

May 9, 2019

Do IAT Scores Explain Racial Gaps?

Lee Jussim  
Rutgers University

Akeela Careem  
Rutgers University

Nathan Honeycutt  
Rutgers University

Sean T. Stevens  
New York University, Stern School of Business

Draft chapter prepared for the 2019 Sydney Symposium of Social Psychology:  
Applications of Social Psychology

## Abstract

In this chapter we critically evaluate the ability of IAT scores to explain real world racial gaps. This is especially difficult to do because there are so many conceptual, psychometric, and validity issues related to the IAT, which we review before tracking the main question of accounting for gaps. That review indicates that the IAT not a clean measure of either implicit biases or of associations. Nonetheless, several meta-analyses have shown that IAT scores predict discrimination to at least a modest extent. To address its ability to account for racial gaps, we work backward. Given the size of some gap, how much of that is likely to be explained by IAT scores? Although we do not answer the question quantitatively, we present a series of heuristic models intended to provide increased clarity about the likely range of possibilities. Alternative explanations for gaps are briefly reviewed, and our models assume that IAT scores can only explain what is left over, which is likely to be only a modest portion of those gaps. We conclude by arguing that, despite its limitations, the IAT should not be abandoned, but that, even after 20 years, much more research is needed to fully understand what the IAT measures and explains.

How much does whatever is measured by implicit association test (IAT) scores explain racial gaps? Our answer is, “probably not much.” How is this even possible, given how much attention implicit bias (the preeminent measure of which is the IAT) has received in scientific journals (e.g., Bertrand & Mullainathan, 2004; Sabin & Greenwald, 2012; van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010), law journals (e.g., Jolls & Sunstein, 2006), the popular press (e.g., Baker, 2018; Mundy, 2017; Szalavitz, 2010), organizational anti-bias interventions (e.g., Department of Justice, Office of Public Affairs, 2016), and even U.S. presidential elections (Hensch, 2016)?

In this chapter, we argue that this is indeed possible because even after 20 years of research, the IAT remains poorly understood. We first review definitional, psychometric, interpretive, construct validity, and predictive validity issues regarding the IAT. We then present a series of heuristic models that may be useful for attempting to understand the likely the role of IAT scores in accounting for important inequalities.

### **Definitional Issues**

There is no widely-agreed upon consensus as to what “implicit bias” means. Numerous articles use the term without defining it at all (e.g., Capers, Clinchot, McDougle, & Greenwald, 2017; Glaser & Finn, 2013; Gonzalez, Steele, & Baron, 2017; Jost, 2019; March & Graham, 2015; Penner et al., 2016; Rae, Newheiser, & Olson, 2015; Teachman & Brownell, 2001; Walker & Hewstone, 2007; White-Means, Dong, Hufstader, & Brown, 2009; Xiao, Fu, Quinn, Qin, Tanaka, Pascalis, & Lee, 2014). Because these articles use the IAT to measure implicit bias (or refer to studies using it as having measured implicit bias), they appear to presume that it means “whatever is measured by the IAT.”

Considering only cases where a definition is actually provided, the specific content varies so widely as to render it impossible for readers to be sure authors are discussing the same construct:

“Cultural stereotypes may not be consciously endorsed, but their mere existence influences how information about an individual is processed and leads to unintended biases in decision-making, so called “implicit bias”” (Chapman, Kaatz, & Carnes, 2013, p. 1504).

“Here, we focus on implicit social bias, a measure of how strongly one associates a concept (e.g., pleasant/unpleasant) with one or another social group” (Stanley, Sokol-Hessner, Banaji, & Phelps, 2011, p. 7710).

“Whereas explicit bias is overt and freely expressed, implicit bias may not be consciously acknowledged and operates in more subtle ways. For example, a clinician with implicit bias may unconsciously exhibit negative behavior or poor communication with a black patient, as has been shown in laboratory research” (Blair et al., 2013, p. 44).

Greenwald (2017), at a special NSF conference on controversies surrounding the IAT, provided this as the working definition of implicit bias provided for academics for most of the prior 20 years:

“Introspectively unidentified (or inaccurately identified) effects of past experience that mediate discriminatory behavior.”

Inasmuch as this definition was offered by the developer of the IAT (Greenwald, McGhee, & Schwartz, 1998), this definition warrants close scrutiny. If this is the definition, it is not clear that the IAT measures implicit bias, because:

1. It is a reaction time measure that assesses the difference in time it takes to do two different yet related categorization tasks. Difference in reaction times is not discrimination.
2. It is a single variable. Although any single variable *might* be a mediator of the effect of some variable on some other variable, that is a separate empirical question that cannot be addressed simply by assessing a single variable. We can conduct research providing evidence that “B mediates the effect of A on C,” but we cannot declare “merely by measuring B we have established that it mediates the effect of A on C.” Defining the

measure as “mediation” biases the communication of what the IAT captures by subterranean importation of the unverified assumption that it is involved in discrimination without having to provide any evidence that it actually is.

3. Nor are IAT scores “introspectively unidentified.” People can quite accurately predict their IAT scores,  $r$ 's ranged from about .60 to .70 (Hahn, Judd, Hirsh, & Blair, 2014).

Here is what is left of Greenwald's definition of implicit bias after we remove statements that do not apply to the IAT:

~~“Introspectively unidentified (or inaccurately identified) effects of past experience that mediate discriminatory behavior.”~~

What is left is “effects of past experience.” This may not be pure coincidence as we discuss in the next section on construct validity.

### **Construct Validity**

Exactly what the IAT measures remains muddled. Yet, there are large bodies of scholarship that are plausibly interpreted as presuming that it is a clean measure of unconscious or automatic prejudice. For example, in 1998 Greenwald and Banaji held a press conference unveiling a “... new tool that measures the unconscious roots of prejudice” (Schwarz, 1998). In 2013, Banaji & Greenwald declared in their book *Blindspot* “...the automatic White preference expressed on the Race IAT is now established as signaling discriminatory behavior.” Many, perhaps most, scholars who have published research using the IAT to assess implicit associations regarding demographic groups once interpreted it similarly (e.g., Jolls & Sunstein, 2006; Jost et al., 2009; McConnell & Leibold, 2001). However, rather than two decades of research converging on clear conclusions regarding exactly what the IAT measures, the level of disagreement and controversy has escalated.

First, consider what the IAT is supposed to measure: *implicit associations* between concepts in memory. Let's temporarily put aside measurement, psychometric, or predictive validity issues,

and stipulate that it succeeds at doing so. How do two concepts in memory become associated? One likely route is that they co-occur in the world in some way. Ham and eggs, for example, go together in the environment far more than do ham and quantum mechanics. Similarly, at least for North Americans, New York is probably more strongly associated with Yankees than with malaria. This analysis has a kinship to that of one of the earliest critiques of the IAT (Arkes & Tetlock, 2004). They argued that, rather than reflecting prejudice, the race IAT might reflect knowledge of cultural stereotypes.

Our argument, however, goes further. The associations tapped by the IAT may reflect implicit cognitive registration of regularities and realities of the social environment. Consistent with this analysis, IAT response times were faster when stimuli corresponded to ecologically valid base rates (Bluemke & Fiedler, 2009). What about evaluations of whether group membership is good or bad, pleasant or unpleasant (as used in so many IATs)? That depends on where evaluations come from. One possibility is that evaluations of groups come from knowledge of those groups' social conditions. Groups disproportionately living in unpleasant conditions (such as poverty, crime, rundown housing) may be associated with "unpleasant" more than other groups.

Research on stereotype accuracy has repeatedly demonstrated that the racial, gender, and age stereotypes of those who have been studied are often quite accurate in that people's beliefs about groups often correlate moderately to highly with credible measures of real group differences, such as Census data and meta-analysis (Jussim, Crawford & Rubinstein, 2015). People who are better at detecting patterns regarding social groups also more efficiently learn, apply, and update their stereotypes (Lick, Alter, & Freeman, 2018).

Although both work on stereotype accuracy and pattern detection work focused on explicit stereotypes, there are reasons to suspect similar processes underlie implicit associations. The earliest work on implicit cognition long predated social psychological approaches. Reviewing two

decades of that research Reber (1989, p. 219, emphasis added) concluded:

“(a) Implicit learning produces a tacit knowledge base that is ***abstract and representative of the structure of the environment***, (b) such knowledge is optimally acquired independently of conscious efforts to learn; and (c) it can be used implicitly to solve problems and ***make accurate decisions about novel stimulus circumstances***.”

A new theoretical perspective on implicit bias reached conclusions largely compatible with this view (Payne, Vuletich & Lundberg, 2017, p. 242): “...the reason that implicit bias is widespread in general is that the environment has a relatively constant level of disparities and systemic inequalities that repeatedly raise the accessibility of stereotypic concepts.” Although Payne et al. (2017) emphasize bias and racism throughout their paper, in fact, the core claim above is plausibly interpretable as consistent with implicit accuracy. That is, systemic inequalities may not merely “raise the accessibility of stereotype concepts,” they may *create* those stereotypic associations. This is exactly what would be expected to happen if, as Lick et al. (2018) and Reber (1989) indicate, people good at detecting patterns in the environment associate groups with their different and unequal social conditions.

Empirical research has begun to confirm the prediction that implicit associations are responsive to regularities in the social environment. In one series of studies, either African-American or White individuals were paired with images reflecting either intelligence (e.g., a brain) or stupidity (e.g., a paper with an “F” on it; Rubinstein & Jussim, 2019). When the African-American target was paired with smart images and the White target with stupid images, conventionally biased IAT scores (in a control condition without such associations) were not merely eliminated, they reversed (people completed Jamal/Smart and Luke/Stupid more quickly than Jamal/Stupid and Luke/Smart). Also consistent with this perspective, one recent study found that, in addition to results interpretable as racial bias, implicit stereotypes as assessed by the IAT correlated

with geographic crime rates among blacks (Johnson & Chopik, 2018).

The weight of the evidence from multiple sources – explicit stereotypes, the earliest work on implicit cognition, and the recent theoretical and empirical work on sources of IAT scores – all converges on the conclusion that what was left in Greenwald’s definition above is no coincidence. Whether or not IAT scores reflect prejudice or knowledge of cultural stereotypes, they likely capture the individual’s past experiences with groups. After reviewing many other theoretically plausible or empirically demonstrated potential influences on IAT scores, Bluemke & Fiedler (2009, p. 1036) put it this way: “Whether an implicit attitude really underlies a latency difference is a question that cannot be answered from the magnitude of the association index alone.”

Thus, one important area for future research is to more thoroughly test the role of experience and social realities in producing IAT scores. How well do Census data on achievement gaps, wealth and income gaps, and health gaps correlate with IAT scores? Addressing such questions may shed light on the role of social realities in producing the implicit associations captured by the IAT.

But even that analysis presumes that the IAT is a clean measure of implicit associations, an assumption unlikely to be generally true (Fiedler, Messner, & Bluemke, 2006). The quad model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) provided five experiments worth of evidence indicating that the IAT simultaneously reflected: implicit associations, discriminability (ability to identify a correct response), ability to overcome the bias produced by having such associations, and guessing biases (other systematic influences on responses). Because their model yields parameters for each of the four sources of IAT scores, it provided one way to make use of the IAT to assess implicit associations. However, it also showed that a presumption that IAT scores are a clean measure of “implicit associations” (let alone “implicit biases”) was not justified. Additional research assessing and removing artifacts and measurement biases (see Fiedler et al., 2006) could also advance the understanding of what the IAT assesses.



### Psychometric Issues

The internal consistency of IAT scores meets conventional standards of acceptability, and ranges from .60-.90 (LeBel & Paunonen, 2011). Test-retest reliabilities fare worse. Even when tested over only a few weeks, test-retest correlations average about  $r = .4$  (Payne et al., 2017).

For most of its existence, IAT effects have been assessed with  $D$  scores (Greenwald, Nosek, & Banaji, 2003).  $D$  scores are computed separately for each participant as:

$D = IAT_{RAW}/SD_{WI}$ . Raw scores are the mean difference in response times in milliseconds to two IATs (such as African-American/pleasant or white/unpleasant versus African-American/unpleasant or white/pleasant).  $SD_{WI}$  is the within individual standard deviation. Blanton, Jaccard, and Burrows (2015) were the first to point out a computational weakness of this method (p. 431):

“This change [to  $D$  scores] unwittingly produced a situation in which a reduction in trial error (typically something a researcher or practitioner would view as desirable) now results in more extreme bias scores, everything else being equal.”

Consider two people with identical raw reaction time differences (let's say, 100ms.). The first has an  $SD$  of 100; the second an  $SD$  of 25. The first will have  $D = 1$ , the second,  $D = 4$ , even though their levels of bias as indicated by the raw difference in response times are identical.

Furthermore, IAT scores are difference scores, and difficulties with interpreting difference scores have been recognized for decades (e.g., Cronbach & Furby, 1970). Although a thorough review of such difficulties is beyond the scope of this chapter, fortunately, one applied specifically to IAT scores already exists (see Fiedler et al., 2006). For example, correlations of IAT scores with outcomes are difficult to interpret because they may reflect a correlation with the compatible trials (e.g., black/unpleasant v. white/pleasant), a correlation with the incompatible trials (e.g., black/pleasant v. white/unpleasant), or a correlation with the difference. The criterion may also correlate more strongly with the compatible vs. incompatible trials, further muddying the

interpretation of the correlation with the difference. Fiedler et al. (2006) also explained why standardized IAT scores via the  $D$  statistic (Greenwald et al., 2003) do not solve these problems, and identify a slew of other issues with the IAT that derive from using difference scores.

### **The Curious Case of The Doubly-Computed Effect Size**

$D$  scores are computationally similar, but not quite identical to the common effect size, Cohen's  $d$ . This is explicitly acknowledged in the paper recommending  $D$  (Greenwald et al., 2003). This raises the question: Why, if  $D$  is a close relative of  $d$ , do many papers reporting IAT effects further transform the (already transformed from raw scores) effect size  $D$  into yet another effect size,  $d$  (e.g., Cao & Banaji, 2016; Lane, Banaji, Nosek, & Greenwald, 2007)?  $D$  is already a standardized effect size. Although we have never seen a rationale for re-computing a  $d$  from the  $D$ 's in print, the consequence of doing so will be clear: If the standard deviation of the IAT  $D$  scores is less than 1,  $d$  will be larger than  $D$ ; if the  $SD$  of each participant is larger than 1,  $d$  will be smaller than  $D$ . Although doing a comprehensive analysis of this issue is beyond the scope of this chapter, every paper of which we are aware that reported both  $D$ 's and Cohen's  $d$ 's reported  $d$ 's that were larger than the mean  $D$  (e.g., Cao & Banaji, 2016; Lane et al., 2007; Charlesworth & Banaji, 2019). If this is not justified, the literature may be filled with exaggerated estimates of implicit bias effects. Absent a compelling and fully articulated reason to do otherwise, we recommend simply reporting the average  $D$  as *the IAT effect size*, rather than further computing Cohen's  $d$ 's.

### **Interpretations of “Bias”**

Let's temporarily put aside all the other problems involved in interpreting IAT scores and consider them to reflect “bias” as have others for over two decades. IAT scores that differ from 0 have routinely been interpreted as “implicit bias” (Greenwald et al., 1998; Greenwald et al., 2003; Lane et al., 2007). There are two separate problems with this. When researchers use conventional null hypothesis significance tests (as nearly all have when conducting research with the IAT), two

factors contribute to concluding whether an IAT score is “different than 0.” The first is that the score cannot equal 0. The second is statistical power. With sufficiently high power, almost any non-zero value will differ “significantly” from 0; the same IAT effect may significantly differ from 0 in a large sample but not in a small sample. Presumably, however, “bias” (if any) or even implicit association, is a function of strength of association, not statistical power. Is a  $D$  score of .04 really “bias” if obtained in a sufficiently large sample? And if so, is it a bias of sufficient magnitude that it warrants being taken seriously, for either theoretical or applied reasons? If so, we have never seen such arguments articulated. Perhaps for these reasons, Nosek et al. (2007) recommended use of an absolute standard of  $D \geq .15$  for “implicit biases” (or associations) to be considered sufficiently substantial to warrant being taken seriously.

Although Nosek et al.’s (2007) recommendations were definitely an improvement (and, in deference to this tradition, we have used it ourselves, e.g., in Rubinstein et al., 2018 and Rubinstein & Jussim, 2019), it is not obvious that it is well-justified. This issue was first raised as a theoretical possibility on the grounds that the preponderance of positive race IAT scores (80% or more) could indicate that the base rate of positive IAT scores in the populations taking the test was simply much higher than the base rate of racist implicit attitudes (Fiedler et al., 2006).

Subsequent research supported this prediction. Research re-analyzing a slew of studies correlating IAT scores with other measures of bias (behavior, attitudes, evaluations, etc.) found that the IAT is consistently *right-biased* (Blanton, Jaccard, Strauts, Mitchell & Tetlock, 2015). That is, rather than  $D = 0$  corresponding to unbiased responding on other measures,  $D$ ’s corresponding to unbiased responding ranged widely, but consistently fell well above 0 (most  $D$ ’s corresponding to egalitarian responding were between .3 and .6). Although the generality of this pattern is unclear, so is the interpretation of IAT scores as reflecting prejudice or bias. If the right bias pattern found by Blanton et al. (2015) is more general, it means not only is the Nosek et al. (2007)

recommendation of  $D \geq .15$  too conservative, but that it might actually reflect *reverse bias*. For example, if a  $D = .5$  corresponds to racial egalitarianism,  $D = .15$  might correspond to *anti-white bias*! This interpretation is given further credence by the facts that at least one study reporting supposedly anti-African-American IAT scores also found far more anti-white than anti-African-American behavioral discrimination (McConnell & Leibold, 2001; although it took a reanalysis by Blanton, Jaccard, Klick, Mellers, Mitchell & Tetlock, 2009, to reveal this inconvenient finding). Mothers reporting positive IAT scores also reported average patterns of unbiased behavior (e.g., Green et al., 2007; Rachlinski, Johnson, Wistrich, & Guthrie, 2009; Stanley, Sokol-Hessner, Banaji, & Phelps, 2011). Because the average of the predictor corresponds to the average of the outcome in bivariate regression, positive IAT scores corresponded to egalitarian behavior in all of these studies. In order to better understand what IAT scores mean, and especially what scores correspond to egalitarian versus biased beliefs and attitudes, whenever possible research report the  $D$  scores that correspond to egalitarian responding on criteria or outcome measures.

### **Predictive Validity Controversies**

One might argue that the above issues are minor nitpicks if stereotype and prejudice-related IAT's predict discrimination. Unfortunately, this issue is unclear. Dueling meta-analysis have appeared since 2009 (Greenwald, Poelman, Uhlmann, & Banaji, 2009; Forscher, Lai, et al, in press; Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013) yielding predictive validities in the .1 to .3 range. These predictive validities have been described as "socially important," in part, because small effects obtained in short term studies may accumulate over time and across perceivers (Greenwald, Banaji, & Nosek, 2015). On the other hand, small effects do not necessarily actually accumulate, and any conclusion that they do requires empirical evidence; absent such evidence, the "socially important" argument has been criticized as requiring more than a vague reference to what is, in essence, little more than a compound interest formula

(Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015).

We can imagine two areas of productive future research on predictive validity issues. First, can conditions be identified a priori under which IAT scores more strongly predict discrimination? Given the history of IAT conclusions being overstated, the registered replication format will probably be necessary to convince skeptics that conditions under which strong effects are systematically claimed to occur were actually predicted a priori rather than HARKed (Kerr, 1998). Second, do the discriminatory effects predicted by IAT scores actually accumulate over time, as predicted by Greenwald et al. (2015)? Or, like most self-fulfilling prophecies, are they more likely to dissipate than to accumulate (Jussim, 2017)? Addressing this question may not only help design interventions to ameliorate some sources of inequality, by studying their role in naturally-occurring applied settings, it may theoretically advance our understanding of the nature and social effects of implicit beliefs and attitudes.

### **IAT, Implicit Bias, and Racial Gaps**

Given the swirl of uncertainties about what the IAT measures, and its seeming modest ability to predict discrimination, it is probably scientifically premature to even speculate on the role of whatever it measures in producing large-scale social inequalities. Nonetheless, some clearly have:

These studies testify to the enduring historical (and psychological) legacies of racism, sexism, and other forms of group-based inequality and they suggest, not so surprisingly after all, that these legacies have left behind at least some residue in our hearts and minds.” (Jost et al., 2009, p.64)

*“Conclusion 7: Implicit race attitudes (automatic race preferences) contribute to discrimination against African-American Americans. ... a sizable collection of studies summarized in a 2009 journal publication made it clear that there now exists a substantial*

body of evidence that automatic White preference—as measured by the Race IAT—predicts discriminatory behavior even among people who fervently espouse egalitarian views. This evidence is far too substantial to ignore....” (Banaji & Greenwald, 2013, at position 88% of epub).

Neither quote addresses *how much* implicit biases contribute to inequalities. The tone of both strongly suggests quite a lot, but both stop short of stating so explicitly. Thus, in the remainder of this chapter, we review theory and evidence to consider the likely extent of such contributions.

Before continuing, a note on nomenclature. In the remainder of this paper we use the term “IAT scores” to refer to “whatever is measured by the IAT.” It would be awkward to keep repeating “whatever is measured by the IAT.” If there was a simple interpretation of IAT scores, say as “unconscious prejudice” or “implicit bias,” we would just use the term for the underlying construct. However, because what is captured by IAT scores is unclear, using a simple term such as “implicit bias” is not justified. The term “IAT scores” suffers no such uncertainties.

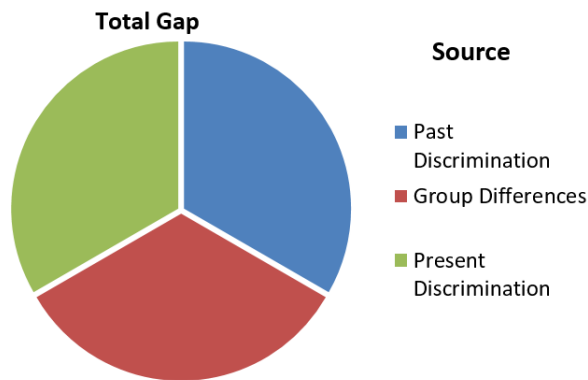
What, then, is scientifically known about the role of IAT scores in producing large scale social inequities? To attempt to understand this, we start from the phenomena that IAT scores might ultimately explain. A wide range of gaps consistently show that, in the U.S., White people to be advantaged over African-American people in income, wealth, higher levels of education, and who live longer (e.g., Arias & Xu, 2018; Fontenot, Semega, & Kollar, 2018; Ryan & Bauman, 2016). In order to try to get some clarity on plausible contributions of IAT scores to gaps such as these, we consider their possible effects in the context of other possible causes of gaps.

To this end, Figure 1 presents a heuristic pie chart of potential contributors to racial inequalities. The pie represents the total gap. Figure 1 identifies three broad sources: Past discrimination, group differences, and discrimination in the present. Although it depicts the contribution of each type of discrimination as equal to the others, this is just for heuristic purposes.

The actual contribution of each may never be knowable, which is why we default to the simple heuristic of treating them as equal. Unless one wishes to argue that one of the three sources of gaps identified here is trivial (which seems highly implausible), whether it contributes 20% or 33% or 50% to current gaps is not important with respect to our analysis. In other contexts when one is uncertain, identifying relevant predictors, weighting them all equally, and adding has proven to be an effective and robust approach to making decisions (Dawes, 1979).

Figure 1:

Where Do These Gaps Come From?



### Past Discrimination

In Figure 1, the blue section represents the portion of a gap due to past racial discrimination. There is ample evidence for the long reach of past discrimination. For example, after analyzing the relationship of proportion of slaves in 1860 to income inequality within counties of 42 U.S. states in 2000, Bertocchi & Dimico (2014) concluded (p. 208) that: "... the legacy of slavery still plays a major role in the US economy and society, since the past use of slave labor persistently affects current inequality..." Similarly, regarding the long reach of slavery and Jim Crow, Loury (1998) concludes: "... for some three centuries now, the communal experience of the slaves and their descendants has been shaped by political, social, and economic institutions that, by any measure

must be seen as oppressive” (p. 41).

Modern IAT scores cannot possibly explain slavery, Jim Crow, or the long legacy of their ugly aftermaths. Furthermore, there was nothing “implicit” about such blatant laws and norms. Even if whatever is measured by IAT scores did cause discrimination in the past, such effects would *still* be considered past discrimination with respect to 2019. IAT scores in the present cannot possibly explain effects of past discrimination, because causality cannot run backwards in time. Thus, to whatever extent past discrimination has produced effects that manifest as modern gaps, modern IAT scores cannot possibly account for such effects.

### **Group Differences in the Present**

In Figure 1, the red section represents reasons groups can differ unrelated to discrimination in the past or present; and, as such, cannot be explained by IAT scores. For example, higher sickle cell anemia rates among American African-Americans than Whites (Centers for Disease Control and Prevention, 2019) may contribute to mortality differences. A culture of honor may be more likely to explain violence among African-Americans than among Whites (Felson & Pare, 2010). There are also a wide range of other differences in beliefs, attitudes, values, habits and other characteristics that could, at least possibly, contribute to racial gaps in wealth and income (Hughes, 2018). Those that result from discrimination do not belong in this section; those that do not result from discrimination do belong in this section.

### **Discrimination in the Present**

**Institutional Discrimination.** IAT scores have the potential to explain individual discrimination in the present. However, not all discrimination in the present is individual. Institutional discrimination refers to laws, rules, norms, and policies of governments and other organizations that might have intentional or unintentional effects of discriminating. Any policy or practice that disproportionately disadvantages people of low SES constitutes institutional racism in

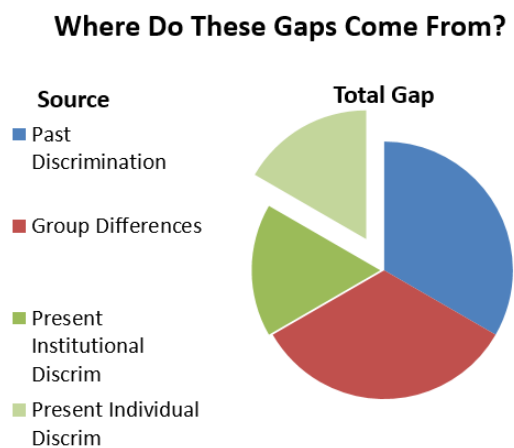


the U.S. because African-Americans tend to be disproportionately poor. States that fund schools through property taxes advantage wealthier communities. Congressional districts that are gerrymandered to reduce African-American political influence constitute institutional racism. Even requiring lawyers in court cases can be viewed as a form of institutional racism, if it is harder for people from low SES backgrounds (disproportionately African-American) to hire good lawyers.

This situation is reflected in Figure 2, which is identical to Figure 1, with one exception. “Discrimination in the present” has been divided into two components: 1. Institutional discrimination (shown in dark green); and 2. Individual discrimination (shown in light green).

Figure 2:

***Individual Discrimination* is a Subset of Total Discrimination**

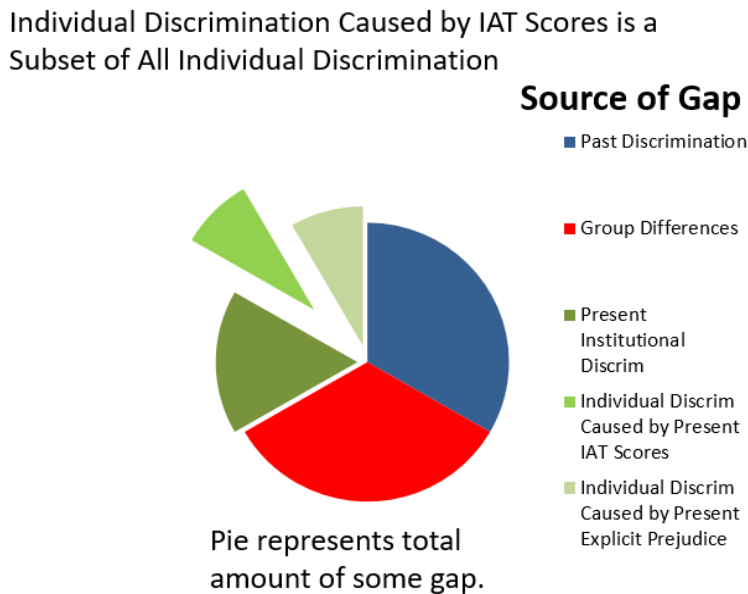


IAT scores can only cause the portion of gaps represented by individual acts of discrimination. Gerrymandering to reduce African-American influence is clearly an explicit, not implicit process. Requiring lawyers in court reflects legal practices developed over centuries in European countries with little or no African presence so the idea that they might reflect implicit racial biases is highly implausible. Nearly all forms of institutional discrimination reflect either intentional biases (e.g., Jim Crow, redlining) or practices that developed for reasons other than race (lawyers, paying for public schools via property taxes).

We know of no evidence linking IAT scores to institutional discrimination. Nonetheless it is possible to hypothesize some role of individual IAT scores in creation of some degree of institutional discrimination. Perhaps, for example, IAT scores reflecting strong associations of African-Americans with unpleasantness, crime, or violence cause individuals to favor politicians claiming to be tough on crime, who then pass laws leading to disproportionate incarceration of African-Americans. Such effects would not be included in the dark green because they start with the IAT scores of individuals (and, as such, would be represented by the light green section for individual discrimination shown in Figure 2).

**Individual discrimination caused by explicit prejudice versus IAT scores.** Of course, not all individual discrimination is necessarily caused by IAT scores. At most, IAT scores reflect implicit associations and perhaps implicit prejudice. Although a review of the literature on links between explicit prejudice and discrimination is beyond the scope of this chapter, much evidence exists demonstrating that explicit racial prejudice predicts discrimination (see e.g., the meta-analysis of 57 studies by Talaska, Fiske, & Chaiken, 2008). To the extent that explicit prejudice causes discrimination, it is contributing to racial gaps. However, explicit prejudice is not what is measured by IAT scores. Therefore, the “Present Individual Discrimination” pie slice in Figure 2 needs to be divided further, as shown in Figure 3, in which the portion of a gap explained by IAT scores is pulled out of the pie.

Figure 3:



**How much individual level discrimination is there in the present?** Whereas there is clear evidence that discrimination at the individual level occurs, exactly how much occurs or how much it explains disparities remains unclear (see Pager & Shepherd, 2008, for a review).

Research on discrimination at the individual level is characterized by conflicting results. Field and audit studies often find evidence of individual level racial discrimination. For example, Quillian et al.'s (2017) meta-analysis revealed that, during the hiring process, White applicants consistently received more interview callbacks than African Americans and Latinos over a 25-year period. Further, hiring discrimination had not declined for African Americans over time. Likewise, Pager et al. (2009) investigated the low-wage labor market in New York City and found that Black applicants were half as likely as equally qualified Whites to receive a callback for an interview or a job offer. Furthermore, Black applicants with no criminal record fared no better than White applicants with a criminal record.

On the other hand, Heckman (1998) argued that audit and field studies on individual discrimination may be less informative than they seem. When the implications of discrimination

audit studies were critically evaluated, Heckman concluded that such studies may overestimate the actual level of discrimination on several grounds:

1. “Discrimination at the individual level is different from discrimination at the group level” (p. 102).
2. “The impact of market discrimination is not determined by the ... average level of discrimination among firms” (p. 102).
3. The real world level of actual discrimination is determined by “... where ethnic minorities or women actually end up buying, working, or borrowing.” (p. 102).
4. “Purposive sorting within markets eliminates the worst forms of discrimination (p. 103).

“Purposive sorting” refers to the idea that people can shift their efforts to contexts (jobs, banks) where they are not likely to be victimized by discrimination. As a result, after reviewing the evidence from discrimination audit studies up to that time, Heckman (1998, p. 105) concluded, “Only a zealot can see evidence in these data of pervasive discrimination in the U.S. labor market.”

Heckman’s (1998) arguments, however, are largely theoretical. More recent empirical data suggest that Blacks and African Americans do not target or avoid particular job types. Instead, they tend to cast a wider net in their job search than similarly situated Whites (Pager & Pedulla, 2015). Thus, having to exert this extra effort is itself an effect of discrimination, and is unlikely to be detected in audit studies. Thus, Heckman’s analysis (1998) may have underestimated the extent to which audit studies capture discrimination (for a more extensive review, see Pager, 2007).

Nonetheless, some empirical studies find little or no discrimination. One audit study sent out 9000 resumes and did not find evidence of racial discrimination in seven major U.S. cities, across six occupational categories (Darolia, Koedel, Martorell, Wilson, & Perez, Arce, 2016). In a similar vein, 17 unpublished studies finding no net bias against African Americans were recently uncovered (Ziggerell, 2018). These samples were nationally representative and included a total of over 13,000

respondents, 2,781 of whom were African American. There was no evidence of bias on a variety of outcomes, including willingness to provide resources, evaluations of competence, and criminal culpability, were assessed.

Some of the clearest evidence of nondiscrimination comes from, surprisingly, work assessing the extent to which the IAT predicts discrimination. First, however, it is important to recognize that the IAT might correlate with some measure of discrimination, even in the absence of net anti-black bias. For example, let's say there is no anti-black bias in some study; bias outcomes range from anti-white to egalitarian. IAT scores could still positively correlate with this outcome, if low IAT scores corresponded to anti-white bias, and high scores to egalitarianism. This is not hypothetical; it is exactly what McConnell & Liebold (2001) found, although this pattern was not reported in their paper. Instead, the pattern was reported in a re-analysis of their data by Blanton et al. (2009).

Another study found no overall racial discrimination among doctors (Green et al., 2007). The study has been cited almost 1000 times according to Google Scholar, and is widely considered a classic in the "implicit bias" literature. For example, it is one of the ten studies "no manager should ignore" according to Jost et al. (2009). Thus, it may be best for us to quote the paper directly to report the largely ignored fact that, overall, they found no bias: "However, participants were **equally likely** to recommend thrombolysis [treatment for blood clots] for black (52%) and white (48%) patients [who complained of chest pain]..." (p. 1234-1235).

Other research assessing the extent to which IAT scores predict discrimination have found similar results indicating egalitarian behavior. For example, a study of trial judges found that: "The judges' determinations [sentencing] were not influenced by race" (Rachlinski, Johnson, Wistrich, & Guthrie, 2009, p. 1214). Yet another study, an experimental laboratory investigation into economic decisions found that "There was no significant difference between the mean [monetary] offers to

black ( $\mu = \$3.74$ ,  $SD = \$1.99$ ) and white ( $\mu = \$3.75$ ,  $SD = \$1.72$ ) partners ...” (Stanley, Sokol-Hessner, Banaji, & Phelps, 2011, p. 7713).

To be clear, we have not performed a comprehensive meta-analysis of IAT predictive validity studies assessing net levels of racial discrimination. Our prediction is that, overall, such research would probably find non-zero net racial discrimination, despite the presence of studies finding no net discrimination and at least one finding net anti-white discrimination (McConnell & Leibold, 2001; Blanton et al., 2009). Regardless, such a meta-analysis (of level of overall racial bias, not predictive validity of the IAT) would be valuable on its merits as yet another way to gain insights into the extent of discrimination.

It is possible that empirical studies may underestimate the amount of discrimination occurring. However, it is a fact that some research finding no discrimination has been published, but, that the narratives around that research have focused on the predictive validity of the IAT rather than the evidence of net nonbias. It is also a fact that forensic work uncovered 17 previously unpublished studies finding no bias; this raises questions about how many other studies finding null bias results remain unpublished.

Thus, it is distinctly possible that there is less individual level discrimination in the present than indicated by our Figure 3. If so, that would warrant the revised heuristic estimate shown in Figure 4.

**Figure 4:**

**Source of Gap:  
If There is Only a Modest Level of Individual Discrimination  
There is Even Less that IAT Scores Can Explain**

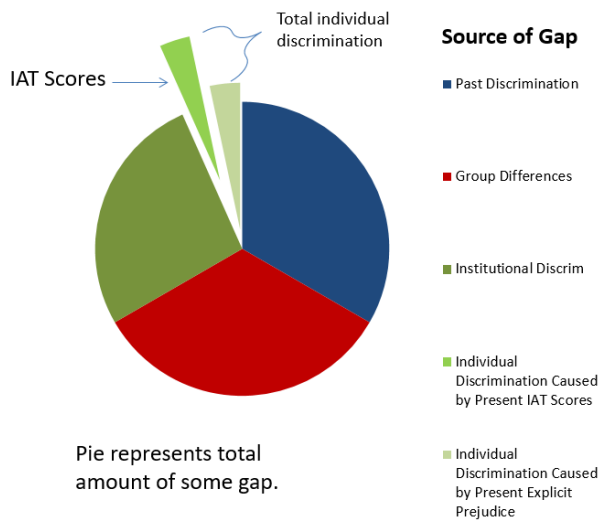


Figure 4 is the same as Figure 3, except that the total proportion of the gap due to individual discrimination has been reduced. It is possible that both Figures 3 and 4 overestimate the contribution of IAT scores to individual discrimination, because they are insufficiently subdivided. Individual discrimination in Figures 3 and 4 is only accounted for by prejudice (either implicit or explicit). Other factors, such as conformity to social norms, can also explain individual acts of discrimination (e.g., Duckitt, 1992). If Figures 3 and 4 fail to capture all sources of individual level discrimination, the proportion capable of being attributed to IAT scores would shrink further, if those other sources were adequately represented. Thus, Figures 3 and 4 probably have a greater risk of overestimating than of underestimating the role of IAT scores in producing individual level discrimination.

Of course, our analysis here was heuristic, not empirical. No one knows how much IAT scores contribute to gaps between African Americans and Whites in the U.S. Discrimination surely plays some role in those gaps; indeed, if one includes the effects of past discrimination and current

discrimination, all of our models, despite their heuristic nature, are consistent with that total contribution being quite substantial. Regardless, our approach may provide some insight into the *potential for IAT scores* to contribute to those gaps, even if that potential cannot be precisely quantified. Unless one assumes that all or most other potential contributors to gaps – the legacy of slavery and Jim Crow, institutional discrimination, explicit prejudice – are trivial, we do not see how the conclusion can be avoided that the role of IAT scores in producing gaps is likely to be highly limited.

### **Conclusions**

Should the IAT be scrapped? No, we are not suggesting that. Like most useful tools, from knives to planes and computers to pesticides, it can be misused. The IAT has many imperfections: dubious construct validity, low test-retest reliability, various other psychometric oddities, its effect size has often been computed in a way that appears to exaggerate its size, it has been almost universally misinterpreted and misrepresented as measuring “implicit bias” when, by Greenwald’s (2017) own definition, it does not do so, its predictive validity has been found to be modest at best, and even if we ignore all that, its ability to account for inequality in the present is likely to be limited.

This does not mean the IAT is useless. We could generate an equally long list for how cars have been misused. It can be constructively used for both self-insight and research mainly by treating the IAT as what it is – at best, a useful but imperfect measure of implicit *associations*. Each of those imperfections, however, can be a starting point for the next phase of research on and with the IAT. Whether and when such associations reflect norms, cultural stereotypes, social realities, or prejudice remains unclear after over 20 years of research on the IAT. Because so much emphasis was placed on the IAT as a measure of “implicit bias” or “unconscious prejudice,” many alternative explanations for its effects received only modest attention. Intensified empirical



attention to those alternatives is sorely needed.

Similarly, the IAT is not necessarily even a pure measure of implicit associations, and may reflect response biases, skill sets, and more. It is also possible that they may reflect different sources in different contexts, and some admixture in yet others. These possibilities, too, can serve as inspirations for future research. It is of course possible that in some contexts and under some circumstances, implicit associations *do* mostly reflect bias and such biases contribute to real world inequalities. That, however, requires more evidence than a mere reaction time difference between two IAT's.

## References

- Arias, E., & Xu, J. Q. (2018). United States life tables, 2015. *National Vital Statistics Reports*, 67(7).
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “Would Jesse Jackson 'fail' the Implicit Association Test?” *Psychological Inquiry*, 15(4), 257-278.  
[https://doi.org/10.1207/s15327965pli1504\\_01](https://doi.org/10.1207/s15327965pli1504_01)
- Baker, A. (2018, July 18). Confronting Implicit Bias in the New York Police Department. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/07/15/nyregion/bias-training-police.html>
- Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. New York, NY: Delacorte Press. (Epub).
- Bertocchi, G., & Dimico, A. (2014). Slavery, education, and inequality. *European Economic Review*, 70, 197-209.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, 94(4), 991–1013.
- Blair, I. V., Steiner, J. F., Fairclough, D. L., Hanratty, R., Price, D. W., Hirsh, H. K., Wright, L. A., Bronsert, M., Karimkhani, E., Magid, D. J., & Havranek, E. P. (2013). Clinicians’ implicit ethnic/racial bias and perceptions of care among Black and Latino patients. *Annals of Family Medicine*, 11(1), 43-52. doi: 10.1370/afm.1442
- Blanton, H., Jaccard, J., & Burrows, C. N. (2015). Implications of the Implicit Association Test D-Transformation for Psychological Assessment. *Assessment*, 22(4), 429–440.  
<https://doi.org/10.1177/1073191114551382>

- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*(3), 567-582. <https://doi.org/10.1037/a0014665>
- Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *Journal of Applied Psychology, 100*(5), 1468-1481. <https://doi.org/10.1037/a0038379>
- Bluemke, M., & Fiedler, K. (2009). Base rate effects on the IAT. *Consciousness and Cognition, 18*(4), 1029-1038. <https://doi.org/10.1016/j.concog.2009.07.010>
- Boutwell, B.B., Nedelec, J.L., Winegard, B., Shackelford, T., Beaver, K.M., et al. (2017). The prevalence of discrimination across racial groups in contemporary America: Results from a nationally representative sample of adults. *PLOS ONE, 12*(8), e0183356. <https://doi.org/10.1371/journal.pone.0183356>
- Cao, J. & Banaji, M. R. (2016). Base rate principle and fairness principle. *Proceedings of the National Academy of Sciences, 113*(27), 7475-7480. <https://doi.org/10.1073/pnas.1524268113>
- Capers, Q., Clinchot, D., McDougle, L., & Greenwald, A. G. (2017). Implicit Racial Bias in Medical School Admissions. *Academic Medicine, 92*(3), 365–369. <https://doi.org/10.1097/ACM.0000000000001388>
- Centers for Disease Control and Prevention (2019). Data and statistics on sickle cell disease. Retrieved from <https://www.cdc.gov/ncbddd/sicklecell/data.html>.
- Chapman, E.N., Kaatz, A. & Carnes, M. (2013). Physicians and implicit bias: How doctors may unwittingly perpetuate health care disparities. *Journal of General Internal Medicine, 28*(11), 1504-1510. <https://doi.org/10.1007/s11606-013-2441-1>

- Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science, 30*(2), 174–192.  
<https://doi.org/10.1177/0956797618813087>
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*(4), 469-487.  
<https://doi.org/10.1037/0022-3514.89.4.469>
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*(1), 68-80. <https://doi.org/10.1037/h0029382>
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2016). Race and gender effects on employer interest in job applicants: new evidence from a resume field experiment. *Applied Economics Letters, 23*(12), 853-856.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist, 34*(7), 571-582.
- Department of Justice, Office of Public Affairs. (2016, June 27). Department of Justice Announces New Department-Wide Implicit Bias Training for Personnel [Press Release]. Retrieved from <https://www.justice.gov/opa/pr/department-justice-announces-new-department-wide-implicit-bias-training-personnel>
- Duckitt, J. H. (1992). Psychology and prejudice: A historical analysis and integrative framework. *American Psychologist, 47*(10), 1182-1193.
- Felson, R.B., & Pare, P. (2010). Gun cultures or honor cultures? Explaining regional and race differences in weapon carrying. *Social Forces, 88*(3), 1357-1378.

- 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*(1), 74-147.  
<https://doi.org/10.1080/10463280600681248>
- Fontenot, K., Semega, J., & Kollar, M. (2018). Income and poverty in the United States: 2017. *U.S. Census Bureau, Current Population Reports*, P60-263.
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B.A. (in press). A meta-analysis of procedures to change implicit measures. *Journal of Personality & Social Psychology*. <https://doi.org/10.31234/osf.io/dv8tu>
- Glaser, J., & Finn, C. (2013). How and Why Implicit Attitudes Should Affect Voting. *PS: Political Science & Politics, 46*(3), 537–544. <https://doi.org/10.1017/S1049096513000656>
- Gonzalez, A. M., Steele, J. R., & Baron, A. S. (2017). Reducing Children’s Implicit Racial Bias Through Exposure to Positive Out-Group Exemplars. *Child Development, 88*(1), 123–130.  
<https://doi.org/10.1111/cdev.12582>
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine, 22*(9), 1231–1238.  
<https://doi.org/10.1007/s11606-007-0258-5>
- Greenwald, A. G. (2017). *Twenty years of research on implicit social cognition*.  
 Presented at the September 2017 National Science Foundation Conference on Implicit Bias.
- Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology, 108*(4), 553-561. <http://dx.doi.org/10.1037/pspa0000016>

- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216. <https://doi.org/10.1037/0022-3514.85.2.197>
- 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. <https://doi.org/10.1037/a0015575>
- Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143*(3), 1369-1392.  
<http://dx.doi.org/10.1037/a0035028>
- Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives, 12*, 101-116.
- Hensch, M. (2016, September 26). Clinton: We must fight “implicit bias.” *The Hill*. Retrieved from <https://thehill.com/blogs/ballot-box/presidential-races/297939-clinton-i-think-implicit-bias-is-a-problem-for-everyone>
- Hughes, C. (2018, July 19). Black American Culture and the Racial Wealth Gap. *Quillette*. Retrieved from <https://quillette.com/2018/07/19/black-american-culture-and-the-racial-wealth-gap/>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581–592.
- Johnson, D. J., & Chopik, W. J. (2019). Geographic variation in the black-violence stereotype. *Social Psychological and Personality Science, 10*(3), 287–294.  
<https://doi.org/10.1177/1948550617753522>

- Jolls, C., & Sunstein, C. R. (2006). The Law of Implicit Bias. *California Law Review*, *94*, 969.
- Jost, J. T. (2019). The IAT Is Dead, Long Live the IAT: Context-Sensitive Measures of Implicit Attitudes Are Indispensable to Social and Political Psychology. *Current Directions in Psychological Science*, *28*(1), 10–19. <https://doi.org/10.1177/0963721418797309>
- Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J. & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39-69.  
<https://doi.org/10.1016/j.riob.2009.10.001>
- Jussim, L. (2017). Accuracy, bias, self-fulfilling prophecies, and scientific self-correction. *Behavioral and Brain Sciences*, *40*. <https://doi.org/10.1017/S0140525X16000339>
- Jussim, L., Crawford, J. T., & Rubinstein, R. S. (2015). Stereotype (in)accuracy in perceptions of groups and individuals. *Current Directions in Psychological Science*, *24*(6), 490–497.  
<https://doi.org/10.1177/0963721415605257>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196-217.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit Measures of Attitudes* (pp. 59–102). New York: Guilford Press.
- LeBel, E. P., & Paunonen, S. P. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*(4), 570–583. <https://doi.org/10.1177/0146167211400619>

- Lick, D. J., Alter, A. L., & Freeman, J. B. (2018). Superior pattern detectors efficiently learn, activate, apply, and update social stereotypes. *Journal of Experimental Psychology: General*, *147*(2), 209-227. <https://doi.org/10.1037/xge0000349>
- Loury, G. C. (1998). An american tragedy: The legacy of slavery lingers in our cities' ghettos. *The Brookings Review*, *16*(2), 38-42. <https://doi.org/10.2307/20080781>
- March, D. S., & Graham, R. (2015). Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Processes & Intergroup Relations*, *18*(1), 89–103. <https://doi.org/10.1177/1368430214542256>
- 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>
- Mundy, L. (2017, April). Why Is Silicon Valley So Awful to Women? *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2017/04/why-is-silicon-valley-so-awful-to-women/517788/>
- Nelson, T. D. (2016). *Handbook of Prejudice, Stereotyping, and Discrimination: 2nd Edition*. New York: NY: Psychology Press.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., ... Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 1–53. <https://doi.org/10.1080/10463280701489053>
- 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>



- Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for future research. *Annals of the American Academy of Political and Social Science*, 609, 104-133.
- Pager, D. & Padulla, D. S. (2015). Race, self-selection, and the job search process. *American Journal of Sociology*, 120, 1005-1054.
- Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, 34, 181-209.
- Pager, D., Western, B., & Bonikowski, B. (2009). Discrimination in a low-wage labor market: A field experiment. *American Sociological Review*, 74, 777-799.
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28(4), 233-248.  
<https://doi.org/10.1080/1047840X.2017.1335568>
- Penner, L. A., Dovidio, J. F., Gonzalez, R., Albrecht, T. L., Chapman, R., Foster, T., ... Eggly, S. (2016). The Effects of Oncologist Implicit Racial Bias in Racially Discordant Oncology Interactions. *Journal of Clinical Oncology*, 34(24), 2874–2880.  
<https://doi.org/10.1200/JCO.2015.66.3658>
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870-10875.
- Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, 84(3), 1195-1246.

- Rae, J. R., Newheiser, A.-K., & Olson, K. R. (2015). Exposure to Racial Out-Groups and Implicit Race Bias in the United States. *Social Psychological and Personality Science*, 6(5), 535–543. <https://doi.org/10.1177/1948550614567357>
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219-235. <https://doi.org/10.1037/0096-3445.118.3.219>
- Rubinstein, R. S., Jussim, L., & Stevens, S. T. (2018). Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology*, 76, 54-70. <https://doi.org/10.1016/j.jesp.2017.11.009>
- Rubinstein, R. S., & Jussim, L. (2019). Stimulus pairing and statement target information have equal effects on stereotype-relevant evaluations of individuals. Manuscript submitted for publication.
- Ryan, C.L., & Bauman, K. (2016). Educational attainment in the United States: 2015. *U.S. Census Bureau, Current Population Reports*, P20-578.
- Sabin, J. A., & Greenwald, A. G. (2012). The Influence of Implicit Bias on Treatment Recommendations for 4 Common Pediatric Conditions: Pain, Urinary Tract Infection, Attention Deficit Hyperactivity Disorder, and Asthma. *American Journal of Public Health*, 102(5), 988–995. <https://doi.org/10.2105/AJPH.2011.300621>
- Schwarz J. (1998, September 29). Roots of unconscious prejudice affect 90 to 95 percent of people, psychologists demonstrate at press conference. *University of Washington News*. Retrieved from <http://www.washington.edu/news/1998/09/29/roots-of-unconscious-prejudice-affect-90-to-95-percent-of-people-psychologists-demonstrate-at-press-conference/>

- Stanley, D. A., Sokol-Hessner, P., Banaji, M. H., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, *108*(19), 7710–7715. <https://doi.org/10.1073/pnas.1014345108>
- Szalavitz, M. (2010, October 11). The Authentic Self: How Do You Know If You're 'Really' Racist or Sexist? *Time*. Retrieved from <http://healthland.time.com/2010/10/11/seeking-the-authentic-self-how-do-you-know-if-youre-really-racist-or-sexist/>
- Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimizing racial discrimination: Emotions, not beliefs, best predict discrimination in a meta-analysis. *Social Justice Research*, *21*(3), 263-296.
- Teachman, B. A., & Brownell, K. D. (2001). Implicit anti-fat bias among health professionals: Is anyone immune? *International Journal of Obesity*, *25*, 1525–1531. <https://doi.org/10.1038/sj.ijo.0801745>
- UCLA Office of Equity, Diversity, & Inclusion. (n.d.). Implicit Bias – UCLA Equity, Diversity & Inclusion. Retrieved May 4, 2019, from <https://equity.ucla.edu/know/implicit-bias/>
- van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap. *American Educational Research Journal*, *47*(2), 497–527. <https://doi.org/10.3102/0002831209353594>
- Walker, P. M., & Hewstone, M. (2007). The influence of social factors and implicit racial bias on a generalized own-race effect. *Applied Cognitive Psychology*, *22*(4), 441-453. <https://doi.org/10.1002/acp.1382>

- White-Means, S., Dong, Z., Hufstader, M., & Brown, L. T. (2009). Cultural competency, race, and skin tone bias among Pharmacy, Nursing, and Medical students: Implications for addressing health disparities. *Medical Care Research and Review*, 66(4), 436–455.  
<https://doi.org/10.1177/1077558709333995>
- Xiao, W. S., Fu, G., Quinn, P. C., Qin, J., Tanaka, J. W., Pascalis, O., & Lee, K. (2014). Individuation training with other-race faces reduces preschoolers' implicit racial bias: A link between perceptual and social representation of faces in children. *Developmental Science*, 18(4), 655-663. <https://doi.org/10.1111/desc.12241>
- Zigerell, L. J. (2018). Black and White discrimination in the United States: Evidence from an archive of survey experiment studies. *Research & Politics*.  
<https://doi.org/10.1177/2053168017753862>