Completing a Race IAT increases implicit racial bias

Ian Hussey & Jan De Houwer

Ghent University, Belgium

Word limit: intro & discussion = 1500, methods & results = 2500

Word count: intro & discussion = 667+1004=1671, methods & results = 1720

Author Note

Corresponding author: Ian Hussey, Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, Gent 9000, Belgium. Email: ian.hussey@ugent.be. Funding provided by Ghent University grants 01P05517 to IH and BOF16/MET\_V/002 to JDH.

Abstract

Over seven million participants have now taken the Implicit Association Test online in order to assess their implicit racial attitudes. Although typically used as an assessment tool, results from four pre-registered experiments (*N* = 940) demonstrated that completing a Race IAT serves to change the very attitudes it seeks to assess. Results indicate that negative automatic racial evaluations of White participants towards Black people increased on two different implicit evaluative measures (Single-Category IAT and Affect Misattribution Procedure) but not on a measure of automatic violence biases (Police Officer’s Dilemma task). These findings not only have important implications for the Race IAT, but also for many other forms of psychological assessment: that by measuring, we often perturb the system that we wish to understand.

*Keywords*: Implicit racial bias; implicit social cognition; implicit association test

Completing a Race IAT increases implicit racial bias

More than seven million Implicit Association Tests (IATs) have been completed on the Project Implicit website (implicit.harvard.edu; Xu, Nosek, & Greenwald, 2014), and the subject of implicit bias has become so mainstream that it served as a topic in recent US presidential debates (Yudkin & Van Bavel, 2018). Research has increasingly asked whether implicit racial biases can be changed via intervention (Lai et al., 2014), with much of this work indicating that those biases are remarkably resistant to change over time (Lai et al., 2016, although see Vuletich & Payne, 2019). Strangely, however, research on implicit bias has never asked if the mere act of measuring such bias serves to change it. This is in stark contrast to work in other domains (e.g., whether asking about suicide serves to increase suicide risk; Blades, Stritzke, Page, & Brown, 2018). Such ‘assessment-as-intervention’ effects are plausible, given that psychological testing provides individuals with new experiences that influence their subsequent behavior. It may be that, akin to Heisenberg’s (1958) observer effect in physics, the mere act of measurement perturbs the system under investigation (for a review of measurement such ‘assessment-as-intervention’ effects in psychology see French & Sutton, 2010).

Preliminary evidence suggests that this may be true for the IAT. For instance, previous work shows that the mere act of completing an IAT containing neutral stimuli can lead to the formation of novel implicit and explicit evaluations towards those stimuli (Hussey & De Houwer, 2018). In that work, the authors argued that these evaluations emerged due to analogical learning taking place within the task itself. Specifically, the act of completing an IAT takes the structure of an analogy, such that there are two pairs of categories on the left (e.g., flowers and positive) and two on the right of the screen (e.g., insects and negative). Across IAT’s blocks, the concept categories (e.g., flowers/insects) swap locations, so that in total, flowers and insects share the same response key and location an equal number of times with positive and negative stimuli. In our previous research, we found that the IAT’s requirement to categorize stimuli into one of these four categories led people to respond analogically (e.g., “flowers are to insects as positive is to negative”). Interestingly, when the insects category was swapped for a novel and neutral one (e.g., Chinese characters), these characters were later evaluated negatively because “flowers are to Chinese characters as positive is to negative”. Equally, when the flowers category was swapped for Chinese characters the latter was evaluated positively (for an in-depth discussion of the analogical nature of the IAT effect see Hussey & De Houwer, 2018).

The current work takes this idea and asks if the very same process plays out when existing social category (i.e., racial) biases are assessed. In other words, are highly ingrained racial biases augmented by the merely act of completing a Race IAT? It may be that hen White participants have to categorize positive and negative stimuli together with pictures of Black and White faces within a Race IAT, they align the two dimensions analogously (e.g., “White people are to Black people as positive is to negative”) so that any existing positive in-group evaluation (i.e., of White people) leads to the inference that the out-group (i.e., Black people) is negative.

This prediction is consistent with previous research that the mere act of completing a Race IAT can influence subsequent behaviour: Vorauer (2012) demonstrated that when White students completed a Race IAT prior to an intergroup interaction, their non-White (Canadian Aboriginal) interaction partner left the exchange feeling less positively regarded (relative to when they completed a Control IAT, or when their interaction partner was White). Whereas Vorauer’s was interested in changes in behaviour in subsequent interracial interactions, we address the question of whether the completing an IAT changes the same implicit evaluative biases that the task is intended to assess. Although Vorauer’s implied mechanisms (e.g., caution, inhibition) certainly differ from those that we put forward here (i.e., analogical learning within the IAT), however the two are not mutually exclusive. For instance, increased *interpersonal* caution and inhibition may be a downstream consequence of the *intrapersonal* analogical learning that plays out within the IAT itself.

Across four pre-registered experiments, we asked participants to first complete either a Race IAT (intervention) or a flowers-insects IAT (control) and then administered additional measures of racial bias. In Experiments 1 and 2 we examined if exposure to a Race IAT influences implicit and explicit evaluative racial bias on the Single-Category IAT, Affect Misattribution Task, and self-report ratings. An ‘assessment-as-intervention’ effect emerged and was subsequently replicated in Experiment 3, demonstrating the robustness of this effect. In Experiment 4, we sought to examine if prior exposure to a Race IAT would also influence other forms of racial behavior such as violence-based bias (using the POD). The ‘assessment-as-intervention’ effect was not found to generalize from evaluative to non-evaluative domains.

# Method

All inclusion and exclusion criteria, data collection stopping rules, analytic strategies and code for their implementation were pre-registered. Pre-registrations, data and code for both measures and analyses are available on the OSF ([osf.io/7pbjq](https://osf.io/7pbjq/)). Detailed descriptions of each measure and the full results of all models are available in the Supplementary Materials.

## Samples

Given that the majority of past race IATs were conducted using Internet samples (see Xu et al., 2014), participants for all studies were recruited online using [prolific.ac](http://www.prolific.ac), and experiments were completed in the participant’s web browser. Participants provided informed consent prior to participation in all experiments. In order to form homogenous racial in- and out-groups between participants and stimuli, we only White participants, and dependent variables employed images of Black people. Sample sizes were selected and pre-registered based on availability of resources. Power analyses were not conducted due to the relative difficulty of specifying expectancies for the larger number of parameters involved in mixed effects models (although see Green & MacLeod, 2016).

**Procedure and measures.** Listed by order of completion, participants completed the Modern Racism Scale (McConahay, 1986), either a (intervention) Race IAT or (control) Flowers-Insects IAT, and a second single-category measure of automatic behaviour towards Black people that varied between studies. They then provided self-report ratings of the images of Black faces used in the behavioural tasks using a 1-to-7 Likert scale (“very negative” to “very positive”). The presence of any differences between the conditions could therefore be attributed to the influence of the IAT’s use of racial or non-racial stimuli.

All IAT task parameters followed the recommendations of a methodological review the IAT (Nosek, Greenwald, & Banaji, 2005). The Race IAT used the same stimuli that are typically employed on the well-known Project Implicit website since 2002 (Xu et al., 2014). The Flowers-Insects IAT was identical to this except that the concept categories were changed from racial exemplars to images of flowers and insects (Greenwald, McGhee, & Schwartz, 1998).

Changes in automatic behaviour due to completing the IAT were assessed using a subsequent task: either the Single-Category IAT (Karpinski & Steinman, 2006), Affect Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005), or Police Officer's Dilemma task (Correll, Hudson, Guillermo, & Ma, 2014). In all cases, we employed single-category variants of these tasks that included images of Black people (but not White people) so as to provide a procedurally non-relative measures of implicit evaluations of the racial out-group only. All other details of each measure followed typical practices for these widely used measures and were based on the result of methodological reviews where possible (Correll et al., 2014; Nosek et al., 2005). The Single-Category IAT (SC-IAT: Experiment 1) provided a measure of automatic evaluations towards Black people through the relative speed of categorization of Black faces when mapped to the same key as either positive or negative words (Karpinski & Steinman, 2006). The Affect Misattribution Procedure (AMP: Experiments 2 & 3) provided a measure of automatic evaluation of Black people through the misattribution of the valence of prime stimuli (Black faces vs. neutral primes) to target stimuli (Chinese characters; see Payne et al., 2010 for previous use of the single-category AMP). Both the SC-IAT and AMP employed the same images of Black faces as the Race IAT. Finally, the Police Officer’s Dilemma task (POD: Experiment 4) captured automatic violence bias by presenting participants with images of Black men who are either armed with a gun or unarmed and requires them to emit a “shoot” or “don’t shoot” response under time pressure. Previous studies have demonstrated a greater propensity to shoot images of Black men relative to White therefore providing a measure of racial violence bias (Correll et al., 2014).

# Results

## Experiment 1

148 participants met inclusion and exclusion criteria (*M*age = 32.1, *SD* = 11.1; 47 women, 98 men, 3 identified using a non-binary category or provided no response), 11 participants were excluded. In line with our pre-registered data analysis plan, we employed linear mixed effects modeling of reaction times on the SC-IAT. This approach increases power by taking all 120 reaction times from each participant into account, but also accomplishes the same goals as the more common strategy of *D* scoring (Greenwald, Nosek, & Banaji, 2003). Reaction time was entered as the dependent variable, SC-IAT block, IAT condition and their interaction were entered as fixed effects, modern racism score was entered as a fixed-effect covariate, and participant was entered as a random effect. As hypothesized, SC-IAT effects differed between the two IAT conditions (i.e., there was an interaction effect between SC-IAT block and IAT condition in the prediction of reaction times), *B* = 4.459, 95% CI [1.030, 7.887], β = 0.017, 95% CI [0.004, 0.029], *p* = .011, with participants in the Race IAT condition demonstrating more negative implicit evaluations of the Black people on the SC-IAT than the Flowers-Insects IAT condition. For the sake of reader familiarity, scoring the SC-IAT data using the *D* score metric revealed that the differences between the conditions were small (*M*race = -0.02, *M*control = 0.04, SDs = 0.33; Cohen’s *d* = 0.17). No differences in self-report ratings were found between the IAT conditions, *B* = -0.04, 95% CI [-0.18, 0.10], β = -0.04, 95% CI [-0.17, 0.09], *p* = .560.

## Experiment 2

213 participants met inclusion and exclusion criteria (*M*age = 35.8, *SD* = 12.1; 103 women, 108 men, 2 identified using a non-binary category or provided no response), 19 participants were excluded. AMP effects were analyzed using a logistic mixed-effects model with AMP ratings as the dependent variable, AMP prime type (Black faces vs. a neutral grey square), IAT condition, and their interaction as fixed effects, modern racism score as a fixed-effect covariate, and participant as a random effect. As hypothesized, AMP effects differed between the two IAT conditions, OR = 0.92, 95% CI [0.90, 0.95], *p* < .001, with participants in the Race IAT condition demonstrating more negative implicit evaluations of Black people than the Flowers-Insects IAT condition. Again, no difference in self-report ratings were found between the IAT conditions, *B* = 0.10, 95% CI [-0.02, 0.22], β = 0.10, 95% CI [-0.01, 0.21], *p* = .089.

## Experiment 3

Next, in order to assess the replicability of the effect on implicit and explicit measures of evaluation, we conducted an exact self-replication of Experiment 2 (i.e., using the AMP as the dependent variable). 333 participants met inclusion and exclusion criteria (*M*age = 36.2, *SD* = 12.6; 178 women, 156 men), 43 participants were excluded. As hypothesized and consistent with Experiment 2, AMP effects differed between the two IAT conditions, OR = 0.94, 95% CI [0.92, 0.97], *p* < .001, with participants in the Race IAT condition demonstrating more negative implicit evaluations of the Black people than the Flowers-Insects IAT condition. Self-report ratings of Black people’s faces were more positive when they previously completed a Race IAT than a Flowers-Insects IAT, *B* = 0.08, 95% CI = [0.08, 0.28], β = 0.17, 95% CI [0.08, 0.25], *p* = .004.

## Experiment 4

246 participants met inclusion and exclusion criteria (*M*age = 36.1, *SD* = 11.5; 152 women, 91 men, 3 identified using a non-binary category or provided no response), 48 participants were excluded. Effects on the Police Officer's Dilemma task are typically analyzed using multiple metrics. We selected the three most common on the basis of a recent meta-analysis (Correll et al., 2014): differential reaction times between trial types (armed vs. not armed), response sensitivity (*d*' scores: the ability to accurately discriminate armed from unarmed individuals), and response bias (*c* scores: propensity towards “shoot” responses relative to a “don’t shoot” responses using the metric). Each was analyzed using a linear mixed-effects model that compared scores between IAT conditions while controlling for self-reported racism. No evidence of the key effect was found on the three metrics (reaction times: interaction between trial type and IAT condition, *B* = 0.983, 95% CI [-0.298, 2.264], β = 0.010, 95% CI [-0.003, 0.023], *p* = .133; sensitivity (*d*’ metric): main effect for IAT condition, *B* = 0.02, 95% CI [-0.10, 0.14], β = 0.02, 95% CI [-0.10, 0.15], *p* = .741; response bias (c metric): main effect for IAT condition, *B* = 0.004, 95% CI [-0.034, 0.042], β = 0.013, 95% CI [-0.113, 0.139], *p* = .840). Self-report ratings of Black people’s faces were more positive when they previously completed a Race IAT than a Flowers-Insects IAT, *B* = 0.19, 95% CI = [0.08, 0.31], β = 0.18, 95% CI [0.08, 0.28], *p* < .001.

## Meta analyses

Meta analyses were used to integrate results across experiments and produce maximally powered conclusions. Restricted maximum likelihood models via the metafor R package were used (Viechtbauer, 2010). Differences between conditions were found on the AMP (Experiments 2 & 3: OR = 0.94, 95% CI [0.92, 0.96], 95% CR [0.91, 0.96], *p* < 0.000000002), whereby participants exposed to the Race IAT evaluated black people more negatively on a subsequently completed AMP. Differences between conditions were found on the self-report ratings (Experiments 1-5: B = 0.12, 95% CI [0.02, 0.21], 95% CR [-0.07, 0.30], p = .021), whereby participants exposed to the Race IAT evaluated black people more positively on the rating scales (see SOM-R).

## Structural validity

These results raise the question of whether differences between the IAT conditions represent ‘genuine’ changes in implicit and explicit bias (i.e., the latent variable) or merely the outcome task demonstrating poorer measurement properties as a result of previously completing the Race IAT (i.e., representing changes in the observed variable only). This was examined by assessing measurement invariance between the two conditions. For the current purposes, tests of metric and scalar invariance between the control and Race IAT conditions were most relevant (see SOM-R). All measures demonstrated good internal consistency and similar configural fit between groups. Critically, all measures passed both metric and scalar invariance between the two IAT conditions, with the exception of the SC-IAT, which failed marginally (see Table 1; note that cut-off values are to be used with caution and are less black-and-white than in hypothesis testing). This suggests that, in general, scores on the outcome measures appear to reflect scores on the latent variables equivalently between the intervention and control IAT conditions. Any differences between the conditions can therefore be attributed to ‘genuine’ changes in implicit and explicit evaluative bias.

Table 1. *Tests of structural validity*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Reliability | | Configural invariance | | Metric invariance | | Scalar invariance | |
|  | α | ω*h* | CFI | RMSEA | ΔCFI | ΔRMSEA | ΔCFI | ΔRMSEA |
| SC-IAT | 0.71 | 0.75 | 1.000 | 0.000 | 0.000 | 0.000 | -0.001 | 0.018\* |
| AMP | 0.82 | 0.83 | 0.977 | 0.127 | 0.002 | -0.036 | -0.008 | -0.001 |
| POD | 0.79 | 0.80 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Self-report ratings | 0.93 | 0.93 | 0.959 | 0.065 | 0.005 | -0.011 | 0.003 | -0.007 |
| *Notes*: SC-IAT = Single-Category Implicit Association Test; AMP = Affect Misattribution procedure; POD = Police Officer’s Dilemma task; α = Cronbach’s alpha; ω*h* = McDonald’s Omega hierarchical; ΔCFI/RMSEA = change in CFI/RMSEA model fit indices between nested models; \*Marginally failed test using the cut-offs suggested by Cheung & Rensvold (2002: ΔCFI ≥ -0.015 & ΔRMSEA ≤ 0.01). Measurement invariance was assessed between the two IAT conditions (control vs. Race IAT). | | | | | | | | |

# Discussion

Across multiple studies we found that the mere act of completing a Race IAT influences both implicit and explicit racial evaluations, suggesting that the IAT changes the very implicit biases it seeks to assess. We refer to these as ‘assessment-as-intervention’ effects. This finding was observed when the outcome measure did (SC-IAT) or did not (AMP) share procedural similarities with the IAT, suggesting that such outcomes are not merely carryover effects from completing a highly similar task twice. A successful replication also demonstrated the effect’s robustness. By controlling for differences in self-reported baseline racism, our analyses also ensured that these outcomes were not due to differences between the intervention and control groups. Finally, tests of measurement invariance suggest that these differences represent genuine changes in the underlying latent variable (racial bias).

Several other findings emerged, some of which were unexpected. First, whereas the IAT changed evaluative biases, it did not change violence biases towards the racial out-group (Police Officer’s Dilemma task). This represents an important and informative constraint on the generality of the effect. Two possibilities present themselves: either (a) the alterations to racial biases produced by the IAT are confined to the same type of bias (evaluations), or (b) implicit racial evaluations are not correlated with racial violence biases. Although the former seems more likely than the latter, in retrospect we realized that there is, to the best of our knowledge, no work that directly compares performance on a Race IAT to that of the POD. We therefore carried out additional analyses testing for such a relationship (see SOM-R) and found that xxx. As such, XX.

Second, we also found that self-report ratings of Black individuals were often more positive in the Race IAT condition, suggesting that ‘assessment-as-intervention’ effects acted in opposite directions across implicit and explicit measures. One explanation for this is that participants in the intervention condition initially encountered a context (Race IAT) where they had to make comparative automatic racial evaluations (i.e., to both Black and White people). Thereafter they had to emit further racial responses in a relatively automatic way (either on the AMP, SC-IAT, or POD). Yet their counterparts in the control condition never encountered comparative racial evaluations, as the outcome tasks only ever assessed (implicitly or explicitly) evaluative biases towards Black people.

It may be that the comparative nature of the IAT sensitized people to the idea of racial comparisons - especially seeing as participants were White and being asked to evaluate Black people. When later given an opportunity to emit evaluations of the racial out-group in a slow, intentional, and deliberate manner, overcompensatory responses may have led them to emit relatively more positive evaluations than those in the control condition where there was no comparative racial context established by the IAT. This would explain why the Race IAT had an impact at the implicit level in one direction and at explicit level in the other. Critically, however, while the direction of the IAT’s impact at the explicit level was opposite to what was predicted, the presence of such an impact in any direction represents an ‘assessment-as-intervention’ effect. This still serves to underscore our core message: that the act of completing a Race IAT changes the racial evaluations it seeks to assess.

At this point, one may ask what implications our findings have for interpreting and using the IAT as a measure of (racial) bias. On the one hand, they highlight an important and previously undetected source of influence on implicit and explicit racial evaluations (i.e., that the act of measuring perturbs the system). On the other hand, this increase in racial bias is relatively small and is unlikely to be the sole driver of the IAT effect itself. There are many other important sources, perhaps most importantly people’s racially biased environments and contexts (see Vuletich & Payne, 2019). Yet, as the creators of the task point out, even statistically small effects can have a societally large impact when implemented at large scale (Greenwald, Banaji, & Nosek, 2015). Although the impact may be small for any given individual, hundreds of thousands of Race IATs are completed each year and may be altering as well as assessing racial bias.

It is also unlikely that these changes in racial bias are long-lived. Growing work shows that multiple interventions intentionally designed to change (implicit) racial bias work in the short-term and yet are largely ineffective in the long-term (Lai et al., 2016; Forscher et al., 2019). Thus the effects observed in our studies - which represent smaller and unintentional changes in racial bias due to measurement - may not produce long-lasting effects within individuals across time.

Finally, it is important that the ‘assessment-as-intervention’ issue not be seen as one unique to the IAT or be considered in a vacuum. Such effects are largely unstudied and rarely acknowledged in psychological science (cf. Blades et al., 2018; French & Sutton, 2010). The ‘assessment-as-intervention’ effects evident on the IAT seem to be small in statistical terms. Yet given the rarity of research on this phenomenon, it is difficult to gauge whether the ‘assessment-as-intervention’ effect associated with the Race IAT is small or large compared to other (implicit) measures of racial bias. Indeed, little is known about the magnitude and severity of this effect in psychological measurement more generally. This is unfortunate given the ubiquity of assessment within psychological science and the potential implications of this effect on our results. Future work needs to examine ‘assessment-as-intervention’ effects in implicit measures in particular, and psychological measurement more generally.

In short, there is good reason to believe that (a) completing a Race IAT serves to change the same evaluations its seeks to assess; (b) the IAT has a differential impact on implicit versus explicit evaluative measures; and (c) this impact influences evaluative rather than other (violence) form of racial bias. More generally, our findings lend support to the idea that the IAT serves to train as well as test for biases, and that training may occur due to the analogical nature of the IAT’s structure (e.g., ‘White is to Black as positive is negative’; see Hussey & De Houwer, 2018).

# Author Contributions

IH developed the study concept, performed data collection, conducted the analysis, and interpreted the results. Both authors contributed to the study design. IH drafted the manuscript and JDH provided comments on revisions.

# 

# Acknowledgements

Thanks to Sean Hughes for his useful comments on earlier versions of the manuscript.

# Open Practices Statement

All experiments were preregistered and provide open materials and data. These can be accessed at [osf.io/7pbjq](https://osf.io/7pbjq/).

References

Blades, C. A., Stritzke, W. G. K., Page, A. C., & Brown, J. D. (2018). The benefits and risks of asking research participants about suicide: A meta-analysis of the impact of exposure to suicide-related content. *Clinical Psychology Review*. <https://doi.org/10.1016/j.cpr.2018.07.001>

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. <https://doi.org/10.1207/s15327752jpa4803_13>

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. <https://doi.org/10.1207/S15328007SEM0902_5>

Correll, J., Hudson, S. M., Guillermo, S., & Ma, D. S. (2014). The Police Officer’s Dilemma: A Decade of Research on Racial Bias in the Decision to Shoot. *Social and Personality Psychology Compass*, *8*(5), 201–213. <https://doi.org/10.1111/spc3.12099>

Forscher, P. S., Lai, C. K., Axt, J., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A Meta-Analysis of Procedures to Change Implicit Measures. Preprint: <https://doi.org/10.31234/osf.io/dv8tu>

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? *British Journal of Health Psychology, 15,* 453–468. <https://doi.org/10.1348/135910710X492341>

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (1st ed., pp. 668–681). New York, NY: Oxford University Press.

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*(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*(4), 553–561. <https://doi.org/10.1037/pspa0000016>

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>

Heisenberg, W. (1958). *Physics and philosophy*.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>

Hussey, I., & De Houwer, J. (2018). The Implicit Association Test as an analogical learning task. *Experimental Psychology*, Accepted manuscript. <https://doi.org/10.31234/osf.io/yug9h>

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32. <https://doi.org/10.1037/0022-3514.91.1.16>

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. <https://doi.org/10.1037/a0036260>

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016. <https://doi.org/10.1037/xge0000179>

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 91–125). San Diego, CA: Academic Press.

Nosek, B. A. (2007). Implicit–Explicit Relations. *Current Directions in Psychological Science*, *16*(2), 65–69. <https://doi.org/10.1111/j.1467-8721.2007.00477.x>

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin*, *31*(2), 166–180. <https://doi.org/10.1177/0146167204271418>

Payne, 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*(3), 277–293. <https://doi.org/10.1037/0022-3514.89.3.277>

Payne, K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology*, *46*(2), 367–374. <https://doi.org/10.1016/j.jesp.2009.11.001>

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

Vorauer, J. D. (2012). Completing the Implicit Association Test Reduces Positive Intergroup Interaction Behavior. *Psychological Science*, *23*(10), 1168–1175. <https://doi.org/10.1177/0956797612440457>

Vuletich, H. A., & Payne, B. K. (2019). Stability and Change in Implicit Bias. *Psychological Science,* Advance online publication. <https://doi.org/10.1177/0956797619844270>

Xu, F. K., Nosek, B. A., & Greenwald, A. G. (2014). Psychology data from the Race Implicit Association Test on the Project Implicit Demo website. *Journal of Open Psychology Data*, *2*(1). <https://doi.org/10.5334/jopd.ac>

Yudkin, D. A., & Van Bavel, J. (2018, January 20). The Roots of Implicit Bias. *The New York Times*. Retrieved from <https://www.nytimes.com/2016/12/09/opinion/sunday/the-roots-of-implicit-bias.html>