

The Implicit Association Test in Introductory Psychology Textbooks: Blind Spot for Controversy

Psychology Learning & Teaching


1–13

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14757257211055200

journals.sagepub.com/home/plj**Jared M. Bartels** 

Department of Psychological Science, William Jewell College

Patricia Schoenrade

Department of Psychological Science, William Jewell College

Abstract

The Implicit Association Test (IAT) has been widely discussed as a potential measure of “implicit bias.” Yet the IAT is controversial; research suggests that it is far from clear precisely what the instrument measures, and it does not appear to be a strong predictor of behavior. The presentation of this topic in Introductory Psychology texts is important as, for many students, it is their first introduction to scientific treatment of such issues. In the present study, we examined twenty current Introductory Psychology texts in terms of their coverage of the controversy and presentation of the strengths and weaknesses of the measure. Of the 17 texts that discussed the IAT, a minority presented any of the concerns including the lack of measurement clarity (29%), an automatic preference for White people among African Americans (12%), lack of predictive validity (12%), and lack of caution about the meaning of a score (0%); most provided students with a link to the Project Implicit website (65%). Overall, 82% of the texts were rated as biased or partially biased on their coverage of the IAT. The implications for the perceptions and self-perceptions of students, particularly when a link to Project Implicit is included, are discussed.

Keywords

implicit association test, IAT, introductory psychology, a. g. Greenwald, m. r. Banaji

Corresponding author:

Jared M. Bartels, Department of Psychological Science, William Jewell College, Liberty, MO 64068, USA.

Email: bartelsj@william.jewell.edu

The Implicit Association Test (IAT; Greenwald et al., 1998) measures automatically activated evaluative associations that form implicit attitudes, attitudes that are consciously inaccessible yet, presumably, capable of impacting one's actions. Since the seminal publication of the IAT, millions of people have taken the tests (e.g., Race IAT) online (Sleek, 2018) and more than 2,500 peer-reviewed journal articles involved use of the IAT (Greenwald et al., 2019). The IAT has spawned a lucrative implicit bias training industry with companies spending thousands for daylong bias training seminars (Miller, 2015). In 2015 alone, Google spent 150 million on such training. Given these statistics and the obvious fervor with which it has been embraced, one would assume the IAT is on firm psychometric and theoretical ground. Yet critics have raised numerous, substantive questions about the measure. A comprehensive review of these issues is beyond the scope of this article and interested readers should consult Blanton and Jaccard (2008) and Mitchell and Tetlock (2017). Rather, our focus is on questions we deem both central to the viability of the IAT as a measure of implicit bias and pedagogically important for introductory psychology students learning about the IAT. These questions center around two chief issues: what the IAT measures (or does not measure) and what it predicts (or fails to predict).

What Is the IAT Measuring?

The title of the press release introducing the IAT (Schwarz, 1998) reads, "*Roots of unconscious prejudice* [emphasis added] affect 90 to 95 percent of people." Then, in the first sentence of the same release, the author notes that the IAT measures the "unconscious roots of prejudice." This may appear to be a trivial distinction (the difference between unconscious prejudice and prejudice) but it is not, as Banaji and Greenwald (2013) note:

This [that the race IAT does not predict overt prejudice] is why we answer no to the question "Does automatic White preference mean 'prejudice'?" The Race IAT has little in common with measures of race prejudice that involve open expressions of hostility, dislike, and disrespect." (p. 94)

After completing the test on the Project Implicit website, respondents are informed about their preferences (e.g., "Your responses suggested a strong automatic preference for White people over Black people."). The results obviously speak to a relative preference, but do they suggest negativity towards Black people? Might there be alternative interpretations? Consider the following interpretation proposed by Mitchell and Tetlock (2017):

It may be that Whites respond more slowly when Latonya requires responding to the good key than when Betsy requires responding to the good key. However, Whites may nevertheless have a positive attitude toward African Americans, albeit not as positive as toward members of their own race. A relative difference in RT [reaction times] between two target sets does not necessarily imply hostility or prejudice toward either group. (p. 267)

Similarly, Brendl et al. (2001) note:

The second problem with the IAT is that it is difficult to interpret the results. The test itself gives at best a relative measure of one target set against another. However, in contradiction to this constraint of relativity, the results of the IAT are often interpreted as reflecting an implicit prejudice for one group over another. The problem with this interpretation is that...prejudice connotes a negative attitude toward a group. When two groups are measured relative to one another, it is possible for one group to be preferred to another without the second group being evaluated negatively. (p. 771)

Studies by Andreychik and Gill (2012) raise additional questions about the interpretation of IAT scores. The authors found that a strong automatic preference for White people (i.e., IAT effect) can reflect empathy for Black people (rather than prejudice) among those who endorse external explanations (e.g., slavery and segregation) for the plight of Black people. Brendl et al. (2001) found an IAT effect for White names and non-words (study 1) similar to the typical IAT effect with White names/pictures and Black names/pictures. The typical IAT effect is interpreted as prejudice against Black people, but familiarity rather than prejudice better accounts for the IAT effect with non-words.

If the IAT is as a measure of prejudice (or anything more than dispassionate associations) against Black people, one would not expect studies to produce an IAT effect among Black participants, yet they do (e.g., Nosek et al., 2002). Arkes and Tetlock (2004) explain:

If we assume that the thousands of African Americans who took the Web-based IAT are not prejudiced against their own race, then these data strongly suggest that culturally stereotypic associations, which they do not endorse, are responsible for this result. (p. 262)

Similarly, Correll et al. (2002) found all participants (White and Black) showed a propensity to shoot an armed target in a videogame if the target was Black. Such results suggest that one cannot take for granted that a measure of implicit bias including the IAT assesses hostility or even negative feelings toward Black people.

Adding to the uncertainty are questions about whether the IAT assesses a stable disposition as opposed to attitudes greatly influenced by situational factors. This is a fair question given the poor test-retest reliability that may suggest it is measuring an unstable construct influenced by situational factors like anxiety over appearing racist (Blanton & Jaccard, 2008). Measurement error and systematic method variance further contribute to instability. Given this, Payne et al. (2017) and Schimmack (2021) have made the case that IAT is best viewed as a state rather than trait measure of associations. Payne et al. (2017) concluded, “the evidence for situational variability is stronger than the evidence for chronic individual differences” (p. 306).

What Does the IAT Predict?

Given these concerns about the meaning of IAT scores, it is crucial that IAT results, if indicative of prejudice, predict meaningful prejudicial behavior. The *Blindspot* authors suggest that an automatic White preference is “now established as signaling discriminatory behavior” (Banaji & Greenwald, 2013, p. 47). Several meta-analyses have addressed questions about the predictive validity of the IAT. While Greenwald et al. (2009) found an average correlation of .27 between the IAT and criterion measures, more recent meta-analyses have found much lower correlations between the IAT and such measures. Oswald et al. (2013) not only found lower correlations between the IAT and a host of criterion measures ($r = .14$), but found particularly unimpressive correlations ($r = .07$) between the IAT and microbehaviors (e.g., body posture in interracial interaction). The importance of this finding cannot be overstated, as Greenwald and Krieger (2006) note that evidence to the contrary (i.e., Greenwald et al., 2009) suggests that “implicit attitudinal biases are *especially* [emphasis added] important in influencing nondeliberate or spontaneous discriminatory behaviors” (p. 961). Additionally, the IAT was no better than explicit measures at predicting behaviors and was weakly correlated with explicit measures of bias ($r = .14$). Based on these findings, Oswald et al. (2013) concluded, “Any distinction between the IATs and explicit measures is a distinction that makes

little difference, because both of these means of measuring attitudes resulted in poor prediction of racial and ethnic discrimination” (p. 183). Though not focusing exclusively on the IAT, a meta-analysis by Forscher et al. (2019) found change in implicit bias had a minimal effect on either explicit attitudes or behavior.

A study by McConnell and Leibold (2001) is held up as an example of the IAT’s ability to predict discriminatory behavior or microbehavior. This study is detailed in the aforementioned book *Blindspot* (Banaji & Greenwald, 2013) and Greenwald and Krieger (2006) suggest that the results, along with those of another study like it, “reveals that implicit bias may affect interviews in ways that can disadvantage Black job applicants” (p. 962). However, a closer inspection of the study and conclusions by the authors suggests that its stature is not warranted. Blanton and Jaccard (2008) not only critiqued the study noting several methodological shortcomings, but reanalyzed the data to examine their impact. In the McConnell and Leibold study, participants interacted with a White and Black experimenter before and after completing the IAT. Observers of the interactions rated the behavior of participants across sixteen dimensions (only 5 of the 16 were significantly correlated with IAT scores). These dimensions were combined to form three molar categories of interaction quality, friendliness, curtness or abruptness, and general comfort level. None of the three, by themselves, were correlated with IAT scores. Moreover, inter-rater reliability between the observers was well below conventional standards (i.e., .70) with reliabilities of .48, .43, and .53 for the three abovementioned categories. Closer inspection of the sample reveals an outlier in terms of age (i.e., a 50-year-old). The significance of this is that slower processing speed associated with age contaminates scores producing false positive IAT effects. When Blanton and Jaccard (2008) removed this outlier and reanalyzed the correlation between IAT scores and molar ratings, it was no longer significant. More significantly, the majority of the participants in the McConnell and Leibold actually acted *more* positively toward the Black experimenter than the White experimenter.

Perhaps the most conspicuous indictment of the predictive validity of the IAT is the pair of contradictory results obtained by McConnell and Leibold (2001) and Shelton et al. (2005); McConnell and Leibold (2001), aforementioned criticisms notwithstanding, reported that IAT effects were predictive of more negative interactions between White participants and a Black experimenter. Shelton et al. (2005) used similar methods but found quite different results. These researchers found IAT effects predicted more positive perceptions of the interactions between these participants and a Black partner. In terms of the predictive validity of the IAT, Oswald et al. (2015) conclude:

Sixteen years and hundreds of studies after the IAT was introduced (Greenwald et al., 1998), we know that the IAT reliably produces distributions of scores that are said by many IAT researchers to reveal large reservoirs of implicit bias against racial and ethnic minorities, but researchers still cannot reliably identify individuals or subgroups within these distributions who will or will not act positively, neutrally, or negatively toward members of any specific in-group or out-group. (p. 569)

Even if one were to discount the aforementioned criticisms, the inconsistency between implicit and explicit attitudes, along with the likely gravity of learning that one possesses an automatic White preference and all that accompanies it, prompted the *Blindspot* authors to provide forewarning while introducing readers to the IAT. For example, after inviting readers to take the Race IAT, the authors provide the following, “A note of caution: If you prefer not to risk discovering a result different from the one you predicted, you might want to avoid this IAT” (Banaji & Greenwald, 2013, p. 42). Further caution is recommended later as the authors state:

Understanding the real discomfort that may be involved in coming to terms with one's IAT results has obliged us to think twice before encouraging people to try an IAT that might expose a heretofore unknown - and unwelcome - dissociation...For those who would rather not know about their hidden biases, it may be best to let sleeping dissociations lie. (p. 102)

If one decides to take the "risk," then it might be somewhat reassuring, assuming the result of an automatic White preference, that the result does not suggest that one is prejudiced. Again, the *Blindspot* authors note: "...we answer no to the question 'Does automatic white preference mean 'prejudice?'" The Race IAT has little in common with measures of race prejudice that involve open expressions of hostility, dislike, and disrespect" (Banaji & Greenwald, 2013, p. 52).

The purpose of the present study is to examine the extent to which the abovementioned criticisms of the IAT are represented in introductory psychology textbooks. The focus on texts for Introductory Psychology was particularly important because, for many students, this course forms a significant part of the initial introduction to the field. Students often anticipate that Introductory Psychology, perhaps more than some other courses, will help them probe important questions facing society, and add to self-understanding. Consequently, they are likely to be drawn to sections relevant to issues current in the news, and perhaps to assess themselves based in part upon what they read. Discussion of the IAT found within an Introductory Psychology text may well contribute to students' perceptions of their own weaknesses and strengths. Presumably with these possibilities in mind, authors of Introductory Psychology texts are, as Ferguson et al. (2018) observe, faced with two goals that may at times conflict. Findings and conclusions must be presented with scientific accuracy, while texts must also be kept interesting, offering clear and applicable 'takeaways.' Griggs and Whitehead (2014) similarly note a tension between the goal of current scientific accuracy and the wish to keep a text at similar length in successive editions. While it is necessary to simplify scientific presentations at the introductory level, doing so may come at a cost of suggesting that conclusions are clearer, more generalizable, and/or less contested than is the case. Psychological scientists, enthusiastic about the field, often wish to demonstrate to students that the work is useful, applicable, and offers hope for constructive change. These conflicting pressures could result in the Introductory Psychology student jumping to conclusions that, unbeknownst to the student, are based on assumptions that are at best poorly supported.

Method

We designed categories to observe within the IAT coverage in a given text from the concerns described in the present review. We endeavored to focus on probable impact on the reader. The first three categories were focused on the presentation of the IAT in the context of the psychological science endeavor. Specifically, we noted whether there was any discussion of uncertainty or confusion as to what the instrument actually measures. We further noted whether authors mentioned the finding that 40% of African American respondents demonstrate an 'automatic white preference,' a result for which bias against an outgroup cannot be assumed to account (Arkes & Tetlock, 2004). We also examined whether authors described the questions surrounding predictive validity of the IAT.

Next, since a link to Project Implicit would presumably increase the likelihood that a reader would complete an IAT, we noted the presence or absence of such a link. A related issue was how a reader may interpret a score, given the possible negative implications regarding one's "implicit" attitudes. Thus, we included a category indicating whether there was any warning or caution regarding interpretation of the results. As we have noted, the test's creators explicitly warn

against interpreting results as reflecting prejudice (Banaji & Greenwald, 2013), recommending the term “automatic white preference” (in the case of the Race IAT). They do, however, suggest that a white-preference result “is now established as signaling discriminatory behavior” (p. 47). (As explained here, this relationship is now very much in question.)

Additionally, we categorized each text section using a system applied by Ferguson et al. (2018), rating each text’s coverage of the IAT as “Good,” “Partially Biased,” or “Biased.” Specifically, textbooks that provided a fair, accurate coverage of controversies surrounding the ability of the IAT to measure implicit prejudice and predict behavior were “Good.” Textbooks that acknowledged IAT controversy but failed to provide elaboration, focusing on the “side” or interpretation of the proponents and authors of the measure were considered “Partially Biased.” Finally, textbooks that failed to even acknowledge the controversy, thus providing a decidedly one-sided presentation of the IAT were labeled as “Biased.”

Introductory Psychology textbooks were identified using VitalSource[®], an online textbook evaluation service. Only texts copyrighted 2018 or later were included, and the most recent edition of each text was used. We gathered 20 textbooks. Each text was searched for possible references to the IAT using the search terms “IAT,” “Implicit Association Test,” “prejudice,” “implicit bias,” and “implicit associations.” If these terms failed to produce results we searched for IAT content using the names of the authors, Greenwald and Banaji. Of the 20 texts identified, 17 included some reference to the IAT. All pages including any discussion of the IAT were reviewed. Working independently, we applied each of the categories to the IAT, using both rubrics. There was generally good agreement, with percentages ranging from 76% to 94%. Any inconsistent ratings were resolved via discussion.

Results

Acknowledgement of Lack of Clarity About What Is Being Measured

Only five of the twenty texts (29%) reviewed appeared to include at least some discussion of the uncertainty about what the IAT actually measures. In one particularly detailed discussion, authors pointed out that there is controversy over this question, that it is possible that familiarity with stereotypes drives the associations, and that the test does not correlate well with explicit measures of prejudice. The remaining four were not as detailed, and 13 contained no acknowledgement at all of the unclarity problem.

African American Scores

Only two textbooks (12%) mentioned automatic white preferences among African Americans taking the Race IAT. While textbooks obviously note the prevalence of the IAT effect among White participants, African American scores speak to the notion that the IAT measures outgroup hostility as opposed to, for example, awareness of cultural stereotypes. The two texts that did mention the scores of African Americans acknowledged that this finding is problematic for the interpretation of scores proffered by IAT proponents. The omission of this important finding is troubling but not surprising given the lack of discussion of confusion over what the IAT measures. Interestingly, one textbook seemed to be suggested that African Americans tend to show the opposite pattern typical of Caucasians, yet failed to quantify the prevalence of such a pattern.

Predictive Validity

Only two of the seventeen textbooks (12%) mentioned the questionable record of the IAT in terms of predictive validity. These two texts fully present the notion that the IAT is controversial and note that one reason for this controversy is its poor record of predicting behavior. Interestingly, more textbooks suggest that the IAT *does* predict behavior, than do acknowledge the controversy surrounding the predictive capability of the IAT. One author notes that IAT scores can influence many types of behavior while another notes that while it doesn't predict overt discrimination, it measures discrimination (presumably implicit).

Link to Project Implicit and Caution Regarding Interpretation of Scores

Eleven of 17 textbooks (65%) included a link to the project implicit website, an invitation to try the IAT. Yet, not one of the eleven textbooks including the link offered any explicit warning or note of caution about interpreting the score. Notably, two of the three books that were categorized as providing "good" coverage (discussed below) *did not* include a link. Perhaps awareness of criticisms of the IAT created concerns among authors about promoting assessment among students. Ironically, one text with coverage rated as good, that did not include a link, provided a fairly clear statement as to why they did provide the link. The author notes that there is insufficient support for the notion that the IAT can detect racial bias or predict bias.

It should be noted that this does leave one textbook as having been evaluated as good, yet not providing an explicit warning about interpreting scores. One might consider extensive coverage of criticisms a sufficient warning or caution. Though not an unreasonable conclusion, we were looking for more direct warning similar to those provided by the *Blindspot* authors.

Bias in Textbooks

Of the 17 textbooks that included coverage of the IAT, only 3 (18%) were evaluated as "good" or provided fair and accurate coverage of the controversies surrounding the IAT. Seven of the textbooks (41%) were evaluated as providing partially biased coverage. Texts in this category contained acknowledgement of controversy, but the controversies were not explored in detail resulting in what was effectively a one-sided presentation. One textbook well illustrated the generosity with which we evaluated bias. The author did acknowledge that scholars have "raised concerns," yet provided relatively extensive coverage of the IAT absent any elaboration of these "concerns," concluding that the IAT measures "implicit racial prejudice." The seven remaining textbooks (41%) were evaluated as providing biased coverage of the IAT. Thus, they did not acknowledge the substantial criticism surrounding the IAT and instead presented the measure as if it were an undisputed measure of implicit prejudice.

Discussion

The results of the present textbook analysis are troubling. Fundamental questions concerning the IAT include the extent to which reaction time differences represent attitudes that are distinct (from explicit attitudes), stable and meaningful (in terms of predicting behavior). These issues, by and large, are not sufficiently addressed in Introductory Psychology textbooks. Only five of the seventeen textbooks we evaluated acknowledged the lack of clarity in terms of what the IAT measures. Though this is not a recently-raised concern regarding the IAT, the most significant

work regarding such questions, from our perspective, was likely not available for review for most, if not all, of the textbook authors included in our analysis (i.e., published from 2018–2021). Schimmack (2021) reanalyzed data from several meta-analyses finding that roughly 20% of the variance in racial attitudes was explained by the Race IAT, there was little distinction between the performance of implicit and explicit attitudes (i.e., poor discriminant validity), and, given the latter, poor incremental validity for the IAT over explicit attitudes. Thus, previous studies reporting a weak or nonsignificant correlation between implicit (as measured by the IAT) and explicit attitudes may have been interpreted as evidence of discriminant validity, but Schimmack demonstrated that the low correlations are better explained by measurement error. Thus, the evidence from Schimmack suggests that the race IAT isn't actually measuring unconscious attitudes but is rather an implicit tool to measure racial attitudes, albeit an unreliable tool with poor validity. Schimmack explains:

Researchers should avoid terms such as implicit attitude or implicit preferences, that make claims about constructs simply because attitudes were measured with an implicit measure. Even Greenwald and Banaji (2017) are trying to avoid writing about implicit constructs, suggesting that it is time to put the ideas that the IAT measured unconscious processes and that individuals harbor hidden implicit attitudes to rest. (p. 409)

Even fewer textbooks, two of seventeen, noted African American scores on the IAT, that is studies reporting an automatic White preference among African Americans (e.g., Nosek et al., 2002). Such results obviously raise fundamental questions about what IAT scores represent. Even if an author wishes to avoid discussing the implications for test validity, this finding is of practical importance, given the frequency with which authors included a link to the IAT. African American students giving the IAT a try might be troubled by results suggesting an automatic White preference and, thus, find solace in the finding that this is neither uncommon nor pathological.

Conspicuously absent from most of the Introductory Psychology textbooks in the current sample was any mention of the questionable predictive validity of the IAT. Most textbooks make reference to the near absence of overt discrimination in contemporary American society, but like the creators of the IAT, suggest that prejudice, now implicit, remains an issue. Few textbooks offer a counter to this claim, yet the results of the Oswald et al. (2013) meta-analysis cast doubt on it. Oswald and colleagues note:

“The present results call for a substantial reconsideration of implicit-bias-based theories of discrimination at the level of operationalization and measurement, at least to the extent those theories depend on IAT research for proof of the prevalence of implicit prejudices and assume that the prejudices measured by IATs are potential drivers of behavior” (pp. 183–184). They go on, “The tremendous heterogeneity observed within and across criterion categories [e.g., microbehaviors like body posture to interpersonal behavior like expressed preferences in an intergroup interaction] indicates that how implicit biases translate into behavior – if they do at all – appear to be complex and hard to predict” (p.185).

Undergraduate students are asked to process and recall a tremendous amount of information through reading, and some report that ‘active learning’ opportunities and demonstrations included in the text are helpful in accomplishing these tasks. Reading a section on assessment, the reader may be inclined to interpret an IAT score as a measure of prejudice. The IAT’s designers themselves explicitly caution against this: “This is why we answer no to the question “Does automatic White preference mean ‘prejudice’?”” (Banaji & Greenwald, 2013, p. 52). The question as an

Table 1. Cautions about IAT score interpretation from Schimmack (2021).

“The low construct validity of IATs raises some concerns about the use of the IAT for the assessment of individual attitudes” (p. 411).
“Over a million visitors of the Project Implicit website have received feedback about their IAT scores with the interpretation that they harbor some potentially unconscious racial biases against African Americans...Given the modest validity of the race IAT and the lack of evidence for discriminant validity, it seems problematic that respondents are not informed about the possibility that their test scores are likely to be invalid.” (p. 411)
“Given the poor validity, Schimmack demonstrates that a potential average score on the IAT could represent an ‘extreme pro-White bias,’ but also an extreme ‘pro-Black bias’” (p. 411).
“The IAT is an implicit measure of attitudes with varying validity. It is not a window into people’s unconscious feelings, cognitions, or attitudes” (p. 412).

Introductory Psychology student might understand it is placed into even sharper focus by the concerns over poor predictive validity described by Oswald et al. (2015).

Given these cautions, it would seem particularly important that, when inviting readers (via a link) to complete the IAT, as a majority of these textbooks did (65%), a similar warning about interpretation of the scores be included in summary discussion. It is reasonable to suggest that there are a number of Introductory Psychology students who think they have recently learned that they are in fact racially prejudiced when this is far from clear. We would argue that if textbook authors are going to continue to encourage students to take the IAT, they consider including some version of the cautions noted in Table 1.

Though we do not want to belabor the point, we feel there are real consequences for the negligence on the part of some authors. Singal (2017) notes that one could reasonably argue that, “Harvard should not be administering the test in its current form, in light of its shortcomings and its potential to mislead people about their own biases” (para. 8). Perhaps textbook authors might also consider the following before encouraging students to take the IAT:

Over and over and over and over, the IAT, a test whose results don’t really mean *anything* for an individual test-taker, has induced strong emotional responses from people who are told that it is measuring something deep and important in them. This is exactly what the norms of psychology are supposed to protect test subjects against. (Singal, 2017, para. 71)

It is, in our estimation, unacceptable to provide a link to a test whose results the creators acknowledge can be very upsetting, and then provide not only no warning about this but no mention of criticism to provide context and nuance to the potentially jarring results. One might forgive a White student for exhibiting substantial fragility in the face of such undue chastisement. In this regard, textbook authors should do better.

There is little doubt that the IAT has garnered profound support and attention due to a number of factors beyond chaste scientific curiosity. Diangelo (2018), in *White Fragility*, while perpetuating misunderstanding about implicit bias (e.g., “Whites are unconsciously invested in racism”), provides an illustrative example of politicized science:

The tension between the noble ideology of equality and the cruel reality of genocide, enslavement, and colonization had to be reconciled. Thomas Jefferson (who himself owned hundreds of enslaved people) and others turned to science. Jefferson suggested that there were natural differences between the races

and asked scientists to find them. If science could prove that black people were naturally and inherently inferior...there would be no contradiction between our professed ideals and our actual practices. There were, of course, enormous economic interests in justifying enslavement and colonization. Race science was driven by these social and economic interests, which came to establish cultural norms and legal rulings that legitimized racism and the privileged status of those defined as white. (p. 16)

Diangelo's account is instructive here. To what extent is implicit race science used in the service of explaining the contradiction between racial disparities in outcomes (e.g., housing and income) in spite of the near absence of explicit racial prejudice? To what extent has implicit race science been driven by social and economic interests? As Singal (2017) notes, in an article critical of the IAT, there is a case to be made "that the IAT went viral not for solid scientific reasons, but simply because it tells such a simple, pat story about how racism works and can be fixed" (para. 8). Singal also notes that revealing one's first IAT scores has become a custom, a practice that allows one to signal to others that you are doing your part to combat racism by, in part, acknowledging your role in it. Mitchell and Tetlock (2006) note that IAT proponents have called for "greater use of affirmative action programs and other 'fair measures' to counter and change implicit biases against women and minorities" (pp. 1026–1027). Moreover, Mitchell and Tetlock (2017) note, "Greenwald and his colleagues have sought to influence how courts and juries think about prejudice and discrimination" (p. 169).

The absence of IAT criticism in textbooks is difficult to explain. Across numerous controversial research topics (e.g., media violence) in introductory textbooks, Ferguson et al. (2018) found biased coverage to be the rule not the exception. In all but one case, the authors found less than 20% of textbooks on any given issue presented it in an unbiased fashion with several issues under 10%. For example, while almost all of the textbooks reviewed discussed antidepressants, only 12.5% provided an unbiased presentation. Similarly, Bartels (2019) found a lack of coverage of issues concerning antidepressant efficacy in abnormal psychology textbooks. Bartels speculated that authors may be reluctant to address criticisms given the potential impact that this knowledge may have on students dealing with their own depression. Yet, as we have argued, the concern with IAT coverage is quite the opposite. Specifically, it is a lack of information that may lead to confusion and misunderstanding about one's implicit biases. Thus, we are forced to consider an explanation advanced by Ferguson et al. (2018) in light of the textbook bias they encountered:

Given that the misinformation contained generally hewed toward presenting contested research as more consistent, generalizable to socially relevant phenomena and higher quality than it was, we believe that these errors are consistent with an indoctrination, however, unintentional, into certain beliefs or hypotheses that may be 'dear' to a socio-politically homogenous psychological community. (p. 579)

Expressing concern about the politicization of the IAT, Singal (2017) notes: "If it is politically palatable to embrace the IAT and the nationwide search for our inner bias, then to criticize the test is to be on the wrong side of the progressive conversation about race" (para. 79). According to Singal (2017), Banaji, co-creator of the IAT, goes further than suggesting that one not embracing the IAT is on the wrong side of the conversation, noting their mental health status may be in question: "questioning the IAT is not something normal, well-adjusted people do" (para. 80).

Authors must resist such threats and acknowledge their blind spots for an IAT narrative that falls apart under close scrutiny. Ferguson et al. (2018) note that when students can easily access information on the internet that contradicts the textbook narrative, the credibility of psychological science suffers. We should promote a critical analysis of controversial topics among students of psychological science. The laudable goal of reducing prejudice and discrimination does not

justify blind support for poor science. Critical thinking demands allegiance to the science and blind spots to political agendas, ideology, and popular conceptualizations of “social justice”.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Jared M. Bartels  <https://orcid.org/0000-0003-2309-7178>

References

*Textbook examined in the present study

- Andreychik, M. R. & Gill, M. J. (2012). Do negative implicit associations indicate negative attitudes? Social explanations moderate whether ostensible “negative” associations are prejudice-based or empathy-based. *Journal of Experimental Social Psychology, 48*(5), 1082–1093. <https://doi.org/10.1016/j.jesp.2012.05.006>
- Arkes, H. R. & Tetlock (2004). Attributions of implicit prejudice or “would Jesse Jackson ‘fail’ the implicit association test?”. *Psychological Inquiry, 15*(4), 257–278. http://doi.org/10.1207/s15327965pli1504_01
- Banaji, M. R. & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. Bantam.
- Bartels, J. M. (2019). Burying the lede: The questionable efficacy of antidepressants in abnormal psychology textbooks. *Scholarship of Teaching and Learning in Psychology, 5*(1), 90–95. <http://dx.doi.org/10.1037/stl10000137>
- Blanton, H. & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology, 34*(1), 277–297. <http://doi.org/10.1146/annurev.soc.33.040406.131632>
- *Bordens, K. S., Gerow, J. R. & Gerow, N. K. (2020). *Essentials of psychology* (4th). Academic Media Solutions.
- Brendl, C. M., Markman, A. B. & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the implicit association test. *Journal of Personality and Social Psychology, 81*(5), 760–773. <http://doi.org/10.1037//0022-3514.81.5.760>
- *Cacioppo, J. T., Freberg, L. A. & Cacioppo, S. (2022). *Discovering psychology: The science of mind* (4th). Cengage.
- *Ciccarelli, S. K. & White, J. N. (2020). *Psychology* (6th). Pearson.
- *Coon, D., Mitterer, J. O. & Martini, T. S. (2019). *Introduction to psychology: Gateways to mind and behavior* (15th). Cengage.
- Correll, J., Park, B., Judd, C. M. & Wittenbrink, B. (2002). The police officer’s Dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 82*(6), 1314–1329. <https://doi.org/10.1037/0022-3514.82.6.1314>
- Diangelo, R. J. (2018). *White fragility: Why it’s So hard for white people to talk about racism*. Beacon Press.
- *Feist, G. J. & Rosenberg, E. L. (2019). *Psychology: Perspectives and connections* (4th). McGraw-Hill.
- *Feldman, R. (2019). *Understanding psychology* (14th). McGraw-Hill.
- Ferguson, C. J., Brown, J. M. & Torres, A. V. (2018). Education or indoctrination? The accuracy of introductory psychology textbooks in covering controversial topics and urban legends about psychology. *Current Psychology, 47*(3), 574–582. <https://doi.org/10.1007/s12144-016-9539-7>
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G. & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*(3), 522–559. <https://doi.org/10.1037/pspa0000160>
- *Gazzaniga, M. (2018). *Psychological science* (6th). Norton.

- *Gray, P. O. & Bjorklund, D. F. (2018). *Psychology* (8th). Worth.
- Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, 72, 861–871. <https://doi.org/10.1037/amp0000238>
- Greenwald, A. G., Brendl, M., Cai, H., Charlesworth, T., Cvencek, D., Dovidio, J. F. & Wiers, R. W. (2019, March 26). *The implicit association test at age 20: What is known and what is not known about implicit bias*. faculty.washington.edu. <http://faculty.washington.edu/agg/pdf/IAT%20at%20age%2020.with%20title%20page.26Mar2019.pdf>
- Greenwald, A. G. & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94(4), 945. <https://doi.org/10.2307/20439056>
- 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(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <http://doi.org/10.1037/a0015575>
- *Griggs, R. A. & Jackson, S. L. (2020). *Psychology: A concise introduction* (6th). Worth.
- Griggs, R. A. & Whitehead, G. (2014). Coverage of the Stanford prison experiment in introductory social psychology textbooks. *Teaching of Psychology*, 41(4), 318–324. <http://doi.org/10.1177/0098628314549703>
- Gross, J. J., Schmader, T., Martin Head, B. & Anderson, A. K. (2019). *Interactive psychology: People in perspective*. Norton.
- *Hockenbury, S. E. & Nolan, S. A. (2018). *Psychology* (8th). Worth.
- *Huffman, K., Dowdell, K. & Sanderson, C. A. (2018). *Psychology in action* (12th). Wiley.
- *King, L. (2019). *Experience psychology* (4th). McGraw-Hill.
- *Kosslyn, S. M. & Rosenberg, R. S. (2020). *Introducing psychology: Brain, person, group* (5.1). FlatWorld.
- Lilienfeld, S. O., Lynn, S. J. & Namy, L. L. (2018). *Psychology: From inquiry to understanding* (4th). Pearson.
- McConnell, A. R. & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5), 435–442. <https://doi.org/10.1006/jesp.2000.1470>
- Miller, D. (2015, July 13). *Tech companies spend big money on bias training – but it hasn't improved diversity numbers*. The Conversation. <https://theconversation.com/tech-companies-spend-big-money-on-bias-training-but-it-hasnt-improved-diversity-numbers-44411>.
- Mitchell, G. & Tetlock, P. E. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal*, 67, 1023–1121.
- Mitchell, G. & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 164–195). Wiley.
- *Morris, C. G. & Maisto, A. A. (2019). *Understanding psychology* (12th). Pearson.
- *Myers, D. G. & DeWall, C. N. (2018). *Psychology* (12th). Worth.
- *Nevid, J. S. (2022). *Essentials of psychology: Concepts and applications* (6th). Cengage.
- Nosek, B. A., Banaji, M. R. & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, 6(1), 101–115. <https://doi.org/10.1037/1089-2699.6.1.101>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171–192. <https://doi.org/10.1037/a0032734>
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, 108(4), 562–571. <https://doi.org/10.1037/pspa0000023>
- Payne, B. K., Vuletic, H. A. & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systematic prejudice. *Psychological Inquiry*, 28(4), 233–248. <https://doi.org/10.1080/1047840x.2017.1335568>

- *Schacter, D., Gilbert, D., Nock, M. & Wegner, D. (2020). *Psychology* (5th). Worth.
- Schimmack, U. (2021). The implicit association test: A method in search of a construct. *Perspectives in Psychological Science*, 16(2), 396–414. <https://doi.org/10.1177/1745691619863798>
- Schwarz, J. (1998, September 29). *Roots of unconscious prejudice affect 90 to 95 percent of people, psychologists demonstrate at press conference*. UW News. <http://www.washington.edu/news/1998/09/29/roots-of-unconscious-prejudice-affect-90-to-95-percent-of-people-psychologists-demonstrate-at-press-conference/>.
- Shelton, J. N., Richeson, J. A., Salvatore, J. & Trawalter, S. (2005). Ironic effects of racial bias during inter-racial interactions. *Psychological Science*, 16(5), 397–402. <https://doi.org/10.1111/j.0956-7976.2005.01547.x>
- Singal, J. (2017, January 11). *Psychology's favorite tool for measuring racism isn't up to the job*. The Cut. <https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html>.
- Sleek, S. (2018, January 31). *The bias beneath: Two decades of measuring implicit associations*. Association for Psychological Science – APS. <https://www.psychologicalscience.org/observer/the-bias-beneath-two-decades-of-measuring-implicit-associations>.
- *Wade, C., Tavis, C., Sommers, S. R. & Shin, L. M. (2018). *Invitation to psychology* (7th). Pearson.

Author Biographies

Jared M. Bartels is an Associate Professor of Psychological Science at William Jewell College where he teaches Developmental Psychology, Abnormal Psychology, and Introduction to Psychology. His research interests include achievement motivation, self-regulated learning, and the history of psychology.

Patricia Schoenrade is a Professor of Psychological Science at William Jewell College where she teaches courses in Experimental Psychology, Social Psychology, and Introduction to Psychology.