

Many Labs 5: Registered Replication of Payne, Burkley, and Stokes (2008), Study 4



Charles R. Ebersole¹, Luca Andrichetto², Erica Casini³, Carlo Chiorri², Anna Dalla Rosa⁴, Filippo Domaneschi⁵, Ian R. Ferguson^{6,7}, Emily Fryberger⁸, Mauro Giacomantonio⁹, Jon E. Grahe⁸, Jennifer A. Joy-Gaba⁶, Eleanor V. Langford¹, Austin Lee Nichols¹⁰, Angelo Panno¹¹, Kimberly P. Parks¹, Emanuele Preti³, Juliette Richetin³, and Michelangelo Vianello⁴

¹Department of Psychology, University of Virginia; ²Department of Educational Science, University of Genova; ³Department of Psychology, University of Milano-Bicocca; ⁴Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova; ⁵Department of Antiquity, Philosophy, and History, University of Genova; ⁶Department of Psychology, Virginia Commonwealth University; ⁷Department of Psychology, New York University; ⁸Department of Psychology, Pacific Lutheran University; ⁹Department of Social & Developmental Psychology, Sapienza University of Rome; ¹⁰Institute of Social and Economic Research, Duy Tan University; and ¹¹Department of Human Science, European University of Rome

Abstract

To rule out an alternative to their structural-fit hypothesis, Payne, Burkley, and Stokes (2008) demonstrated that correlations between implicit and explicit race attitudes were weaker when participants were put under high pressure to respond without bias than when they were placed under low pressure. This effect was replicated in Italy by Vianello (2015), although the replication effect was smaller than the original effect. In the current investigation, we examined the possibility that the source of a study's sample moderates this effect. Teams from eight universities, four in the United States and four in Italy, replicated the original study (replication $N = 1,103$). Although we did detect moderation by the sample's country, it was due to a reversal of the original effect in the United States and a lack of the original effect in Italy. We discuss this curious finding and possible explanations.

Keywords

attitudes, direct replication, reproducibility, open data, open materials, preregistered

Received 1/31/18; Revision accepted 9/27/19

Attitude researchers frequently distinguish between two classes of attitudes, explicit and implicit attitudes. Explicit attitudes are those that can be consciously recalled and deliberately reported. Implicit attitudes are evaluative associations that exist largely outside of conscious control (Devine, 1989; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). These two classes are generally believed to be related but distinct constructs (Nosek & Smyth, 2007) and are often reliably but modestly correlated (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). The strength of these relations varies

by topic, being moderated by factors such as the social context of and individual involvement with the attitude object (Nosek, 2005; Nosek et al., 2007).

Another potentially important factor determining correlations between implicit and explicit attitudes (hereafter, *implicit-explicit correlations*) is the way in which these attitudes are assessed. Explicit attitudes are

Corresponding Author:

Charles R. Ebersole, University of Virginia, Department of Psychology, 485 McCormick Rd., Charlottesville, VA 22904
E-mail: cebersole@virginia.edu

Advances in Methods and
Practices in Psychological Science
1–7

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2515245919885609

www.psychologicalscience.org/AMPPS



generally measured through one or several self-report survey questions with Likert-type response options. Implicit attitudes are often measured through indirect responses, such as ratings of neutral stimuli after the attitude object of interest has been primed (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Payne, Cheng, Govorun, & Stewart, 2005) or reaction times on categorization tasks pairing attitude objects with positive and negative words (e.g., Greenwald, McGhee, & Schwartz, 1998). These differences in measurement could dampen implicit-explicit correlations by introducing construct-independent error.

Payne, Burkley, and Stokes (2008) investigated this possibility. Their prediction, which they labeled the *structural-fit hypothesis*, was that implicit-explicit correlations would increase as the methods of their measurement became more similar. They found support for this account in three studies in which implicit race attitudes, assessed using the Affect Misattribution Procedure (AMP; Payne et al., 2005), correlated more strongly with more procedurally similar explicit race attitude measures (e.g., explicit evaluations of the AMP priming stimuli using the same response scale) than with less similar measures (e.g., the Modern Racism Scale; McConahay, 1983). On the basis of these results, the authors suggested that the relatively weak implicit-explicit correlations reported in the literature were due, in part, to procedural differences between measures of implicit attitudes and measures of explicit attitudes.

In their fourth study, Payne et al. (2008) sought to rule out an alternative explanation for their results: that structurally similar implicit and explicit measures do not successfully distinguish between the two types of attitudes. To assess this possibility, the researchers included a manipulation of social pressure, telling participants either to freely report their attitudes on the measures (low pressure) or to be vigilant against racial bias when responding (high pressure). The introduction of pressure to respond without bias led to less biased explicit responses, weakening the implicit-explicit correlation in the high-pressure condition relative to the low-pressure condition. This suggested that structurally similar implicit and explicit measures did still assess the two classes of attitudes independently.

This latter study was the focus of one replication in the Reproducibility Project: Psychology (RP:P; Open Science Collaboration, 2015; Vianello, 2015). Although the moderation effect of social pressure on implicit-explicit correlations was replicated from the standpoint of statistical significance ($p = .045$), the effect was much weaker than what was observed in the original study (original $\Delta R^2 = .09$; replication $\Delta R^2 = .016$). One reason for the attenuated effect could be differences introduced by the change in the source of the sample. The original study was conducted at a university in the

southeastern United States; the RP:P replication was conducted in Italy. Differences in culture or differences in reactions to pressure to not be biased might moderate this effect and account for the difference in its strength. The current investigation examined that possibility.

Disclosures

Preregistration

This study's design and plans for confirmatory analyses were preregistered on the Open Science Framework (<https://osf.io/4f5zp/>).

Data, materials, and online resources

All materials, data, and code are available on the Open Science Framework (<https://osf.io/wxd4g/>). Additional analyses and key passages of the study materials are available in the Supplemental Material (<http://journals.sagepub.com/doi/suppl/10.1177/2515245919885609>).

Reporting

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

Ethical approval

Data were collected in accordance with the Declaration of Helsinki. The research was approved by the institutional review boards at the data-collection sites.

Method

Sample and power analyses

We recruited participants at universities in the United States and Italy. We replicated Vianello's (2015) power analysis, using the effect size for the focal replication effect from the original study ($\Delta R^2 = .09$, $f^2 = .098$) and $\alpha = .05$. This analysis indicated that at each site we would need 135 participants to achieve 95% power. In order to maximize power to detect the effect, we attempted to recruit this sample size at a minimum of three collection sites within each geographic region. There were no planned exclusions.

To calculate power for our planned mixed-effects analyses (to account for data nested within different collection sites), we simulated a data set with the minimum target sample size, re-creating the implicit-explicit correlations Payne et al. (2008) observed in the United States for the low-pressure ($r = .71$) and high-pressure ($r = .31$) conditions and re-creating the implicit-explicit correlations Vianello (2015) observed in Italy for the

low-pressure ($r = .63$) and high-pressure ($r = .40$) conditions. Power simulations with $\alpha = .05$ (1,000 iterations using the *simr* package; Green & MacLeod, 2016)¹ indicated that the minimum target sample would result in a study well powered to detect the focal replication effect in the aggregate sample (99.40% power, 95% confidence interval, CI = [98.70%, 99.78%]) and in the U.S. sample (99.00% power, 95% CI = [98.17%, 99.52%]). Reflecting the smaller effect found by Vianello (2015), this analysis indicated that the minimum target sample would yield much less power to detect the focal replication effect in the Italian sample (40.90% power, 95% CI = [37.83%, 44.02%]).

Finally, we conducted power simulations to determine our likely power to detect a reliable difference in the focal replication effect between geographic regions. These simulations indicated that the minimum target sample would likely be underpowered to detect a significant difference between the U.S. and Italian samples (37.60% power, 95% CI = [34.59%, 40.69%]), probably because the original and replication studies found statistically reliable results in the same direction. Thus, although the minimum target sample likely provided both adequate power for testing the focal replication effect and relatively precise effect estimates in each sample for the broader Many Labs 5 project (Ebersole et al., 2020, this issue), statistical conclusions about geographic region as a moderator should be taken with caution from this single study and may be best interpreted descriptively.

A total of 1,130 participants provided at least one response. Of those, 1,105 provided enough responses to be included in the analyses. In an unplanned exclusion, 2 additional participants were dropped for claiming to be able to read the Chinese pictographs that were meant to be neutral stimuli. The final sample consisted of 1,103 participants recruited from four universities in the United States ($n = 558$) and four universities in Italy ($n = 545$). The U.S. sample had an average age of 19.59 years ($SD = 2.05$); 61.9% of these participants were female. The sample's racial-ethnic distribution was as follows: 54.0% White, 11.7% Black, 13.2% Asian American, 9.7% Latino or Hispanic, 0.9% Native Hawaiian or other Pacific Islander, 0.4% Native American or Alaskan Native, 7.4% multiracial, and 2.7% other. The Italian sample had an average age of 21.76 years ($SD = 3.05$); 67.6% of these participants were female. This sample's racial-ethnic distribution was as follows: 86.2% White, 0.9% Black, 0.2% Asian, and 12.7% other.

Materials and procedure

Participants in the United States and Italy completed a procedure intended to directly replicate Payne et al.'s

(2008) Study 4.² They completed the study in a lab space at individual computers. Participants were first randomly assigned to either the low-pressure or the high-pressure condition. They read a passage about race relations that either encouraged them to express their honest opinions, even if those opinions were not "politically correct" (low-pressure condition) or encouraged them to provide their opinions while keeping in mind that all people are susceptible to racial biases (high-pressure condition; see the Supplemental Material available online for the full wording of these passages).

Participants then completed a version of the AMP that contained two blocks of trials: one that elicited indirect ratings (assessing implicit attitudes) and one that elicited direct ratings (assessing explicit attitudes). On indirect-rating trials, participants first saw one of three types of primes: a Black face, a White face, or a gray square, which served as a neutral prime. The face primes were 12 different pictures of Black men and 12 different pictures of White men, each showing only the individual's face displaying a neutral expression. The prime appeared in the center of the screen for 100 ms. It was followed by a blank screen for 100 ms, and then a Chinese pictograph appeared for 100 ms. Following the pictograph, a patterned mask of black and white noise appeared on the screen. Participants rated the pleasantness of the pictograph using a scale at the bottom of the screen: -2 (*very unpleasant*), -1 (*slightly unpleasant*), $+1$ (*slightly pleasant*), $+2$ (*very pleasant*). They were specifically warned not to let the prime influence their evaluation. After the rating was reported, the next trial began. In this block, participants completed 72 trials (24 with neutral primes, 24 with Black primes, 24 with White primes) in a random order.

On direct-rating trials, participants saw the same series of stimuli (prime, pictograph, and then noise). However, on these trials, participants were asked to rate the primes, not the pictographs. Participants completed 24 trials in this block, one for each face (gray squares were not rated).

The order of the indirect-ratings block and the direct-ratings block was counterbalanced across participants. After participants completed both blocks, they provided demographic information.

Results

Confirmatory analyses: focal replication effect

Following Payne et al. (2008), Study 4, we hypothesized that implicit and explicit attitudes would be more strongly related in the low-pressure condition than in the high-pressure condition. Responses on both

direct- and indirect-rating trials were first recoded to a scale from 1 to 4 (instead of -2 to 2). Indirect ratings from the AMP were scored by taking the difference between the average of indirect ratings on Black-prime trials and the average of indirect ratings on White-prime trials. Direct ratings were scored using the same procedure. Across all samples, participants were relatively unbiased in both their explicit ($M = -0.01$, $SD = 0.39$) and their implicit ($M = -0.02$, $SD = 0.30$) attitudes.

To test the focal replication effect, we constructed hierarchical mixed-effects models (using the *lme4* package; Bates, Mächler, Bolker, & Walker, 2015). In the first step, we regressed fixed effects of indirect rating and social-pressure condition (contrast coded -1 for low pressure and +1 for high pressure) on direct rating, with random effects of indirect rating and condition nested within collection site. The model failed to converge, so we removed the random effect of indirect rating. In the second step, we added the critical interaction term, Indirect Rating \times Condition) to the model. The fixed effects in this model, and the comparison between Steps 1 and 2, conceptually replicate the model used to test the focal replication effect in the original and replication studies. The random effects account for differences introduced by the nesting of our data, which was not a feature of either the original or the replication study. The addition of the interaction did not improve the model, $\chi^2(1, N = 1,103) = 2.07$, $p = .150$, pseudo- $R^2 = .002$. That is, the relation between implicit and explicit attitudes did not differ reliably between participants in the low-pressure condition and participants in the high-pressure condition.

Confirmatory analyses: moderation by country

Next, we examined whether the focal interaction was moderated by the country in which the data were collected. If the key phenomenon is less robust in Italy than in the United States, it is possible that the lack of an interaction effect in the total sample is due to a relatively weak effect in half of the sample. To test this possibility, we added a third step to the hierarchical mixed-effects model. This step contained a three-way interaction of indirect rating, condition, and country (contrast coded -1 for the United States and 1 for Italy) as a fixed effect. This interaction term did improve the model, $\chi^2(1, N = 1,103) = 9.20$, $p = .002$, pseudo- $R^2 = .008$.

To better understand this three-way interaction, we applied the hierarchical mixed-effects model for testing the focal replication effect to the samples from each country separately. In the Italian samples, the interaction between indirect rating and condition did not

reliably predict direct rating, $\chi^2(1, N = 545) = 0.76$, $p = .382$, pseudo- $R^2 = .001$. However, in the U.S. samples, this interaction did significantly improve the model, $\chi^2(1, N = 558) = 12.21$, $p = .0005$, pseudo- $R^2 = .022$. However, unlike in the original study, this interaction was driven by a stronger relation between indirect and direct ratings in the high-pressure condition ($r = .58$, 95% CI = [.50, .66]) than in the low-pressure condition ($r = .47$, 95% CI = [.37, .55]). Figure 1 displays the effect size for this interaction at each data-collection site. We found the same pattern of results using Vianello's (2015) ratio scoring method (see the Supplemental Material for full details of this analysis and Table S2 for site- and condition-level analyses).

Exploratory analyses: results within target groups

The lack of pro-White bias in the attitude measures (given our largely White samples), as well as the reversal of the target replication effect in the United States prompted some concern about the results. In an attempt to gain more insight into the measures and pattern of results, we examined the data from White and Black participants separately, given that those participants belonged to the focal groups used in the AMP. Within the U.S. sample, 299 participants self-identified as White (non-Hispanic and non-Latino); 65 participants self-identified as Black. Black participants demonstrated more pro-Black explicit attitudes ($M = -0.24$, $SD = 0.39$) than did White participants ($M = -0.01$, $SD = 0.32$), $d = -0.71$, 95% CI = [-0.98, -0.44], and also demonstrated more pro-Black implicit attitudes ($M = -0.15$, $SD = 0.32$) than did White participants ($M = -0.01$, $SD = 0.24$), $d = -0.53$, 95% CI = [-0.80, -0.26].

Next, we examined the relations between direct and indirect ratings among White and Black participants. Among Black participants, implicit and explicit attitudes were strongly correlated ($r = .60$, 95% CI = [.41, .73]), and there was little difference between the low-pressure condition ($r = .61$, 95% CI = [.34, .79]) and the high-pressure condition ($r = .59$, 95% CI = [.30, .78]). White participants demonstrated a weaker implicit-explicit correlation overall ($r = .38$, 95% CI = [.28, .48]). However, compared with Black participants, they showed a greater between-conditions difference in this correlation, which was generally weaker in the low-pressure condition ($r = .33$, 95% CI = [.18, .46]) than in the high-pressure condition ($r = .47$, 95% CI = [.33, .59]). Among White participants, there was little variation in either implicit or explicit attitudes between the two social-pressure conditions (all d s between -0.15 and 0.05; see Table S3). Overall, it appears that the reversal of the expected effect in the United States was driven by

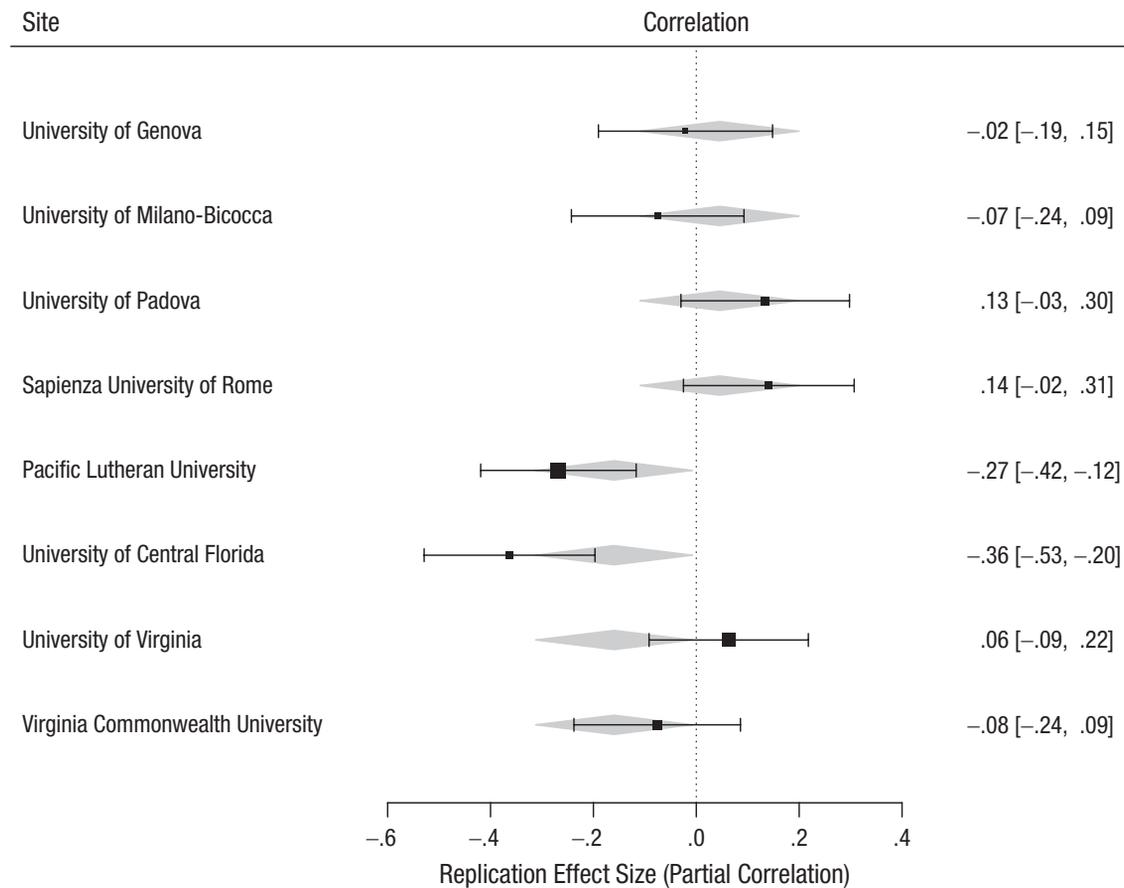


Fig. 1. Replication results by collection site. Each row summarizes the effect size (presented as a partial correlation with its 95% confidence interval) for the focal replication effect at one of the collection sites. The size of the point estimates (the squares) is an inverse function of the model weights. Positive effect sizes indicate effects consistent with the direction of the effect in the original study. The gray diamonds represent meta-analytic aggregate effect estimates within the United States and Italy.

White participants. However, this reversal does not seem to have been caused by the social-pressure manipulation increasing or decreasing implicit or explicit attitudes in aggregate.

Discussion

Payne et al. (2008) proposed the structural-fit hypothesis as an explanation for modest observed correlations between implicit and explicit attitudes. In their experiments, structurally similar direct and indirect measures of attitudes produced stronger implicit-explicit correlations than were generally reported in the literature. In the current investigation, we similarly observed robust correlations between implicit and explicit attitudes using structurally similar measures. In this respect, our results are consistent with the structural-fit hypothesis.

However, our results from the United States were inconsistent with the focal replication effect of stronger implicit-explicit correlations under low social pressure

relative to high social pressure. Rather, the current U.S. data support the opposite conclusion, that implicit-explicit correlations are stronger under high relative to low social pressure. This result was certainly unexpected, and it is unclear why our results diverge so strongly from those of both the original study and Vianello's (2015) replication.

A very plausible explanation is that the social-pressure manipulation did not have the intended effect in the current sample. However, why the manipulation's effect would have diverged from its effect in the original study is less clear. The original study's data were collected in 2006, whereas the current replication's data were collected in 2016 and 2017. The climate surrounding race and racial attitudes changed in that time, particularly in the United States. There has been both a recent increase in reported hate crimes in the United States (U.S. Department of Justice, Federal Bureau of Investigation, 2018) and a recent growth in activist groups supporting racial equality (e.g., Black Lives

Matter; Sawyer & Gampa, 2018). These polarizing forces may lead to differing reactions to manipulations like the one used in this study. Individuals with more prejudiced attitudes may show reactance to instructions to not be prejudiced and therefore report their true (more biased) feelings.³ Similarly, individuals with more egalitarian attitudes may react negatively to being told that they need not be politically correct, and thus report more unbiased attitudes. Countervailing forces like these could lead to differing reactions to the same manipulation, producing different patterns of implicit-explicit relations while not producing mean-level differences in attitudes between conditions. We can only speculate as to these possibilities. However, this study may be an example of how participants' interpretations of study materials change as society changes and may illustrate a challenge of conducting direct replications after periods of social change. We have made our data and analysis scripts available to spur further inquiry into these curious results (<https://osf.io/wxd4g/>).

Conclusion

In the current crowdsourced replication project, we sought to replicate a finding from Payne et al.'s (2008) Study 4, as well as test for moderation of this effect by the sample's source (United States vs. Italy). We did find evidence for such moderation, but with an unexpected pattern. Participants in the United States demonstrated a pattern of results reliably in the direction *opposite* the direction of the original finding; participants in Italy demonstrated no reliable effect. The cause of this reversal among U.S. participants is unknown, although changing social and political climate may be one explanation. Regardless, although the current study did not replicate the prior social-pressure effect on relations between implicit and explicit attitudes, the overall pattern of relations does support the structural-fit hypothesis put forth by Payne et al. (2008), as structurally similar attitude measures produced strong implicit-explicit correlations regardless of the sample's source.

Transparency

Action Editor: Daniel J. Simons

Editor: Daniel J. Simons

Author Contributions

C. R. Ebersole conceived the project, conducted the analyses, and drafted the manuscript. All the authors contributed to data collection and revising the manuscript. E. V. Langford and K. P. Parks assisted with supplementary analyses.

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.

Open Practices

Open Data: <https://osf.io/p2mte>

Open Materials: <https://osf.io/ajwre>

Preregistration: <https://osf.io/4f5zp>

All data, analysis scripts, and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/p2mte> and <https://osf.io/ajwre>. The design and analysis plans were preregistered at the Open Science Framework and can be accessed at <https://osf.io/4f5zp>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245919885609>. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Charles R. Ebersole  <https://orcid.org/0000-0002-8607-2579>

Jon Grahe  <https://orcid.org/0000-0001-6970-0863>

Angelo Panno  <https://orcid.org/0000-0002-6516-161X>

Michelangelo Vianello  <https://orcid.org/0000-0002-1530-1469>

Acknowledgments

The authors would like to thank Jordan Axt and Brian Nosek for helpful comments on earlier drafts of this manuscript and Millisecond Software, LLC, for providing software for this research.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245919885609>

Notes

1. The script used to conduct these power analyses is available at <https://osf.io/rb4nj/>. The simulated data used for the power analyses are available at <https://osf.io/y9sr3/>.
2. The script for the RP:P version was created by Vianello from stimuli provided by Keith Payne and was then checked by Payne for fidelity to the original design. This script was translated into English for the U.S. samples.
3. We thank Keith Payne for suggesting this explanation.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi:10.18637/jss.v067.i01
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18
- Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D.-J., Buttrick, N. R., Chartier, C. R., . . . Nosek, B. A. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in*

- Methods and Practices in Psychological Science*, 3, XXX–XXX.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7, 493–498.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test. *Journal of Personality and Social Psychology*, 97, 17–41.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin*, 9, 551–558.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565–584.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test. *Experimental Psychology*, 54, 14–29.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18, 36–88.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, Article aac4716. doi:10.1126/science.aac4716
- Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94, 16–31.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277–293.
- Sawyer, J., & Gampa, A. (2018). Implicit and explicit racial attitudes changed during Black Lives Matter. *Personality and Social Psychology Bulletin*, 44, 1039–1059.
- U.S. Department of Justice, Federal Bureau of Investigation. (2018). *Hate crime statistics, 2017*. Retrieved from <https://ucr.fbi.gov/hate-crime/2017>
- Vianello, M. (2015). *Replication of Payne, Burkley, & Stokes (2008, JPSP, Study 4)*. Retrieved from <https://osf.io/79y8g/>
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126.