'I'm happy to own my implicit biases': Public encounters with the implicit association test

Jeffery Yen[1]* (iD), Kevin Durrheim[2] (iD) and Romin W. Tafarodi[3] (iD)

[1]University of Guelph, Ontario, Canada
[2]University of Kwazulu-Natal, Pietermaritzburg, South Africa
[3]University of Toronto, Ontario, Canada

The implicit association test (IAT) and concept of implicit bias have significantly influenced the scientific, institutional, and public discourse on racial prejudice. In spite of this, there has been little investigation of how ordinary people make sense of the IAT and the bias it claims to measure. This article examines the public understanding of this research through a discourse analysis of reactions to the IAT and implicit bias in the news media. It demonstrates the ways in which readers interpreted, related to, and negotiated the claims of IAT science in relation to socially shared and historically embedded concerns and identities. IAT science was discredited in accounts that evoked discourses about the marginality of academic preoccupations, and helped to position test-takers as targets of an oppressive political correctness and psychologists as liberally biased. Alternatively, the IAT was understood to have revealed widely and deeply held biases towards racialized others, eliciting accounts that took the form of psychomoral confessionals. Such admissions of bias helped to constitute moral identities for readers that were firmly positioned against racial bias. Our findings are discussed in terms of their implications for using the IAT in prejudice reduction interventions, and communicating to the public about implicit bias.

In the two decades since the notion of 'implicit bias' first emerged in social psychology, it has spawned a prolific body of research utilizing the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998). Although the paradigm is beset by psychometric and conceptual challenges (Arkes & Tetlock, 2004; Blanton & Jaccard, 2006, 2015; Durrheim, 2012), and its validity and meaning continue to be controversial (Blanton, Jaccard, Strauts, Mitchell, & Tetlock, 2015; Fiedler, Messner, & Bluemke, 2006; Gawronski, LeBel, & Peters, 2007), it has significantly influenced the ways in which social psychologists conceive of and study the problem of prejudice. Indeed, while implicit bias can be conceptualized and measured in a variety of ways, it has in many respects become conflated with the IAT. Implicit bias research has also impacted social and public life outside of the academy, inspiring shifts in prejudice reduction strategies and policies. In the United States, in particular, great efforts are being made to incorporate IAT research into state law, workplace practices, health care, law enforcement, and the writing of social policy (Edwards, 2016; Gove, 2011;

*Correspondence should be addressed to Jeffery Yen, Department of Psychology, University of Guelph, Guelph, ON, Canada, N1G 2W1 (email: jyen@uoguelph.ca).

Zestcott, Blair, & Stone, 2016). Implicit bias became a focal point of public discourse about race in the months surrounding the 2008 and 2012 US presidential elections (Dooley, 2015; Gladwell, 2005; Herbert, 2012; Rachlinski & Parks, 2008). More recently, amid a spate of high profile incidents of excessive use of force by police against Black men, and extended discussion of implicit bias in the first US presidential debate of 2016, the IAT and implicit bias appear again to have captured the American public imagination. Even without the interest generated by these events, the IAT has benefited from steadily increasing media coverage and social media discussion around the world. Rough estimates of English-language publications in scholarly journals or books, and newspapers and magazines containing the keywords 'implicit association test' suggest rapid growth in both research activity and popular coverage of the IAT (see Table 1).

The influence of the paradigm is also attributable in part to the IAT's availability, since 1998, for public use through a 'demonstration website' (www.implicit.harvard.edu), which gives test-takers individual test results (a practice which has also been controversial; Nosek, Banaji, & Greenwald, 2002). By early 2012, the test had been taken by people from around the world over 13 million times (Nosek & Riskind, 2012). Given this participation and interest, it is crucial for social psychologists to understand the ways in which members of the public have made sense of this research and the IAT, particularly as they relate to conceptions of the nature of prejudice. Many implicit bias researchers – and the demonstration website itself – refrain from evaluative language and are careful to distinguish their construct from prejudice and discriminatory behaviour *per se.* This is explicit, and it is done to clearly keep all semblance of intention to harm away from the phenomenon they work with (e.g., Spencer, Charbonneau, & Glaser, 2016). Although there are concerns that the public is conflating implicit bias with prejudice (Bazelon, 2016), interpretations of the meaning of a non-negative IAT result show wide variation among researchers themselves. For some, the IAT only measures automatic associations (e.g., Rooth, 2010), or preferences (e.g., Howell & Ratliff, 2017) that might predict things like discriminatory behaviour, while others have used it as a direct measure of implicit stereotypes and prejudice (e.g., Rudman, Ashmore, & Gary, 2001). Rather than view this variation as evidence of widespread misinterpretation of the IAT and implicit bias, we regard this as part of ongoing historical contestation over the nature and meaning of prejudice as a social, psychological and moral problem (Cherry, 2000; Goodman & Rowe, 2014; Wetherell, 2012).

Table 1. Academic works and news articles on the implicit association test between 1998 and 2016

Years	Academic works	Online news articles
1998–2000	64	7
2001–2002	208	7
2003–2004	519	8
2005–2006	847	22
2007–2008	1,500	52
2009–2010	1,880	34
2011–2012	2,985	24
2013–2014	3,400	72
2015–2016 (May)	2,380	70
Total	13,783	296

In this article, we report a qualitative study of public engagement with the IAT and implicit bias in science reporting in mainstream news media. The news media reflect and strongly influence popular understandings of both science and prejudice (Clark & Illman, 2006; Simmons & Lecouteur, 2008). Through a critical discourse analysis of readers' responses to the IAT and implicit bias, we identify common tropes for responding to the IAT and its claims, and the ways in which these concepts figure in lay people's debates about the nature of bias and prejudice. We also examine instances of reflexive engagement with the test and their implications for the use of the IAT in prejudice reduction strategies.

Previous work on lay or folk conceptions of prejudice has been limited mainly to looking at the extent to which they deviate from, or fall short of, our social scientific theory and knowledge. This fits with the prevailing tendency to view people on the whole as possessing a deficient understanding of their own biases and therefore standing in need of corrective education. As Figgou and Condor (2006, p. 219) argue, 'academic social psychologists have tended to treat ordinary social actors as unreflexive formulators of stereotyped views, bearers of prejudiced attitudes, or agents of discriminatory behaviour'. Accordingly, it is often assumed that widespread dissemination of knowledge about the psychological processes underlying prejudice will necessarily contribute to its reduction. Education and 'debiasing' (Lilienfeld, Ammirati, & Landfield, 2009) informed by research will provide ordinary people with greater self-understanding and improve their 'cognitive sophistication' (Duckitt, 2001) on the subject. On this view, the increasing prominence of 'implicit bias' in public discourse is evidence of the successful uptake of the science of prejudice by the general population (Kaslow, 2015). The high media exposure of the IAT has not only increased awareness of implicit bias, but also raised questions about the limits of self-awareness (Nosek & Riskind, 2012) and inspired self-monitoring of unwanted thoughts and efforts to correct them (Monteith, Mark, & Ashburn-Nardo, 2010). Consistent with this perspective, the few studies focusing explicitly on people's reactions to the IAT have done so in the service of reducing prejudice through raising awareness of the pervasiveness of implicit bias (Morris & Ashburn-Nardo, 2009).

We argue that the work that treats lay people as the bearers of prejudice fails to appreciate how ordinary people use and contest scientific explanations and technologies. In their everyday lives, people use and evaluate scientific claims about prejudice as they express their values and identities and manage their relationships and conduct (Wynne, 2006). In this regard, they are not dissimilar from their academic counterparts whose views about prejudice are 'essentially contested' (see Dixon, 2017). Lay conceptions of prejudice and racism may even be well-developed and sophisticated (Figgou & Condor, 2006). In this article, therefore, we move beyond the approach that seeks to use the IAT to educate lay people about their prejudices. Instead, we highlight the more complex interpretive reactions of lay publics as they make sense of the IAT and its findings.

Public discourse and the social psychology of prejudice

Historical studies have shown that academic psychological understandings of prejudice have been shaped by, and have helped to shape, discourses about prejudice circulating in wider society. Theorizing about prejudice has been informed throughout by the pressing social issues of the day (Durrheim, 2014). Post-war research on prejudice in social psychology sought to address the scourge of fascism (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), the problem of racial segregation (Allport, 1979), and later, public opposition to liberal egalitarian reforms, such as affirmative action (Kinder & Sears,

1981). This work often positioned prejudice as a product of irrational personality dynamics or cognitive distortion, and thus prejudice reduction efforts have focused on eliminating cognitive error and facilitating a more objective appreciation of others (e.g., Luguri, Napier, & Dovidio, 2012). This view of prejudice as unreason can be traced back to the moral ideals of Enlightenment liberalism (Billig, 1988).

Prejudiced discourse itself often reflects prevailing social commitments about the nature and acceptability of prejudicial sentiment. In liberal contexts, a general 'taboo on prejudice' holds sway (Billig, 1988). Thus, when people express negative sentiment about racialized others, they are careful to present their views as reasonable, grounded in reality, or supported by evidence (Augoustinos & Every, 2007; Edwards, 2003; Van Dijk, 2000; Wetherell & Potter, 1993) – sometimes deriving from psychological theory itself (Lea, 1996). By means of disclaimers, mitigations, counter-accusations and denials, people can use notions of prejudice as irrational bigotry to defend racial thinking and practices as being *not prejudiced*, while identifying *others* as the real bigots deserving of sanction (Goodman & Rowe, 2014). For example, opponents of asylum in Australia and Europe explicitly link charges of racism with censorship, thus casting supporters of asylum as intolerant of free speech and 'playing the race card' (Burke & Goodman, 2012). Anti-refugee arguments are thus promoted by positioning (liberal) opponents as 'pernicious, oppressive, discriminatory or just plain crazy (the 'loony left')' (Augoustinos & Every, 2010, p. 253). In response, supporters of asylum orient to this positioning by painstakingly avoiding making accusations of racism (Goodman, 2010).

Discursive research of this kind illustrates how struggles over the nature of prejudice can be themselves forms of social action:

> The struggle to define the nature of prejudice is best conceived not as the work of a detached expert who must determine its true underlying nature, but as the objective of people who define prejudice to explain events, persuade and mobilize others, and act in creditable ways. Such definitions of prejudice are resources for identification and action.(Durrheim, Quayle, & Dixon, 2016, p. 19)

Concepts of prejudice and bias thus provide moral framing for recognizing oneself and others in a world of racial inequality; they help to explain what is wrong with the world and what should be done to fix it; and they are useful resources for political mobilization – as was evident in campaigning for Brexit (Durrheim *et al.*, in press) and for the Trump presidency (Engber, 2016). In this work, we treat the deployment of concepts of bias, prejudice, and racism in lay discourse about the IAT as identity performances (Durrheim *et al.*, 2016) and focus analytic attention on the way in which they are constructed and used as interpretative resources for identification and explanation.

The IAT as research instrument and intervention tool

Both the IAT and the concept of implicit bias emerged as the nature of prejudice was being re-examined in social psychology. Theorists argued that in liberal contexts, prejudice was not in decline but instead assuming covert 'modern' forms (e.g., subtle, aversive, benevolent, or symbolic). The figure of the hostile, irrational bigot was giving way to that of the liberal, rational citizen, adept at self-presentation, perhaps internally conflicted, or unaware of their own prejudices. This reformulation intensified the long-standing and now widespread suspicion of self-report measures in social psychology, based on two assumptions: that people have limited insight into their 'true' attitudes, and that self-

report measures are biased by 'social desirability'. The ostensive appeal of implicit measures such as the IAT thus lay in claims to be able to circumvent people's (self-presentational) discourse to expose prejudiced propensities, and it is now widely assumed that they measure relatively stable, 'introspectively inaccessible' attitudes (Banaji & Greenwald, 1994) in a way that is mostly resistant to social desirability biases. Although the validity of these assumptions appears to be equivocal at best, they continue to underlie much research utilizing these measures (Blair, 2002; Calanchini, Sherman, Klauer, & Lai, 2014; Gawronski *et al.*, 2007).

Nevertheless, some researchers have begun to examine people's experiences with the IAT, but this work has focused primarily on how subjects' responsivity poses a challenge to the validity of the IAT as a measurement tool. For example, research by De Houwer, Beckers, and Moors (2007) suggests that test-takers are able to 'fake' results on the IAT using certain test strategies. Studies focused on the reduction of implicit biases have, furthermore, demonstrated their 'malleability' in response to short-term experimental manipulations (Blair, 2002), although there is evidence to suggest that such changes are temporary (Lai *et al.*, 2016). In contrast, a handful of studies have examined the IAT's potential as a tool for *prejudice reduction* (e.g., Devine, Forscher, Austin, & Cox, 2012). Those who have done so argue that the IAT can facilitate prejudice reduction by making people aware of their biases – in particular, through an experience of automaticity – and motivate them to actively control their responses (Greenwald, Nosek, & Banaji, 2003). Although this experience is thought to be a powerful motivator for change (Morris & Ashburn-Nardo, 2009), others have suggested that awareness of bias, at least as provoked by the IAT, may in fact lead people to adopt a 'prevention focus' – to become overly inhibited, cautious, and less efficacious in their interactions and thus appear to others to be more prejudiced than they intend (Vorauer, 2012). Whereas Vorauer (2012) explores how cognitive and self-protective processes might lie behind these reactions, in this article we show how reactivity to the IAT can be understood in relation to prevailing, historically contingent, discourses about psychology, and the nature and meaning of prejudice.

Methods

The data in this article come from a larger project that examined the social construction of prejudice in lay people's discussions about, and the experience of taking, the IAT (Yen, 2013). We chose to analyse coverage of the IAT in *The New York Times* (hereafter, NYT) that occurred around the time of the 2008 US presidential elections, in which 'race' was foregrounded in the public consciousness by Barack Obama's candidacy. This period, roughly between 2008 and 2010, saw a relative spike in media attention and sustained discussion of prejudice and IAT research. Political commentators wondered whether, in a repeat of the 'Bradley effect', Obama's firm lead in pre-election polls would fail to translate into a win at the final ballot (Altman, 2008; Greenwald, Smith, Sriram, Bar-Anan, & Nosek, 2009) because more than 75% of Americans were shown to harbour implicit racial bias. Once Obama was elected, however, the concept of implicit bias and the 'real-world' validity of the IAT were called into question in public discourse. News coverage of the IAT has since increased, but this was the first high profile public discussion of the test (see Table 1). In total, eleven articles on the IAT appeared in the NYT during that period, but only seven allowed reader comments. All seven were selected for our study, comprised of 793 reader comments and 123,254 words (see Table 2).

Table 2. New York Times articles and comments selected for analysis, with word counts

Date published	Article title	Type	Reader comments	Words	Abbreviation
29/10/2008	What? Me biased?	Op-Ed	213	26,842	WMB
07/11/2008	Where have all the bigots gone?	Science	104	12,814	WBG
17/11/2008	A shocking test of bias	Science	72	1,913	STB
18/11/2008	Further reading on unconscious bias	Science	9	14,029	FRB
19/11/2008	How do you measure bias?	Science	19	2,263	HMB
21/02/2009	A nation of cowards	Op-Ed	340	61,441	NOC
01/06/2010	The visible hand	Science	36	3,952	TVH
Totals			793	123,254	

Several other factors motivated our choice of the NYT: As a 'newspaper of record', it has the third largest circulation in the United States and is often chosen for media analysis because of its considerable influence on public opinion (Freeman, 2010; Ismail & Mishra, 2009) and public science knowledge (Marshall, 1998). Finally, it is the third most visited online news service after Reddit.com (first) and CNN.com (second) (Alexa Internet Analytics, n.d.). The NYT is, however, perceived by many to be a liberal-leaning newspaper (Prior, 2013), and readers may, but not necessarily, hold similar political views. Readers of NYTimes.com who post comments are more likely, on average, to be wealthier, more educated, and more socially and politically engaged (Chung, 2008; New York Times, 2009). Despite its US base, 25.5% of its online readers are located outside of the country (Alexa Internet Analytics, n.d.). No information about the racial identification of its readers was available, although some readers voluntarily disclosed this information in their comments.

Analytical approach

Our analysis of reader comments proceeded according to two interpretive strategies. First, the entire corpus was read and reread to derive the themes and narratives, which referred to IAT research and implicit bias. The relative importance of themes was determined not from their frequency, but from the intensity and quality of discussion within them (Jørgensen & Phillips, 2002). On the basis of this first reading, we focused our discursive analysis on a smaller subset of data that referred directly to implicit bias or to experiences of taking the IAT. Roughly 60% (475 out of 793) of the reader comments, or parts of comments, met these criteria.

Subsequently, a more fine-grained critical discourse analysis was undertaken in which we examined these texts to identify how implicit bias was problematized in different accounts; that is, whose problem it was said to be (e.g., the general public), where it was 'located' (e.g., in society, in people's brains, 'deep down'), and what should be done about it. Our analysis was oriented less to the comments' local, interactional significance, and more to the broader social and institutional discourses that made these utterances 'possible' (Wetherell, 1998). We were especially interested in how these accounts were both intelligible and persuasive because of their evocation of a broader historical, interpretive, or discursive context (Laclau, 1993). The micro-social dynamics of naturalistic discussion, and analysis of this interactional context, was therefore not our main focus.

Reader comments comprised some combination of their responses to the news articles, interpretation of implicit bias, experiences of taking the IAT, and/or other readers' comments. Even so, individual comments tended to be self-contained statements or questions that rarely referenced the specific comments of other readers. We have, however, noted details of this context where possible and relevant to specific analytical arguments. We also provide illustrative excerpts of reader comments, identified by the specific article to which they responded (see Table 2), as well as a number indicating their order of appearance. Where online names were given, these are also included with their excerpts.

Analysis

We begin by locating our analysis with reference to features of the news articles which occasioned the reader comments. The NYT's coverage of the IAT was targeted at a non-specialist audience in two types of writing: In op-eds, the IAT and its findings were employed in polemics about current affairs, while in science columns, the IAT and prejudice were more critically discussed in less emotive language. Thematically, the op-ed pieces were concerned with claims of widespread prejudice in the United States. They emphasized contradictions between the values of equality and fairness in the nation's self-image and evidence of widespread implicit racial bias as revealed by the IAT, and they used explicitly psychological language regarding human brains, cognition, and the need for an 'honest' and 'brave' national dialogue on race. The op-eds used provocative headlines (see Table 2) to frame calls to learn from 'the facts' (as revealed by the IAT), and 'have an adult conversation' about 'unconscious bias'. In these pieces, IAT science was presented as uncontroversial and self-evidently true, but the articles anticipated reader scepticism by referring people to the demonstration website where they 'can test for your own unconscious biases' (Kristof, 2008). By contrast, the science articles introduced dissent and questioned the usefulness of IAT science in relation to the looming election, referring to critiques from within psychology. They explicitly questioned its knowledge claims and, like the op-eds, provided links to the IAT website. One notably different piece linked readers to academic journal articles and invited them to make up their own minds (Tierney, 2008).

Overview of analysis

Reader comments revolved primarily around the credibility of IAT science, and accordingly, our analysis distinguishes between comments that accepted its claims as true from those that dismissed or deemed them to be questionable in some way. Each had different implications for how the IAT and implicit bias figured in discussion about who was or was not biased, what the nature and causes of prejudice were, and what should be done about it.

We focus first on sceptical dismissals of implicit bias that effectively minimized its significance in everyday life, and instead, portrayed the concept and the test itself as an expression of ideological bias. Readers were here positioned as more objective, astute observers of intergroup relations and the workings of science, while psychologists were ideologically tainted or myopically focused on academic abstractions. We then turn to an analysis of comments that took implicit bias, and specifically, an IAT result, to heart. These comments confirmed the provocative nature of the test experience, but also suggested that the IAT is construed in ways that run counter to expectations about the benefits of being made aware of one's bias.

Implicit bias as academic abstraction

Readers cast doubt on the reality or significance of implicit bias by contrasting it with 'real', explicit, or extreme forms of prejudice and racism. Here, the phenomena that psychologists were said to be studying – reaction times, split-second decisions, and 'unconscious bias' – were depicted as unimportant in comparison with 'obvious' real-world instances of racism. The abstractions of scientific data were set against the concreteness of 'real life', and it was implied that society was far more complex, and racism more tangible, than was indicated by the IAT and the concept of implicit bias. For example:

> To me the question of whether [unconscious] racism exists is almost irrelevant when 1 in 15 black adults and 1 in 9 black men between 20 and 34 is in jail.(Nick, NOC-1)

> It's time scientists started studying explicit bias instead of playing guessing games with "hidden" bias. Our prisons are full of hate groups of all kinds – a ready resource if properly approached . . . Well, what are you scientists waiting for? The terrorists to get nukes? Go! (Marcella, FRB-1)

> It's a shame so much time is spent pulling apart . . . such tiny bits of data. . . There are many, many examples of actual bias but for some reason there is a shift of interest toward the marginal case where implicit attitudes may – repeat may – affect things. . .(Jonathan, FRB-2)

> More seriously: I am very suspicious of claims (and tests) for unconscious bias and its place in discussions about race. Most serious prejudice shows up in tangible social statistics – for example, . . . incarceration rates, median family wealth. These are matters for public policy. Unconscious bias is not.(Mark S., STB-1)

Citing 'tangible' data (Nick and Mark S.), and referencing 'hate groups' and 'terrorists' (Marcella), these accounts foreground the 'hiddenness' and marginality of implicit bias by comparison with the visible and verifiable. By implication, psychologists appear to be fiddling in academic trivialities, and said to be 'playing guessing games', or 'pulling apart tiny bits of data', while the authors of these comments are positioned as better informed, more in touch with 'actual bias' and better able to identify it.

In this vein, other comments attacked the superficiality of the IAT itself. These critics argued that its mechanistic and reductionist nature could not plumb the depths of their phenomenological experience. Readers argued its 'thin-slicing' approach was unable to do justice to their sense of uniqueness and interiority, based on their own understanding of themselves, their histories, or their daily experiences. Consider the following:

> I was smugly satisfied with my results: no racial . . . bias. [But] how can I, having spoken with less than 10 African people in my life, state confidently that I have no bias? . . . It was an entertaining little study, but I for one don't feel those 10-minute tests pried my mouth open, peered down my throat and asked, "Is there a soul in there?"(Amanda, STB-2)

Amanda expresses concern about the moral weight of an IAT result. She evokes individuality and interiority through her use of the word 'soul' and mocks the IAT as an 'entertaining little study' that barely scratches the surface of her psychological reality. Here, the touchstone for judgement was personal experience, rather than 'tangible social statistics'.

The IAT as ideologically biased

The excerpts above have emphasized the trivial nature and insignificance of their IAT and response time data. Here, IAT research and claims about implicit bias were depicted as politically correct impositions of liberal politics. These critiques suggested that the IAT had been devised by psychologists with vested interests in exposing 'White hypocrisy', and whose assertions of widespread hidden bias were themselves expressions of 'bias'.

> What the test does seem to measure is the degree [to which] those in the academic sphere are willing to fall into *mea culpas* – "Oh, indeed, I am biased, even if I didn't know it." . . . It is not politically acceptable to say that the "bias" against the category "black" supposedly revealed by the test is actually due to the basic design of the test. One should now consider whether the authors of the test should re-examine their own biases in favor of their own work, which has got one of them a plumb (sic) job at Harvard.(CK, STB-3)

> I know when I deal with someone obviously different from me . . . I deliberately try to be more thoughtful and careful about what I say and do towards them. This is both tolerance . . . but also self protection: As a white man I can crack a joke about my white, male, co-worker's Arkansas accent . . . and not get in trouble. But if I do the same towards yyyyyy (fill in your favorite discriminated group) [it] would not be received in the same manner. So is measuring the deltas in my reaction time when I deal with someone different from me "prejudice"? We laugh at the religious for blindly following dogma and dismissing "science". There is as much dogma in this test methodology and the conclusions its backers draw from it.(Luke, FRB-3)

CK and Luke each locate their accounts in a discourse of political-correctness-as-censorship (Goodman & Burke, 2010), positioning themselves as restricted in their self-expression. Luke's critique emphasizes his identification as a White male and the infringement on his fundamental freedoms that this entails (see Cabrera, 2014). Both claim to have unmasked the 'bias' or 'dogma' underlying the ostensibly neutral and scientific IAT. It is psychologists themselves, in these excerpts, who are by implication positioned as biased, unfair, and lacking in objectivity.

A notable, though subtler, feature of such comments was their use of sardonic humour. We read these kinds of account as ironic, knowing displays of reflexivity about how the test might position White people. For instance:

> I'm a white male in his mid-30s, yet I'm good. Even subconsciously! Yes!! "The results of your test are outlined below: Your data suggest no difference in your automatic preferences for White people vs. Black people".(vkm, WMB-1)

> I just took the online test you suggested, and my analysis was that I have a slight preference for black people. I guess that makes me your garden-variety guilty white liberal?(CSK, WMB-2)

We are all biased: Trusting the IAT

In the foregoing accounts, a variety of rhetorical strategies and discursive contexts were mobilized in trivializing the IAT and implicit bias. In these accounts, furthermore, implicit bias was carefully differentiated from 'real prejudice', allowing the concept to be successfully portrayed as yet another example of ivory tower abstraction or liberal attempt at censorship through political correctness.

In contrast, where readers accepted the validity of implicit bias and claims of its pervasiveness, their comments contemplated its nature and source, and the deeply personal implications of its detection. Here, the 'implicitness' of implicit bias connoted

depth, as opposed to superficiality, and implicit bias was accorded the same moral status as 'real prejudice'. In these accounts, the IAT took on the status of a powerful diagnostic, truth-telling technology, provoking a great deal of self-reflection and sometimes confession. Reader comments, particularly in response to taking the IAT, were characterized by language that was distinctly morally and emotionally charged. We have identified these responses, according to their content, emotional tone, and form, as (1) confession, (2) exhortation, and (3) absolution. Each of these responses takes implicit bias to be both self-evidently real and inevitable, and the IAT as a powerful technique for revealing this bias.

Confession.   In these kinds of response, the IAT had revealed something 'shocking' about the intrinsic moral status of the reader, occasioning a searching, disclosing, giving-account-of-oneself. Autobiographical details and experiences were narrated as possible explanations for the IAT result obtained. Such self-scrutinizing confessionals were not limited to 'bad' IAT results (i.e., the revelation of a racial preference), but were also offered, surprisingly, where the reader had been pronounced free of bias. Such self-accounting was also marked by displays or declarations of openness and non-defensiveness. For example:

> I tried that Project Implicit test linked off the article, and actually I tested with a strong bias in favor of blacks. My own race is Southeast Asian, born in the U.S. Why were my results so skewed towards blacks? Is it because I was born in a predominantly black city (though I didn't grow up there)? Is it because I myself am not white? Actually, somewhat more troubling to me is not my results, but that I almost feel proud of them, when my sense of right and wrong tells me I shouldn't be proud of having an anti-white bias, much as I wouldn't be of an anti-black one. After all, in my personal habits and tastes the things I favor are quite often created by whites, though I've always considered my aesthetics "equal-opportunity".(Iris, NOC-1)

The above extract exemplifies the many instances of morally inflected soul-searching induced by taking the test. In attempting to account for her result, the reader here lays bare her thought process, revealing that she is 'troubled' and confused. She questions not only the reason for her test result, but also her feelings about it. Notably, her confessional assumes impartiality is the moral ideal ('I shouldn't feel proud. . .', 'my aesthetics [are] "equal opportunity"'). Consider also the following extracts:

> For the record, I took this test a while ago and I have a slight anti-black bias. . . Although I think of myself as passionately egalitarian, I'm happy to own my implicit biases and glad to be made conscious of them. Someday I hope to be able to take the same test and see how my brain feels about men and women.(Jennifer, NOC-2)

> When I took the test, I found the results interesting, and I am open-minded enough to be introspective and search my soul for bias of which I might have been unaware previously, as even the director of Project Implicit discovered about himself.(Bob H., NOC-3)

The relative equanimity or stoicism with which a potentially damning result from the IAT is received is remarkable, for example, in Jennifer's hope to 'someday' be able to take the IAT again, and her reference to the brain as a separate agency which can 'feel' something.

In each of the above excerpts, introspection and openness to criticism, or, at least, the willingness to examine oneself, are idealized as forms of ethical aspiration. What is interesting about these kinds of self-accounting is the way in which they constitute a deep, introspective, and reflexive self (Foucault, Martin, Gutman, & Hutton, 1988). While each extract can be read as a confession of flawed subjectivity, they can be understood, at the same time, as performative of a morally correct identity. We elaborate further on this point in the discussion below.

Exhortation.   These responses were characterized by sanctimonious injunctions to self-examination, and the accusatory positioning of those who were sceptical of the IAT, as psychologically immature and 'defensive'. The injunction to confront the truth about oneself is here more directly and forcefully articulated, and the validity of anger as a response to IAT research was discounted, and even pathologized in these accounts. Such responses were also marked by utopian visions of fair-minded and objective individuals, and of a 'tolerant' society in which everyone 'minds their own business'. We include several extracts here to illustrate outstanding features of this kind of account:

> Interesting that a number of these posts are angry. Is this the response of defensive people who don't want to get close enough to the truth of something to acknowledge it may have merit? Aren't blameless people more apt to respond with perplexity?(Laura, NOC-4)

> Fascinating research. To me, the amount of anger and defensiveness generated by the topic only serves to underline the fact that there is still a lot of racism bubbling just below the surface of our national psyche.(Maryanne, NOC-5)

> Critics need to decide exactly what they are mad about – I understand how cultural defensiveness can make one throw a hissy-fit in response to being called a bigot. I also think that given the consistency and robustness in the trends, that these critics sort of have to get over it and embrace the broader, very significant, and potentially therapeutic dialogue on the topic.(Averroes, HMB-1)

> We alone have to dig deep into our individual hearts and minds and endeavor to eliminate any undesirable thought and negative attitude towards our fellow human beings.(Mary, NOC-6)

> Being a full human being involves being as free of prejudice of all kinds as possible... We should each go home and look in the mirror and recognize the ultimate Bad Guy – who ultimately must become the Good Guy to be the solution.(John, NOC-7)

These types of comment evoke psychotherapeutic discourses of authentic self-disclosure as evidence of psychological health, maturity, and non-defensiveness (Laura) and invoke notions of psychic surface and depth (Mary and Maryanne). Implicit bias was here conceived as originating from deep inside hearts and minds. As was the case with confessional responses, the prescriptions were clear: 'Health' and good neighbourliness are achieved through self-scrutiny, courage to face the truth, admission of guilt, and

elimination of negative attitudes. Psychological and moral sophistication are here conflated.

Laura and Maryanne find people's anger and IAT research to be 'interesting' and 'fascinating', respectively. They invoke a psychoanalytic discourse of hidden, uncomfortable truths, the angry denial of which itself signifies guilt. They are positioned as seeing from the outside and having greater insight into, this hidden state of affairs; they are thus excluded from the diagnosis they impose on others. This juxtaposition of insight and defensiveness takes on a different, more rationalist inflection in Averroes' comment. Critics of IAT research are 'mad' and have 'hissy-fits' while the commenter coolly observes 'defensiveness' and 'consistency and robustness in the trends'. All three accounts employ language that constructs others' guilt as plainly visible to all.

Finally, Mary and John take up further aspects of psychotherapeutic discourse, but these comments foreground their own ethical responsibility and potential guilt. Although neither comment explicitly takes a morally superior position, each constitutes a display of 'insight' or acknowledgement of guilt, thus modelling the forms of subjectivity desired in therapeutic discourse or confession.

Absolution.   This third category of comments speaks directly to what advocates of the IAT argue is its main prejudice-reducing potential, namely, the experience of their own automatic responses (Greenwald *et al.*, 2003; Vorauer, 2012). Here, readers focused on the inevitability of bias, and appealed to the scientific authority of evolutionary psychology, social cognition, or brain functioning to construct bias as 'just the way we are'. Importantly, such accounts tended to be stoically resigned, abstract, and rational-scientific in tone and content. Consider the following:

> I do not believe we can ever get rid of racism and sexism from within ourselves. No amount of education on the importance of tolerance and equality can trump our biological instincts. (David, WMB-3)

> I can't help [but]... realize that my [prejudiced] attitude was not influenced by any of the reasons that people who don't like blacks can give to justify their feelings, and I think that maybe blacks feel the same way toward us and blame it on the treatment whites gave them for so many years, *when in reality it's an automatic feeling that would be there even if the story was completely different*. Go figure... (John, NOC-9, italics ours)

The uncontrollability of implicit bias is here grounds for a fatalism about prejudice. The second extract displays a degree of bewilderment at how the IAT experience appears to shift the meaning of the concept of prejudice, which is no longer grounded in morally significant 'reasons' or justifications, but is reduced simply to 'automatic feelings'.

Taken together, our analysis of these accounts illustrates how discussions of implicit bias in response to reading about it, or receiving an IAT result, might position readers in morally 'favourable' ways; specifically, that the acknowledgement of implicit bias can be a way of presenting oneself as being not prejudiced, and as tolerant and open-minded. In displaying their willingness to accept their (potential) culpability and announcing their openness to introspection, they take the moral high ground. The efficacy of this positioning can be seen to derive from two closely related discursive practices in Western culture – the psychotherapeutic enactment of self-disclosure and 'insight' and their

associated markers of psychological health (Rose, 1996), and Christian discourses of confession and redemption that emphasize the potential for 'sins of the mind' or conscience (Taylor, 2008). Moreover, readers who commented in these terms were positioned – through references to the 'science of prejudice' – as inherently flawed, rather than malignly antipathetic, and therefore ultimately absolved of blame for their bias.

Discussion

Our analysis demonstrates the ways in which test-takers interpreted, related to, and negotiated the claims of the IAT and implicit bias in relation to socially shared and historically embedded concerns and identities. Rather than focus on the extent to which readers accurately understood or accepted these claims, we sought to highlight the ways in which people performed specific moral identities in debating their validity. Thus, IAT science was dismissed or discredited in accounts that evoked discourses about the marginality of academic preoccupations, and those that helped to position test-takers as targets of an oppressive political correctness and psychologists as liberally biased. Alternatively, the IAT was understood to have powerfully revealed widely and deeply held biases towards racialized others, eliciting accounts and explanations that took the form of psychomoral confessionals. These constituted, we argue, displays of insight and openness that helped to position these readers as 'properly' oriented to what they construed as the implicit moral claims of the IAT.

In this respect, the present study extends discourse analytic work on lay people's constructions of prejudice (Durrheim *et al.*, 2016, in press; Figgou & Condor, 2006; Goodman & Burke, 2010) by shedding light on the ways in which 'implicit bias' and the IAT function in such debates. The discourses analysed here bear some similarity with those reported in the literature; they instantiate the Enlightenment norm against prejudice, and they seek to deny the speaker's prejudice and locate it elsewhere. This was especially true of the sceptical dismissals of implicit bias as revealed by the IAT. However, when an IAT result was understood to have revealed a hidden bias, *admitting to this bias* functioned as a way of identifying oneself as sincerely opposed to prejudice. Such admissions might be understood in relation to the contemporary valorization of confession as a mode of work on the self, a discursive act which has become associated with sincerity, authenticity, and deciphering the truth about oneself (Brooks, 1996; Taylor, 2008). Various benefits accrue to those who disclose themselves, since in speaking of 'what we hide' we not only receive the 'satisfaction of feeling transgressive and progressive' (Taylor, 2008, p. 74), but we may also be positioned as enlightened and morally right. However, Sara Ahmed (2004, para. 1) has observed how such declarations are often 'non-performative', involving a 'fantasy of transcendence in which "what" is transcended is the very "thing" admitted to in the declaration (e.g., if we say that we are racists, then we are not racists, as racists do not know they are racists)'.

Correspondingly, our analysis of how test-takers might be positioned in relation to the IAT has implications for both empirical studies of people's reactions to feedback about their prejudices, and more broadly, for theoretical accounts of prejudice reduction. While some researchers are optimistic about the role that the IAT might play in prejudice reduction strategies because of to its 'consciousness-raising' qualities (Paluck & Green, 2009), others, like Vorauer (2012), have shown that the test experience can 'backfire'. However, negative reactions to having one's biases revealed tend to be regarded as evidence of irrational denial, the so-called 'bias blind spot', the solution to which is,

greater recognition of one's biases and cultivation of more rational cognition (Pronin, Lin, & Ross, 2002). Our findings demonstrate the importance of the interpretive or discursive context in making sense of people's reactions to the IAT and its associated knowledge claims. They concur with, but also help to situate findings such as Vorauer's (2012), since they complicate and call into question the effectiveness of 'awareness' as a goal of prejudice reduction efforts.

Finally, our analysis highlights a persistent dilemma faced by social psychologists who wish to present their findings to the public as amoral, neutral discoveries about psychological processes underlying prejudice, and at the same time make claims about their moral/ethical significance (see Brinkmann, 2009). As our findings suggest, the 'implicitness' of implicit bias – as operationalized in the IAT – can connote for test-takers either its automatic, fleeting, and therefore superficial character, or alternatively, its unconscious, deep, and morally deplorable nature. The problem of bias, as refracted through the lens of the IAT, is thus either easily written off as the result of amoral automatic associations, and inevitable cognitive limitations for which we should not be held responsible, or, on the other hand, as a reprehensible psychological flaw to be dealt with through confession and psychomoral work on the self.

Concluding comments

Our aim in this research has not been to mount a theoretical (e.g., Gawronski *et al.*, 2007) or psychometric (e.g., Fiedler *et al.*, 2006) criticism of implicit bias or the IAT. Rather, our objective has been to draw out the social implications of the science in relation to the changing context of prejudice discourse. The IAT paradigm has lowered the threshold for what might count as prejudice (Arkes & Tetlock, 2004; Durrheim, 2012), challenged received views of moral responsibility for prejudice (Vargas, 2016), and made implicit bias evident to people in unprecedented ways. Our analysis provided a first examination of how this research and technology have begun to function in lay understandings of prejudice and public discourse.

Our selection of discussion and debate from the NYT is certainly not representative of race and prejudice talk among lay people in general. It does, however, suggest an emergent discourse among the liberal left, highlighting especially the way that the idea of implicit bias can inform identity performances in which an anti-racist footing can be gained by *confessing* one's bias. Both the idea of implicit bias and the practice of measuring it can thus impact on the way people think of themselves, others, and their prejudices. They provide tools for talking about prejudice, for moral-psychological work on the self, for explaining social ills, and for mobilizing others to act in the interests of change (Nadan & Stark, 2016).

Acknowledgements

References

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. Oxford, UK: Harpers.

Ahmed, S. (2004). Declarations of whiteness: The non-performativity of anti-racism. *Borderlands*, *3* (2). Retrieved from http://www.borderlands.net.au/vol3no2_2004/ahmed_declarations.htm

Alexa Internet Analytics (n.d.). *The top 500 sites on the web*. Retrieved from http://www.alexa.com/topsites/category/News

Allport, G. W. (1979). *The nature of prejudice*. Reading, MA: Basic Books.

Altman, A. (2008, October 17). The bradley effect. *TIME*. Retrieved from http://www.time.com/time/magazine/article/0,9171,1853294,00.htm

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the implicit association test?". *Psychological Inquiry*, *15*, 257–278. https://doi.org/10.1207/s15327965pli1504_01

Augoustinos, M., & Every, D. (2007). The language of "race" and prejudice: A discourse of denial, reason, and liberal-practical politics. *Journal of Language and Social Psychology*, *26*, 123–141. https://doi.org/10.1177/0261927X07300075

Augoustinos, M., & Every, D. (2010). Editorial: Accusations and denials of racism: Managing moral accountability in public discourse. *Discourse & Society*, *21*, 251–256. https://doi.org/10.1177/0957926509360650

Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7, pp. 55–76). Hillsdale, NJ: Erlbaum Hillsdale.

Bazelon, E. (2016, October 18). How "bias" went from a psychological observation to a political accusation. *The New York Times*. Retrieved from http://www.nytimes.com/2016/10/23/magazine/how-bias-went-from-a-psychological-observation-to-a-political-accusation.html

Billig, M. (1988). The notion of "prejudice": Some rhetorical and ideological aspects. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(1–2), 91–110. https://doi.org/10.1515/text.1.1988.8.1-2.91

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *The American Psychologist*, *61*(1), 27–41. https://doi.org/10.1037/0003-066X.61.1.27

Blanton, H., & Jaccard, J. (2015). Not so fast: Ten challenges to importing implicit attitude measures to media psychology. *Media Psychology*, *18*, 338–369. https://doi.org/10.1080/15213269.2015.1008102

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *The Journal of Applied Psychology*, *100*, 1468–1481. https://doi.org/10.1037/a0038379

Brinkmann, S. (2009). Facts, values, and the naturalistic fallacy in psychology. *New Ideas in Psychology*, *27*, 1–17. https://doi.org/10.1016/j.newideapsych.2007.10.001

Brooks, P. (1996). Storytelling without fear? Confession in law and literature. In P. Brooks & P. Gewirtz (Eds.), *Law's stories: Narrative and rhetoric in the law* (pp. 114–134). New Haven, CT: Yale University Press.

Burke, S., & Goodman, S. (2012). "Bring back Hitler's gas chambers": Asylum seeking, Nazis and Facebook – a discursive analysis. *Discourse & Society*, *23*(1), 19–33. https://doi.org/10.1177/0957926511431036

Cabrera, N. L. (2014). "But I'm oppressed too": White male college students framing racial emotions as facts and recreating racism. *International Journal of Qualitative Studies in Education*, *27*, 768–784. https://doi.org/10.1080/09518398.2014.901574

Calanchini, J., Sherman, J. W., Klauer, K. C., & Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. *Personality & Social Psychology Bulletin*, *40*, 1285–1296. https://doi.org/10.1177/0146167214540723

Cherry, F. (2000). The nature of the nature of prejudice. *Journal of the History of the Behavioral Sciences*, *36*, 489–498. https://doi.org/10.1002/(ISSN)1520-6696

Chung, D. S. (2008). Interactive features of online newspapers: Identifying patterns and predicting use of engaged readers. *Journal of Computer-Mediated Communication*, *13*, 658–679. https://doi.org/10.1111/j.1083-6101.2008.00414.x

Clark, F., & Illman, D. L. (2006). A longitudinal study of the *New York Times* Science Times section. *Science Communication*, *27*, 496–513. https://doi.org/10.1177/1075547006288010

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the implicit association test. *Journal of Experimental Social Psychology*, *43*, 972–978. https://doi.org/10.1016/j.jesp.2006.10.007

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*, 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Dixon, J. (2017). "Thinking ill of others without sufficient warrant?" Transcending the accuracy-inaccuracy dualism in prejudice and stereotyping research. *British Journal of Social Psychology*, *56*(1), 4–27. https://doi.org/10.1111/bjso.12181

Dooley, R. (2015, December 28). Trump explained by neuroscience. *Forbes*. Retrieved from http://www.forbes.com/sites/rogerdooley/2015/12/28/trump-neuroscience/

Duckitt, J. (2001). Reducing prejudice: An historical and multi-level approach. In M. Augoustinos & K. J. Reynolds (Eds.), *Understanding prejudice, racism and social conflict* (pp. 253–272). London, UK: Sage. https://doi.org/10.4135/9781446218877

Durrheim, K. (2012). Implicit prejudice in mind and interaction. In J. Dixon & M. Levine (Eds.), *Beyond prejudice: Extending the social psychology of conflict, inequality and social change* (pp. 179–199). Cambridge, UK: Cambridge University Press.

Durrheim, K. (2014). Two histories of prejudice. In C. Tileagă & J. Byford (Eds.), *Psychology and history: Interdisciplinary explorations* (pp. 205–222). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139525404

Durrheim, K., Okuyan, M., Twali, M. S., García-Sánchez, E., Pereira, A., Portice, J. S., . . . Keil, T. (In press). How racism discourse can mobilize right-wing populism: The construction of identity and alliance in reactions to UKIP's Brexit "Breaking Point" campaign. *Journal of Community and Applied Social Psychology*.

Durrheim, K., Quayle, M., & Dixon, J. (2016). The struggle for the nature of "prejudice": "Prejudice" expression as identity performance. *Political Psychology*, *37*(1), 17–35. https://doi.org/10.1111/pops.12310

Edwards, D. (2003). Analyzing racial discourse: The discursive psychology of mind-world relationships. In H. van den Berg, M. Wetherell, & H. Houtkoop-Steenstra (Eds.), *Analyzing race talk: Multidisciplinary approaches to the research interview* (pp. 31–48). Cambridge, UK: Cambridge University Press.

Edwards, J. (2016, June 28). *Justice Dept. mandates 'implicit bias' training for agents, lawyers*. Retrieved from http://www.reuters.com/article/us-usa-justice-bias-exclusive-idUSKCN0ZD251

Engber, D. (2016, November 11). *Donald trump is a racist. Here's why many Americans don't see him that way*. Retrieved from http://www.slate.com/articles/news_and_politics/politics/2016/11/the_people_who_look_at_trump_and_don_t_see_a_racist.html

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the I, the A, and the T: A logical and psychometric critique of the implicit association test (IAT). *European Review of Social Psychology*, *17*, 74–147. https://doi.org/10.1080/10463280600681248

Figgou, L., & Condor, S. (2006). Irrational categorization, natural intolerance and reasonable discrimination: Lay representations of prejudice and racism. *The British Journal of Social Psychology/the British Psychological Society*, *45*(Pt 2), 219–243. https://doi.org/10.1348/014466605X40770

Foucault, M., Martin, L. H., Gutman, H., & Hutton, P. H. (1988). *Technologies of the self: A seminar with Michel Foucault* (p. 166). Amherst, MA: University of Massachusetts Press.

Freeman, C. (2010). Through a western lens: Portrayals of a 'rising' ASEAN and its member countries in the *New York Times*. *International Communication Gazette*, *72*, 269–285. https://doi.org/10.1177/1748048509356950

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *2*, 181–193. https://doi.org/10.1111/j.1745-6916.2007.00036.x

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. Toronto, ON: Little, Brown and Company.

Goodman, S. (2010). "It's not racist to impose limits on immigration": Constructing the boundaries of racism in the asylum and immigration debate. *Critical Approaches to Discourse Analysis across Disciplines*, *4*(1), 1–17.

Goodman, S., & Burke, S. (2010). 'Oh you don't want asylum seekers, oh you're just racist': A discursive analysis of discussions about whether it's racist to oppose asylum seeking. *Discourse & Society*, *21*, 325–340. https://doi.org/10.1177/0957926509360743

Goodman, S., & Rowe, L. (2014). 'Maybe it is prejudice . . . but it is NOT racism': Negotiating racism in discussion forums about Gypsies. *Discourse & Society*, *25*(1), 32–46. https://doi.org/10.1177/0957926513508856

Gove, T. G. (2011). Implicit bias and law enforcement. *The Police Chief*, *78*, 44–56.

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. https://doi.org/10.1037/0022-3514.74.6.1464

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*, 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy*, *9*(1), 241–253. https://doi.org/10.1111/j.1530-2415.2009.01195.x

Herbert, W. (2012, May 25). *Red mind, blue mind: Are there really any independents?* Retrieved from http://www.huffingtonpost.com/wray-herbert/red-mind-blue-mind-are-th_b_1507434.html

Howell, J. L., & Ratliff, K. A. (2017). Not your average bigot: The better-than-average effect and defensive responding to implicit association test feedback. *The British Journal of Social Psychology*, *56*, 125–145, https://doi.org/10.1111/bjso.12168

Ismail, A., & Mishra, S. (2009). In the eye of the occupier: News coverage of Iraq war 'terrorism' in the *New York Times* and the *Times of India*. *Journalism Practice*, *3*, 456–471. https://doi.org/10.1080/17512780903010892

Jørgensen, M., & Phillips, L. J. (2002). *Discourse analysis as theory and method*. London, UK: Sage. https://doi.org/10.4135/9781849208871

Kaslow, N. J. (2015). Translating psychological science to the public. *The American Psychologist*, *70*, 361–371. https://doi.org/10.1037/a0039448

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, *40*, 414. https://doi.org/10.1037/0022-3514.40.3.414

Kristof, N. D. (2008, October 30). What? Me biased? *The New York Times*, p. A39.

Laclau, E. (1993). Politics and the Limits of Modernity. In T. Docherty (Ed.), *Postmodernism: A reader* (pp. 329–344). London, UK: Harvester Wheatsheaf.

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*, 1001–1016. https://doi.org/10.1037/xge0000179

Lea, S. J. (1996). "That ism on the end makes it nasty": Talking about race with young White South Africans. *South African Journal of Psychology*, *26*(3), 183–190. https://doi.org/10.1177/008124639602600308

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*, 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x

Luguri, J. B., Napier, J. L., & Dovidio, J. F. (2012). Reconstruing intolerance: Abstract thinking reduces conservatives' prejudice against non-normative groups. *Psychological Science*, *23*, 756–763. https://doi.org/10.1177/0956797611433877

Marshall, E. (1998). The power of the front page of *The New York Times*. *Science*, *280*, 996–997. https://doi.org/10.1126/science.280.5366.996

Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations: GPIR*, *13*, 183–200. https://doi.org/10.1177/1368430209353633

Morris, K. A., & Ashburn-Nardo, L. (2009). The implicit association test as a class assignment: Student affective and attitudinal reactions. *Teaching of Psychology*, *37*(1), 63–68. https://doi.org/10.1080/00986280903426019

Nadan, Y., & Stark, M. (2016). The pedagogy of discomfort: Enhancing reflectivity on stereotypes and bias. *British Journal of Social Work*, *47*, 683–700https://doi.org/10.1093/bjsw/bcw023

New York Times (2009). *NYTimes.com demographic profile*. Retrieved from http://www.nytimes.whsites.net/mediakit/pdfs/online/NYTimes_com_ReaderProfile.pdf

Nosek, B. A., Banaji, M., & 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

Nosek, B. A., & Riskind, R. G. (2012). Policy implications of implicit social cognition. *Social Issues and Policy Review*, *6*(1), 113–147. https://doi.org/10.1111/j.1751-2409.2011.01037.x

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*(1), 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality & Social Psychology Bulletin*, *28*, 369–381. https://doi.org/10.1177/0146167202286008

Rachlinski, J. J., & Parks, G. S. (2008, October 30). *Unconscious race bias and Obama's candidacy*. Retrieved from http://www.news.cornell.edu/stories/2008/10/unconscious-race-bias-and-obamas-candidacy

Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*, 523–534. https://doi.org/10.1016/j.labeco.2009.04.005

Rose, N. (1996). *Inventing ourselves: Psychology, power, and personhood*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511752179

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*, 856–868. https://doi.org/10.1037/0022-3514.81.5.856

Simmons, K., & Lecouteur, A. (2008). Modern racism in the media: Constructions of 'the possibility of change' in accounts of two Australian 'riots'. *Discourse & Society*, *19*, 667–687. https://doi.org/10.1177/0957926508092248

Spencer, K. B., Charbonneau, A. K., & Glaser, J. (2016). Implicit bias and policing. *Social and Personality Psychology Compass*, *10*(1), 50–63. https://doi.org/10.1111/spc3.12210

Taylor, C. (2008). *The culture of confession from Augustine to Foucault: A genealogy of the 'confessing animal'*. New York, NY: Taylor & Francis.

Tierney, J. (2008, November 7). Where have all the bigots gone? *The New York Times*. Retrieved from http://tierneylab.blogs.nytimes.com/2008/11/07/where-have-all-the-bigots-gone/

Van Dijk, T. A. (2000). New(s) racism: A discourse analytical approach. In S. Cottle (Ed.), *Ethnic minorities and the media* (pp. 33–49). Buckingham, UK: Open University Press.

Vargas, M. R. (2016). *Implicit bias, responsibility, and moral ecology*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2778922

Vorauer, J. D. (2012). Completing the implicit association test reduces positive intergroup interaction behavior. *Psychological Science*, *23*, 1168–1175. https://doi.org/10.1177/0956797612440457

Wetherell, M. (1998). Positioning and interpretative repertoires: Conversation analysis and post-structuralism in dialogue. *Discourse & Society*, *9*, 387–412. https://doi.org/10.1177/0957926598009003005

Wetherell, M. (2012). The prejudice problematic. In J. Dixon & M. Levine (Eds.), *Beyond prejudice: Extending the social psychology of conflict, inequality and social change* (pp. 158–178). Cambridge, UK: Cambridge University Press.

Wetherell, M., & Potter, J. (1993). *Mapping the language of racism: Discourse and the legitimation of exploitation*. New York, NY: Columbia University Press.

Wynne, B. (2006). Public engagement as a means of restoring public trust in science–hitting the notes, but missing the music? *Community Genetics*, *9*, 211–220. https://doi.org/10.1159/000092659

Yen, J. (2013). *Psychology and the social scientific construction of prejudice: Lay encounters with the Implicit Association Test*. Unpublished doctoral dissertation. Toronto, Canada: University of Toronto.

Zestcott, C. A., Blair, I. V., & Stone, J. (2016). Examining the presence, consequences, and reduction of implicit bias in health care: A narrative review. *Group Processes & Intergroup Relations*, *19*, 528–542. https://doi.org/10.1177/1368430216642029