, University of California, Berkeley
Leif D. Nelson, Haas School of Business, University of California, Berkeley
OSF Project: <https://osf.io/vg7ss/>

A total of 218 students (Professor $n = 116$; Hooligan $n = 102$) were recruited from the Marketing subject pool at Haas and received course credit for participating. Participants were tested in individual cubicles. We used the provided PsychoPy scripts, and were data-blind until after pre-registering the pre-data manuscript. In all other aspects, we followed the official protocol.

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

A total of 121 students (Professor $n = 52$; Hooligan $n = 69$) 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 by the experimenter. We used the provided PsychoPy scripts after translating the contents into Turkish. After completing the computer tasks, participants responded to two brief paper-pencil self-report measures. In all other respects, we followed the official protocol.

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

A total of 168 participants (Professor $n = 86$; Hooligan $n = 82$) were recruited via online advertisements, presentations in classes, and in-person recruitment on campus at Massey

University. Participants were provided with shopping vouchers as compensation for their time. Participants were tested at both the Palmerston North and 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. Firstly, as noted in our pre-lab-specific registration, 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. Secondly, one of our research assistants had not previously had experience conducting a laboratory-based human-subjects study (although she did have experience with other human-subjects research). This research assistant received extra supervision, including several practice runs, to ensure compliance with the experimental protocols. Although our pre-registered 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 (post-exclusion) data from 27 men in the Professor condition and 36 in the Hooligan condition. We obtained usable data from 43 women in the Professor condition, and 34 in the Hooligan condition. An additional two participants began the study but experienced crashes with no recoverable data being produced (not even condition). They are not included in the counts of included or excluded participants.

Monique M.H. Pollmann, Tilburg University

Emiel Kraemer, Tilburg University

Juliette Schaafsma, Tilburg University

OSF Project: <https://osf.io/43fw8/>

A total of 121 students (Professor $n=61$; Hooligan $n=54$; not recorded $n=6$) were recruited from the Communication and Information Sciences subject 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 pre-registered plan that specified 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 either younger than 18, older than 25, or did not provide their age, six participants did not provide their gender, and that the data from several subjects was 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 Project: <https://osf.io/2xw7n/>

A total of 220 participants (Professor $n = 111$; Hooligan $n = 109$) were recruited at Heinrich Heine University and received course credit or a small honorarium for participating. Participants were tested in a room with dividers separating 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 pre-registered plan specified to collect 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. Further, we refrained from promoting data collection especially for male participants after we had reached the desired number of female participants, although the pre-registered plan specified to do so, because this 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.

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

A total of 210 students (Professor n=110; Hooligan n=100) were recruited from the social sciences subject pool at University of Presov and received course credit for participating. Participants were tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts after adapting 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 Project: <https://osf.io/n2974/>

A total of 104 students (Professor n=45, Hooligan n=53, Unknown data failure = 6) were recruited through the SONA subject pool at Rochester Institute of Technology and received course credit for participating. Participants were tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts in the original format and did not differ from the original protocol in any way. Although our pre-registered plan specified that we would test 50 men and 50 women in each condition, we were unable to recruit enough participants after significant computer malfunctions to meet these quotas. We ended data collection of usable data at 49 males (professor n=22, hooligan n=27) and 38 females (professor n=14, hooligan n=24). Imbalances in condition are due to random assignment by software.

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

A total of 152 students (Professor n=67; Hooligan n=85) were recruited from the psychology subject pool at the University of Toronto Scarborough and received course credit for participating. Participants tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts in English. In all other respects, we followed the official protocol. Although our pre-registered plan specified that we would test at least 30 men and 30 women in each condition, we ended with a larger proportion of men compared to women. This meant that we only collected n=28 usable female participants in the Professor condition. While

this number was below our pre-registered target, data collection ended above the participant minimums (n=25) for all conditions.

Michael Schulte-Mecklenbeck, University of Bern, Switzerland and Max Planck Institute for Human Development, Berlin

Evi Ackermann, University of Bern, Switzerland

Geraldine Neeser, University of Bern, Switzerland

OSF Project: <https://osf.io/t47wp/>

A total of 111 students (Professor n=50; Hooligan n=61) were recruited from the psychology subject pool at the University of Bern, Switzerland and received a voucher for a lottery for participating. Participants tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts with 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 Project: <https://osf.io/ah9w2/>

We followed the official protocol and our pre-registered plan. An additional contributor (C. Yang) assisted with data collection.

Kenneth M. Steele, Appalachian State University

Corey M. Magaldino, Appalachian State University

Andrew J. Graves, Appalachian State University

Justin Fisher, Appalachian State University

OSF Project: <https://osf.io/dhgkv/>

A total of 634 students (Professor n=316; Hooligan n=318) were recruited from the psychology subject pool at Appalachian State University and received course credit for participation.

Participants were tested in a room with dividers separating 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, with a goal of 100 men. However, we changed location prior to beginning data collection. 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 for 53 participants, including condition assignment. 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 Project: <https://osf.io/264p5/>

A total of 158 students (Professor n=72; Hooligan n=86) were recruited from the psychology subject pool at the University of Queensland and received course credit or a monetary incentive for participating. Participants were tested in a room with dividers separating participants from one another, such that they could not see each other's screens. We used the most up-to-date PsychoPy script provided by the editors for all testing. 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 Project: <https://osf.io/pwesh/>

A total of 241 students (Professor n = 132; Hooligan n = 109) were recruited from either the Psychology participant pool or from classes at California State University, Bakersfield. Participants received either course credit or five dollars cash for participating. Participants tested 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. 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, ELTE Eötvös Loránd University, Budapest, Hungary
Mark Zrubka, Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
Janos Salamon, Institute of Psychology and Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
Balazs Aczel, Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary
OSF Project: <https://osf.io/ps6fz/>

A total of 269 students (Professor n=130; Hooligan n=139) were recruited from the psychology subject pool at Eötvös Loránd University and received course credit for participating. Participants were tested in a room with dividers separating participants from each other. We used the provided PsychoPy scripts with 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 Project: <https://osf.io/7vahw/>

A total of 220 students (Professor n=113; Hooligan n=107) were recruited from the psychology subject pool at Universidad Nacional de Colombia and received course credit for participating. Participants were tested in a room with dividers in a way that they could not see 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 with the

laboratory at the Universidad de Granada in Spain (for initial and blind back-translation). In all other respects, we followed the official protocol. Although our pre-registered 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 Project: <https://osf.io/92ujp/>

A total of 149 students were recruited from the psychology subject pool at Singapore Management University and received SG\$6 (around US\$4.30) for participating. Participants were tested in a room with dividers separating them from each other. We used the provided PsychoPy scripts in English. Our pre-registered plan specified that we would test 30 men and 30 women in each condition. However, there were not enough male participants after participant exclusion due to program clash and participant exceeding age limit. Hence, only male participants were recruited in later phase of data collection. 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.

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 Project: <https://osf.io/tx7dg/>

A total of 278 students (Professor n=144; Hooligan n=134) were recruited from the psychology, human resources, and occupational therapy subject pools at University of Granada and received course credit for participating. Participants 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 pre-registered official protocol. Although we originally recruited enough participants, after exclusions and computer crashes we ended with usable data from 190 females (professor n = 89; hooligan n = 101) and 61 males, but only 22 of them were randomly assigned to the hooligan condition (professor n = 39).

Robert Zheng, University of Utah
Kevin Greenberg, University of Utah
OSF Project: <https://osf.io/wujkd/>
A total of 122 students (Professor n=61; Hooligan n=61)

Mean age=20.4) were recruited from the Educational psychology and psychology subject pools at University of Utah and received course credit for participating. Participants were tested 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 pre-registered 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.

Unedited Manuscript