

Persons or Situations? Individual Differences Explain Variance in Aggregated Implicit Race Attitudes

James R. Rae & Anthony G. Greenwald

To cite this article: James R. Rae & Anthony G. Greenwald (2017) Persons or Situations? Individual Differences Explain Variance in Aggregated Implicit Race Attitudes, Psychological Inquiry, 28:4, 297-300, DOI: [10.1080/1047840X.2017.1373548](https://doi.org/10.1080/1047840X.2017.1373548)

To link to this article: <https://doi.org/10.1080/1047840X.2017.1373548>



Published online: 30 Nov 2017.



Submit your article to this journal [↗](#)



Article views: 166



View Crossmark data [↗](#)



Persons or Situations? Individual Differences Explain Variance in Aggregated Implicit Race Attitudes

James R. Rae^a and Anthony G. Greenwald^b

^aDepartment of Experimental Psychology, University of Oxford, Oxford, UK; ^bDepartment of Psychology, University of Washington, Seattle, Washington

Payne, Vuletich, and Lundberg (this issue; hereafter PVL) argue that implicit attitudes are “a social phenomenon that passes through the minds of individuals, but exists with greater stability in the situations they inhabit” (p. 236). To support their argument, PVL analyzed data collected via Project Implicit (Xu et al., 2016) and contrasted the high stability in implicit race attitudes for U.S. states between 2007 and 2016 (unweighted $r = .76$) with evidence showing more modest stability on implicit measures for individual respondents ($r = .42$; Gawronski, Morrison, Phillips, & Galdi, 2017). This commentary presents evidence consistent with the widespread assumption that PVL oppose—that implicit attitudes are characteristics of people, almost certainly more so than a property of situations. Specifically, we present evidence that an individual difference variable—racial identity—explains a large proportion of the variability in state-level implicit race attitudes that PVL attributed to situational determinants (p. 233).





The initial report of the Implicit Association Test (IAT; Greenwald, Schwartz, & McGhee, 1998) revealed the measure’s sensitivity to expected differences in associations between subject populations. In the same laboratory situation, Japanese American and Korean American subjects showed implicit preferences favoring their own ethnic identity. Much subsequent research has documented individual difference correlates of implicit attitudes (e.g., right-wing authoritarianism: Cunningham, Nezlek, & Banaji, 2004; race identity: Nosek et al., 2007). Moreover, some individual differences are so reliably assessed by IAT measures that they are used to test capabilities of proposed improvements of the IAT procedure (e.g., political ideology and gender identity; Greenwald, Nosek, & Banaji, 2003; Sriram & Greenwald, 2009).

Although existing evidence demands a conclusion that individual differences explain variance in respondent-level implicit attitudes, do they also explain variance in aggregated attitudes, such as the state-level aggregates that PVL presented in their Figure 1?

Reanalysis: Stability of U.S. State Mean Race IATs Across a Decade Is Due to Individual Differences Among Respondents Within States

PVL reported the stability of U.S. state mean race attitude IAT scores measured a decade apart, in 2007 and 2016 (their Figure 1). PVL found that state-level mean race attitude IATs for 2007 and 2016 were highly correlated at $r = .694$, a result confirmed in our analysis of the same data.¹ Given the variability in state sample sizes (e.g., the five smallest states averaged 521 respondents and the five largest state samples averaged 20,680 respondents), we also computed correlations weighted by the average sample size for each state combining the years 2007 and 2016. As a further minor departure from PVL’s analysis, we included data from District of Columbia. As shown in our Figure 1A, a weighted analysis including data from District of Columbia also produced a strong correlation between state means, $r(49) = .739$. Data points in Figure 1A are shaded by rank among states for the year 2007, with darker shading indicating higher IAT scores (i.e., stronger automatic White preference).

Our analysis focuses on one individual difference variable, racial identity, that is correlated with IAT-measured implicit race attitude in the Project Implicit archival data. Racial identity was coded using *criterion scaling* of the self-report question that asked subjects to choose an appropriate racial self-identifier (response options: “American Indian/Alaska Native,” “East Asian,” “South Asian,” “Native Hawaiian or other Pacific Islander,” “Black or African American,” “White,” “More than one race—Black/White,” “More than one race—Other,” “Other or Unknown”).² Criterion scaling converts a categorical predictor to a quantitative one by assigning to each response category the mean value on the criterion measure (IAT, in this case) for subjects who chose that category. This coding was done separately for each of the 10 years (2007–2016) of data, although results were very similar when criterion scaling was done after combining data across all 10 years. For a categorical variable such as racial identity, criterion scaling can be more useful than

CONTACT James R. Rae  james.rae@psy.ox.ac.uk  Department of Experimental Psychology, University of Oxford, 15 South Parks Road, Oxford, OX1 3UD, UK; Anthony G. Greenwald  agg@u.washington.edu  Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525, USA

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/hpli.

¹The data reported here (<https://osf.io/52qxl/>) and R code (R Core Team, 2014) used in our analyses (<https://osf.io/azpk7/>) are available on the Open Science Framework.

²Response options differed across years. Response options were equated across time by collapsing across response options indicating a multiracial, other, or unknown racial identity.

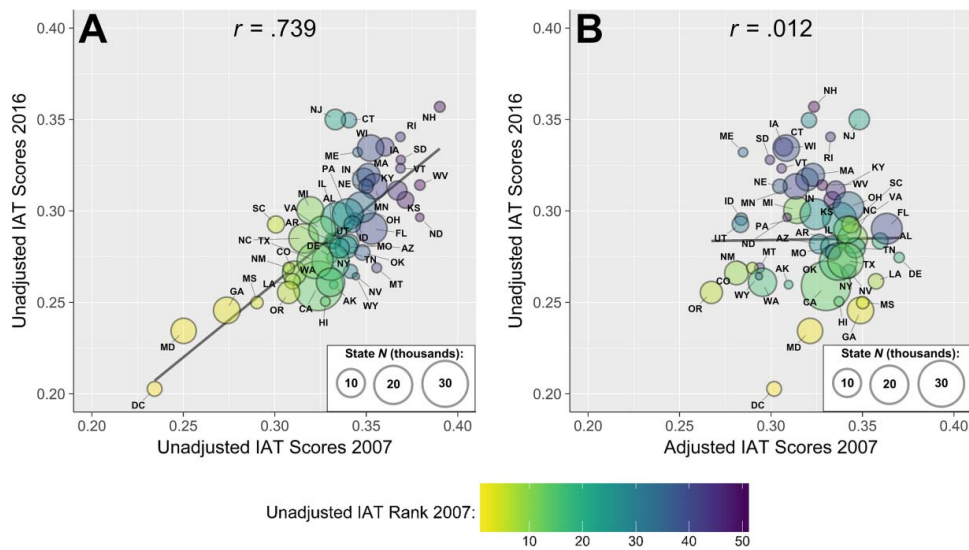


Figure 1. Weighted correlation (by state sample size) between unadjusted state means on the race attitude Implicit Association Test (IAT) in 2007 and 2016 (Panel A) and race-identity-adjusted means for 2007 and unadjusted state means for 2016 (Panel B). *Note.* The size of each data point corresponds to the average state sample size between 2007 and 2016. The color of each data point corresponds to the rank (in ascending order) of state IAT means in 2007.

the familiar contrast coding strategy of dummy coding (Pedhazur, 1997, pp. 501–504). An asset of the criterion-scaling method is that it allows expressing the relationship between the categorical variable and the criterion variable as a correlation coefficient (Levin, Serlin, & Webne-Behrman, 1989). As expected, racial identity (criterion scaled) correlated with scores on the race attitude IAT among all subjects for the year 2007, $r(246,300) = .359$. Surprisingly, this correlation was noticeably weaker in the data for the year 2016, $r(326,380) = .274$.

To test if racial identity explained variance in aggregated state mean IAT scores, we computed adjusted state means for the race attitude IAT for 2007 (using a standard adjustment formula; cf. Pedhazur, 1997, p. 637). The relationship between unadjusted and racial-identity-adjusted means is plotted in Figure 2A. Unexpectedly, this correlation was near zero, $r(49) = .012$. We further tested the consequences of accounting for racial identity in the 2007 data by examining the correlation between racial-identity-adjusted means for 2007 and unadjusted means for 2016. This correlation was also near zero, $r = .012$ (see Figure 1B). More than other results presented in this comment, these observations indicate that the $r = .694$ correlation reported in PVL's Figure 1 does not stand as evidence for systematic effects of situations on implicit race attitudes; it

was wiped out by controlling for a single individual difference variable, racial identity, in the 2007 data.

Results were different for the 2016 data. The correlation of unadjusted versus racial-identity-adjusted state mean IATs for 2016 yielded a substantial positive correlation, $r(49) = .667$ (Figure 2C). After convincing ourselves that we had not miscalculated for either 2007 or 2016, we examined the correlation for an intermediate year, 2012 (Figure 2B), yielding an intermediate value, $r(49) = .209$. Thus, racial identity accounted for a substantial proportion of the variability in state means on the race attitude IAT for each year of data we analyzed, but the explained proportion of variance explained declined through the decade covered by the data set. Although this seems an important observation, its further analysis requires treatment beyond the scope of this comment article.

We next examined the correlation between racial-identity-adjusted state mean IATs for 2007 and racial-identity adjusted state mean IATs for 2016. Figure 3 shows this correlation to be substantial, $r = .615$, which was nevertheless noticeably smaller than the correlation between unadjusted state means for the same two years, $r = .739$ (see Figure 1A). The pattern of light and dark data points, the ordering of adjusted state means varies substantially from the ordering of unadjusted state

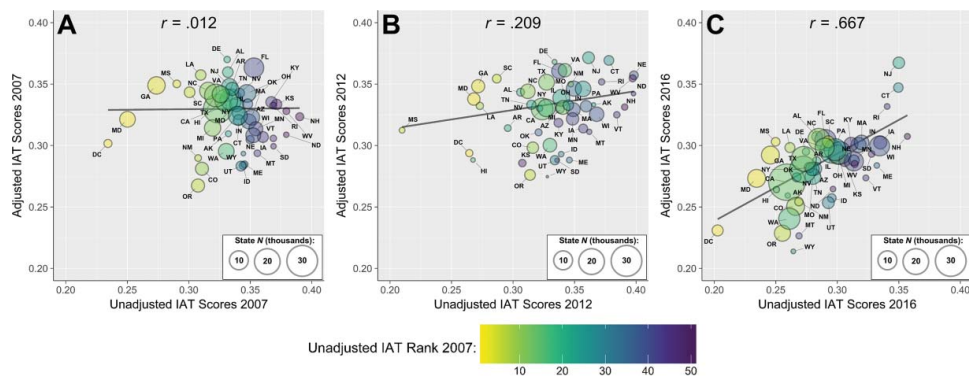


Figure 2. Weighted correlations between unadjusted and race-identity adjusted state means on the race attitude Implicit Association Test (IAT) in 2007 (Panel A), 2012 (Panel B), and 2016 (Panel C). *Note.* The size of each data point corresponds to the average sample size for adjusted and unadjusted means.

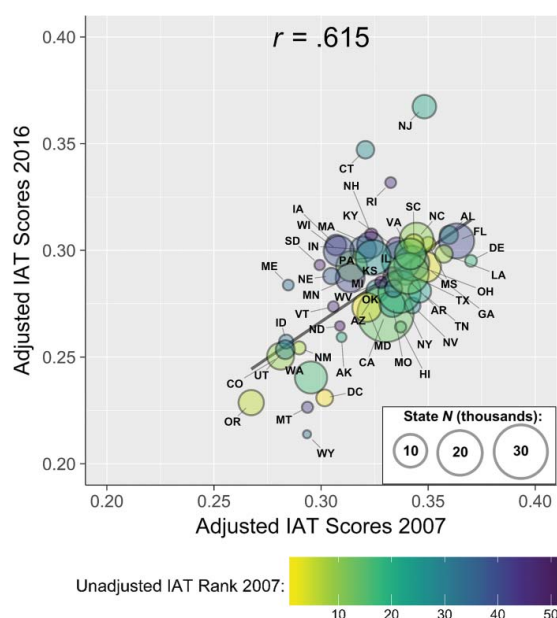


Figure 3. Weighted correlation between race-identity adjusted state means on the race attitude Implicit Association Test (IAT) in 2007 and 2016. *Note.* The size of each data point corresponds to the average state sample size for adjusted means between 2007 and 2016.

means; for example, Maryland had one of the lowest 2007 unadjusted means but an intermediate rank for its 2007 adjusted mean. This result clearly demonstrates that whether states have higher or lower unadjusted means on the race attitude IAT is dictated by the racial identity of the subjects in that state.

None of the other demographic predictors in the Project Implicit data set (e.g., sex, age, religion, political ideology) approached the importance of racial identity as a predictor of the race attitude IAT. The analyses reported in this commentary clearly establish that PVL's Figure 1 is not to be taken as support for their thesis that situation differences are the dominating cause of differences among race attitude IAT scores aggregated at the level of U.S. states. Our Figure 3 analysis indicates that there is variance in state means not accounted for by racial identity but provides no basis for distributing that variance between person and situation influences. Fortunately, there remain as-yet-unexplored possibilities for attempting to explain the unaccounted for shared variance, that is, between the adjusted state means for 2007 and 2016. A remaining task, which we hope PVL will pursue, should be to show how situational variables, such as those reported by the U.S. Census Bureau, can be used in explanatory fashion, parallel to the use of racial identity in this comment. The most convincing analyses will be those showing contributions of situational variables above and beyond known individual difference correlates of implicit attitudes.

Balancing Person and Situation Contributions to IAT Measures

Very few psychologists view human social behavior as being governed exclusively by either person or situation variations. Most consider it reasonable to ask how person and situation variations jointly influence social behavior. PVL nevertheless

declare their strong opposition to the importance of person variations in determining IAT scores ("most of the systematic variance in implicit bias is situational," PVL, this issue, p. 233). Our position that IAT variability reflects a mixture of person and situation influences is based not only on our demonstration that the individual difference dimension of racial identity was responsible for much of the variation in state-level race attitude IAT means in PVL's Figure 1 but also on the following more conceptual considerations.

1. *PVL's Puzzle 1 ("Large and unstable")*: Latency-based measures typically have lower test-retest reliabilities than most self-report measures (e.g., Cunningham, Preacher, & Banaji, 2001). The IAT is generally found to be psychometrically the best among latency-based measures of social-cognitive constructs (e.g., Bar-Anan & Nosek, 2014). Nevertheless, PVL's title for this puzzle accurately indicates that some of the most polarized and pervasive IATs (associated with demographic contrasts such as Black-White, male-female, and young-old) have relatively low test-retest reliability, often less than $r = .50$. Nevertheless, test-retest reliabilities of IAT measures vary widely (see Table 2 in Gawronski et al., 2017).

There are two reasons for variability in IAT test-retest reliability estimates. First, IATs associated with demographic variations derive from widely shared cultural experiences that should shape similar attitudes and stereotypes in many people. To the extent that people have the same attitudes or stereotypes, test-retest correlations will necessarily be relatively weak. A sample of 50% Black and 50% White respondents will have considerably greater variability, and accordingly greater test-retest reliability, than will all-White or mostly White samples (see Rae & Olson, 2017), which characterize most published reports of test-retest reliability of that IAT. Second, few test-retest reliability studies have tested IATs with near-zero means and substantial variability, such as ones assessing consumer brand preferences or political party preferences; these IATs often have test-retest reliabilities in the vicinity of $r = .70$ (e.g., Bar-Anan & Nosek, 2014; Sriram & Greenwald, 2009, for the Brief IAT).

2. *PVL's Puzzle 2 ("Permanent yet unstable")*. PVL characterized this puzzle by juxtaposing (a) modest test-retest reliability of IATs with (b) stable sample mean values of IATs across extended periods ("if one's biases are not stable across a month, how can they be stable across a lifetime?"; PVL, this issue, p. 234). Their statement of this concern indicates that they view situational causation as incompatible with (rather than as coexisting with) person causation, for which they provided no compelling justification.

3. *PVL's Puzzle 3 ("Places and people")*. PVL observe that IAT scores aggregated across people predict behavior better than do individual scores. PVL take this observation to imply that the effect of aggregation is entirely due to situational properties of the geographic region shared by those people. PVL do not acknowledge that aggregation in the form of repeated measurement of the same individual also improves prediction. Repeated administration of IAT measures is rare in research settings but has sometimes been used (Sriram & Greenwald, 2009, for the Brief IAT). Averages of repeated measures result in increased test-retest reliability. Just as it was incorrect to assume that state-level aggregates represented entirely

situational influences, it is similarly unjustified to assume that other aggregations are entirely situational, reflecting no effects of individual differences.

4. *Informative comparison with blood pressure measurement.* The sphygmomanometer is the familiar device used to provide a measure that (a) is subject to multiple situational influences and (b) has no more than modest test–retest reliability but (c) is nevertheless highly valued as an individual difference measure of blood pressure. The well-recognized situational influences include location (hospital vs. home), prior activity (rest vs. exercise), breathing (deep vs. shallow breathing), smoking versus not, or eating versus fasting. In research uses of sphygmomanometers, researchers routinely deal with the situational influences by aggregating across two or more measurements to produce a single observation (e.g., Perati et al., 2010; Stergiou et al., 2002).³

5. *Implication that test–retest reliability is to be interpreted as the square root of percentage of variance explained by an individual difference construct.* PVL assert that “test–retest correlations for implicit measures [averaging] $r = 0.42$ [suggest] that less than 20 percent of the variability in implicit bias can be explained by an individual’s level of implicit bias a few weeks earlier” (p. 234). Their assertion might be taken by a reader to suggest that the percentage of individual-difference construct variance explained by an IAT measure is indexed by the square of the IAT’s reliability (in this case $.42^2 \approx 18\%$). However, the test–retest coefficient *itself* (not its square) corresponds to the percent of construct variance explained (Cohen, Cohen, West, & Aiken, 2003, p. 56).⁴ PVL’s assertion would have been accurate had they stated that 18% of the measure’s variance was predicted by the *measure* obtained a few weeks earlier, rather than the “individual’s level of implicit bias a few weeks earlier” (p. 234).

Conclusion

The present authors agree that IAT measures, like most other psychological measures, have situational influences. These include an effect of order of combined tasks that depends on the number of trials in the block that gives practice in the reversal of sides for two of the IAT’s four categories, enhanced polarization of an IAT measure if it is the very first IAT experienced by a subject, an effect of instructions to respond slowly to the easier of two combined tasks (often used as a faking instruction), and effects of a host of interventions that temporarily alter the cognitive representation of one of the IAT’s four categories. Importantly, there are no such situational interventions that have been established to have durable effects on IAT measures (Lai et al., 2016). The few existing claims of durable effects await convincing empirical confirmations. For PVL, a step toward establishing the validity of situational influences on the race attitude IAT could be for them to identify situational (rather than demographic) variables that predict some of the variance in the racial-identity-adjusted measures plotted in this comment article’s Figure 3.

Acknowledgments

We thank Nils Karl Reimer for helpful input on plotting results of our data analyses.

References

- Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, 46, 668–688. doi:10.3758/s13428-013-0410-6
- Casiglia, E., Tikhonoff, V., Albertini, F., & Palatini, P. (2016). Poor reliability of wrist blood pressure self-measurement at home. *Hypertension*, 68, 896–903. doi:10.1161/HYPERTENSIONAHA.116.07961
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis in the behavioral sciences* (Third Edition). Mahwah, NJ: Erlbaum.
- Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, 30, 1332–1346. doi:10.1177/0146167204264654
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163–170. doi:10.1111/1467-9280.00328
- Gawronski, B., Morrison, M., Phillips, 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., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi:10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216. doi:10.1037/0022-3514.85.2.197
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... & Simon, S. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, 145, 1001–1016. doi:10.1037/xge0000179
- Levin, J. R., Serlin, R. C., & Webne-Behrman, L. (1989). Analysis of variance through simple correlation. *The American Statistician*, 43, 32–34. doi:10.1080/00031305.1989.10475605
- 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, 36–88. doi:10.1080/10463280701489053
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. Belmont, CA: Thomson Learning.
- Perati, G., Stergiou, G. S., Asmar, R., Biko, G., de Leeuw, P., Imai, Y., ... Mancía, G. (2010). European Society of Hypertension Practice Guidelines for home blood pressure monitoring. *Journal of Human Hypertension*, 24, 779–785. doi:10.1038/jhh.2010.54
- Rae, J. R., & Olson, K. O. (2017). Test-retest reliability and predictive validity of the implicit association test in children. *Developmental Psychology*.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, 56, 283–294. doi:10.1027/1618-3169.56.4.283
- Stergiou, G. S., Baibas, N. M., Gantzarou, A. P., Skeva, I. I., Kalkana, C. B., Roussias, L. G., & Mountokalakis, T. D. (2002). Reproducibility of home, ambulatory, and clinic blood pressure: Implications for the design of trials for the assessment of antihypertensive drug efficacy. *American Journal of Hypertension*, 15, 101–104. doi:10.1016/S0895-7061(01)02324-X
- Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K., Bar-Anan, Y., Uman-sky, E., ... Smith, C. (2016). *Project Implicit Demo Website Datasets* [Data file]. Retrieved from <https://osf.io/52qxl/>

³The test–retest reliability reports of averages of eight measures by Stergiou et al. (2002) indicated (applying the Spearman–Brown prophecy formula in reverse) that the test–retest reliability of single measurements of blood pressure in their study ranged from $r = .43$ to $r = .56$.

⁴The statement in the Cohen et al. (2003) text was: “the reliability coefficient equals the proportion of the observed score variable that is true score variance.”