American Psychological Association

Journal of Experimental Psychology: General

0096-3445/19/$12.00

2019, Vol. 148, No. 6, 1022–1040
http://dx.doi.org/10.1037/xge0000623

Establishing Construct Validity Evidence for Regional Measures of Explicit and Implicit Racial Bias

Eric Hehman
McGill University

Jimmy Calanchini
University of California, Riverside

Jessica K. Flake
McGill University

Jordan B. Leitner
Facebook, Menlo Park, California

Large-scale data collection has enabled social scientists to examine psychological constructs at broad, regional levels. However, because constructs and their measures initially operationalized at the individual level may have qualitatively and quantitatively different properties at other levels of analysis, the validity of constructs must be established when they are operationalized at new levels. To this end, the current research presents evidence of construct validity for explicit and implicit racial bias at region levels. Following classic measurement theory, we examine the substantive, structural, and external evidence of construct validity for regional biases. We do so with responses from ~2 million Black and White North Americans collected over 13 years. Though implicit measures typically demonstrate low retest reliability at the individual level, our analyses reveal conventionally acceptable levels of retest reliability at the highest levels of regional aggregation. Additionally, whereas previous meta-analyses find relatively low explicit–implicit correlations at the individual level, the present research uncovered strong explicit–implicit correlations at regional levels. The findings have implications for how we interpret measures of racial bias at regional levels.

Keywords: intergroup dynamics, racial bias, stereotypes, prejudice

Supplemental materials: http://dx.doi.org/10.1037/xge0000623.supp

Over the past decade, large-scale data collection has created new opportunities for social scientists. In the past, data were primarily collected through small, controlled experiments in laboratories on university campuses recruiting undergraduate psychology students as participants. Now, advances in technology facilitate the collection of massive amounts of data from diverse populations and locations. Such vast data open up new opportunities for exploration, theory building, and hypothesis testing.

Recent work using such large-scale approaches has revealed a number of insights into human behavior. For instance, more introverted people prefer mountainous regions (Oishi, Talhelm, & Lee, 2015), personality "fit" with a city is associated with greater self-esteem (Bleidorn et al., 2016), and people tend to migrate toward "ideological enclaves" occupied by others sharing their political values (Motyl, Iyer, Oishi, Trawalter, & Nosek, 2014). In tandem with these empirical findings, complementary theoretical perspectives posit how regional variation in individual-level psychological constructs such as attitudes and personality traits might emerge, persist, and be expressed in diverse outcomes (Oishi & Graham, 2010; Rentfrow, Gosling, & Potter, 2008). These insights would not have been possible without data from broad populations over wide regional areas.

The Need for Ongoing Construct Validation

As researchers expand their work to include different levels of analysis and diverse groups of people, they are faced with critical questions about the validity of their measures. In psychology, constructs are often latent in nature, and cannot be directly observed. To study these latent constructs, we develop measures to assess them and gather evidence that the measures capture the constructs of interest. This process of construct validation is a fundamental part of psychological science (Cronbach & Meehl, 1955), constraining how phenomena are studied and what claims can be made about them.

Eric Hehman, Department of Psychology, McGill University; Jimmy Calanchini, Department of Psychology, University of California, Riverside; Jessica K. Flake, Department of Psychology, McGill University; Jordan B. Leitner, Facebook, Menlo Park, California.

All authors developed the ideas. Eric Hehman and Jordan B. Leitner compiled the data. Eric Hehman and Jordan B. Leitner analyzed the data. All authors wrote the article. We thank Sam Gaertner for his comments on an early versions of this article. This research was partially supported by a SSHRC Insight Development Grant (430-2016-00094) to Eric Hehman and Jessica K. Flake, and an Alexander von Humboldt Research Fellowship to Jimmy Calanchini.

Correspondence concerning this article should be addressed to Eric Hehman, Department of Psychology, McGill University, 845 Sherbrooke Street West, Montreal, Quebec, Canada H3A 0G4. E-mail: eric.hehman@mcgill.ca

A tenet of modern validation theory is that construct validity does not pertain to a measure itself, but to the interpretation of the scores a measure yields (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Validity evidence for a measure of a construct is limited to a specific use or purpose (Kane, 2013). Consequently, when an established measure is used in a new context (e.g., for a different purpose, with a different population), new evidence is needed to assess whether the previously established interpretation is valid in the new context.

The present work focuses on a construct regularly studied at the individual level by psychological scientists: racial bias. The construct validity of racial bias aggregated at regional levels cannot be inferred from evidence of construct validity established at the individual level. Consequently, we respond to this need by exploring construct validity evidence of measures of racial bias, initially developed to measure individual differences in racial bias, at levels of regional aggregation. To study the racial biases of a region is a fundamentally different research endeavor than to study the racial biases of an individual, in that different mechanisms (e.g., psychological, structural) may be involved in racial biases at each level of analysis.

Loevinger (1957) categorized the process of construct validation into three phases, around which we organize the present work: substantive, structural, and external. The substantive phase comprises the theoretical underpinnings of a construct, in which previous literature is used to define it, outline its scope, and describe the necessary content required for reasonably measuring it. The structural phase includes quantitative analyses, examining the psychometric properties of the measure (e.g., factor structure, reliability). The final, external phase focuses on evidence of how the measure relates to other measures of similar constructs and predicts criteria, placing the measure in a larger nomological network (Cronbach & Meehl, 1955).

This process must proceed sequentially. If the theoretical and substantive foundation underlying a measure is tenuous, then any results generated from that measure are suspect. Accordingly, we begin with describing the substantive, structural, and external validity evidence that has previously been established with explicit and implicit racial bias measures at the individual level. Using the construct validity of individual-level racial bias as a starting point, we then review substantive evidence for regional racial bias, and present new structural and external evidence for these measures. Finally, we integrate this evidence with existing theoretical perspectives to provide a foundation for valid interpretation of measures of regional racial bias.

Explicit and Implicit Racial Bias at the Individual Level

Substantive

Though our focus is on implicit and explicit racial bias operationalized to regional levels, we begin by reviewing the extant literature on individual-level racial bias. Explicit attitudes have traditionally been conceptualized as reflecting deliberate mental processes that are available through conscious introspection, and are typically measured using self-report questionnaires (Dovidio, Hewstone, Glick, & Esses, 2010; Gawronski & Bodenhausen, 2006). In contrast, implicit[1] attitudes have been conceptualized as reflecting mental processes that occur unintentionally and outside of conscious awareness, and are often measured relatively indirectly from the speed and/or accuracy of responses rather than from the contents of responses per se (Fazio, Jackson, Dunton, & Williams, 1995; Gaertner & Mclaughlin, 1983; Gawronski, 2009; Greenwald, McGhee, & Schwartz, 1998). For excellent reviews on the nature of implicit and explicit biases, their interpretation at the individual level, and the relationships between them, see Cunningham, Zelazo, Packer, and Van Bavel (2007); Gawronski and Creighton (2013); Greenwald, Banaji, Rudman, Farnham, Nosek, and Mellott (2002); Rydell and McConnell (2006); Strack and Deutsch (2004), and Wilson, Lindsey, and Schooler (2000). Though both explicit and implicit measures have been used to assess a wide variety of attitudes, the present work focuses specifically on explicit and implicit White–Black racial attitudes.

Structural

Explicit measures of individual racial bias, such as feeling thermometers, often demonstrate relatively high retest reliability compared to implicit measures (and cognitive tasks more generally) which often suffer from lower retest reliability (Hedge, Powell, & Sumner, 2018). Various implicit measures have been devised, which vary considerably in their procedures and psychometric properties. Some, such as the affect misattribution procedure (Payne, Cheng, Govorun, & Stewart, 2005), are well-validated and have been used to investigate a variety of attitudes (Cameron, Brown-Iannuzzi, & Payne, 2012). The most widely used and well-validated implicit measure is the Implicit Association Test (IAT; Greenwald et al., 1998), and the Project Implicit demonstration website has collected IAT data from millions of respondents across the globe for nearly two decades. Given the favorable psychometric properties of the IAT relative to most other implicit measures (Bar-Anan & Nosek, 2014), and the wealth of available IAT data, the present research focuses on the IAT as our operationalization of implicit racial bias.

Retest reliability. In the context of racial bias, various investigations into the retest reliability of the IAT have revealed reliabilities of r = .42 across a 2-month interval of measurement (Gawronski, Morrison, Phills, & Galdi, 2017), r = .31 across a 2-week interval (Cunningham, Preacher, & Banaji, 2001), and r = .45 across a 1-hr interval (Bar-Anan & Nosek, 2014). In contrast, the retest reliability of explicit racial bias measures is typically relatively high, for example, r = .78 (Gawronski et al., 2017). Retest reliability is assumed to indicate the extent to which a measure does not change over time. Consequently, such large differences in the reliabilities of each measure of bias have hindered attempts not only to understand the qualitative nature of each form of bias, but also how they are related to one another. One interpretation of these differences in retest reliability is that ex-

[1] In this literature there is some conflicting use of terminology regarding implicit measures. In the present work, we use the term "implicit" to refer to indirect measures, in contrast to explicit or direct measures. Additionally, because responses on a measure are not synonymous with the mental construct the measure is intended to assess, throughout this article we differentiate between tools developed to assess implicit attitudes ("implicit measures") and the underlying latent constructs being measured ("implicit bias").

plicit bias is a stable, trait-like construct but implicit bias is a state-like construct that is not stable over even short amounts of time. Another interpretation of these differences is that a substantial proportion of variance in implicit bias scores reflect measurement error. Consistent with the latter interpretation, latent variable modeling has identified substantial measurement error in implicit measures (Cunningham, Nezlek, & Banaji, 2004; Cunningham et al., 2001).

Explicit–implicit correlations. Meta-analyses and reviews routinely report a relatively narrow range of correlations between explicit and implicit measures of racial bias: r = .14 (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013), r = .24 (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), r = .25 (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), r = .31 (Nosek et al., 2007), r = .35 (Bar-Anan & Nosek, 2014). One interpretation of these relatively low explicit–implicit correlations is that the relationship between these two measures should be much stronger if both were capturing the same construct. Indeed, from the perspective of dual-representation models (e.g., Wilson et al., 2000), low explicit–implicit correlations are taken as evidence that the two measures capture distinct constructs.

However, low correlations do not provide unambiguous evidence that explicit and implicit racial bias are distinct constructs. Correlations between measures are inextricably linked to the reliability of each, because reliability provides an upward bound on the extent to which measures can correlate (Nunnally, 1970; Spearman, 1904). Consequently, relatively low explicit–implicit correlations could be an artifact of the low reliability of implicit measures, rather than evidence that they are distinct constructs.

External

The external phase of construct validity for individual-level explicit and implicit racial bias focuses on relationships with other variables. External validity is primarily demonstrated through examining how explicit and implicit racial bias predict relevant behavioral outcomes.

Meta-analyses have generally found that individual-level explicit and implicit racial bias predict outcomes to a similar extent, though a small degree of independent contributions are observed. Greenwald and colleagues' meta-analysis (Greenwald et al., 2009) revealed that implicit racial bias explained 4% of additional variance in outcomes beyond what is explained by explicit racial bias, and that explicit racial bias explained .08% of additional variance in outcomes beyond what is explained by implicit racial bias. Oswald, Mitchell, Blanton, Jaccard, and Tetlock (2013) reanalyzed these data and obtained similar estimates, with implicit bias uniquely explaining 2% of additional variance in outcomes, and explicit bias uniquely explaining .9% of additional variance. A recent and larger meta-analysis by Kurdi et al. (2018) came to a similar conclusion, that the predictive validity of the IAT of intergroup discrimination domains was relatively small, and that explicit and implicit measures of racial bias provide unique and roughly equivalent incremental validity.

Explicit and Implicit Racial Bias at Regional Levels

To our knowledge, there has been no systematic investigation to date into the construct validity of explicit and implicit racial bias beyond the individual level. It may seem intuitive that the validity of an individual-level construct would be similar at other levels (e.g., county, state, nation), but this is not necessarily true. *The Standards* advises that construct validation is necessary whenever measures are used for different purposes than those for which they were initially validated. Additionally, the ecological fallacy (Selvin, 1958), also known as Simpson's paradox, indicates that relationships between variables can differ across levels of analysis. Examining data in a clustered fashion can reveal different relationships at different levels of aggregation, which could influence multiple dimensions of construct validity. We illustrate such a possibility schematically in Figure 1. In this fictional example, suppose that we are interested in the relationship between income and number of car accidents. When examining this pattern at one regional level (e.g., between states), we find that state income is positively related to car accidents. However, when examining this pattern at a different regional level (e.g., within states), a different relationship emerges, such that income is negatively related to car accidents.

The ecological fallacy is not proof that a relationship at one level of analysis does not persist to other levels. Instead, it demonstrates the hazard of *assuming* correspondence across levels of analysis. Consequently, the extent to which the well-established characteristics of explicit and implicit bias observed at the individual level persist across other levels remains an open question. Importantly, when we aggregate bias, the unit of measurement shifts from the individual to the region (e.g., county, state). Different mechanisms may underpin the racial biases of individuals (e.g., social desirability concerns) versus the racial biases of regions (e.g., residential segregation) and, therefore, properties of explicit and implicit bias at these levels may also differ. Thus, establishing substantive predictions, structural evidence, and relationships with external variables in the three phases of construct validity is required for any researcher seeking to understand bias (or any other construct) at regional levels (Stapleton, Yang, & Hancock, 2016).

Proposed Interpretation of Regional Racial Bias

Building on the existing evidence, we believe that regional aggregates of implicit and explicit racial biases of individuals

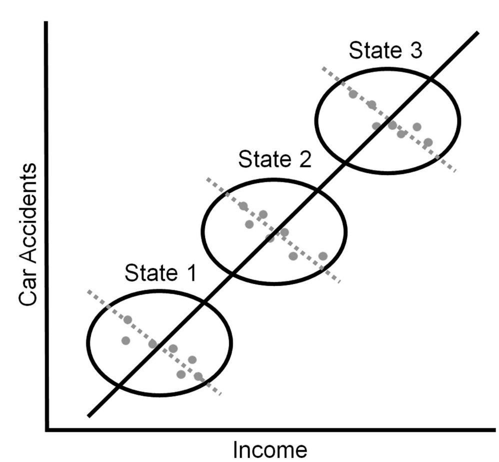

Figure 1. Schematic illustration of the ecological fallacy.

are best conceptualized as two distinct measures capturing a single latent factor reflecting an average, or collective, psychological predisposition of one subgroup (e.g., White people) toward another subgroup (e.g., Black people). We hypothesize that the level of this construct varies regionally, and is both a cause and consequence of various regional characteristics. We present evidence for interpreting these measures as racial bias, including their relationship at regional levels and how they contribute to a nomological network and predict theoretically relevant outcomes.

The Present Research

Adopting this construct validation approach, we begin by defining the three levels of analysis upon which we focus: county, core-based statistical area (CBSA), and state. At each of these levels, we will examine explicit and implicit racial bias for Black and White people separately. We first describe the nature of the data. In a substantive phase, we then review theory regarding regional racial bias. In a structural phase, we examine the psychometric properties of regional explicit and implicit bias, and examine their relationship. Finally, in an external phase, we describe existing literature on the relationship between regional explicit and implicit racial bias, and provide evidence of how regional explicit and implicit racial bias relate to additional theoretically relevant outcomes.

Source of Data

We calculated racial bias at the different regional levels by aggregating individual-level data from Project Implicit (see Figure 2).

Project Implicit is a nonprofit demonstration website that has been measuring explicit and implicit bias over the Internet since 2002 for education and research purposes. In a publicly available dataset, 6,624,119 unique session IDs were recorded from voluntary respondents for measures of racial bias toward White and Black people over 13 years (Xu, Nosek, & Greenwald, 2014). We examined these data at three different levels of regional aggregation: county, CBSA, and state. CBSAs are defined by the U.S. Office of Management and Budget as areas of at least 10,000 people and adjacent areas that are socioeconomically linked with an urban center by commuting. We included Washington, DC and excluded territories. These three levels of resolution are commonly adopted by researchers examining phenomena across geographies.

Respondents to the Project Implicit demonstration website reported their explicit attitudes by completing two feeling thermometers in which they separately rated how warm or cold they felt toward Black and White people (0 = *very cold* to 10 = *very warm*). Explicit bias was calculated by subtracting responses on the Black feeling thermometer from responses on the White feeling thermometer, consistent with past regional work (Hehman, Flake, & Calanchini, 2018; Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016a, 2016b). This operationalization of explicit bias

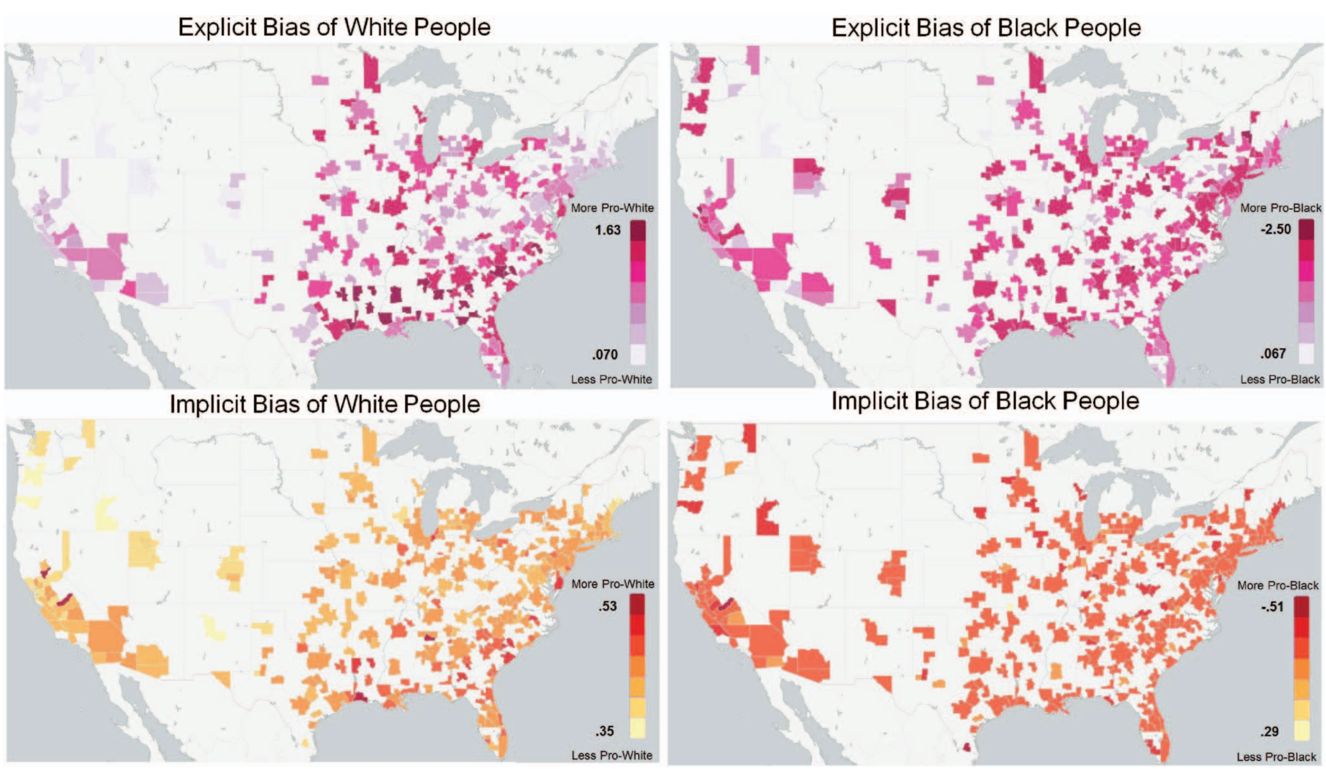

Figure 2. Average explicit bias of White and Black people of each core-based statistical area included in the present analyses (averaged across 2002–2015). Scale range differs for each group. Color determined by Jenks optimization. See the online article for the color version of this figure.

provided the largest number of respondents for analysis relative to other measures of explicit bias available in the Project Implicit dataset. Additionally, several meta-analyses of individual-level bias have found larger implicit-explicit correlations when both measures were operationalized relatively rather than absolutely (Kurdi et al., 2018; Oswald et al., 2013). Implicit-explicit correlations were greater when both measures compared behavior or judgment toward the IAT's two contrasted groups (i.e., target concepts) versus when one measure contrasts two groups but the other measure focuses only on one group. Consequently, operationalizing explicit racial bias as the relative difference between two feeling thermometers does not introduce a methodological artifact that could potentially suppress correlations between implicit and explicit regional racial bias.

Implicit racial bias was assessed with a Black/White racial prejudice IAT (Greenwald et al., 1998), a speeded dual-categorization task in which respondents simultaneously categorized social targets (i.e., pictures of Black and White people) and attributes (i.e., pleasant and unpleasant words) by timed computer-key press. The speed with which people respond to one set of target-attribute pairings (e.g., White-Good and Black-Bad) relative to the other set of pairings (e.g., White-Bad and Black-Good) is thought to reflect the strength with which the target categories are associated with one versus the other attribute category. Implicit bias was calculated according to the recommended *D* scoring algorithm (Greenwald, Nosek, & Banaji, 2003). Both the explicit and implicit bias measures were calculated such that more positive values represent more positive attitudes toward White relative to Black people.

We used data only from respondents who were based in the U.S., were either Black or non-Hispanic White, had geographic information available, and had both explicit and implicit racial bias data available. We additionally excluded respondents with response latencies faster than 300 ms on 10% or more of trials, as recommended by Greenwald and colleagues (Greenwald et al., 2003). These criteria left 1,461,861 White and 272,088 Black people across 3,098 counties, 414 CBSAs, and 50 states plus Washington, DC who completed these measures between 2002 and 2015.

Project Implicit respondents visit the website voluntarily and, thus, do not constitute a sample that is representative of the U.S population. That said, we can examine the extent which this Project Implicit sample corresponds to the U.S. population based on the demographic information Project Implicit collects. The Project Implicit respondents are younger (median = 23.0) relative to the general public (median = 38.1), and more likely to be female (59.4% vs. 49.8%). Previous research has found that the percentages-by-region of Black and White respondents in the Project Implicit sample strongly correlate with local racial demographics (r = .910, $p < .0001$, 95% CI [.878, .935]; Hehman et al., 2018). This strong correlation indicates that the proportions of people from different racial groups in the Project Implicit sample covary with the racial proportions of the regions in which they are located, but is not definitive evidence that this sample is representative of the U.S. population. That is, though the proportions correlate, the mean proportions might differ substantially. In the General Discussion section we return to the issue, and discuss how (un)representativeness of our data might influence the generalizability of our findings.

Section 1: Substantive

Because regions are comprised of individuals, it may seem logical to assume that region- and individual-level bias have the same theoretical underpinnings. From this perspective, regional bias may reflect the aggregate of individual-level biases of that region. For example, social impact theory (Latané, 1981) posits that local clustering of attitudes and beliefs can occur when individuals engage in repeated social interactions. However, regional bias may not reflect *simply* the aggregate of individual biases. Instead, several perspectives (e.g., Oishi & Graham, 2010; Rentfrow et al., 2008) propose recursive relationships in which individual-level attitudes and beliefs become manifest in social structures (e.g., laws, institutions) that, in turn, influence the attitudes and beliefs of the individuals in that region. Not only can such recursive relationships perpetuate biases, but they can also lead to emergent phenomena that are qualitatively distinct from the sum of their individual inputs (Smaldino, 2014).

As an additional challenge to assumptions of correspondence between region- and individual-level bias, Payne, Vuletich, and Lundberg (2017) recently proposed a novel conceptualization of implicit racial bias. Building upon the perspective that biases can be perpetuated through social structures, they conceptualize implicit bias as a relatively stable property of contexts, rather than individuals. This interpretation is consistent with the low retest reliabilities observed in the literature (Bar-Anan & Nosek, 2014; Cunningham et al., 2001; Hofmann et al., 2005), in that individual-level implicit bias should vary across situations, rather than exist as a stable property of the individual.

Not only do individual and regional racial bias potentially differ in terms of underlying mechanisms, but research investigating individual versus regional racial bias also differs for methodological reasons. To date, research on individual racial bias has been complicated, in part, by the low retest reliability of implicit measures, relative to conventionally accepted standards for reliability (i.e., $\alpha \geq .70$) as well as relative to the reliability of explicit measures (e.g., Bar-Anan & Nosek, 2014; Cunningham et al., 2001; Gawronski et al., 2017). The extent to which one measure can correlate with another is limited by the reliability of each measure (Nunnally, 1970; Spearman, 1904). Consequently, relatively low observed correlations between individual implicit racial bias and behavioral outcomes, and between individual implicit and explicit racial bias, (e.g., Greenwald et al., 2009; Oswald et al., 2013) are not unambiguous evidence of low correspondence between constructs; instead, these correlations may be downwardly biased because of the low reliability of individual implicit racial bias measures. In contrast, we expect our investigation into the construct validity of regional racial bias to be less affected by issues of reliability because aggregate measures of any construct are inherently more reliable than individual measures (Rushton, Brainerd, & Pressley, 1983). Error is associated with any measurement, but with aggregation errors are averaged away, allowing for the true magnitude of underlying relationships to be observed.

Section 2: Structural

In the structural phase of examining the construct validity of regional racial bias, we first focus on retest reliability, before moving to examine how implicit and explicit bias are related to one another at different levels of regional aggregation. The IAT *D* score is derived from an algorithm in which response latencies are aggregated by block, transformed, and subtracted from one another (Greenwald et al., 2003). In terms of racial bias, the resulting difference score is operationalized to reflect relative preference for White versus Black people. Because we were interested in contrasting the structural validity of explicit and implicit bias at regional levels, we also examined explicit bias conceptualized as a difference score between attitudes toward White and Black people. Relative explicit measures correspond more strongly with IAT *D* scores than do absolute measures of bias (Hofmann et al., 2005), and so we considered this an appropriate comparison. That said, one limitation of this operationalization of explicit bias is that difference scores generally remove reliable variance, and have lower reliability than do absolute measures (Cohen, Cohen, West, & Aiken, 2013).

Several well-powered studies have examined the retest reliability of explicit and implicit bias at the individual level (Gawronski et al., 2017; Nosek, Greenwald, & Banaji, 2005). However, research examining retest reliability at regional levels is limited. One challenge of examining retest reliability is the assumption that the construct under study is stable during the time interval. Schmidt and Nosek (2010) examined changes over time in U.S. respondents' racial bias operationalized at the national level and found little to no change over the course of 2.5 years, suggesting retest reliability could be a useful indicator of reliability of bias as a region-level construct. However, their analytic approach focused on the U.S. as a whole, and included individuals of all races, potentially occluding interregion and interracial variability. For instance, country-level mean bias would remain the same if Texas had a sharp rise in bias while California had a correspondingly sharp decrease. Further, even substantial variability in the biases of Black Americans over time might be masked because Black people constitute a relatively small percentage of the U.S. population. Very recently, Payne et al. (2017) found that implicit bias operationalized at the state level demonstrates high retest reliability (rs ~ .6–.7). In the present section on the structural validity of regional racial bias, we extend this examination of retest reliability to both explicit and implicit bias for different-sized regional units, and for White and Black people separately.

Analytic Approach

We aggregated explicit and implicit racial bias each year at the regional unit. Though time is continuous, binning the data over the 13-year collection period provided an intuitive way to examine year-to-year variability in bias. To quantify retest reliability, we estimated Pearson's correlation coefficients between each subsequent year (e.g., correlating Year X with Year X + 1) across all 13 years. Because we were interested in the average retest reliability across all time points, we then averaged these correlations. We calculated these correlations separately for explicit and implicit racial bias, and for White and Black people, at each regional level of analysis (see Figure 3). Our sample of White Project Implicit respondents is significantly larger than our Black sample. Sample size influence metrics of reliability, so we report two parallel sets of analyses: one on the full data set, and another using size-equated samples. Data and code for all analyses is available here: https://osf.io/3jz6x/.

Full Sample Results

White explicit. At the county level, average retest reliability for White people's explicit racial bias across all 13 years was low (M_r = .058, SD = .040). Reliability was higher when aggregated at the CBSA level (M_r = .275, SD = .151) and much higher at the state level (M_r = .865, SD = .054). In other words, retest reliability for White people's explicit racial bias is essentially zero at the county level, meaning that a county's explicit racial bias in one year does not account for any variance in explicit racial bias in the subsequent year, on average. However, retest reliability consistently improved as the region increased in size, such that a state's explicit racial bias accounts for 74.8% of the variance in explicit racial bias measured in the subsequent year, on average.

White implicit. The average retest reliability of White people's implicit racial bias similarly improved as the regions became larger. Again, reliability was low at the county level (M_r = .025, SD = .032), but was higher at CBSA (M_r = .171, SD = .113) and state levels (M_r = .693, SD = .156). This state-level retest estimate replicates the findings reported by Payne et al. (2017).

Black explicit. The retest reliability of Black people's explicit racial bias did not improve as dramatically as did White people's explicit racial bias as regions became larger. Reliability was consistently low at the county (M_r = .050, SD = .042), CBSA (M_r = .056, SD = .090), and state levels (M_r = .203, SD = .280). Even at the highest level of aggregation, Black people's explicit racial bias at the state level in one year explained only 4.1% of the variance in explicit racial bias measured in the subsequent year, on average.

Black implicit. Similarly, the retest reliability of Black people's implicit racial bias was consistently low across county (M_r = .032, SD = .040), CBSA (M_r = .029, SD = .050), and state levels (M_r = .171, SD = .211).

Sample-Size Equated Results

In the full-sample analyses, White people's explicit and implicit racial biases at all levels of analyses were more reliable than those of Black people. Taken at face value, this pattern of results suggests that the regional racial biases of White people are more reliable over time than the regional racial biases of Black people. Yet reliability depends in part on sample size, and there are far more White than Black people in our sample. For instance, of the 1,461,861 White and 272,088 Black Project Implicit respondents in our sample, there were on average 28,664 White versus 5,335 Black people per state, and this discrepancy persists at all regional levels. Because the total number of Black people in our sample was 15.7% of the number of White people in our sample, we controlled for sample size by randomly sampling 15.7% of the White people in each regional unit into a smaller dataset, thereby creating a new dataset equivalent in size with the dataset from Black people. Then, we recalculated retest reliability of the explicit and implicit racial bias of White people at each regional level.

White explicit. The average retest reliability of White people's explicit racial bias still improved with larger units of aggregation, but the magnitude of correlations was noticeably lower in

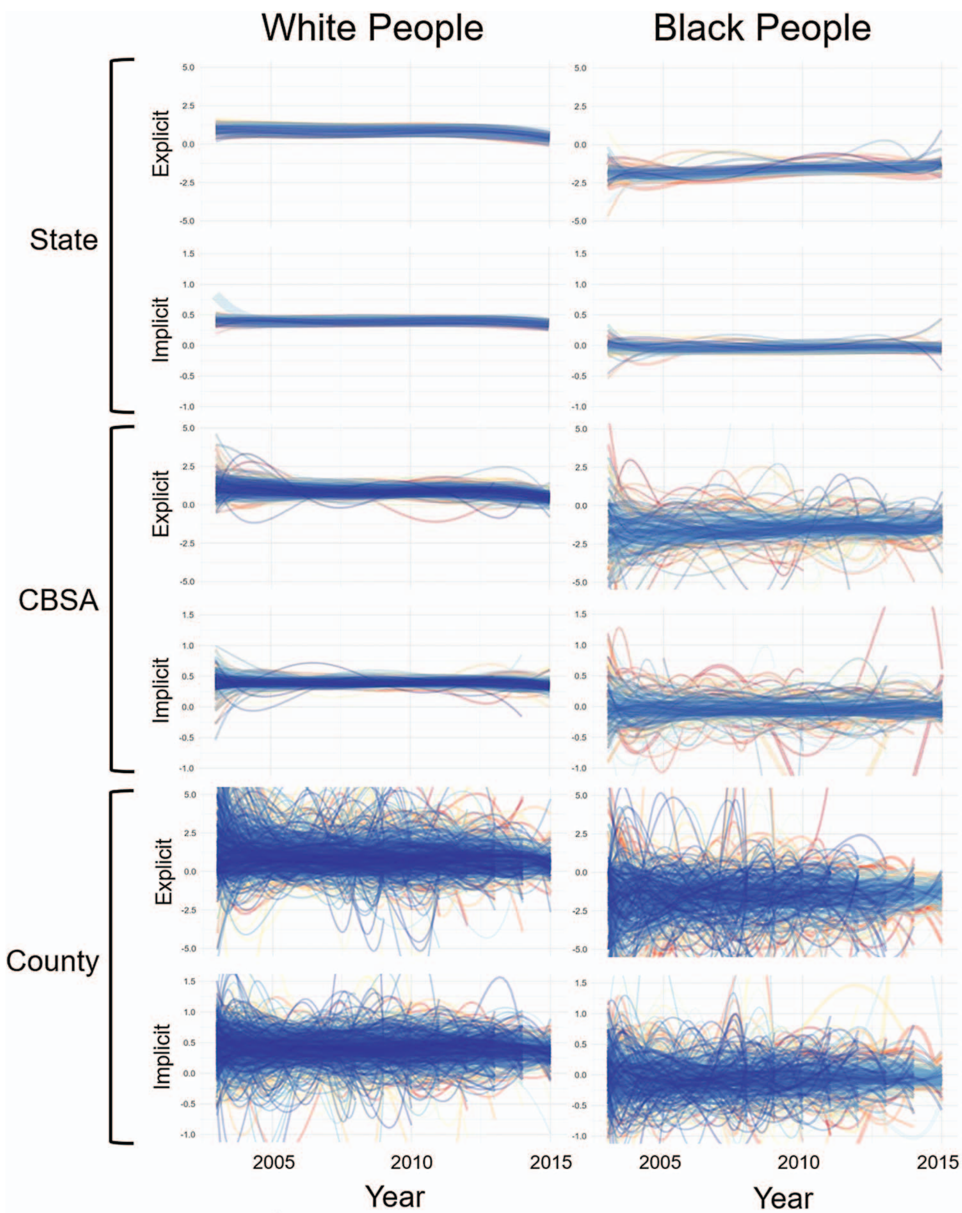

Figure 3. Variation in mean explicit and implicit racial bias over time in at different regional levels for the full samples of White and Black Project Implicit respondents. Each line represents a distinct regional unit (i.e., individual counties in the county analysis; individual core-based statistical areas (CBSAs) in the CBSA analysis; individual states in the state analysis). See the online article for the color version of this figure.

this smaller sample than in the full sample. Reliability was low at the county (M_r = .030, SD = .030), higher at the CBSA (M_r = .131, SD = .069), and highest at the state level (M_r = .504, SD = .284).

White implicit. A similar pattern was evident with White people's implicit racial bias: The magnitude of correlations was lower in this smaller sample than in the full sample. Again, reliability was low at the county (M_r = .017, SD = .029), higher at the CBSA (M_r = .071, SD = .054), and highest at the state level (M_r = .212, SD = .261).

Retest Discussion

In this structural phase of examining the construct validity of regional explicit and implicit racial bias, we examined retest reliabilities separately for White and Black Project Implicit respondents at various levels of analysis. Across different regional levels, the year-to-year retest reliabilities of the explicit and implicit racial biases of both White and Black people was relatively low, in comparison with individual level explicit and implicit racial bias (Bar-Anan & Nosek, 2014; Cunningham et al., 2001; Gawronski et al., 2017), as well as in comparison to conventionally accepted levels of retest reliability (i.e., $r \geq .7$). Only at the state level did the regional explicit and implicit racial bias reach conventionally acceptable retest reliability, and only in the case of White people.

Additionally, these analyses highlight the importance of sample size when examining psychometrics such as retest reliability. What initially appeared to be a robust effect of greater regional racial bias retest reliability among White relative to Black respondents was subsequently revealed to largely be due to sample size differences. Nevertheless, given that retest reliability depends on sample size, the large retest reliabilities in the full White data set ($r_{explicit} = .865$, $r_{implicit} = .693$) offers a glimpse of the level of reliability that is possible for measures of racial bias, given a sufficiently large sample.

Relationships Between Explicit and Implicit Bias

We next tested the relationships between explicit and implicit bias at different regional levels.

Analytic approach and description of data. Five-thousand bias-corrected bootstrapped correlations at the county, CBSA, and state levels estimated the relationship between explicit and implicit bias for Black and White people separately. Data and code for all analyses is available here: https://osf.io/3jz6x/. The distributions of White and Black people's explicit and implicit bias are plotted in Figure 4, which generally indicate that there is greater variance in Black than White people's biases, as two-tailed F tests confirm. The explicit bias of Black people was consistently more variable than that of White people at each level of geography (all Fs $>$ 2.02, all ps $< .015$). The same was true for the implicit bias of Black relative to White people at the county and CBSA levels (all Fs > 5.13, all ps $< .001$). However, at the state level, the implicit bias of Black people was only marginally more variable than the implicit bias of White people, $F = 1.74$, $p = .052$, 95% CI of ratio of variances [.994, 3.053]. Taken together, the explicit and implicit biases of Black people are more variable than those of White people at each regional level.

Results for White people. We began by estimating the individual-level correlation between Project Implicit respondents' explicit and implicit racial bias. For White people, the correlation between explicit and implicit bias was $r = .215$, 95% CI [.214, 217], which is consistent with, but on the lower end of, what has been observed in previous research (e.g., Bar-Anan & Nosek, 2014; Greenwald et al., 2009; Hofmann et al., 2005; Nosek et al., 2007; Oswald et al., 2013). Turning to the relationship between regional explicit and implicit bias, we observed a similar correlation at the county level, $r = .267$, 95% CI [.181, .323], but stronger correlations at the CBSA, $r = .772$, 95% CI [.722, .812], and state levels, $r = .846$, 95% CI [.730, .908]. In other words, a state's

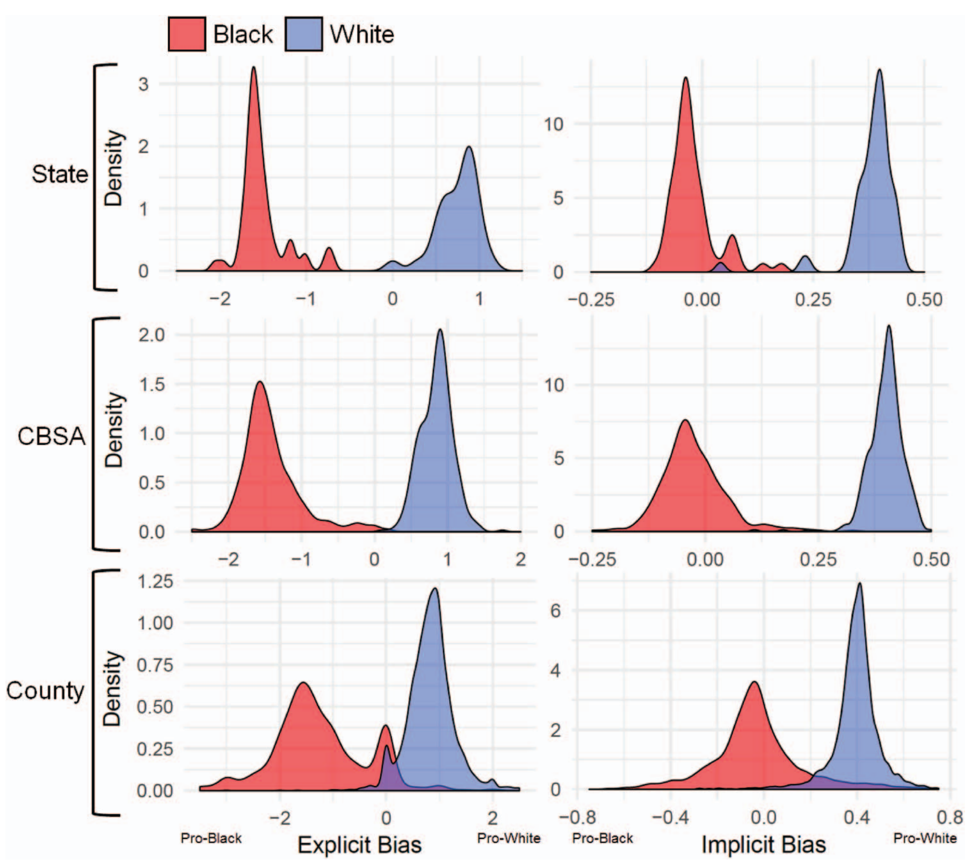

Figure 4. Density plot distributions of Black and White people's explicit and implicit bias at county, core-based statistical area, and state levels. See the online article for the color version of this figure.

implicit bias explains 71.6% of variance in its explicit bias (see Figure 5).

Though our goals were primarily descriptive, we used two-tailed Fisher r-to-z transformations to compare correlation coefficients across levels of analysis. Because aggregation increases the reliability of a measure (Rushton et al., 1983), and reliability constrains potential correlations with other measures (Nunnally, 1970), we had anticipated that correlations might increase with greater levels of aggregation. Our analyses generally supported this prediction. The relationship between explicit and implicit bias was stronger at the county than individual level (z = 3.07, p = .002), and at the CBSA than county level (z = 14.32, p < .001), though only descriptively larger at the state than CBSA level (z = 1.42, p = .156). Taken together, the relationship between White people's explicit and implicit bias was strengthened as the level of regional aggregation increased.

Results for Black people. The relation between the explicit and implicit bias of Black people at different levels of geography demonstrated a similar, but not identical, pattern to that of White people. The individual-level correlation between explicit and implicit bias was low, r = .149, 95% CI [.145, 153], and remained low at the county level, r = .216, 95% CI [.148, .286], as well as

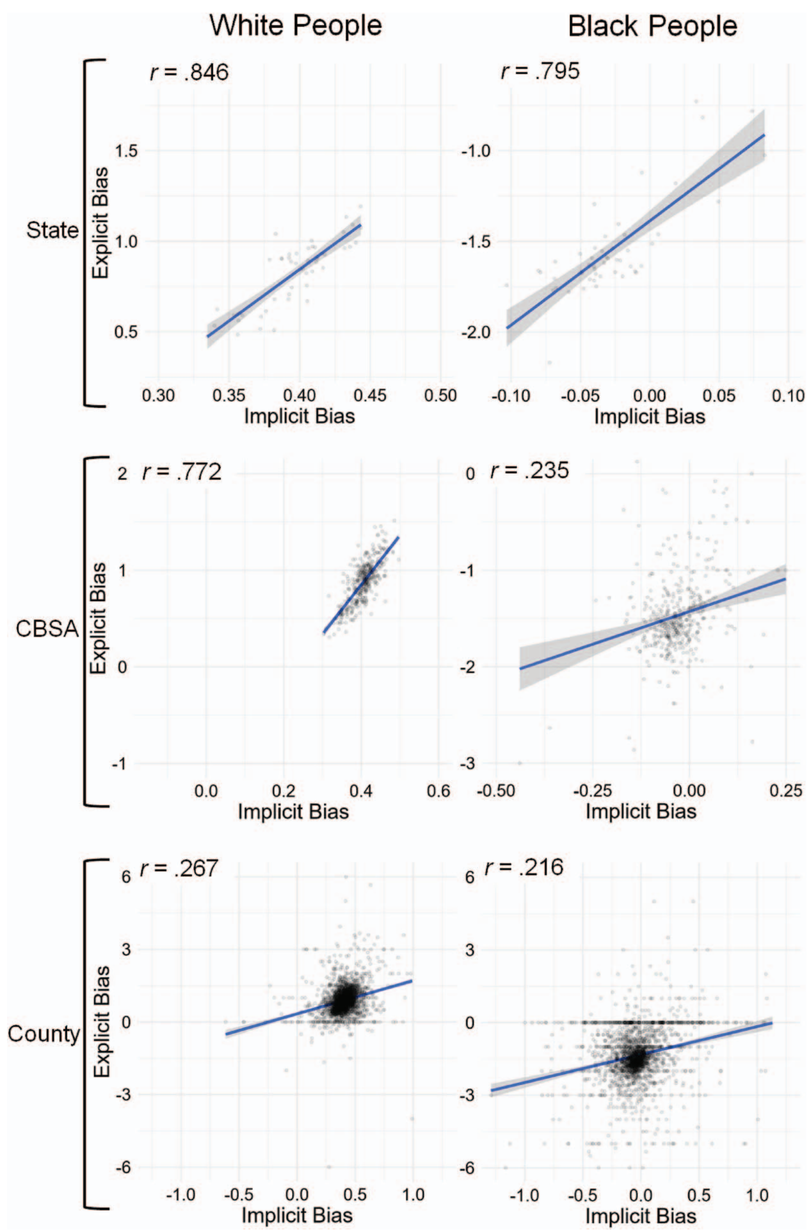

Figure 5. The relationship between explicit and implicit bias at different regional levels for Black and White people. At each regional level, the absolute values of the x- and y-axes differ, but the range of the axes represented for White and Black people is equivalent to allow for comparison. CBSA = core-based statistical area. See the online article for the color version of this figure.

the CBSA level, r = .235 95% CI [.041, .402]. However, at the state level, the relationship between explicit and implicit bias for Black people was of a comparable magnitude to that of White people, r = .795, 95% CI [.667, .872].

Again, we used two-tailed Fisher r-to-z transformations to compare these correlation coefficients. Similar to White people, Black people's explicit and implicit correlations were higher at the county than individual level (z = 3.27, p = .001), but CBSA and county levels were not different (z = .370, p = .711). The explicit–implicit correlation was significantly stronger at the state than CBSA level (z = 5.54, p < .001). Taken together, Black people's explicit–implicit correlation also increased with level of aggregation, though not as consistently as did explicit–implicit correlations among White people.

Randomly assigning geography. Aggregation necessarily increases the reliability of a measure (Rushton et al., 1983). Additionally, the upper limit on how strongly two measures can correlate is a function of the reliability of each (Nunnally, 1970; Spearman, 1904). Consequently, one explanation for the strong correlations observed between explicit and implicit bias for both Black and White people, especially at the state level, is that this relationship is an artifact of aggregation. However, a further consideration of the results reported above reveal this is unlikely to be the case: at each level of regional analysis there are far more White people aggregated within each regional unit than Black people, yet we observe similar correlations between explicit and implicit bias for Black and White people (at least, at the county and state levels). If these results were solely an artifact of aggregation, we would expect consistently stronger correlations for White than Black people across all levels of analysis due solely to the greater number of White people in the sample.

Nevertheless, we further probed whether the strong explicit–implicit correlations at larger level of analysis reflect an artifact of aggregation or, alternately, coherent regional constructs. To this end, we reanalyzed these data with respondents' regional locations randomly assigned, with ns per region corresponding to the ns observed in the actual data. Because the correlations between explicit and implicit bias were already low at the county level, we did not include counties in this analysis. Thus, we aggregated these randomly assigned respondents at the levels of the CBSA and state, and once again estimated bias-corrected bootstrapped correlations between explicit and implicit bias. If large explicit–implicit correlations are simply an artifact of aggregation, then we should expect the pattern of increasing correlations observed previously to replicate when geography is assigned randomly. However, if CBSA and state level bias instead reflects coherent regional constructs—for example, if the biases of residents within a region correspond to the biases of their neighbors, or if bias operationalized regionally reflects something about the region per se—then we should expect weaker relations between explicit and implicit bias when geography is randomly assigned than when aggregation reflects respondents' true locations.

As shown in Figure 6, the correlation between explicit and implicit bias when aggregated randomly was much weaker than when aggregated according to respondents' true locations (see Figure 5). Specifically, two-tailed Fisher r-to-z transformations reveal the relationship between White people's explicit and implicit bias at the CBSA level when location was randomly assigned, r = .144, 95% CI [.025, .260] is much weaker than when aggregation reflected true locations, r = .772, z = 12.62, p < .001. Similarly, the correlation between White people's explicit and implicit bias at the state level when location was randomly assigned, r = −.097, 95% CI [−.434, .254] is much weaker (and not different from zero) than when aggregation reflected true locations, r = .846, z = 6.56, p < .001. Correlations for Black people demonstrate the same pattern of results. The relationship between Black people's explicit and implicit bias at the CBSA level when location was randomly assigned was not significantly different than zero, r = .121, 95% CI [−.008, .247], but was only a slightly weaker relationship than when aggregation reflected true locations, r = .235, z = 2.67, p = .091, likely because this relationship was already somewhat low. The correlation between Black people's explicit and implicit bias had a much sharper reduction at the state-level when location was randomly assigned (and not different from zero), r = .045, 95% CI [−.351, .394] than when aggregation reflected true locations, r = .795, z = 5.09, p < .001. Together, these results indicate that true geography matters, and suggest that biases operationalized at the regional level reflects cohesive regional constructs.

Explicit–Implicit Relationships Discussion

Explicit and implicit racial bias at regional levels are generally positively associated with one another, and the strength of this association increases with the level of aggregation, with particularly strong relationships at the state level. This relationship is not an artifact of aggregation, as this relationship is diminished when location is randomly assigned. We interpret this result as strong evidence that these measures of explicit and implicit regional racial bias are capturing a phenomenon that is geographically situated and varying across regions. This result additionally highlights a methodological concern for studying regional racial biases. Any research attempting to model regional outcomes must necessarily aggregate implicit and explicit bias to the same regional level. The results of this section reveal that multicollinearity between implicit and explicit bias will become an increasingly large problem at larger regional levels, and must be dealt with. We return to this issue and its multiple interpretations in the General Discussion section.

Section 3: External

In the external phase of the process of construct validation, we aimed to place regional explicit and implicit bias in a larger nomological network, demonstrating how these variables relate to external variables. We first summarize previous research that has examined explicit and implicit bias at regional levels.

Previous Research

To our knowledge, at the time of writing there were six published articles that examined explicit and implicit racial bias together in predicting regional outcomes. We organize our review according to level of analysis.

At the county level, Leitner and colleagues (2016a, 2016b) examined the relationship between explicit and implicit regional racial bias and health outcomes. In an initial investigation, they found that county-level explicit but not implicit bias predicted

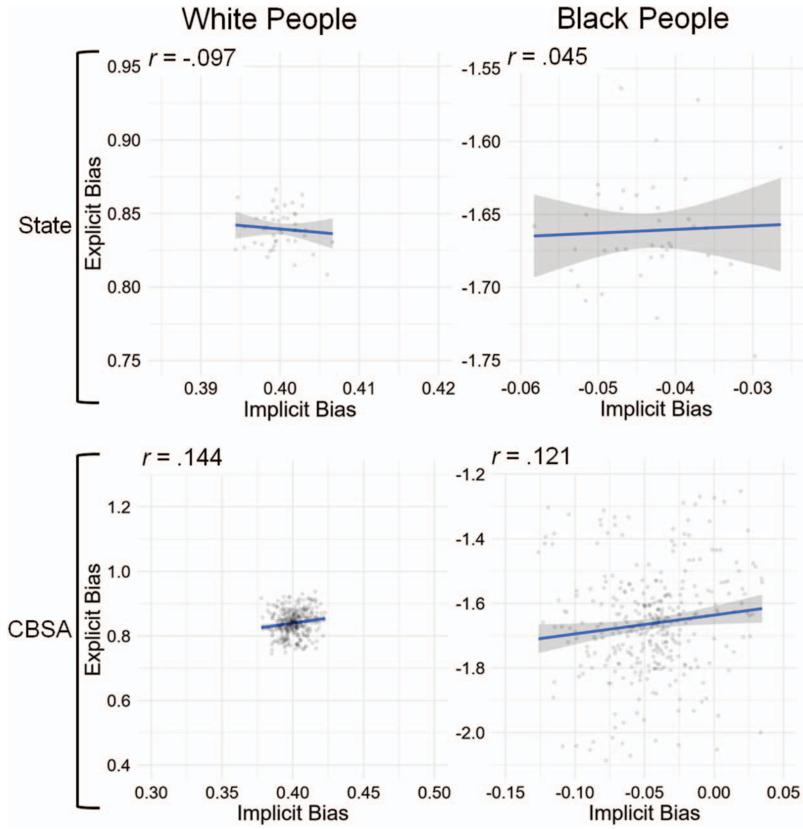

Figure 6. Correlations between explicit and implicit bias with regional unit randomly assigned. Compare these panels with those in Figure 5. At each regional level, while the absolute values of the *x*- and *y*-axes differ, the range of the axes represented for White and Black people is equivalent to allow for comparison. CBSA = core-based statistical area. See the online article for the color version of this figure.

disparities between Black and White people in terms of access to affordable health care and death from circulatory disease (Leitner et al., 2016a). Importantly, the primacy of explicit over implicit bias only emerges when both measures are included in the same model; when modeled separately, explicit and implicit bias were each significantly related to mortality disparities. In a subsequent investigation, Leitner and colleagues found that the implicit biases of Black residents of a given county predicted Black mortality rates in that county, but the explicit biases of White residents predicted White mortality rates (Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016b). Similarly, Orchard and Price (2017) examined the relationship between county-level explicit and implicit racial bias and racial disparities in birthing outcomes and infant health. Both county-level explicit and implicit racial biases predicted these birthing and infant disparities, but the effects were stronger for explicit bias.

At the CBSA level, Hehman, Flake, and Calanchini (2018) examined the relationship between explicit and implicit racial bias and disproportionate use of lethal force by police. They found that White people's, but not Black people's, implicit bias predicted disproportionate killing of Black people by police relative to their presence in the population. Explicit bias (marginally) predicted disproportionate use of lethal force, but implicit bias was descriptively the stronger predictor.

Finally, at the state level, Rae, Newheiser, and Olson (2015) examined the relationship between racial bias and the proportion of White versus Black residents per state. They found that larger proportions of Black people in a given state were associated with stronger levels of both explicit and implicit in-group bias among both White and Black people. Additionally, Leitner and colleagues examined the relationship between racial bias and state-level Medicaid spending (Leitner, Hehman, & Snowden, 2018). Black people disproportionately benefit from Medicaid programs, and Leitner and colleagues' analyses reveal that states with higher levels of either explicit or implicit racial bias spent less per Medicaid enrollee.

Relationships With Other Variables

In this section, we examine external validity on two dimensions: convergent and discriminant validity. Convergent validity will be demonstrated to the extent that our operationalizations of regional racial bias correspond with other, theoretically relevant outcomes, and discriminant validity will be demonstrated to the extent that our operationalizations of regional racial bias do not correspond with theoretically unrelated outcomes. To establish convergent validity, we examined relationships between regional racial bias and two external outcomes: racially charged Internet searches, and racial attitudes as measured in a separate, representative dataset. To establish discriminant validity, we examined relationships between regional racial bias and birth rates.

Racially charged search rates. Stephens-Davidowitz and colleagues (Chae et al., 2015; Stephens-Davidowitz, 2014) introduced the Racially-Charged Search Rate, which reflects the proportion of Google searches made between 2004 and 2007 in a region that contain the racial slur "nigger" and its plural.[2] Racially-Charged Search Rates are negatively associated with the share of votes President Obama received by region (Stephens-Davidowitz, 2014), and with greater Black–White disparities in mortality rate (Chae et al., 2015). These findings dovetail with those of Leitner and colleagues, whose investigations into mortality disparities were based on Project Implicit data (Leitner et al., 2016a, 2016b).

Rae et al. (2015) reported that Project Implicit estimates of implicit racial bias were strongly positively associated with Racially Charged Search Rates. Building upon their findings, we extend this analysis to the explicit bias of White and Black people to shed additional light on explicit and implicit racial bias as regional constructs. Stephens-Davidowitz aggregated search data at the level of the "designated market areas" (Stephens-Davidowitz, 2014), which do not directly correspond to either counties or CBSAs. Consequently, we aggregated both Racially-Charged Search Rates and explicit and implicit bias from Project Implicit at the state level. Then, we estimated relations between Racially-Charged Search Rates and explicit and implicit bias, separately, using 5,000 bias-corrected and accelerated two-tailed bootstrapped correlations. Because the race of people searching Google for racial slurs is unknown, we correlated Racially-Charged Search Rates separately with the biases of Black and White Project Implicit respondents.

Consistent with findings from Rae et al. (2015), state-level Racially-Charged Search Rates correlated strongly with the state-level implicit racial biases of White Project Implicit respondents, $r = .795$, 95% CI [.688, .866]. Extending previous work, the state-level explicit racial biases of White Project Implicit respondents demonstrated a similarly strong relationship with state-level Racially-Charged Search Rates, $r = .693$, 95% CI [.528, .803]. For Black Project Implicit respondents, state-level Racially Charged Search rates correlated strongly, but negatively, with the state-level implicit racial biases, $r = -.478$, 95% CI [$-.643$, $-.257$], but the correlation between search rates and explicit biases was not different from zero, $r = -.299$, 95% CI [$-.525$, .006].

Pew attention to racial issues. One limitation of the present research, as well as much of the research reviewed thus far, is that it relies exclusively on data from Project Implicit, and the Project Implicit sample is not representative of the North American population. To address this limitation, and further establish the external validity of regional racial bias as operationalized in the present research, we examined the relationship between regional racial bias in the Project Implicit sample and racial attitudes in a nationally representative sample, "2016 Racial Attitudes in America Survey" collected by Pew Research Center (http://www.pewsocialtrends.org/dataset/2016-racial-attitudes-in-america-survey/).

Our analyses aggregated, at the state level, White people's responses to the question: "In general, do you think there is too much, too little, or about the right amount of attention paid to race and racial issues in our country these days?" (Pew, 2017). We recoded responses such that higher scores reflect stronger endorsement of the sentiment that race receives too much attention. Consistent with previous research demonstrating a positive relationship between colorblindness (i.e., a perspective that minimizes the importance of race) and intergroup bias (Hehman et al., 2012; Riek, Mania, & Gaertner, 2006; Verkuyten, 2005), we predicted that regions in which race is perceived as receiving too much attention will also demonstrate higher levels of racial bias.

Relying on 5,000 bias-corrected and accelerated two-tailed bootstrapped correlations, state-level White racial attention attitudes correlated positively with both the state-level implicit, $r = .366$, 95% CI [.143, .552], and explicit, $r = .479$, 95% CI [.186, .660] racial biases of White Project Implicit respondents. Conversely, state-level White racial attention attitudes correlated negatively with the state-level implicit racial biases of Black Project Implicit respondents, $r = -.339$, 95% CI [$-.559$, $-.099$]. The relationship was in the same direction, but weaker and not different from zero, for the state-level explicit racial biases of Black Project Implicit respondents, $r = -.239$, 95% CI [$-.649$, .113].

Birth rates. Above, we report convergent validity evidence that aggregated regional bias might be interpreted in a way consistent with racial bias by showing correlations with outcomes that might be theoretically expected to relate to racial bias. Complementing this approach, we now examine discriminant validity, which aids the interpretation of regional bias as racial bias by showing it is not correlated with outcomes that would presumably be unrelated to racial bias. Consequently, we analyzed whether state-level racial biases of Black and White Project Implicit respondents were associated with birth rates. We are unaware of existing evidence or theoretical frameworks relating overall birth rates to racial bias, so if the Project Implicit data reflects regional racial bias, we should not expect it to relate to birth rates.

Overall state-level 2017 birth rates were obtained from the Centers for Disease Control's WONDER database (United States Department of Health & Human Services, 2018), calculated as the number of births divided by total population. Five-thousand bias-corrected and accelerated two-tailed bootstrapped correlations indicated that state-level birth rates were not correlated with White Project Implicit respondents' state-level implicit, $r = -.110$, 95% CI [$-.369$, .134], or explicit racial biases, $r = .173$, 95% CI [$-.173$, .453], nor were they correlated with Black Project Implicit respondents' implicit, $r = -.232$, 95% CI [$-.560$, .155] or explicit racial biases, $r = .058$, 95% CI [$-.348$, .416].

Discussion

This external phase of construct validity provides initial evidence of the predictive validity of explicit and implicit racial bias operationalized at regional levels. The diverse outcomes predicted by implicit and/or explicit bias in previous research—related to health, law enforcement, and intergroup contact—are meaningful and societally significant, and help to situate regional explicit and implicit racial bias in a larger nomological network. That said, six articles do not constitute a sufficient basis from which to draw strong conclusions about the predictive validity of explicit and implicit bias at the regional level.

Supplementing the published literature on regional racial bias and behavioral outcomes, our own analyses of regional explicit

[2] Previous work has gone into greater detail as to why people are searching for these terms (Stephens-Davidowitz, 2014).

and implicit racial bias, as operationalized in the Project Implicit sample, reveal theoretically expected relationships with racially charged Google searches and attitudes about the amount of attention given racial issues in a nationally representative sample. Moreover, our analyses did not find a relationship between regional racial biases and an outcome theoretically unrelated to racial attitudes (i.e., birth rates). Taken together, these results demonstrate convergent and discriminant validity for the measures of explicit and implicit regional racial bias used by Project Implicit. The strong correlations presented here between Project Implicit estimates and other sources (e.g., Google search rates), as well as relationships with racial attention attitudes as measured by a representative sample, suggests that an interpretation of Project Implicit's measures as tapping a latent construct of regional racial bias has strong evidence. Additionally, across the six studies reviewed and two analyses presented here, explicit and implicit bias tend to be equally strong predictors of the outcomes of interest in terms of zero-order correlations (Hehman et al., 2018; Leitner et al., 2016a, 2016b; Orchard & Price, 2017). Taken together, the present analyses supplement previous work by constructing a nomological network of related constructs, demonstrating additional relationships between regional explicit and implicit racial bias and other outcomes.

General Discussion

Advances in large-scale data collection have presented new opportunities for the study of racial bias. However, explicit and implicit bias measures were initially developed and validated at the individual level. When established measures are used in a new context, new validity evidence is needed to support interpretations (Kane, 2013). Indeed, to study the racial biases of a region is a fundamentally different research endeavor than to study the racial biases of an individual. Therefore, in substantive, structural, and external phases (Loevinger, 1957), the present work sought to advance construct validity evidence of measures of racial bias at the region level.

In the substantive phase, we reviewed previous theory indicating that regional bias may not reflect simply the aggregate of individual biases, but instead may reflect emergent and/or qualitatively distinct phenomena. In a subsequent structural phase, we report the retest reliability of the explicit and implicit biases of Black and White people at various levels of regional aggregation, and examine the relationships between explicit and implicit bias. And finally, in an external phase, we establish relationships between regional explicit and implicit racial bias and other outcomes. Taken together, this validity evidence represents an important and necessary first step in understanding regional racial bias as a macropsychological construct. Below, we discuss the results of each phase of construct validation in greater detail, then synthesize the findings in terms of their implications for understanding regional racial bias.

Substantive Phase

Regions are comprised of individuals, so regional bias could simply reflect the aggregate of the individual-level biases within a given region. However, various theoretical perspectives propose recursive relationships between individuals and properties of the region (e.g., laws, institutions, housing patterns) that may lead to emergent phenomena that are qualitatively distinct from the sum of individual inputs (Oishi & Graham, 2010; Rentfrow et al., 2008). Moreover, aggregation increases the reliability of a measure (Rushton et al., 1983), and the magnitude of relationships between measures is a function of the reliability of each measure (Nunnally, 1970; Spearman, 1904). Consequently, the relationships observed in the present research among measures of regional racial bias, and between measures of racial bias and other outcomes, should not be downwardly biased by measurement reliability.

Structural Phase

As structural evidence of the construct validity of regional racial bias, we examined retest reliability for the explicit and implicit biases of Black and White people. Though it may be tempting to compare regional reliabilities to individual reliabilities, these constructs are at conceptually distinct levels of analysis, and one does not necessarily inform the other (Selvin, 1958). Instead, individual retest reliability reflects how much variance in a given person's responses at one measurement time is accounted for by their responses at a subsequent measurement time, whereas regional reliability reflects how much variance in a sample of people's responses at one measurement time is accounted for by the responses of another sample of people from the same region at a subsequent measurement time. Moreover, extant research has examined retest reliability of explicit and implicit bias over relatively short periods of time, ranging from 1 hr (Bar-Anan & Nosek, 2014) to 2 months (Gawronski et al., 2017). In contrast, the present research examines the stability of regional biases over the span of years.

Aggregated at the state level, explicit bias retest reliability matched that previously observed at the individual level, and implicit bias retest reliability far exceeded that previously observed at the individual level (Gawronski et al., 2017). These results are consistent with the findings of previous large regional work (Payne, Vuletich, Lundberg, 2017; Schmidt & Nosek, 2010) demonstrating that racial bias remains relatively stable over time, but also provide a more nuanced perspective on the nature of this stability. Namely, while *mean* country-wide bias of White people is reliable over time (as illustrated in Figure 3), there was substantial year-to-year variability in the bias of White people at the CBSA and county level, and at all levels for Black people. Lower reliability is evident in three distinct patterns: for smaller versus larger regional units, for implicit versus explicit bias, and for Black versus White people.

In general, retest reliability increased when aggregating at higher regional levels, but still remained relatively low. For instance, even at the level of aggregation for which Black people's reliability was highest (i.e., the state level), a state's explicit bias score in 1 year only explained 4% of next year's score, on average. Regional implicit bias regularly had lower levels of retest reliability than explicit bias for both Black and White people across all regional levels, a result consistent with previous research on individual implicit and explicit bias (Gawronski et al., 2017). At the county level, a region's level of implicit bias in one year explained *at most* .03% of the variance in implicit bias measured in the subsequent year, on average. A similar story emerged for both the regional explicit and implicit biases of Black people: at no level of aggregation was a year meaningfully informative of the next year's level of bias. That said, sample size was largely responsible for the differences in reliability between Black and White people's biases. When sample sizes were equated, differences between the reliabilities of Black and White people's regional racial biases disappeared (see Table 1).

Substantively, these results reveal that though the *patterns* of bias (i.e., means, variance) are different for Black versus White people, the psychometric properties of bias (i.e., reliability; explicit–implicit correlation) for each group are largely the same. On the point of reliability, there were several reasons we might have expected the biases of Black people to be more variable over time than the biases of White people. For instance, because White people are a numerical majority in American society, the average Black person encounters White people far more frequently than the average White person encounters Black people. Consequently, the relative frequency of intergroup contact for Black people versus White people might cause Black people's biases to vary more so than the biases of White people (e.g., Pettigrew & Tropp, 2006). Similarly, high-profile interracial events, such as a White police officer shooting an unarmed Black child, might impact the racial biases of Black people more strongly than those of White people. However, in the present research, we do not find that Black people's racial biases, either explicit or implicit, are more reliable over time than White people's racial biases. Instead, retest reliabilities were generally low for both Black and White people at all (size-equated) regional levels of analysis.

At least two interpretations of low regional racial bias retest reliabilities are possible. To the extent that regional racial bias is a stable construct over time, then these low retest reliabilities reflect a large amount of measurement error. This interpretation is consistent with some theoretical models of individual bias which postulate that biases are a result of associations learned over a lifetime, and fairly immutable to change (e.g., Baron & Banaji, 2006; Wilson et al., 2000). However, this interpretation is inconsistent with the reliability benefits that should be expected to come from regional aggregation (Rushton et al., 1983). In contrast, to the extent that aggregation minimizes measurement error, then these low retest reliabilities suggest that regional racial bias fluctuates dramatically over time. This interpretation is consistent with Payne et al.'s (2017) proposal that racial bias (or, at least, implicit racial bias) reflects a property of the situation rather than of the individual. The present research cannot distinguish between these possibilities, and preferring one interpretation over another hinges on one's theoretical perspective.

Our examination also revealed strong positive relationships between explicit and implicit regional racial bias, for both Black and White people, that increase in magnitude as regional units increase in size. At the individual level we find fairly low correspondence between explicit and implicit racial bias ($r = \sim.2$), but at the state level explicit and implicit bias correspond very strongly ($r = \sim.8$). In other words, one measure of bias explains 60–70% of the variance in the other. Moreover, this correspondence is not an artifact of aggregation. When respondents' geography was randomly assigned, the relationship between measures of bias substantially decreased relative to when geography reflected respondents' true locations. The finding that the explicit and implicit bias of randomly clustered individuals does *not* correlate provides evidence that these measures are capturing meaningful psychological constructs that vary systematically across regions. Though there is clearly a great deal of individual variability in bias within any region, there is also meaningful variation across regions.

External Phase

In our final section we reviewed extant research examining regional explicit and implicit racial bias. Only a handful of studies to date have examined regional racial biases, so we supplemented this review by demonstrating relationships between regional racial bias and two outcomes that should theoretically be related to regional racial biases: racially charged Internet searches, and racial attention attitudes. Additionally, we demonstrated discriminant validity by finding no relationship between regional racial bias and an outcome that is theoretically unrelated: birth rates. Taken together, we believe this pattern of results is consistent with an interpretation of regional aggregates of both explicit and implicit racial biases as two measures of a broader latent construct: regional racial bias. Moreover, these findings provide further evidence of the utility of aggregating racial biases at region levels to predict relevant outcomes of serious societal significance. Some of the outcomes linked with regional bias in previous research—related to health, intergroup contact, law enforcement, and lifestyle choices—are difficult to study in the lab for practical and ethical reasons. For example, police use of force is hard to study in practice because these are relatively rare events, and cannot be ethically recreated in a laboratory with any degree of ecological validity. Thus, operationalizing bias at region levels provides a powerful tool for social scientists interested in such meaningful but challenging phenomena.

Table 1
Estimates of Retest Reliability and the Correlation Between Implicit and Explicit Bias for White and Black People at Different Regional Levels of Analysis

	Retest reliability						Implicit-explicit correlation	
	White		White (Size-Equated)		Black			
Unit	Implicit	Explicit	Implicit	Explicit	Implicit	Explicit	White	Black
Individual	.31–.45*	.78*					.215	.149
County	.025	.058	.017	.030	.032	.050	.267	.216
CBSA	.171	.275	.071	.131	.029	.056	.772	.235
State	.693	.865	.212	.504	.171	.203	.846	.795

Note. CBSA = core-based statistical area.
* Indicates estimates reported in the literature, and not from the present analyses.

Broader Conclusions

Improved predictive ability at regional versus individual levels. Taken together, the results of our investigation into the construct validity of regional racial bias suggest a number of important implications. One is that aggregating bias regionally can be a useful way to predict important real-world outcomes. Though there are numerous examples of individual racial bias predicting behavioral outcomes, meta-analyses reveal relatively low correspondence between both explicit and implicit racial bias and individual behavior (Greenwald et al., 2009; Oswald et al., 2013). To the extent that a given measure reflects random noise mixed with an underlying signal, averaging over a greater number of observations cancels out the noise but leaves the signal intact (Rushton et al., 1983). Functionally, by averaging over an increasing number of observations at different regional levels, we increase the reliability and precision of our estimates of racial bias for each region. Consequently, more precise estimates and less error may increase the predictive validity of racial bias measures at the regional relative to the individual level—much like the fictional analytic method of psychohistory in Isaac Asimov's *Foundation* series, which increases in precision as the scale of prediction increases. The handful of studies published to date examining regional racial bias and behavior are suggestive evidence of this possibility, (descriptively) demonstrating much larger effects than do meta-analyses of individual racial bias and behavior.

Alternatively, to the extent that regional and individual racial biases are distinct constructs, regional biases may reveal patterns that do not exist at the individual level. Payne et al. (2017) bias of crowds perspective supports this possibility: if implicit racial bias reflect a stable property of the situation rather than of the individual, then we should expect larger effects for regions than for individuals because we have appropriately calibrated the level of our psychological construct with the level of our outcome to be predicted. Taken together, the present research suggests that regional analyses have the potential to be a powerful analytic tool, but additional research is necessary to further establish the utility of this approach.

The relationship between regional explicit and implicit bias. The present research offers novel insight into the qualitative natures of explicit and implicit bias operationalized at regional levels in two ways. One is by examining to what extent regional explicit and implicit racial bias similarly predict outcomes, and the other is by directly examining the relationship between the two measures of racial bias.

In terms of predicting outcomes, the reviewed evidence and present analyses consistently indicate that explicit and implicit bias generally predict the same outcomes when both are entered into multiple regression models (Hehman et al., 2018; Leitner et al., 2016b; Orchard & Price, 2017). The same pattern of results is observed in zero-order correlations between outcomes and explicit and implicit bias. Moreover, when both forms of bias are entered simultaneously in a model, sometimes explicit bias is the better predictor (e.g., Leitner et al., 2016a), but other times implicit bias is the better predictor (e.g., Hehman et al., 2018).

One possible reason for the inconsistent predictive superiority of one measure of regional bias over the other is that, as the present research reveals, explicit and implicit bias are highly correlated at region levels. In a regression model, such high collinearity indicates that there is a relatively small portion of *non*-overlapping variance with which to predict an outcome. Consequently, one implication of explicit–implicit collinearity in regional racial bias research is that the predictive superiority of one measure over the other may be driven by random fluctuations in the data and, thus, not necessarily reflect meaningful variation. In other words, the distinction between explicit versus implicit predictive superiority may not be theoretically meaningful at region levels. That said, we cannot rule out yet-unobserved moderators to explain when explicit and implicit racial bias as *distinct regional level constructs* uniquely predict outcomes.

The present research also offers novel insight into the qualitative natures of explicit and implicit racial bias operationalized at regional levels by directly examining the relationship between the two measures. As the level of aggregation increases, the magnitude of the relationship between explicit and implicit racial bias similarly increases. At the state level, the relationship between explicit and implicit racial bias for Black and White people is $r = .846$ and .795. The most parsimonious interpretation of these strong correlations is that regional explicit and implicit racial bias are likely different measures of a *single* phenomenon (i.e., regional racial bias). However, these findings should not be interpreted to indicate that individual explicit and implicit racial biases are also different measures of a single phenomenon. On one hand, individual explicit–implicit correlations may be artificially suppressed by the low reliability of implicit measures, and only through aggregation does the true relationship between explicit and implicit racial bias emerge. However, on the other hand, racial bias may have different underlying mechanisms at the individual versus region levels, and different levels of analysis are conceptually distinct. Consequently, future research is necessary in order to draw stronger conclusions about the relationship between individual and regional racial bias.

Proposed Causal Model of Regional Bias

Extending from extant theory and the evidence presented here, we propose a recursive causal relationship between regional racial biases and regional outcomes. To the extent that a region is characterized by a relatively high level of racial bias, then those residing in that region should have higher levels of racial bias than those residing in regions with lower levels of racial bias. Accordingly, outcomes and behaviors consistent with higher levels of racial bias should also be more common in regions that are relatively more racially biased. Critically, all individuals in a region do not need to be racially biased for biased outcomes to occur. For example, a nonbiased person may behave in biased ways because their friends, neighbors, or bosses expect, reward, or model such behavior. Additionally, the racial biases of individuals may become instantiated as properties of a region, such as residential, retail, educational, and legal institutions, which over time in turn produce biased outcomes without the active input of any single individual. Any biases built into these institutions may be slower and more difficult to change than the attitudes of individual citizens. Whatever the source of outcomes in a region, these outcomes can in turn reinforce and perpetuate the racial biases of a region. For example, to the extent that residents of a region usually see members of certain racial groups living in impoverished neighborhoods, working low-status jobs, failing out of schools, and profiled as suspects or defendants in local media, residents' preexisting negative racial associations will be maintained or strengthened. By capitalizing on data that is longitudinal in nature, examining the causal pathways of this proposed bidirectional relationship between regional racial biases and regional outcomes will be possible in future research.

Limitations

The present research is limited in several ways. For example, we focus solely on racial bias, so any conclusions drawn from our findings are limited to this domain. Explicit and implicit measures have been used to collect information on other psychological constructs besides racial bias, such as self-esteem and stereotyping (Greenwald & Banaji, 1995), and to a wide variety of attitude objects ranging from consumer brands to political candidates (Graham, Haidt, & Nosek, 2009; Greenwald & Farnham, 2000; Maison, Greenwald, & Bruin, 2001; Rudman, Ashmore, & Gary, 2001). Moreover, other psychological constructs have been operationalized at regional levels, such as personality (e.g., Rentfrow et al., 2013) and religiosity (e.g., Gebauer, Paulhus, & Neberich, 2013). The extent to which these other constructs correspond at the individual and regional levels remains an important and interesting avenue for future research.

Another limitation of the present work is that we focused only on the biases of Black and White people. We did this in large part because Black-White racial dynamics are highly salient in North American society. However, the present research offers no insight into whether the differences revealed here between Black and White people's explicit and implicit racial biases replicate in other groups. This represents a fruitful direction for future research.

The present research is also limited in that it focuses solely on the IAT as a measure of implicit bias. Other implicit bias measures, such as the affect misattribution procedure (Payne et al., 2005), are well-validated and widely used to study a variety of attitude objects (Cameron et al., 2012). However, to our knowledge, the IAT is the only implicit measure that has been used to study attitudes operationalized at region levels. Because implicit measures vary in their structures (e.g., stimuli presented concurrently vs. sequentially) and demands (e.g., attend to all vs. some stimuli), different implicit measures necessarily reflect the contributions of different mental processes (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne et al., 2010). Consequently, future research that seeks to use other measures to study regional attitudes should also include evidence of construct validity at that level of analysis.

Similarly, we have used a limited measure of explicit racial bias, based on two thermometer ratings. This may not be the best measure of explicit racial bias. Recent research has found that a single item directly asking about attitudes toward White people relative to Black people has the highest correlation with the IAT (Axt, 2017). To assess the robustness of the results we report here against specific operationalizations of explicit racial bias, we reanalyzed implicit–explicit correlations using a single-item measure instead of the thermometer difference score. The results of these analyses are nearly identical to the ones reported above and, importantly, lead to the same conclusion: that correlations between implicit and explicit racial bias increase as level of analysis increases. Consequently, we conclude that these two operationalizations of explicit bias are functionally identical in the context of the present research. A table presenting the correlation between these two measures is available in the online supplementary materials. Yet both these measures of bias are extremely short, and may be missing out on unique and important content captured in more extensive measures. Future research should examine our conclusions with more comprehensive measures.

An additional limitation of the present research is that the sample visiting Project Implicit is not representative of the general North American population. Consequently, any conclusions drawn from these data do not necessarily generalize to the population at large. Of course, this limitation is also true of lab-based research, which has for decades primarily relied on university undergraduates who differ from the general population on a wide variety of dimensions (e.g., Sears, 1986). That said, previous published research examining disparities in health care (Orchard & Price, 2017), policing (Hehman et al., 2018), mortality (Leitner et al., 2016a, 2016b), and other outcomes as reviewed in the External section, indicate that the biases of Project Implicit respondents are associated with important society-level outcomes. Consequently, perhaps the more pertinent question is not whether the Project Implicit sample is representative, but why this sample predicts these outcomes—outcomes which, statistically speaking, Project Implicit visitors were not likely to have directly participated in. Yet the present research also presents novel evidence regarding the representativeness of the Project Implicit sample, finding positive associations between the biases reported by Project Implicit respondents and a measure tapping race-related attitudes in a nationally representative sample collected by Pew Research Center. To our knowledge, this is the first time responses from Project Implicit data has been linked with nationally representative samples (for racial attitudes, see Ofosu, Chambers, Chen, & Hehman, 2019 for anti-gay attitudes). Though these findings do not prove that the Project Implicit sample is representative of the U.S. population, they provide evidence that data from the Project Implicit sample perform like representative data, at least in some contexts. Future research should continue to examine similarities and differences between the Project Implicit sample and other, representative data sets.

Finally, an issue not fully resolved by the current research is whether regional racial bias can be interpreted to reflect the same latent construct at different levels of aggregation. In other words, is state-level regional bias the same thing as county or CBSA level regional bias? On the one hand, an infinite number of constructs corresponding to infinite levels of regional aggregation is certainly not parsimonious. However, on the other hand, this question maps onto an issue identified by geographers—the modified aerial unit problem—which posits that ". . . when spatial data are aggregated, the results are conditional on the spatial scale at which they are conducted" (Manley, 2014, p. 1157). To the extent that a regional construct is related to how space is parsed, the causes and consequences of regional racial bias likely vary across levels of aggregation. For example, the present research revealed equivalent racial bias retest reliabilities for Black and White people at the county and CBSA levels when sample sizes were equated. However, at the state level, sample-size equated retest reliability for explicit racial bias was descriptively higher for White versus Black people (rs = .504, .203, respectively). This pattern of results suggests that there is something qualitatively different about racial bias at the state level compared to the county and CBSA levels, which may be a fruitful direction for future research. That said, whether the meaning of the construct itself—for example, as reflecting racial bias—also varies by level of aggregation remains a question for future research.

Recommendations

The present work highlights the very large number of respondents necessary to reliably estimate regional racial bias. Only the state-level estimates of White people's bias approached acceptable levels of reliability, and in these analyses there were on average 28,664 respondents in each regional unit. At the CBSA level, which averaged 3,531 respondents per regional unit, reliabilities were below conventionally accepted levels. Researchers seeking to examine regional biases over time will need very large samples, and at the state level, these numbers currently exist only for White people. Because low reliability can artificially suppress relationships between variables, researchers should be appropriately cautious when conducting regional analyses in order to avoid Type II error—though, of course, this is also true at the individual level.

Due to the strong relationship between explicit and implicit racial bias at region levels, we recommend examining explicit and implicit racial bias as predictors in separate statistical models when examining regional outcomes. Otherwise, the strong relationship between the two forms of bias will introduce collinearity into a model. Consequently, it is not clear whether any significant effects would reflect truly unique, theoretically meaningful variance or random fluctuations in the data. This is particularly important for state-level analyses, in which the explicit–implicit relationship is quite high, but should be considered for other regional levels of analysis as well.

Conclusion

In summary, the present research is the first to investigate the construct validity of regional explicit and implicit racial bias of Black and White people. As social scientists continue to accumulate data from increasingly large and diverse samples, new opportunities will arise to explore questions that cannot be investigated in the context of the laboratory. The process of validating newly conceived macropsychological constructs is critical to interpreting any results from these explorations, and can provide new insight into established findings. The present research offers a promising first step in understanding racial bias on a regional scale.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Joint Committee on Standards for Educational and Psychological Testing.

Axt, J. R. (2017). The best way to measure explicit racial attitudes is to ask about them. *Social Psychological & Personality Science*.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect measures of social cognition. *Behavior Research Methods, 46,* 668–688. http://dx.doi.org/10.3758/s13428-013-0410-6

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17,* 53–58. http://dx.doi.org/10.1111/j.1467-9280.2005.01664.x

Bleidorn, W., Schönbrodt, F., Gebauer, J. E., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2016). To live among like-minded others: Exploring the links between person-city personality fit and self-esteem. *Psychological Science, 27,* 419–427. http://dx.doi.org/10.1177/0956797615627133

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. http://dx.doi.org/10.1177/1088868312440047

Chae, D. H., Clouston, S., Hatzenbuehler, M. L., Kramer, M. R., Cooper, H. L. F., Wilson, S. M., . . . Link, B. G. (2015). Association between an internet-based measure of area racism and black mortality. *PLoS ONE, 10,* e0122963. http://dx.doi.org/10.1371/journal.pone.0122963

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlational analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge. http://dx.doi.org/10.4324/9780203774441

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,* 469–487. http://dx.doi.org/10.1037/0022-3514.89.4.469

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

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. http://dx.doi.org/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. http://dx.doi.org/10.1111/1467-9280.00328

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25,* 736–760.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping, and discrimination: Theoretical and empirical overview. In *The Sage handbook of prejudice, stereotyping, and discrimination* (pp. 3–29). London, UK: Sage. http://dx.doi.org/10.4135/9781446200919.n1

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. http://dx.doi.org/10.1037/0022-3514.69.6.1013

Gaertner, S. L., & Mclaughlin, J. P. (1983). Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics. *Social Psychology Quarterly, 46,* 23–30. http://dx.doi.org/10.2307/3033657

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology, 50,* 141–150. http://dx.doi.org/10.1037/a0013848

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.

Gawronski, B., & Creighton, L. A. (2013). Dual process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282–312). New York, NY: Oxford University Press.

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. http://dx.doi.org/10.1177/0146167216684131

Gebauer, J. E., Paulhus, D. L., & Neberich, W. (2013). Big Two personality and religiosity across cultures: Communals as religious conformists and agentics as religious contrarians. *Social Psychological & Personality Science, 4,* 21–30. http://dx.doi.org/10.1177/1948550612442553

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96,* 1029–1046. http://dx.doi.org/10.1037/a0015141

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27. http://dx.doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25. http://dx.doi.org/10.1037/0033-295X.109.1.3

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79,* 1022–1038. http://dx.doi.org/10.1037/0022-3514.79.6.1022

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,* 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. http://dx.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,* 17–41. http://dx.doi.org/10.1037/a0015575

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50,* 1166–1186. http://dx.doi.org/10.3758/s13428-017-0935-1

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. http://dx.doi.org/10.1177/1948550617711229

Hehman, E., Gaertner, S. L., Dovidio, J. F., Mania, E. W., Guerra, R., Wilson, D. C., & Friel, B. M. (2012). Group status drives majority and minority integration preferences. *Psychological Science, 23,* 46–52. http://dx.doi.org/10.1177/0956797611423547

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385. http://dx.doi.org/10.1177/0146167205275613

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50,* 1–73. http://dx.doi.org/10.1111/jedm.12000

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., . . . Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*. Advance online publication. http://dx.doi.org/10.1037/amp0000364

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36,* 343–356. http://dx.doi.org/10.1037/0003-066X.36.4.343

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016a). Blacks' death rate due to circulatory diseases is positively related to Whites' explicit racial bias. *Psychological Science, 27,* 1299–1311. http://dx.doi.org/10.1177/0956797616658450

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016b). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine, 170,* 220–227. http://dx.doi.org/10.1016/j.socscimed.2016.10.007

Leitner, J. B., Hehman, E., & Snowden, L. R. (2018). States higher in racial bias spend less on disabled Medicaid enrollees. *Social Science & Medicine, 208,* 150–157. http://dx.doi.org/10.1016/j.socscimed.2018.01.013

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694. http://dx.doi.org/10.2466/pr0.1957.3.3.635

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32,* 1–9.

Manley, D. (2014). Scale, aggregation, and the modifiable areal unit problem. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of regional science* (pp. 1157–1171). Heidelberg, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-23430-9_69

Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2014). How ideological migration geographically segregates groups. *Journal of Experimental Social Psychology, 51,* 1–14. http://dx.doi.org/10.1016/j.jesp.2013.10.010

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology, 31,* 166–180. http://dx.doi.org/10.1177/0146167204271418

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. http://dx.doi.org/10.1080/10463280701489053

Nunnally, J. (1970). *Introduction to psychological measurement*. New York, NY: McGraw-Hill.

Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences, 116,* 8846–8851. http://dx.doi.org/10.1073/pnas.1806000116

Oishi, S., & Graham, J. (2010). Social ecology: Lost and found in psychological science. *Perspectives on Psychological Science, 5,* 356–377. http://dx.doi.org/10.1177/1745691610374588

Oishi, S., Talhelm, T., & Lee, M. (2015). Personality and geography: Introverts prefer mountains. *Journal of Research in Personality, 58,* 55–68. http://dx.doi.org/10.1016/j.jrp.2015.07.001

Orchard, J., & Price, J. (2017). County-level racial prejudice and the Black-White gap in infant health outcomes. *Social Science & Medicine, 181,* 191–198. http://dx.doi.org/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. http://dx.doi.org/10.1037/a0032734

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Affect Misattribution as Implicit Measurement, 89,* 277–293. http://dx.doi.org/10.1037/0022-3514.89.3.277

Payne, B. K., Hall, D. L., Cameron, C. D., Bishara, A. J., Payne, B. K., Hall, D. L., . . . Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychology Bulletin, 36,* 1397–1408. http://dx.doi.org/10.1177/0146167210383440

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,* 1–64. http://dx.doi.org/10.1080/1047840X.2017.1335568

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90,* 751–783. http://dx.doi.org/10.1037/0022-3514.90.5.751

Pew. (2017). *2016 Racial Attitudes in America Survey*. Washington, DC: Pew Research Center. Retrieved from https://www.pewsocialtrends.org/dataset/2016-racial-attitudes-in-america-survey/

Rae, J. R., Newheiser, K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the United States. *Social Psychological & Personality Science, 6,* 535–543. http://dx.doi.org/10.1177/1948550614567357

Rentfrow, P. J., Gosling, S. D., Jokela, M., Stillwell, D. J., Kosinski, M., & Potter, J. (2013). Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. *Journal of Personality and Social Psychology, 105,* 996–1012. http://dx.doi.org/10.1037/a0034434

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science, 3,* 339–369. http://dx.doi.org/10.1111/j.1745-6924.2008.00084.x

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 10,* 336–353. http://dx.doi.org/10.1207/s15327957pspr1004_4

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

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity. *The Principle of Aggregation, 94,* 18–38.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995–1008. http://dx.doi.org/10.1037/0022-3514.91.6.995

Schmidt, K., & Nosek, B. (2010). Implicit (and explicit) racial attitudes barely changed during Barack Obama's presidential campaign and early presidency. *Journal of Experimental Social Psychology, 46,* 308–314. http://dx.doi.org/10.1016/j.jesp.2009.12.003

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515–530.

Selvin, H. C. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology, 63,* 607–619. http://dx.doi.org/10.1086/222356

Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *The Behavioral and Brain Sciences, 37,* 243–295. http://dx.doi.org/10.1017/S0140525X13001544

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15,* 72–101. http://dx.doi.org/10.2307/1412159

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41,* 481–520. http://dx.doi.org/10.3102/1076998616646200

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics, 118,* 26–40. http://dx.doi.org/10.1016/j.jpubeco.2014.04.010

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220–247. http://dx.doi.org/10.1207/s15327957pspr0803_1

United States Department of Health and Human Services. (2018). *Natality public-use data 2007–2017*. Washington, DC: United States Department of Health and Human Services.

Verkuyten, M. (2005). Ethnic group identification and group evaluation among minority and majority groups: Testing the multiculturalism hypothesis. *Journal of Personality and Social Psychology, 88,* 121–138. http://dx.doi.org/10.1037/0022-3514.88.1.121

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101–126. http://dx.doi.org/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*. Advance online publication. http://dx.doi.org/10.5334/jopd.ac

Received October 21, 2017
Revision received January 27, 2019
Accepted April 3, 2019 ■