

Stability and Change in Implicit Bias

**Heidi A. Vuletich^{ID}** and **B. Keith Payne**

Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

Psychological Science
2019, Vol. 30(6) 854–862
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797619844270
www.psychologicalscience.org/PS



Abstract

Can implicit bias be changed? In a recent longitudinal study, Lai and colleagues (2016, Study 2) compared nine interventions intended to reduce racial bias across 18 university campuses. Although all interventions changed participants' bias on an immediate test, none were effective after a delay. This study has been interpreted as strong evidence that implicit biases are difficult to change. We revisited Lai et al.'s study to test whether the stability observed reflected persistent individual attitudes or stable environments. Our reanalysis ($N = 4,842$) indicates that individual biases did not return to preexisting levels. Instead, campus means returned to preexisting campus means, whereas individual scores fluctuated mostly randomly. Campus means were predicted by markers of structural inequality. Our results are consistent with the theory that implicit bias reflects biases in the environment rather than individual dispositions. This conclusion is nearly the opposite of the original interpretation: Although social environments are stable, individual implicit biases are ephemeral.

Keywords

implicit bias, bias of crowds, implicit attitudes, IAT, prejudice, open data

Received 5/17/18; Revision accepted 1/22/19

Are implicit biases indelible attitudes ingrained through years of exposure to society's prejudice? Or are they momentary patterns of semantic activation, as changeable as a stream of thought? Questions of stability and change are important for understanding the nature of implicit bias and for prospects of reducing unintended discrimination.

Implicit biases are automatic associations with social groups (Fazio & Olson, 2003; Gawronski & Bodenhausen, 2006). They are considered biases because different associations are linked via social stereotypes to different groups. Even though an association between a group and stereotyped concepts does not imply intentional animosity, implicit biases have been theorized to be a source of (perhaps unintentional) discriminatory treatment (Bargh, 1999; Devine, 1989; Greenwald & Banaji, 1995). Early theories assumed that implicit biases were difficult to change (Bargh, 1999; Devine, 1989; Wilson, Lindsey, & Schooler, 2000). Later research, however, found that implicit tests were malleable in response to mental imagery, shifts in attention and goals, salient exemplars, and other interventions (Blair, 2002; Dasgupta, 2013; Lai, Hoffman, & Nosek, 2013; Payne & Gawronski, 2010). A meta-analysis of procedures intended to change

implicit bias found evidence that scores were malleable when measured on immediate tests, but few studies assessed changes after a delay (Forscher et al., 2017).

The most comprehensive evidence of stability comes from a large-scale experimental study (Lai et al., 2016). Nine interventions were examined using a longitudinal design. Data were collected from 18 university campuses. This study included a pretest measure of implicit race bias, followed by an intervention phase with an immediate posttest, and a follow-up measure completed up to a few days later. The study had a large sample, experimental control, and a longitudinal design, making it the highest quality test of stability and change currently available.

Results of the study indicated that all nine interventions were effective on the immediate test, yet none of the interventions produced a lasting effect after 1 to 2

Corresponding Author:

Heidi A. Vuletich, University of North Carolina at Chapel Hill,
Department of Psychology and Neuroscience, 250 E. Cameron Ave.,
Campus Box #3270, Chapel Hill, NC 27713
E-mail: heidiv@unc.edu

days. These effects have been interpreted as strong evidence that implicit biases are rigid. As Lai and colleagues concluded, “These findings are a testament to how the mind’s prejudices remain steadfast in the face of efforts to change them” (p. 1014). According to this view, individuals’ attitudes snap stubbornly back to their baseline after a short delay.

In this article, we suggest an alternative interpretation of the findings of Lai and colleagues (2016). Rather than reflecting rigid attitudes, the absence of intervention effects following a delay may instead reflect highly malleable individual attitudes, constrained by the stability of social environments. Our reasoning is based on the bias-of-crowds model of implicit bias (Payne, Vuletich, & Lundberg, 2017).

The bias-of-crowds model posits that implicit bias is driven by the cognitive accessibility of concepts linked to social categories. Accessibility refers to the likelihood that a piece of information will be retrieved and used in later processing (Fazio & Williams, 1986; Higgins, 1996; Srull & Wyer, 1979). The accessibility of associations can vary both chronically (as a feature of the person) and situationally (as an aspect of the environment). However, reviews of the literature suggest that measures of implicit bias tend to be temporally unstable (Gawronski, Morrison, Phillips, & Galdi, 2017) and produce small correlations with outcomes as individual-difference measures (Cameron, Brown-Iannuzzi, & Payne, 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). In contrast, aggregate levels of implicit bias across cities, counties, or nations tend to be highly stable and produce large associations with aggregate-level racial disparities (Hehman, Flake, & Calanchini, 2018; Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016; Orchard & Price, 2017). This evidence suggests that accessibility may vary systematically as a function of situations rather than as a result of individual dispositions.

The bias-of-crowds model also posits that implicit biases are largely transient at the level of individuals. For an individual, the concepts most accessible at any moment depend on countless factors, from shared cultural stereotypes to unshared experiences such as recent media exposures and fleeting thoughts. However, when independent observations are aggregated, they function like the wisdom-of-crowds phenomenon in which partial knowledge distributed among many individuals gives rise to stable and accurate aggregate estimates (Surowiecki, 2004). When individual measures of implicit bias are aggregated, the randomly distributed transient influences will be averaged away, whereas shared environmental influences will be sharpened. As a result, the average level of implicit bias in an environment converges on an accurate estimate of

the level of cultural stereotypes and structural inequalities in that environment.

Evidence consistent with the structural-inequality account was reported in a study linking geographical differences in average levels of implicit bias to geographical differences in historical slavery (Payne, Vuletich, & Brown-Iannuzzi, *in press*). The study compared the proportion of the population enslaved, according to the 1860 census, with county-level implicit race bias in more than 1,400 counties from the Project Implicit database (Xu, Nosek, & Greenwald, 2014). Much research suggests that economic dependence on slavery motivated a range of cultural, legal, economic, and ideological reactions aimed at justifying slavery and, later, maintaining the racial hierarchy (Acharya, Blackwell, & Sen, 2018; Anderson, 2016; Rothstein, 2017). Those structural inequalities were hypothesized to be associated with present-day implicit bias. As we predicted, states and counties that were more dependent on enslaved labor in 1860 displayed higher levels of implicit bias more than 150 years later. The association between slave populations and implicit race bias was partially mediated by markers of structural inequalities, including present-day residential segregation and intergenerational mobility.

These findings suggest that aggregate levels of implicit bias reflect structural inequalities in the environment, consistent with the bias-of-crowds model. Applying the model to understand the results of Lai and colleagues suggests that the lack of intervention effects after a delay may not reflect stable individual attitudes, as originally concluded. In a reanalysis of the data, we hypothesized that individual implicit-bias scores would not return systematically to their preintervention level. Instead, we expected individual scores to fluctuate from one time point to another, producing low stability at the individual level. In contrast, we expected the campus-level averages for the posttest to return to levels similar to those at the campus-level pretest. To test whether these mean levels reflect structural inequalities, we examined the association between campus-level means and several markers of structural inequalities.

If our interpretation is correct, it would substantially change the conclusion from Lai et al.’s study. Far from being rigid attitudes, implicit biases could be like the weather: If you want it to change, simply wait a day or two. Average environmental levels of bias, in contrast, may be more like climate. Changes are slow, but even small changes may have immense consequences.

Method

To test our interpretation, we reanalyzed the data from Study 2 by Lai et al. (2016; data can be found at

<https://osf.io/lkp6b/>), who measured implicit racial bias using an implicit association test (IAT). As noted in the original article, the sample was highly powered to detect even very small effects ($N = 4,842$). A participant whose Time 3 score was 7 standard deviations above the mean was excluded from all reported analyses. Results were identical when this participant was included. Critically for our purposes, the researchers collected data from 18 university sites.¹ The nine interventions were randomly assigned across all campuses, so they were uncorrelated with locations. We took advantage of this nested design to compare the stability of individual attitudes and university means across time.

We examined structural inequalities as a potential source of campus-level stability. There is no consensus on how to measure structural inequalities, but researchers typically use geographically based objective measures of present or historical inequalities (Bailey et al., 2017; Krieger, 2012). Obtaining campus-specific measures raises additional challenges because some of the metrics commonly used (such as residential segregation or disparities in income, education, or employment) may not meaningfully apply to students in a university context. We identified three measures meant to capture historical and current inequalities that could plausibly affect the members of university communities today. Because the original data were collected in 2014, we used the closest available data to this year and specify the date for each data source when possible.

Some of the most visible contemporary signs of historical racism are monuments in public spaces. According to the bias-of-crowds model, visible displays of institutional inequalities play a critical role in cuing stereotypic associations. To measure the public display of structural inequality, we coded each campus for whether a Confederate monument was publicly displayed (1 = yes, 0 = no). Data were retrieved from a *Chronicle of Higher Education* database that includes data compiled by both journalists and crowd-sourcing (Bauman & Turnage, 2017). Monuments at The University of Texas at Austin and The University of North Carolina at Chapel Hill were on display in 2014 but were later removed; they are retained in the data to match the year of data collection.

The second structural inequality marker we used was faculty diversity. Underrepresentation of minority faculty may be a visible signal of institutional inequalities. Conversely, a diverse faculty may reflect efforts to actively foster a diverse and inclusive university community. To measure faculty diversity, we coded the percentage of full-time faculty who were non-White at each university on the basis of 2015 data (the closest available to 2014) in a database maintained by *The Chronicle of Higher Education* (2017).

The third marker of structural inequality was a measure of campus-specific social mobility. The data were retrieved from a large-scale study that estimated social mobility for nearly all universities in the United States (Chetty, Friedman, Saez, Turner, & Yagan, 2017). This measure reflects the percentage of students whose parents occupied the poorest income quintile and who made it to the top quintile in adulthood.² The data are based on the 2014 earnings of former college students who attended college during the early 2000s. Although this measure is not race specific, the association between race and socioeconomic status means that low social mobility has a disproportionate impact on members of minority groups.

We collected data on students' SAT scores at each university (we used the 75th percentile because the median was not available) and each university's admissions selectivity (number admitted divided by the number of applications) as control variables. The data, collected in 2017, were accessed from the U.S. Department of Education's Integrated Postsecondary Education Data System (National Center for Education Statistics, 2019). These variables were chosen to help separate measures of institutional inequality from overall selectivity.

Results

As reported by Lai et al., average IAT scores decreased from Time 1 to Time 2 and then rebounded at Time 3. Table 1 displays the descriptive statistics for each campus. To better understand sources of stability and change, we tested three primary hypotheses. The first hypothesis was that individual-level implicit-bias scores would show low stability over time. The second hypothesis was that campus-level means, in contrast, would show high stability. The third hypothesis was that campus-level implicit bias would be associated with markers of structural inequalities.

As displayed in Figure 1, the person-level test-retest correlation between Time 1 and Time 3 was $r(4839) = .25$, $p < .001$, bootstrapped 95% confidence interval (CI) = [.22, .27]. A possible reason for the low test-retest stability is that the interventions may have affected individuals in heterogeneous ways, thus changing their rank orders over time. To test this possibility, we examined the stability of scores among participants in the control condition, who received no intervention. The association between Time 1 and Time 3 scores was identical to those in the larger sample, $r(462) = .25$, $p < .001$, bootstrapped 95% CI = [.15, .34]. The low stability observed here contrasts with the interpretation that individual biases are difficult to change, because individual scores did not generally return to pretest levels. Individual scores changed a great deal, although not systematically.

Table 1. Descriptive Statistics for Campus-Level *D* Scores on the Implicit Association Test (IAT) at Times 1 and 3

Campus	IAT Time 1			IAT Time 3		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Iona College	0.59	0.50	113	0.49	0.41	94
New York University	0.58	0.40	258	0.45	0.44	248
Purdue University	0.60	0.39	126	0.52	0.41	123
Rutgers University	0.52	0.39	207	0.42	0.38	169
San Diego State University	0.55	0.41	326	0.41	0.41	282
Tulane University	0.56	0.42	146	0.45	0.38	142
University of Arkansas	0.59	0.40	165	0.46	0.41	153
University of California, Davis	0.52	0.41	320	0.39	0.43	310
University of California, Irvine	0.50	0.41	361	0.39	0.43	324
University of Florida	0.61	0.42	480	0.48	0.40	429
University of Nebraska–Lincoln	0.53	0.43	492	0.46	0.40	420
University of North Carolina at Chapel Hill	0.62	0.40	323	0.54	0.38	307
University of Texas, Austin	0.63	0.39	135	0.44	0.43	122
University of Virginia	0.59	0.39	769	0.44	0.42	752
University of Virginia's College at Wise	0.55	0.41	86	0.39	0.41	83
Virginia Commonwealth University	0.61	0.41	114	0.49	0.41	107
University of Wisconsin–Madison	0.57	0.39	583	0.46	0.43	561
Yale University	0.58	0.41	266	0.46	0.43	239

Next, we examined the role of environments in creating stable means. To the extent that stability is found in local environments, we expected Time 3 campus means to be predicted by Time 1 campus means. As displayed in Figure 2, the bivariate association was strong, $r(16) = .72$, $p < .001$, bootstrapped 95% CI = [.47, .99].

We further investigated the role of individuals and environments simultaneously by estimating a multilevel model to account for the nested structure of the data. Individual scores (Level 1) were nested within university (Level 2). IAT scores at Time 1 were campus-mean

centered to separate within- and between-campus effects and were used to predict raw Time 3 scores (Raudenbush & Bryk, 2002). Mean-centering scores by campus yielded an unbiased estimate of the individual-level association between Time 1 scores and Time 3 scores, holding campus effects constant. To obtain an estimate of the campus effect, we included site means at Level 2 as a predictor of Time 3 intercepts. The equations for the Level 1 (person-level) model were as follows:

$$\text{IAT Time 3} = \beta_{0j} + \beta_{1j}\text{T1_IATcmc}_{ij} + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2).$$

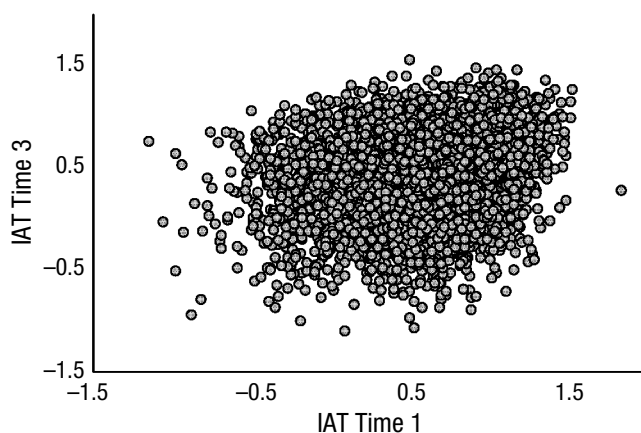
The equations for the Level 2 (campus level) model were as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{T1_site_IAT}_j + \mu_{0j}$$

$$\mu_{0j} \sim N(0, \tau_{00})$$

$$\beta_{1j} = \gamma_{10},$$

where γ is the fixed effect, r is the residual, μ is the random effect, N is the total number of units at Level 1,

**Fig. 1.** Test-retest associations between individual *D* scores from the implicit association test (IAT).

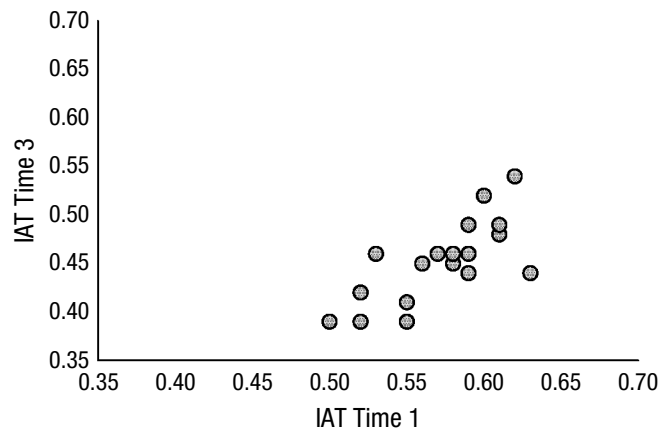


Fig. 2. Test–retest associations between university-mean *D* scores from the implicit association test (IAT).

σ^2 is the Level 1 residual variance, and τ_{00} is the variance of the random effect. T1 refers to Time 1, T3 to Time 3, cmc to campus-mean centered, *i* to individual, and *j* to campus.

The university-level means at Time 1 were strongly predictive of means at Time 3, $\gamma_{01} = 0.87$, $p < .001$, 95% CI = [0.44, 1.29]. This effect was much larger than the stability of individual scores across time, $\gamma_{10} = 0.25$, $p < .001$, 95% CI = [0.22, 0.27]. As can be seen by the confidence intervals, the coefficient for campus means was significantly different from the coefficient for individual scores. Given these comparisons, the high stability in means observed by Lai et al. does not appear to result primarily from individuals reverting to their original attitudes following the intervention. Instead, the stability appears to result from university means returning with high fidelity to earlier university means.

A critical difference between subject-level and campus-level scores is that campus-level scores are aggregated across many individuals. The intraclass correlation coefficient was .01, meaning that campuses accounted for approximately 1% of the variability in individual scores. This small association is consistent with our reasoning that the effect of contexts on individual implicit-test scores may be small and fleeting, and the context-based signal is revealed only when noise is reduced by aggregation across individuals.

Did this aggregation reveal the level of implicit bias on each campus or spuriously create the observed relationship? To examine the effect of aggregation alone, we randomly reassigned all person-level observations to 18 new nominal groups of the same size as those in the original sample, using sampling without replacement (for a similar approach, see Hehman, Calanchini, Flake, & Leitner, 2018). Then, we estimated the multi-level model and repeated this randomization and analysis procedure a total of 100 times.³ If the stable means that we observed in the original sample were caused

by mere aggregation, we would expect the means of the randomized groups to also be stable. But if aggregation instead reveals systematic variance at the group level, then randomizing the groups should reduce the group-level effect. In fact, the campus-level coefficient was greatly reduced, average $\gamma_{01} = 0.26$, 95% CI = [0.21, 0.31]. Interestingly, this coefficient approximated the same value as the person-level effect (0.25). Whereas our original analysis revealed a large campus-level effect and a small person-level effect, randomly assigning the persons to nominal groups left only systematic effects at the person level. The large reduction of the campus-level effect suggests that there was indeed a campus-specific signal that was revealed by aggregation.

Our final analysis was intended to shed light on the nature of the campus-specific signal. We hypothesized that implicit bias would be associated with markers of structural inequalities. As can be seen in Table 2, Time 1 implicit bias was significantly correlated with the three structural-inequality measures. IAT means were higher on campuses with a Confederate monument displayed, $r(16) = .64$, $p = .005$, bootstrapped 95% CI = [.39, .89], and lower on campuses with more faculty diversity, $r(16) = -.48$, $p = .043$, bootstrapped 95% CI = [-.91, -.12], and with greater social mobility, $r(16) = -.61$, $p = .007$, bootstrapped 95% CI = [-1.00, -.25].

To create a more robust measure of structural inequalities, we constructed a composite index by reverse-scoring faculty diversity and mobility so that higher values reflect less diversity and poorer mobility. Then, we averaged the (standardized) variables into a structural-inequality index (Cronbach's $\alpha = .71$). The association between Time 1 implicit bias and the structural-inequality index was strong, $r(16) = .73$, $p = .001$, bootstrapped 95% CI = [.47, 1.00] (see Fig. 3). When we controlled for campus SAT scores and admissions selectivity, the structural-inequality coefficient remained significant and large, $\beta = 0.67$, $p = .006$, 95% CI = [0.65, 0.69].

Discussion

Evidence that the effects of implicit-bias interventions are short-lived has been interpreted as showing that individuals' implicit biases are difficult to change (Lai et al., 2016). We reanalyzed the data by considering both individuals and environments as potential sources of stability. Our analyses suggest that at the individual level, implicit biases were far from permanent. Average bias at the follow-up test, however, was well predicted by the preexisting average at each university. The pre-existing level of implicit bias, in turn, was associated with measures of structural inequality. Consistent with the bias-of-crowds model, fleeting biases at the individual level revealed stable and meaningful estimates of environmental bias when aggregated.

Table 2. Correlations Among Campus-Level Measures of Implicit Bias and Markers of Structural Inequalities

Variable	IAT Time 1	IAT Time 2	IAT Time 3	Confederate monument	Faculty diversity	Social mobility
IAT Time 1	—	.37 [.07, .70]	.72* [.47, .99]	.64* [.39, .89]	-.48* [-.91, -.12]	-.61* [-1.00, -.25]
IAT Time 2		—	.43 [.15, .71]	.17 [-.31, .65]	-.35 [-.75, .02]	-.12 [-.57, .27]
IAT Time 3			—	.37 [.03, .74]	-.29 [-.75, .14]	-.63* [-.92, -.37]
Confederate monument				—	-.41 [-.76, -.05]	-.38 [-.70, -.06]
Faculty diversity					—	.56* [.18, .99]
Social mobility						—

Note: Values in brackets are 95% confidence intervals. IAT = implicit association test.

* $p < .05$.

From a traitlike view of implicit bias, the combination of low stability in individual attitudes with high stability in means is puzzling. But the bias-of-crowds model provides a natural explanation for the return to preexisting averages after a delay, despite low stability in individual attitudes.

The strong campus-level stability is unlikely to be an artifact of aggregation. The same degree of aggregation did not produce high stability when individuals were randomly reshuffled into nominal groups. Instead, aggregation revealed meaningful differences among campus contexts. Those differences were systematically associated with measures of structural inequality, providing further evidence that campus-level biases are meaningful.

Aggregating repeated observations of the same subjects could potentially produce stable estimates of

individual attitudes by revealing chronic accessibility effects. In practice, stable individual scores may require up to a dozen observations per subject (Kurdi & Banaji, 2017). Our core claim in this article is that individual-level implicit bias is not a rigid attitude. The need for many observations to obtain a stable individual measure is consistent with that claim, because it suggests that implicit bias must not be as rigid as was once thought.

Our reinterpretation has fundamental implications for the nature of implicit bias. Most theories of implicit bias assume that it is a property of individuals, such as an attitude, belief, or trait. The context-based perspective, in contrast, suggests that implicit bias is a social phenomenon that passes through individuals like “the wave” passes through fans in a stadium. Rather than a property of individuals, it may more properly be considered a property of social contexts. Most theories assume that a person’s implicit bias is difficult to change. The context-based view, in contrast, suggests that a person’s level of implicit bias is transient and can change as often as the context changes.

Most theories assume that individuals with high levels of bias are substantially more likely to discriminate than those with low levels of bias. The context-based view, in contrast, suggests that certain contexts encourage discrimination more than others, largely independently of the individual decision makers passing through those contexts. In these ways, the context-based view advanced by the bias-of-crowds model reverses core theoretical assumptions about the nature of implicit bias.

Our reinterpretation also has practical implications for reducing discrimination. Namely, changing the social environment may be more effective than changing individual attitudes. Environmental interventions might take either of two forms. Temporary interventions, such as the ones tested by Lai and colleagues,

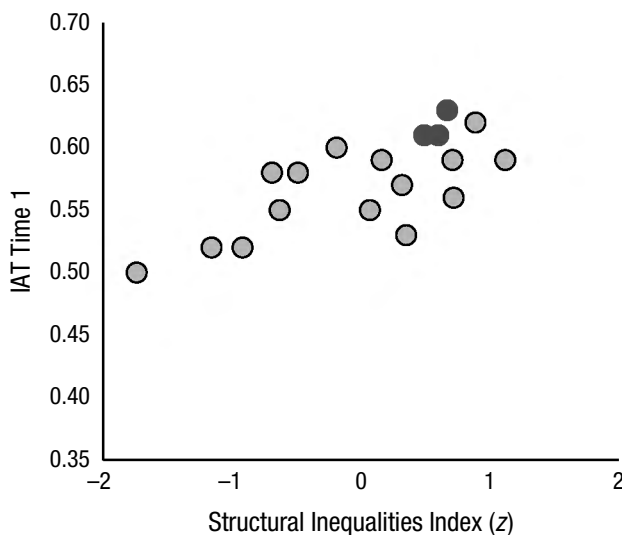


Fig. 3. Average implicit bias for each campus plotted against an index of structural inequalities. IAT = implicit association test.

may be powerful if they are targeted at the time when important decisions are being made. For example, cuing decision makers to think about counterstereotypical thoughts or affirming egalitarian values immediately before making hiring or admissions decisions may reduce unintended bias in those decisions. Rather than changing attitudes, this strategy may modify the concepts that are most accessible in the decision situation. A second approach is to change social environments in more lasting ways. Our findings suggest the hypothesis that removing environmental cues of inequality, such as Confederate monuments, may reduce aggregate implicit bias. Likewise, increasing faculty diversity at universities, or diversity in an organization's leadership more generally, may produce sustained changes in institutional bias.

These ideas remain to be tested, because this correlational design could not establish the causal role of the campus contexts. Nor can it rule out selection effects. However, if the effects were primarily driven by attributes of the students that attracted them to particular colleges, then we would expect to see relatively large person-level effects. Nonetheless, research that randomly assigns individuals to different contexts is an important next step for understanding the causal role of contexts in cuing bias. Another limitation of this study is that, despite the large sample, having only 18 campuses provides relatively low power for comparisons at the campus level.

In the reported research, we measured implicit bias using the IAT, which is only one of multiple implicit tests available and has psychometric limitations (Fiedler, Messner, & Bluemke, 2006). Future research should examine the relative importance of person and context effects using other measures, such as the affect-misattribution procedure (Payne, Cheng, Govorun, & Stewart, 2005) or sequential priming (Fazio, Jackson, Dunton, & Williams, 1995).

Despite these limitations, our research suggests that a core claim in much research on implicit bias may need to be revised. Far from being a rigid attitude, implicit bias is highly transient at the individual level but stable for social contexts. If these findings are confirmed in future research, they suggest that the source of stable implicit bias—and the opportunity for change—is to be found in the places and people around us.


Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

Both authors analyzed the data, wrote the manuscript, and approved the final manuscript for submission.

ORCID iD

Heidi A. Vuletich  <https://orcid.org/0000-0002-8614-4333>

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was funded in part by a National Science Foundation Graduate Fellowship and the Paul and Daisy Soros Fellowships for New Americans (both awarded to H. A. Vuletich) and a National Science Foundation Research Grant (awarded to B. K. Payne).

Open Practices



All data have been made publicly available via the Open Science Framework and can be accessed at osf.io/ng9k8. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619844270>. This article has received the badge for Open Data. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

Notes

1. The original article described 17 sites because the researchers combined two campuses of the University of Virginia, but we separated these because they were in different counties.
2. Social-mobility statistics were not available for Tulane University. However, county-level mobility data were available, and the two were highly correlated across the other campuses, $r(15) = .70, p = .002$. We therefore used county-level mobility to impute campus-level mobility for Tulane.
3. Data and associated code are available at osf.io/ng9k8.

References

- Acharya, A., Blackwell, M., & Sen, M. (2018). *Deep roots: How slavery still shapes southern politics*. Princeton, NJ: Princeton University Press.
- Anderson, C. (2016). *White rage: The unspoken truth of our racial divide*. New York, NY: Bloomsbury Publishing.
- Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., & Bassett, M. T. (2017). Structural racism and health inequities in the USA: Evidence and interventions. *The Lancet*, 389, 1453–1463. doi:10.1016/S0140-6736(17)30569-X
- Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York, NY: Guilford Press.
- Bauman, D., & Turnage, C. (2017, August 22). We're tracking Confederate monuments. Tell us what's on your campus. *The Chronicle of Higher Education*. Retrieved

- from <https://www.chronicle.com/article/We-re-Tracking-Confederate/240967>
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261. doi:10.1207/S15327957PSPR0603_8
- Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, 16, 330–350. doi:10.1177/1088868312440047
- Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility report cards: The role of colleges in inter-generational mobility* (NBER Working Paper No. 23618). Retrieved from <https://www.nber.org/papers/w23618.pdf>
- The Chronicle of Higher Education. (2017, June). *Race, ethnicity, and gender of full-time faculty at more than 3,700 institutions* [Data file]. Retrieved from <https://www.chronicle.com/interactives/faculty-diversity>
- Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and self-concept. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 233–279). San Diego, CA: Academic Press. doi:10.1016/B978-0-12-407236-7.00005-X
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18. doi:10.1037/0022-3514.56.1.5
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027. doi:10.1037/0022-3514.69.6.1013
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327. doi:10.1146/annurev.psych.54.101601.145225
- Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude–perception and attitude–behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology*, 51, 505–514. doi:10.1037/0022-3514.51.3.505
- 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. doi:10.1080/10463280600681248
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2017). *A meta-analysis of procedures to change implicit measures*. Retrieved from PsyArXiv: psyarxiv.com/dv8tu
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692–731. doi:10.1037/0033-2909.132.5.692
- Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, 43, 300–312. doi:10.1177/0146167216684131
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27. doi:10.1037/0033-295X.102.1.4
- 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, 17–41. doi:10.1037/a0015575
- Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2018). *Establishing construct validity evidence for region-level measures of explicit and implicit racial bias*. Manuscript submitted for publication.
- Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological & Personality Science*, 9, 393–401. doi:10.1177/1948550617711229
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York, NY: Guilford Press.
- Krieger, N. (2012). Methods for the scientific study of discrimination and health: An ecosocial approach. *American Journal of Public Health*, 102, 936–944.
- Kurdi, B., & Banaji, M. R. (2017). Reports of the death of the individual difference approach to implicit social cognition may be greatly exaggerated: A commentary on Payne, Vuletich, and Lundberg. *Psychological Inquiry*, 28, 281–287. doi:10.1080/1047840X.2017.1373555
- Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social & Personality Psychology Compass*, 7, 315–330. doi:10.1111/spc3.12023
- 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. doi:10.1037/xge0000179
- Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine*, 170, 220–227. doi:10.1016/j.socscimed.2016.10.007
- National Center for Education Statistics. (2019). *The Integrated Postsecondary Education Data System*. Retrieved from <https://nces.ed.gov/ipeds/use-the-data>
- Orchard, J., & Price, J. (2017). County-level racial prejudice and the Black-White gap in infant health outcomes. *Social Science & Medicine*, 181, 191–198. doi:10.1016/j.socscimed.2017.03.036
- 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, 171–192. doi:10.1037/a0032734
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as

- implicit measurement. *Journal of Personality and Social Psychology*, 89, 277–293. doi:10.1037/0022-3514.89.3.277
- Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1–15). New York, NY: Guilford Press.
- Payne, B. K., Vuletich, H. A. & Brown-Iannuzzi, J. L. (in press). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences, USA*.
- 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, 233–248. doi:10.1080/1047840X.2017.1335568
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.
- Rothstein, R. (2017). *The color of law: A forgotten history of how our government segregated America*. New York, NY: Liveright Publishing Corp.
- Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, 37, 1660–1672. doi:10.1037/0022-3514.37.10.1660
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY: Random House.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126. doi:10.1037/0033-295X.107.1.101
- Xu, K., Nosek, B., & Greenwald, A. G. (2014). Psychology data from the Race Implicit Association Test on the Project Implicit Demo website. *Journal of Open Psychology Data*, 2(1), Article e3. doi:10.5334/jopd.ac