Journal of Personality and Social Psychology

On the Relationship Between Indirect Measures of Black Versus White Racial Attitudes and Discriminatory Outcomes: An Adversarial Collaboration Using a Sample of White Americans

Jordan R. Axt, Paul Connor, Suzanne Hoogeveen, Cory J. Clark, Michelangelo Vianello, Joanna N. Lahey, Adam Hahn, Jeffrey To, Richard E. Petty, Thomas H. Costello, Gregory Mitchell, Philip E. Tetlock, and Eric Luis Uhlmann

Online First Publication, March 2, 2026. https://dx.doi.org/10.1037/pspa0000480

CITATION

Axt, J. R., Connor, P., Hoogeveen, S., Clark, C. J., Vianello, M., Lahey, J. N., Hahn, A., To, J., Petty, R. E., Costello, T. H., Mitchell, G., Tetlock, P. E., & Uhlmann, E. L. (2026). On the relationship between indirect measures of Black versus White racial attitudes and discriminatory outcomes: An adversarial collaboration using a sample of White Americans. *Journal of Personality and Social Psychology*. Advance online publication. https://dx.doi.org/10.1037/pspa0000480

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

https://doi.org/10.1037/pspa0000480

On the Relationship Between Indirect Measures of Black Versus White Racial Attitudes and Discriminatory Outcomes: An Adversarial Collaboration Using a Sample of White Americans

Jordan R. Axt[1], Paul Connor[2], Suzanne Hoogeveen[3], Cory J. Clark[4, 5], Michelangelo Vianello[6], Joanna N. Lahey[7], Adam Hahn[8], Jeffrey To[1], Richard E. Petty[9], Thomas H. Costello[10], Gregory Mitchell[11], Philip E. Tetlock[4], and Eric Luis Uhlmann[12]

[1] Department of Psychology, McGill University
[2] Meta Platforms, Inc., Menlo Park, California, United States
[3] Department of Methodology and Statistics, Utrecht University
[4] The Wharton School, School and the College of Arts and Sciences, University of Pennsylvania
[5] Division of Social Sciences, New College of Florida
[6] Department of Philosophy, Sociology, Education and Applied Psychology, School of Psychology, Università degli Studi di Padova
[7] Department of Public Service and Administration, Bush School of Government and Public Service, Texas A&M University
[8] Department of Psychology, University of Bath
[9] Department of Psychology, The Ohio State University
[10] Department of Social and Decision Sciences, Carnegie Mellon University
[11] School of Law, University of Virginia
[12] Department of Organizational Behavior, INSEAD

The idea that racial prejudice contributes to discrimination not only deliberately but also in a more automatic fashion has been one of the most prominent topics in social psychological research in the past 30 years. Much of the evidence for theories of automatic prejudice stems from the use of indirect measures of implicit attitudes, yet meta-analyses give differing estimates regarding the predictive validity of such measures. The present adversarial collaboration provides a test of the relationships between prominent measures of implicit racial attitudes and discriminatory behavior using a set of established lab-based paradigms among a sample of White Americans ($N = 2{,}114$). Using structural equation models that can account for measurement error, frequentist and Bayesian multiverse analyses confirmed that White Americans' performance on indirect measures correlate modestly with these behavioral outcomes, and explain unique variance (∼2.5%) beyond direct, self-report measures of racial attitudes. At the same time, self-report measures exhibited greater predictive and incremental validity than indirect measures (explaining ∼45% of the variance) despite behavioral measures of discrimination displaying weak internal reliability. Results provided some support for greater predictive and incremental validity for indirect measures among participants scoring relatively low on measures of executive function and motivation to control prejudice. These results lend themselves to both relatively optimistic and pessimistic interpretations concerning scientific and practical significance. All collaborators agree that the best path forward is collaborative and focused on the generalizability of implicit racial attitudes to high-accountability organizational settings.

Benedek Kurdi served as action editor.

Jordan R. Axt https://orcid.org/0000-0003-3944-8331

Jordan R. Axt, Paul Connor, and Suzanne Hoogeveen contributed equally to this article. Deidentified data, materials, and analysis code are available on the Open Science Framework at https://osf.io/wsu4q/. Study design and analyses were preregistered at https://osf.io/gjxu5. The authors declare no competing interests. The authors thank Li Li for their assistance in organizing submissions of this work and facilitating the collaboration among authors.

Jordan R. Axt played a supporting role in methodology and writing–original draft and an equal role in investigation, project administration, supervision, and writing–review and editing. Paul Connor played a lead role in conceptualization, investigation, and methodology, a supporting role in writing–original draft, and an equal role in project administration and writing–review and editing. Suzanne Hoogeveen played a lead role in data curation, formal analysis, software, validation, and visualization and a supporting role in project administration and writing–review and editing. Cory J. Clark played a supporting role in project administration and an equal role in conceptualization, funding acquisition, investigation, methodology, supervision, and writing–review and editing. Michelangelo Vianello played a supporting role in conceptualization, data curation, formal analysis, methodology, software, visualization, and writing–review and editing. Joanna N. Lahey played a supporting role in conceptualization, methodology, and writing–review and editing. Adam Hahn played a supporting role in conceptualization, formal analysis, methodology, and writing–review and editing. Jeffrey To played a supporting role in investigation, methodology, and writing–review and editing. Richard E. Petty played a supporting role in conceptualization, funding acquisition, methodology, and writing–review and editing. Thomas H. Costello played a supporting role in data curation, formal analysis, and writing–review and editing. Gregory Mitchell played a supporting role in conceptualization, funding acquisition, methodology, writing–original draft, and writing–review and editing. Philip E. Tetlock played a lead role in funding

continued

Statement of Limitations

Despite a large sample and multiple indirect, direct, and behavioral measures, the present study has several limitations. The generality of our conclusions is constrained through using a sample from a specific demographic (White American adults) and focusing measures on only one form of bias (toward White vs. Black people). Another limitation is the exclusive use of lab methods rather than including outcomes collected in field or more naturalistic settings. Relatedly, most of the behavioral measures of discrimination, despite being drawn from prior work, suffered from poor reliability and did not create evidence of anti-Black discrimination on average. Addressing these limitations, ideally through adversarial collaboration, will be critical for future research on this topic.

Keywords: implicit social cognition, racial attitudes, prejudice, discrimination, adversarial collaboration

Supplemental materials: https://doi.org/10.1037/pspa0000480.supp

What is the nature of contemporary racial prejudice? According to one influential perspective, racial prejudice has largely "gone underground," such that many people who endorse egalitarian values still experience negative reactions to racial groups that leak out in discriminatory behavior (Devine, 1989; Gaertner & Dovidio, 1986; Greenwald & Banaji, 1995; Greenwald & Lai, 2020).

Evidence that racial prejudice manifests unintentionally is provided by indirect measures, which rely on indicators such as reaction times to capture biased reactions without asking participants to directly report their attitudes (Fazio & Olson, 2003; Gawronski et al., 2020; Payne & Lundberg, 2014; Uhlmann et al., 2012; see Park et al., 2025). Here, we distinguish between a measurement procedure (a specific methodology) and the measure's targeted construct. Direct measures are completed under conditions with relatively high levels of participant control and are largely rooted in self-report. For instance, some of the direct measures used in the present work ask participants to self-report how strongly they associate Black and White people with either positivity or negativity. We call the construct assessed by these direct measures "explicit racial attitudes."

In contrast, indirect measures use tasks in which participants have limited control over the influence of an underlying construct (such as automatic racial biases) on the measurement outcome. Though there are many forms of indirect measurement (see De Houwer & Moors, 2010; Petty & Cacioppo, 1981; Vargas et al., 2004), the most common approaches in the contemporary literature use response latencies to infer the strength of "automatic associations" between concepts and valence (e.g., between Black people and negative concepts), though this too is contested (De Houwer, 2014). The development of such indirect measures was rooted in prior research on implicit memory (Greenwald & Banaji, 1995), drawing from studies revealing how automatic processes contributed to deliberative judgments in ways that were unavailable to introspection or effortful control (Jacoby & Dallas, 1981; Jacoby et al., 1992).

One example of an indirect measure is the evaluative priming task (EPT). In an EPT looking to assess implicit racial attitudes, Black and White faces briefly flash before participants categorize valenced stimuli as "Good" or "Bad" (Fazio et al., 1986, 1995). If Black faces facilitate the categorization of negative words as bad, and inhibit categorization of positive words as good, compared with the pattern of effects after exposure to White faces, this is interpreted as a negative implicit attitude toward Black relative to White people. Similarly, in the Implicit Association Test (IAT; Greenwald et al., 1998), faster reaction times when categorizing "Black" + "Bad" and "White" + "Good" stimuli with the same response key (compared with the reverse pairing) are interpreted as a negative implicit attitude toward Black versus White people. We call the construct assessed by these indirect measures "implicit racial attitudes" and refer to "negative implicit racial attitudes" when responses on such measures indicate that certain racial groups are more strongly tied to the concept of negativity (or positivity) than another racial group.

The concept of implicit racial attitudes has been widely influential in legal contexts (Greenwald & Krieger, 2006) and the diversity training industry (Onyeador et al., 2021). The IAT in particular is frequently used in teaching and corporate trainings, some of which have been mandated by law (Casad et al., 2013; Hagiwara et al., 2020) and has been completed millions of times online by volunteer participants via the *Project Implicit* demonstration website (Charlesworth & Banaji, 2019).

Along with this widespread impact, research on implicit attitudes has been the target of considerable criticism (Arkes & Tetlock, 2004; Blanton & Jaccard, 2006; Blanton et al., 2009; Mitchell & Tetlock, 2006; Schimmack, 2021). Some critiques have focused on contested terms, such as whether biases on reaction time tasks that appear to be learned from the broader culture can be characterized as racial prejudice (Arkes & Tetlock, 2004). Other critiques have been empirical, raising questions about whether indirect measures are sufficiently reliable (Bosson et al., 2000), tap the intended construct (Kurdi et al., 2021; Schimmack, 2021), and exhibit convergent validity (Bosson et al., 2000; Olson & Fazio, 2003).

acquisition, a supporting role in investigation, methodology, project administration, and writing–review and editing, and an equal role in conceptualization. Eric Luis Uhlmann played a lead role in writing–original draft, a supporting role in formal analysis and writing–review and editing, and an equal role in conceptualization, funding acquisition, investigation, methodology, project administration, and supervision.

Correspondence concerning this article should be addressed to Jordan R. Axt, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada, or Suzanne Hoogeveen, Department of Methodology and Statistics, Utrecht University, Sjoerd Groenmangebouw, Padualaan 14, 3584 CH Utrecht, The Netherlands, or Eric Luis Uhlmann, Department of Organizational Behavior, INSEAD, 1 Ayer Rajah Avenue, Singapore 138676, Singapore. Email: jordan.axt@mcgill.ca or s.hoogeveen@uu.nl or eric.luis.uhlmann@gmail.com

However, perhaps the most important criticism of research in implicit social cognition has been that indirect measures are weak and inconsistent predictors of discriminatory behavior (Oswald et al., 2013). On this point, a series of meta-analyses have provided differing estimates. Greenwald et al. (2009) and Cameron et al. (2012) reported modest correlations between indirect measures and relevant behaviors ($r = .24$ and $r = .28$, respectively). However, Oswald et al. (2013) and Kurdi et al. (2019) observed lower estimates ($r = .14$ and $r = .10$, respectively).

Importantly, meta-analyses have also highlighted that the relationship between indirect measures and discriminatory behaviors may depend on methodological features of studies (Greenwald et al., 2009; Kurdi et al., 2019). The relationship has been strongest when behavioral measures involve the relative treatment of Black versus White people, when subjectively coded correspondence between measures is high (Ajzen, 1985; Ajzen & Fishbein, 1977) and when relating indirect measures to behaviors is the primary focus of the research (Kurdi et al., 2019). When such moderators are combined, the average indirect measure-behavior correlations have been estimated to reach $r = .37$ (Kurdi et al., 2019). However, uniting the most successful moderators within a pool of relatively underpowered studies (mean post hoc power of .27; Kurdi et al., 2019) increases the likelihood of publication bias driving results. It is therefore important to examine how well indirect measures of implicit racial attitudes perform in large-scale preregistered tests under conditions designed to be as theoretically optimal as possible.

The Present Adversarial Collaboration

Given the conflicting results and interpretations within the literature on indirect measures of racial attitudes, the predictive validity of indirect measures is an ideal candidate for adversarial collaboration (Tetlock & Mitchell, 2009a). Adversarial collaboration seeks to reduce the influence of researchers' biases and build consensus by having disputant scholars work with each other (e.g., Clark & Tetlock, 2023; Mellers et al., 2001). The present investigation brings together prior proponents and critics of indirect measures, plus relatively neutral parties. Together, we codesigned the study to provide a rigorous, transparent test of our competing predictions. See additional online material (https://osf.io/7jtmb) for more information about the theoretical perspectives and diverging predictions of the collaborators on this project.

We identified a set of lab-based measures of explicit and implicit racial attitudes as well as behavioral measures of racial discrimination using an extensive literature search and expert consultation. This process involved placing particular emphasis on accounting for the reliability and validity of measures. In addition, we used structural equation modeling (SEM) to model implicit attitudes, explicit attitudes, and discriminatory behavior as latent constructs, and to account for measurement error in each construct's observed indicators. Notably, while SEM has been used to account for measurement error in direct and indirect measures (e.g., Axt et al., 2020; Bar-Anan & Vianello, 2018; Carpenter et al., 2023; Cunningham et al., 2001, 2004), it has to our knowledge not been used to account for error in discrimination outcome measures, despite such error being a potentially important factor attenuating previously observed associations (Kurdi et al., 2019).

Competing Accounts of Indirect Measures and Discrimination

We sought to adjudicate between eight competing perspectives concerning the prevalence and predictive validity of indirect measures of racial attitudes. The strongest claims from the implicit attitude perspective are what we call the *pervasive bias* position (e.g., Jost et al., 2009). According to this view, racial bias is robust, operating automatically and unintentionally. Further, this bias contributes to discriminatory behavior (e.g., Correll et al., 2002; Greenwald et al., 2003; Hekman et al., 2010; Kubota et al., 2013; Lahey & Oxley, 2021; Peyton & Huber, 2021; Rudman & Lee, 2002; Ziegert & Hanges, 2005), though this claim may only apply to more naturalistic real-world settings and not generalize to laboratory measures. Still, from a view that anticipates anti-Black discrimination to appear in more controlled, laboratory settings, favoritism toward White over Black people should emerge on both indirect measures and measures of behavior, with indirect measures being associated with behavioral outcomes. Further, indirect measures should exhibit incremental validity by predicting discriminatory outcomes over and above self-reported racial attitudes (i.e., direct measures).

A more nuanced perspective is what we call the *incremental validity* account. This view argues that anti-Black discrimination may not emerge on lab-based behavioral measures, but will emerge on indirect measures, which are less susceptible to self-presentation concerns. Indirect measures are still expected to predict variance in discriminatory behavior above-and-beyond direct measures. This pattern has occurred in prior work, such as in research linking IAT scores to decisions concerning economic trust (Stanley et al., 2011) or in hypothetical admissions tasks (Axt et al., 2016). Further, anti-Black discrimination should still be observable in participants who score highest in bias on indirect measures of implicit racial attitudes.

Dual-process theorizing offers a more complex set of predictions (e.g., Fazio, 1990; Gawronski & Bodenhausen, 2006), positing moderating effects of the controllability of the behavior as well as the motivations of the participant (Dunton & Fazio, 1997). According to the MODE model (Fazio, 1990), individuals low in motivation to control prejudice should exhibit positive correlations between indirect measures and measures of discrimination. However, those highly motivated to respond without prejudice will correct for their automatic biases when possible, showing no such relationship and may even overcompensate by discriminating in favor of Black targets (Dunton & Fazio, 1997). On this *motivated correction* account, behavioral biases are a function of interactions between implicit attitudes and personally endorsed motivations.

Other dual-process theorizing posits that indirect measures should predict less controlled behaviors, while direct measures should predict more controlled acts (e.g., Dovidio et al., 1997; Friese et al., 2008; Gawronski & Bodenhausen, 2006; Perugini et al., 2010), aligning with past research that has found indirect measures exhibited higher predictive validity for the behavior of participants with lower executive control (e.g., Hofmann, Friese, et al., 2008) or under conditions of limited control (e.g., Hofmann, Gschwendner, et al., 2008). This is often termed the *moderation account* (for summaries, see Friese et al., 2008; Perugini et al., 2010). While our collaboration focused on discriminatory outcomes (e.g., biased hiring decisions) that are in principle controllable, we included two measures that provide partial tests of the moderation account. Specifically, we measured participants' self-reported distraction during the behavioral tasks, since behavior of more distracted participants should be better predicted by indirect measures. In addition, we included an established measure of executive control, an automated span task (Unsworth et al., 2005), testing whether indirect measures better predict the discriminatory behavior of participants who do worse on the span task (Hofmann, Friese, et al., 2008).

The most skeptical perspective is what we call the *false positives* position. Though past meta-analyses report positive but small predictive validity correlations (Greenwald et al., 2009; Oswald et al., 2013; most recently $r = .10$ in Kurdi et al., 2019), work in other fields finds that meta-analytic estimates based on the published literature are on average three times larger than replication effect sizes (Kvarven et al., 2020). The small average estimated relationship of $r = .10$ between indirect measures and behavior (Kurdi et al., 2019) may be a product of the publication filter (Fanelli, 2010) combined with researcher degrees of freedom (Wagenmakers et al., 2012), though the Kurdi et al. (2019) meta-analysis found no evidence of publication bias in studies concerning the association between indirect measures and intergroup behavior. Still, one possible position is that the collection of predictive validity studies for indirect measures of racial attitudes constitutes a "null field" (Ioannidis, 2005) with a true overall estimate near zero. If so, the previously reported relationship between indirect measures and behaviors should not emerge within our preregistered analyses. Previously observed discriminatory outcomes, such as relatively less resource allocation to Black targets (Kubota et al., 2013), should fail to replicate as well. Although none of the present adversarial collaborators endorse the false positives perspective, it is worth testing (and potentially disconfirming).

Another critical position is that indirect measures predict discriminatory behavior but have *limited value-add* above and beyond self-report measures of explicit racial attitudes (e.g., Schimmack, 2021). One reason this could occur is if indirect measures are highly process-impure and tap a mix of individual-level attitudes, propositions, and exposure to cultural knowledge (Arkes & Tetlock, 2004). Direct measures are more influenced by propositional thinking (Gawronski & Bodenhausen, 2006) and assumedly less contaminated by extraneous factors, such that they are superior assessments of racial attitudes (Corneille & Gawronski, 2024). This perspective anticipates that indirect measures may correlate significantly with behavioral measures of discrimination, but the indirect measures will not exhibit incremental validity after controlling for scores on direct measures.

Confirmatory and Exploratory Tests

We tested four key research questions in a large sample of White Americans. Research Questions 1 and 2 focused on the presence or absence of racial bias within our indirect and behavioral measures. Research Questions 3 and 4 focused on the predictive and incremental validity of indirect measures, asking whether latent implicit racial attitudes significantly predicted latent tendencies to racially discriminate, without (predictive validity) and with (incremental validity) controlling for latent explicit racial attitudes.

To allow for informative null results, each research question encompassed two versions. "A" versions tested for the existence of nonzero effects via traditional null hypothesis significance tests comparing observed effects to null hypotheses of zero. "B" versions were tested when the corresponding "A" version failed to find support. Specifically, their aim of the "B" version analyses was to test for the effect to be within a bound of practical nonsignificance, comparing observed effects to prespecified equivalence bounds (Lakens et al., 2018). For equivalence tests, the chosen bounds were by necessity subjective and were selected to reflect a balance between (a) the desire for informative tests in which bounds represent conservative estimates of expected effects; (b) the desire for high statistical power, with each equivalence test achieving 95% power to reject effects outside the selected bounds; and (c) agreement among collaborators that these represent meaningful thresholds (e.g., for negligible vs. nonnegligible predictive validity of indirect measures).

> *Research Question 1A:* Do White Americans display pro-White/Anti-Black racial bias on indirect measures?

> *Research Question 1B:* Do White Americans display pro-White/Anti-Black racial bias on indirect measures weaker than an effect of $d = 0.1$?

> *Research Question 2A:* Do White Americans display pro-White/Anti-Black racial bias on behavioral discrimination measures?

> *Research Question 2B:* Do White Americans display pro-White/Anti-Black racial bias on behavioral discrimination measures weaker than an effect of $d = 0.1$?

> *Research Question 3A:* Among White Americans, do latent implicit racial attitudes predict latent tendencies to racially discriminate, without controlling for latent explicit racial attitudes (predictive validity)?

> *Research Question 3B:* Among White Americans, is the relationship between latent implicit attitudes and latent tendencies to racially discriminate without controlling for latent explicit racial attitudes weaker than a standardized implicit attitude → discrimination path coefficient of $\beta = 0.15$?

> *Research Question 4A:* Among White Americans, do latent implicit racial attitudes predict latent tendencies to racially discriminate over and above latent explicit racial attitudes (incremental validity)?

> *Research Question 4B:* Among White Americans, is the relationship between latent implicit racial attitudes and latent tendencies to racially discriminate after controlling for latent explicit racial attitudes weaker than a standardized implicit attitude → discrimination path coefficient of $\beta = 0.15$?

Table 1 summarizes the key predictions of the competing theories, which differ for many of the analyses associated with the research questions above. The table also summarizes the potential moderating effects of additional measures of individual differences included based on our expert consultations, such as explicit attitude strength (Petty & Krosnick, 1995) and study savviness (i.e., whether effects differ among participants with prior experience completing psychology studies). We treated this array of moderator tests as exploratory considering many cases where the expected empirical pattern was unclear or multiple patterns might be anticipated (see Table 1 for more information). Note that Research Question 1 is not covered in Table 1

Table 1
Directional Predictions of the Competing Theories of Implicit Attitudes and Behaviors

Theoretical perspective	Research question			
	RQ2: Racial bias on behavioral measures	RQ3: Predictive validity of IMs	RQ4: Incremental validity of IMs	Moderators of predictive and incremental validity of IMs
Pervasive bias account	Pro-White/anti-Black bias in behaviors, at least in naturalistic real-world settings.	Pro-White/anti-Black bias on IMs predicts pro-White/anti-Black bias on behavioral measures.	Pro-White/anti-Black bias on IMs predicts pro-White/anti-Black bias on behavioral measures after controlling for explicit racial attitudes.	No directional prediction.
Incremental validity account	No directional prediction.	Pro-White/anti-Black bias on IMs predicts relatively more favorable behaviors toward White than Black targets.	Pro-White/anti-Black bias on IMs predicts relatively more favorable behaviors toward White than Black targets after controlling for explicit racial attitudes.	No directional prediction.
Dual process: moderation account	Pro-White/anti-Black bias most likely to emerge for decision makers low in executive control, or who report being distracted or fatigued on relatively less controlled decisions.	No directional prediction; unclear a priori how controlled or uncontrolled the behaviors measured will be overall.	No directional prediction; unclear a priori how controlled or uncontrolled the behaviors measured will be overall.	IMs will predict behavior more effectively for people low in executive control and distracted participants. The predictive validity, incremental validity, and behavioral discrimination predictions of the pervasive bias approach should be supported for participants low in executive control and participants who are distracted during the behavioral task.
Dual process: motivated correction	Pro-White/anti-Black bias most likely to emerge for participants' low (rather than high) in motivation to respond without prejudice. Those high in such motivations may exhibit less bias, no bias, or pro-Black/anti-White bias.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in motivation to be unprejudiced.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in motivation to be unprejudiced.	More positive implicit attitudes toward White relative to Black people predicts more favorable behaviors toward White than Black targets, but more so for participants low (rather than high) in motivations to respond without prejudice. For those high in such motivations, this correlation should be attenuated.
False positives account	No significant or substantial bias on behavioral measures.	No significant or substantial correlation in any direction.	No significant or substantial variance in behavior explained by IMs after controlling for explicit racial attitudes.	No moderating effects of individual differences on the predictive or incremental validity of IMs.
Limited value-add account	No directional prediction.	No directional prediction.	No significant or substantial variance in behavior explained by IMs after controlling for explicit racial attitudes.	Considering further individual differences will not lead to strong predictive or incremental validity for IMs after controlling for explicit attitudes.
Attitude strength account	No directional prediction.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in explicit attitude strength.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in explicit attitude strength.	More positive implicit attitudes toward White relative to Black people predicts more favorable behaviors toward White than Black targets, but more so among participants low (rather than high) in explicit attitude strength. For those high in attitude strength, this correlation should be attenuated, absent, or reversed.

(*table continues*)

Table 1 (*continued*)

Theoretical perspective	Research question RQ2: Racial bias on behavioral measures	RQ3: Predictive validity of IMs	RQ4: Incremental validity of IMs	Moderators of predictive and incremental validity of IMs
Study savviness account	Pro-White/anti-Black bias most likely to emerge for participants' low (rather than high) in study savviness. Those high in study savviness may exhibit less bias, no bias, or pro-Black/anti-White bias.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in study savviness.	Same prediction as the pervasive bias account, but more so for participants low (rather than high) in study savviness.	More positive implicit attitudes toward White relative to Black people predicts more favorable behaviors toward White than Black targets, but more so among participants low (rather than high) in study savviness. For those high in savviness, this correlation should be attenuated, absent, or even reversed.

Note. In this table, "no significant or substantial" should be read as a prediction that tests of "A" versions of the relevant research questions will be nonsignificant, while tests of "B" versions (i.e., equivalence tests) will be significant. IMs = indirect measures; RQ2, RQ3, RQ4 = Research Questions.

because all theoretical perspectives are compatible with (i.e., not contradicted by) pro-White/Anti-Black racial attitudes on the four indirect measures used.

The present collaboration is not the last word on the relationship between indirect measures of implicit racial attitudes and discriminatory behavior. No single study could hope to fully resolve this debate regarding a vast literature involving many measurement approaches (Uhlmann et al., 2012) and discrimination paradigms (Kurdi et al., 2019). Rather, we aimed to introduce rigorous, transparent, and high-quality empirical evidence into this discussion. Our adversarial collaboration also offered the opportunity for the respective sides of the ongoing debate about the predictive validity of indirect measures to update preexisting beliefs and perhaps move closer to consensus.

Method

Participants

Participants were recruited through Connect, an online source of participants managed by CloudResearch. The study was available to Connect participants who (a) identified as White and non-Hispanic, (b) were citizens or residents of the United States, (c) were at least 18 years old, and (d) had a prior approval rate of at least 95%. To increase representativeness, we stratified the sample to match population proportions in terms of gender and political affiliation (see Supplemental for details). We planned to recruit 2,400 participants to meet our target sample size of 2,000 (anticipating exclusions following attention checks and preregistered criteria for each measure and attention check, detailed in additional online material at https://osf.io/7jtmb). In total, data were collected from 2,313 participants. Data collection was stopped slightly early due to an inability to recruit additional participants willing to complete a two-session study. These difficulties in finishing data collection also meant that there were fewer participants, and especially male participants, relative to our targeted sample demographics.

Of the 2,313 participants who started the survey, 32 were duplicates (participants who completed the study multiple times), three started but did not finish Session 1, and 124 completed Session 1 but did not return for Session 2 (5.4% attrition). Among the remaining 2,154 unique participants who completed both sessions, 40 were excluded for failing the preregistered attention check, resulting in a final analytic sample of 2,114 (56.5% female, 43.2% male, 0.3% nonbinary, and 0.05% preferred not to say, $M_{\text{age}} = 41.9$, $SD = 13.6$). Participants reported the following political affiliations: 29.8% democrat, 29.6% republican, 40.1% independent, and 0.5% other/unknown. Participants were compensated \$6 per session, plus a bonus of \$6 for participants completing both sessions (\$18 total). Note that other preregistered exclusion criteria (e.g., related to response latencies or error rates on specific measures) led only to task-level data removal, not exclusion of participants from the overall data set.

Design

Participants were instructed that the study was aimed at measuring different cognitive abilities, attitudes, and beliefs and that they would complete two sessions across multiple days (within the same week), each taking approximately 45 min. In the first session, participants completed four behavioral measures of racial discrimination in randomized order, as well as suspicion probes, measures of executive function, situational fatigue, and demographics. In the second session, participants completed four indirect measures of racial attitudes, followed by four direct measures, each in randomized order, and finally measures of motivations to control prejudice and attitude strength.

We viewed a two-session design as necessary to avoid a prohibitively long single-session study, though this decision likely inflated the correlations between measures completed in the same session (e.g., those between direct and indirect measures of racial attitudes). If anything, including direct and indirect measures in the same study session then provides a more conservative test of the incremental predictive validity of our indirect measures.

Notably, we employed a fixed order of measures (though randomizing within each type of measure): the behavioral measures were always completed initially within the first session, while the second session always started with the indirect followed by the direct measures. The decision to place behavioral measures first followed from our belief that completing obviously race-related measures like the IAT or self-reported racial attitudes would sensitize participants to the purpose of the study, reducing the chances that indirect measures would predict behavior.

In addition, the decision to place indirect measures at the beginning of the second session arose from concerns that direct measures would further heighten the salience of race, which could distort responses on the EPT given prior work finding that responses on some indirect measures may shift when race is made more salient beforehand (Olson & Fazio, 2003). Placing indirect measures at the beginning of the second session came from an attempt to minimize the influence of participant fatigue; since these measures are more cognitively taxing than self-report measures, completing them earlier in the session could help minimize participant dropout.

Finally, moderator variables were collected in a fixed order, with savviness, situational fatigue, and executive control assessed at the end of the first session while attitude strength and motivation to control prejudice were collected at the end of the second session. This order stemmed from a priority to place the savviness measure immediately after the behavioral measures, since we were concerned that suspicion would distort responses on Session 1's behavioral measures more so than on Session 2's racial attitude measures.

Specifically, our focus was on participant suspicion distorting behavioral responses. We believed that many participants would report suspicion that the study concerned race after completing any of our direct or indirect measures of racial attitudes, so assessing savviness after those measures would not be informative about whether suspicion was present during the subtler behavioral measures. Including the suspicion probe later in the study design would have then precluded tests of suspicion as a moderator and put the study-savviness account at a disadvantage. Similarly, we were interested in measuring cognitive resources at the time discriminatory decisions were made. The placement of the remaining moderator variables (attitude strength and motivations to respond without prejudice) in the second session was for the practical reason of the greater amount of time remaining in Session 2 versus Session 1, yet arbitrary on theoretical grounds.

Discrimination Measures

To identify relevant studies that included behavioral measures of discrimination, we reviewed the references of Kurdi et al. (2019), Oswald et al. (2013), Greenwald et al. (2009), and Cameron et al. (2012), and searched for additional articles published between 2019 and July 26th, 2022, using the terms "anti-Black laboratory discrimination experiment." We also contacted experts in the field to solicit suggestions.

This search resulted in an extensive list of prior methods. To narrow this field, we applied a number of criteria (see Supplemental Materials for full list of measures considered). In particular, we sought measures that (a) were deployable online, (b) elicited overall evidence of anti-Black discrimination in prior work, (c) had been linked with indirect measures in prior work, and (d) provide an individual-level index of participants' relative tendencies to discriminate (thus enabling us to use each measure as indicators of anti-Black discriminatory behavior in our SEM).

We selected four measures, each meeting at least three of the criteria. The set of racial discrimination measures included an ultimatum game (UG; Kubota et al., 2013), a trust game (TG; Stanley et al., 2011), a version of the judgment bias task (JBT; Axt et al., 2016), and a resume evaluation task (RET; Lahey & Oxley, 2021). These four measures, two concerning selection and two concerning cooperation, come with rich theoretical backgrounds for why target race may influence responses. For instance, past work has found greater tolerance or support for biased resource allocations that favor one's ingroup (Diekmann et al., 1997), and comparable effects could emerge among participants in this work who must make allocation decisions to racial ingroup versus outgroup members. Similarly, prior studies on the topic of aversive racism suggest that decision-makers are inclined to favor racial ingroup members when selection judgments are difficult or ambiguous (Dovidio & Gaertner, 2000; Dovidio & Gaertner, 2004). We further believed that the behavioral measures tap basic social orientations toward members of the relevant racial groups and mirror socially important decisions. We provide additional details on each measure in the Supplemental Materials.

UG. In the UG participants decided whether to accept or reject splits of $10 offered by interaction partners. If participants accepted a split, they and their partner kept their respective shares. If they rejected a split, both received nothing. The measure aimed to capture the extent to which participants were willing to accept a monetary loss in order to punish partners for insufficient offers. Interaction partners consisted of men and women (40 Black, 80 White, 40 from other races) displayed via upper-body photographs. Although analyses were based on responses to Black and White targets, the targets from other races (and unequal proportions) were included to create believable demographics of the target pool. In pretesting, all Black and White interaction partners were categorized as the intended race by at least 80% of raters. All interaction partners were also matched across racial groups on prior ratings of perceived age and attractiveness. Interaction partners were described as previously recruited participants who made offers in an earlier data collection session. Participants were instructed that two of their decisions would be randomly selected to be realized after the completion of the study, at which point both they and their partners would receive their earnings for those two decisions.[1] In prior research using this method, participants were more willing to accept offers from White versus Black partners (Kubota et al., 2013). Individual differences in this bias were correlated with race IAT scores ($r = .42$), after controlling for racial biases in points of indifference on the task (i.e., race-based differences in where an offer was equally likely to be accepted or rejected). In our version, Black and White targets were separately and randomly paired with matching distributions of offers (40 unique offers matching the distribution used by Kubota et al., 2013; $M = \$1.94$, $SD = \$0.99$, range = $0.00–$3.80; each offer matched twice with White targets). The targets from other races were paired with 40 unique offers matching the Black/White distributions in mean and variance.

TG. In the TG, participants chose whether or not to "trust" the same set of interaction partners used in the UG. In each trial, participants selected how much out of $3.00 they wished to share with a partner (again displayed via upper-body photograph), with the possible choices $0.00, $0.50, $1.00, $1.50, $2.00, $2.50, or $3.00. Ostensibly, shared money was then quadrupled, and interaction partners were given the power to either reshare half of the larger sum or keep it all. Participants therefore knew that any money they shared could be doubled (if their partner reshared) or lost (if their partner defected). Partners were again described as previously recruited participants who made their decisions to share or defect in a prior data collection session, and participants were again instructed that two decisions would be randomly selected to be realized after completion of the study. Images used in the TG were the same as those in the UG, since stimuli for both tasks were described as coming from people recruited in a public place. Prior research found

[1] Participants were instead paid a flat bonus for completing these measures. This deception was approved by the relevant ethics board in our study review protocol.

that while participants were similarly willing to share money with Black and White interaction partners, participants with higher pro-White race IAT scores shared greater amounts of money with White versus Black partners (r = .41 in Stanley et al., 2011).

JBT. In the JBT, participants acted as hiring managers and made decisions to accept or reject job applicants (16 Black, 32 White, 16 other races) based on displayed profiles. By design, half of the profiles in each racial group were higher quality and half were lower quality, based on four metrics: science proficiency, math proficiency, reference quality, and interview score. In addition to scores on each metric, profiles displayed AI-generated professional headshots of candidates, obtained from multiple sources and matched across racial groups on prior ratings of age, attractiveness, positive affect, and grooming.

The JBT has previously revealed overall discrimination along numerous dimensions, such as physical attractiveness (Axt et al., 2018), but has produced pro-Black biases in decisions to accept applicants into academic honors societies (Axt et al., 2016). Despite failing to show anti-Black discrimination, the JBT was included given evidence of its relatively strong psychometric properties and reliable correlations with a Brief IAT measuring implicit racial attitudes (meta-analytic r = .16 in Axt et al., 2016). To make the task more relevant to participants, the context was shifted from academic decisions to personnel selection. As in the standard JBT, participants were instructed to accept approximately half of the applicants and viewed each profile one by one in randomized order before making decisions. A pilot study using this adapted JBT suggested it performed similarly to the standard JBT, with participants showing an overall pro-Black bias on selection decisions but with JBT scores correlating weakly with IAT scores (r = .16) and a direct measure of racial attitudes (r = .15; see Pilot 3 in Supplemental Materials).

RET. In the RET, participants acted as hiring managers and evaluated the resumes of Black, White, and other race applicants for an administrative role. These resumes were developed based on prior research where participants provided lower ratings of hireability for Black versus White resumes, specifically among younger applicants (<45 years). Resume evaluations were associated with indirect measures (albeit measures of automatic age bias in a nonlinear, quadratic relationship; Lahey & Oxley, 2021). Each resume was prerated on perceived competence, trainability, and hireability, but with race cues removed (i.e., names blacked out). These ratings were high in reliability (Cronbach's α = .92) and were averaged to form overall resume evaluation scores. Using these scores, we selected two groups of six resumes each that had equally strong ratings. We then assigned these groups Black- and White-coded names (three female and three male names each). Each name was pretested among US participants for perceived race in prior research (Barlow & Lahey, 2018; Crabtree et al., 2023). Pretesting found that the 12 names used in primary analysis had an average correct racial classification rate of 85% (minimum 69%), with 10 names having a correct classification rate above 80%. Eight additional "dummy" resumes were also included to increase external validity; these remaining resumes were assigned White- (3), Hispanic- (3), and Asian-coded (2) names (four male, four female). Before making evaluations, participants viewed each resume in a randomized order.

Indirect Measures

When selecting indirect measures, we considered instruments' widespread adoption, psychometric properties, convergence with other indirect measures, and prior evidence of predictive validity. We also sought methodological diversity to minimize shared method variance. Based on these criteria, we selected the Black–White IAT (Greenwald et al., 1998) and EPT (Fazio et al., 1986), as well as the Affect Misattribution Procedure (Payne et al., 2005), and Black and White Single-Category IATs (Karpinski & Steinman, 2006).

In one sense, these indirect measures could be considered measures of behavior since each forms an index of racial attitudes based on participant reactions (e.g., the speed at which computer keys are pressed; De Houwer, 2019). While we acknowledge that the distinction between indirect measures and behavioral outcomes is a fine one, the selected indirect measures sought to assess Black–White racial attitudes more generally, compared to behavioral measures that looked to assess preferences within a specific context (e.g., in evaluating resumes).

Stimuli and Category Labels. The EPT, IAT and SC-IAT each used the labels "Good" and "Bad" for valenced word stimuli. The IAT and SC-IATs used the labels "Black People" and "White People" for Black and White targets, and the AMP used the labels "Pleasant" and "Unpleasant" for Chinese character stimuli. Tasks requiring word stimuli (IAT, EPT, SC-IAT) used as good words "friend," "smiling," "adore," "joyful," "pleasure," "friendship," "happy," and as bad words "bothersome," "poison," "pain," "nasty," "dirty," "hatred," "rotten," and "horrific." These were all classed as high-performing words on indirect measures in prior research (Axt et al., 2021). For Black and White target images the IAT, SC-IAT, and AMP used a set of 12 cropped Black and White faces of male and female athletes commonly employed in indirect measures (Nosek & Banaji, 2001). The EPT used 48 images of Asian, Black, Latino, and White faces (six female and six male for each race) from the Chicago Face Database (CFD; Ma et al., 2015). In all tasks, category labels appeared at the top left and right corners of the screen, and the "E" and "I" keys were used to categorize targets to left and right categories, respectively.

Implicit Association Test. The IAT used a seven-block procedure with 120 critical trials (Nosek et al., 2007). In each trial, participants categorized words or images appearing on their screens as quickly as possible with computer key presses. In Block 1 (practice block), participants categorized Black or White faces. In Block 2 (another practice block), participants categorized good or bad words. Blocks 3 and 4 (test blocks) combined the first two blocks, for example, by grouping Black faces and bad words with one key and White faces and good words with the other key. Blocks 5, 6, and 7 (practice, test, and test blocks, respectively) were the same as Blocks 1, 3, and 4, but the key assigned to racial categories switched sides. Block order was randomized across participants.

Single-Category Implicit Association Test. The SC-IAT consisted of two separate tasks, a Black SC-IAT and a White SC-IAT. Each had four blocks and 192 critical trials (Bar-Anan & Nosek, 2014). Like the IAT, participants categorized words or images as quickly as possible. Unlike the IAT, participants categorized just three categories of stimuli: good words, bad words and either Black (in the Black SC-IAT) or White (in the White SC-IAT) targets. In Blocks 1 and 2, targets were grouped with either the good or bad words on one key. In Blocks 3 and 4, targets were grouped with the opposite valenced-words on one key. Block order within each SC-IAT (i.e., blocks pairing with positive or negative valence words) was randomized. Black and White SC-IATs were completed sequentially but their order was also randomized.

EPT. The EPT used a seven-block procedure and 192 critical trials (Fazio et al., 1995). In Blocks 1 and 2 (described as a practice and test block, respectively), participants categorized good or bad words appearing on their screens as quickly as possible, within a time limit of 1,750 ms. Before each word appeared, the prime "+++" appeared in the center of the screen for 315 ms, followed by an interstimulus gap of 135 ms. In Blocks 3 and 4, participants memorized 16 Asian, Black, Latino, and White female and male CFD faces, each presented for 315 ms once in each block. Block 5 was a memory test intended only to bolster the cover story, in which participants were shown 20 faces and asked to judge whether each face was previously presented (10 faces from Blocks 3 and 4, 10 novel CFD faces). In Block 6 (the test block), participants were told the previous two tasks were being combined to assess the independence of processes involved in word recognition and memory. This block proceeded as Blocks 1 and 2, but with 48 novel Asian, Black, Latino, and White CFD faces appearing as primes in place of "+++." Participants tried to categorize words as quickly as possible, but simultaneously tried to memorize the faces appearing prior to the words. Block 7 consisted of a second memory test intended to bolster the cover story, with participants shown 20 faces (10 of the primes from block 6 plus 10 novel CFD faces) and asked to judge whether each was previously presented.

Affect Misattribution Procedure. The AMP used a four-block procedure with 120 critical trials (Bar-Anan & Nosek, 2014). Each trial in Blocks 1 (a practice block), 2, 3, and 4 (test blocks) followed the same pattern. First, a Black or White face appeared for 75 ms, followed by a 125 ms interstimulus gap, followed by a picture displaying a Chinese character for 100 ms, followed by a mask of visual noise. Participants were asked to ignore the faces and judge whether the Chinese character was more or less visually pleasing than average.

Direct Measures

Direct Preference Measures. We included a set of five direct preference measures (DPMs) capturing self-reported preferences and relative feelings about Black versus White people (Axt, 2018). Item 1 was a multiple choice question measuring relative preference for Black compared with White people ("Which statement best describes you?" with seven response options ranging from *I strongly prefer Black people to White people* to *I strongly prefer White people to Black people*). Items 2 and 3 were multiple choice questions measuring liking of White and Black people ("On average, how much do you like White/Black people?" with seven response options ranging from *strongly like* to *strongly dislike*). Items 4 and 5 were sliders measuring positivity of feelings of White and Black people ("How negative or positive are your feelings toward White/Black people?" With labels of *extremely positive* and *extremely negative* at the slider endpoints and *neutral* at the midpoint).

ANES Measures. We included a measure of pro-White/anti-Black stereotyping from the American National Election Survey (ANES; Payne et al., 2010). This measure consisted of six multiple choice items, with responses constructed to represent a range from positive to negative attitudes and beliefs about Black people (e.g., response options for the item "How often have you felt sympathy for Black people?" are *always*, *most of the time*, *about half the time*, *once in a while*, and *never*).

Prejudice Index. Bobo and Kluegel's (1993) Prejudice Index (PI) consists of 10 multiple choice questions capturing stereotypical beliefs about Black and White people. Items 1–5 refer to Black people (e.g., "Do Black people tend to be violence prone or do they tend not to be prone to violence?" with response options *very violence prone*, *moderately violence prone*, *slightly violence prone*, *neutral*, *slightly not violence prone*, *moderately not violence prone*, *very not violence prone*). Items 6–10 repeat items 1–5 but refer to White people.

Symbolic Racism. The Symbolic Racism Scale (SRS; Henry & Sears, 2002) consists of eight multiple choice questions measuring attitudes and beliefs about Black people (e.g., "It's really a matter of some people not trying hard enough; if Black people would only try harder they could be just as well off as White people" with response options *strongly agree*, *somewhat agree*, *neither agree nor disagree*, *somewhat disagree*, and *strongly disagree*).

In general, our direct measures showed less methodological diversity than our indirect measures, as each relied on Likert scales with similar numbers of response options. Greater amounts of shared method variance likely inflate the observed correlations between our direct measures.

Moderating Variable Measures

Finally, we included several individual differences measures of potential moderating factors.

Motivation to Control Prejudice. Whether or not there is a direct effect of implicit racial attitudes on behaviors may depend on the motivation levels of the participant (Fazio, 1990). We measured individual differences in such orientations using Dunton and Fazio's (1997) Motivation to Control Prejudice Reactions (MCPR) scale (e.g., "In today's society it is important that one not be perceived as prejudiced in any manner.").

Executive Function. To measure executive function, we used an automated operation span task (AOST). Participants solve a series of basic math problems, for example, "$(1^*2) + 1 = ?$," while simultaneously memorizing a sequence of letters. The AOST exhibits strong internal consistency and test–retest reliability (Unsworth et al., 2005) and in prior research has been associated with the predictive validity of indirect measures, with individuals scoring relatively lower on the AOST exhibiting stronger relationships between indirect measures and behaviors (Hofmann, Friese, et al., 2008).

Situational Fatigue. To measure state cognitive fatigue, we used three items from Shirom and Melamed's (2006) Burnout Questionnaire, adapted to measure temporary state fatigue: "I feel drained," "My mind feels unfocused right now," "Right now, it would take a lot of effort for me to focus on something."

Attitude Strength. Influential work on attitude–behavior relations finds that strongly held self-reported attitudes predict behavior more effectively than weaker attitudes (Petty & Krosnick, 1995). This relationship implies that indirect measures may fail to show incremental validity for individuals with strongly held explicit racial attitudes.[2] We included six items to measure two subfacets of attitude strength: subjective ambivalence (AS-SA; e.g., "Rate the extent to

[2] A related possibility is that attitude strength is associated with higher intercorrelations between explicit and implicit attitudes due to increased elaboration (Axt & Roy, 2025; Nosek, 2005). Measures of implicit racial attitudes could then have strong indirect effects on behavior via explicitly self-reported attitudes when the latter are strongly held. Testing this perspective was not part of our preregistered analysis plan but is reported in the Online Supplemental (see the Supplemental Results section).

which you feel conflicted regarding your attitudes about race"; Priester & Petty, 1996) and attitude certainty (AS-AC; e.g., "How certain are you in your attitudes about race?"; Petrocelli et al., 2007).

Study Savviness. Contemporary research participants may be relatively wise to experimental manipulations and able to guess the study's purpose, modifying their responses to provide data with little relation to their true attitudes and biases (Tierney et al., 2020). In our primary measure of suspicion, a free response item asked, "What did you think this study was about?" This open-ended item was presented immediately after the behavioral measures and on a separate page from the subsequent items. Free responses mentioning race, ethnicity, bias, discrimination, or synonyms of these words were coded as reflecting high suspicion. The following page included follow-up items prompting participants to think about the role of race in their judgments. A numeric rating item asked, "During the tasks you just completed, did the race of the target person (Black or White) influence you in any way?" (Response options ranged from 1 = *definitely not*, to 6 = *yes, definitely*). A follow-up free response item inquired "If yes, please explain how and why it influenced you in your own words."

Our procedure for assessing savviness was adapted from funneled debriefings used in the literature on awareness of influence (Bargh & Chartrand, 2000; Uhlmann et al., 2011). As supplementary indicators of likely study savviness, participants reported the number of research studies they had taken part in before, whether they had completed a similar study previously, and whether they had taken a psychology course before (Tierney et al., 2020).

Transparency and Openness

We report how we determined our sample size, all data exclusions, and all measures. All de-identified data,[3] materials and analysis code are available on the Open Science Framework (OSF) at https://osf.io/wsu4q/ (Connor et al., 2025). The study design and analysis were preregistered (see https://osf.io/gjxu5). We note any deviations from the preregistration.

Analytic Strategy

Given the scope of the project, we incorporated multiple analytic strategies to ensure transparency and assess the robustness of findings. First, we preregistered a set of confirmatory analyses that directly test the four primary hypotheses on racial bias (implicit and behavioral) and the predictive and incremental validity of indirect measures on discriminatory behavior. For these we used field-standard frequentist analyses—one-sample t tests and SEM. Given the anticipated noise in the indirect and behavioral measures, we opted for an SEM that automatically accounts for the relative reliability of the measures by weighing their contribution by the amount of shared variance (i.e., signal) in the data. SEMs were estimated with *lavaan* (Rosseel, 2012), using the robust maximum likelihood method to protect estimates from potential multivariate nonnormality with the Yuan–Bentler scaled chi-squared statistic and robust standard errors (MLR; Yuan & Bentler, 2000). All SEMs included mean structures (Little, 1997) and missing data were handled via full information maximum likelihood estimation (Anderson, 1957; Arbuckle et al., 1996). All frequentist significance tests relied on a traditional α threshold of .05.

Second, we conducted a Bayesian multiverse analysis (Hoogeveen et al., 2023; Steegen et al., 2016) that directly compares the evidence for the different theoretical perspectives given various plausible constellations of the data (e.g., operationalization of variables) and modeling decisions (e.g., inclusion of specific measures in the SEM). Because the multiverse was not preregistered, we applied a blinding strategy (Dutilh et al., 2019; MacCoun & Perlmutter, 2015) to avoid bias in path specification. Third, we explored evidence for further moderators and predictive value of each of the measures separately in a Bayesian linear regression framework, applying model averaging to guard against spurious effects from the large number of predictors. Fourth, to address measurement noise, we fit hierarchical Bayesian models for each of the indirect and behavioral measures to quantify the ratio between reliable interindividual differences in racial bias effects and trial noise (cf. Rouder & Mehrvarz, 2024). Finally, to examine boundary conditions, we explored the conditions under which anti-Black discrimination emerged and further assessed the incremental validity of indirect over direct measures in predicting discriminatory behavior. All analyses were performed in R (Version 4.3.3; R Core Team, 2024; specific packages are reported in the relevant sections).

Results

Statistical Power

To estimate statistical power for the single-sample t tests (RQ1A and RQ2A) and their associated equivalence tests (RQ1B and RQ2B), we used functions from the MESS (Ekstrøm, 2022) and TOSTER (Caldwell, 2022; Lakens, 2017) R packages to examine whether our targeted sample size of $N = 2000$ provided adequate power for primary analyses. Results suggested that all planned t tests had >95% power to detect effects of Cohen's $d = 0.1$ or larger. Associated equivalence tests had >95% power to reject effects outside equivalence bounds of Cohen's $d = -0.1$ and 0.1.

For tests of predictive and incremental validity (RQ3A and RQ4A) and their associated equivalence tests (RQ3B and RQ4B), we conducted simulation-based bootstrapped power analyses (1,000 iterations each) using the faux (DeBruine, 2023) and lavaan (Rosseel, 2012) R packages. We began by simulating a population of data with standard normal distributions of each of the observed variables.

In these simulated data, relationships between indicators were governed by six parameters: indirect measure-criterion correlations (icc), indirect–direct measure correlations (idc), direct measure-criterion correlations (dcc), indirect measure-indirect measure correlations (iic), direct measure–direct measure correlations (ddc), and criterion-criterion correlations (ccc). We based estimates of icc (.10), idc (.13), and dcc (.14) values using Kurdi et al.'s (2019) figures. For idc (0.28) and ddc (.47) values, we used figures from data published by Bar-Anan and Nosek (2014). No study we are aware of has measured participants on multiple behavioral racial discrimination tasks, so for the ccc we split the difference between the iic and eec estimates (.375). Results of the simulation-based bootstrapped power analysis with a sample size of $N = 2000$

[3] Due to the structure of the raw data (involving multiple tasks and repeated sessions), full deidentification of the raw data could not be guaranteed without compromising reproducibility. Therefore, we only make available (i) the complete preprocessing script and (ii) the fully de-identified preprocessed data set. The original raw data can be made available to researchers upon reasonable request to author S. Hoogeveen.

suggested that our SEM-based tests (RQ3A and RQ4A) would have >95% power to detect both the predictive and incremental validity of indirect measures.

To estimate power for the corresponding equivalence tests (RQ3B and RQ4B), we adjusted icc values until the relevant effects were null. Results suggested that equivalence tests would have >95% power to reject standardized implicit attitude → discrimination latent paths effects outside the equivalence bounds of $\beta = -0.15$ and 0.15.

Descriptive Statistics, Correlations, and Reliability

The descriptive statistics (raw means and standard deviations) of each of the primary 12 measures can be found in Table 2 and the bivariate correlations and distributions are shown in Figure 1. For the indirect and behavioral measures, the table includes both the descriptives per race condition as well as the contrast effect. In these measures, positive values reflect pro-White/anti-Black bias and negative values reflect pro-Black/anti-White bias.

Figure 2 additionally shows the individual anti-Black bias estimates for each of the indirect and behavioral tasks based on hierarchical Bayesian models with trials nested in participants. These figures highlight two main points: first, while the indirect tasks show anti-Black bias on average (see statistical tests in the next section), none show convincing evidence for the majority of participants, as seen in the percentages of credible effects (orange lines) in Figure 2. The behavioral measures, in contrast, show anti-White bias on average, yet even less indication of substantial effects within participants: for a large majority of participants, there is simply no clear effect.

Second, the hierarchical estimates reflect the relative amount of stable individual differences (between-subject variation) versus measurement noise (within-subject variation). Combining the ratio of between- and within-variability with the number of trials in each task, we calculated the reliability of the experimental tasks (Rouder & Mehrvarz, 2024). As shown in Table 2, we found satisfactory levels of reliability for the IAT ($r = .88$), the SC-IAT ($r = .77$), and the AMP ($r = .75$). The EPT had very low reliability ($r = .01$). Though the use of online data collection and a lengthy study session may have contributed to the high measurement error of the EPT in this investigation, these results align with prior results showing low reliability of EPT measures in other contexts (Axt et al., 2024; Bosson et al., 2000; Cunningham et al., 2001; Koppehele-Gossel et al., 2020; Olson & Fazio, 2003). For the behavioral measures, the TG was the only behavioral task that showed acceptable reliability ($r = .95$). By the psychometric standards applied to self-report scales, the UG ($r = .21$), JBT ($r = .36$), and RET ($r = .02$) all proved unreliable as measures of individual differences in tendencies for discriminatory behavior (see Online Supplemental for number of trials needed to reach acceptable reliability). These weak reliability estimates are also reflected in the factor loading findings in the SEM as reported below, with the less reliable tasks showing lower factor loadings on the respective latent constructs.

Participant Suspicion of Study Purpose

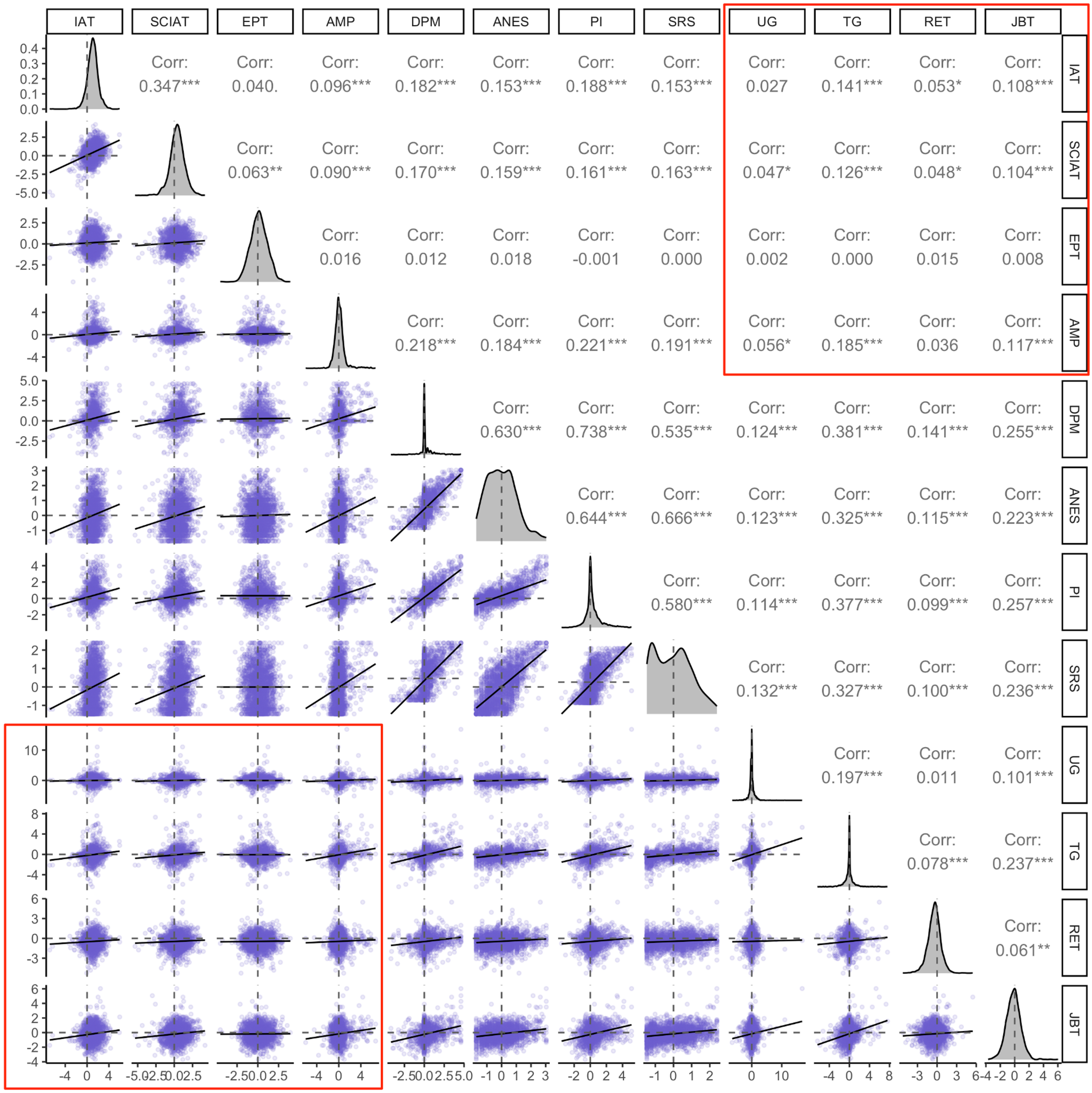
Our preregistration detailed specific terms (race, bias, discrimination, etc.) that needed to be mentioned during the free response item to categorize a participant as aware of the study's purpose. Coding the free response and the follow-up item "when did you decide what this study was about?" suggested that 9.1% of

Table 2
Descriptives Statistics and Reliability for the Primary Measures

Task	Anti-Black bias effect	Reliability	White-positive, Black-negative blocks	Black-positive, White-negative blocks
	M (*SD*)		*M* (*SD*)	*M* (*SD*)
Indirect measures				
IAT	0.467 (0.458)	0.884	0.753 (0.180)	0.901 (0.230)
SC-IAT	0.189 (0.467)	0.776	0.660 (0.136)	0.682 (0.144)
AMP	0.019 (0.150)	0.754	0.622 (0.156)	0.603 (0.168)
EPT	0.071 (0.510)	0.013	0.610 (0.116)	0.612 (0.116)
Direct measures				
DPM	1.503 (7.526)	0.920		
ANES	2.116 (0.838)	0.867		
PI	0.397 (1.182)	0.805		
SRS	2.343 (0.953)	0.928		
Behavioral measures				
TG	−0.020 (0.326)	0.953	1.498 (0.778)	1.517 (0.807)
UG	−0.087 (8.049)	0.194	0.591 (0.295)	0.594 (0.300)
JBT	−0.027 (0.161)	0.358	0.530 (0.148)	0.557 (0.176)
RET	−0.227 (0.507)	0.022	4.910 (0.845)	5.138 (0.834)

Note. The bias effect column gives the values per measures based on the scoring as described in the Methods section. The last two columns give the raw scores per condition for the indirect and behavioral measures and relate to either response times (IAT, SC-IAT, EPT), decision rates (AMP, UG, JBT), or scale ratings (TG, RET). IAT = Implicit Association Test; SC-IAT = Single-Category IAT; AMP = affect misattribution procedure; EPT = evaluative priming task; ANES = American National Election Survey; DPM = direct preference measures; PI = Prejudice Index; SRS = Symbolic Racism Scale; UG = ultimatum game; TG = trust game; JBT = judgment bias task; RET = resume evaluation task.

Figure 1
Bivariate Correlations and Distributions of All Direct, Indirect, and Behavioral Measures



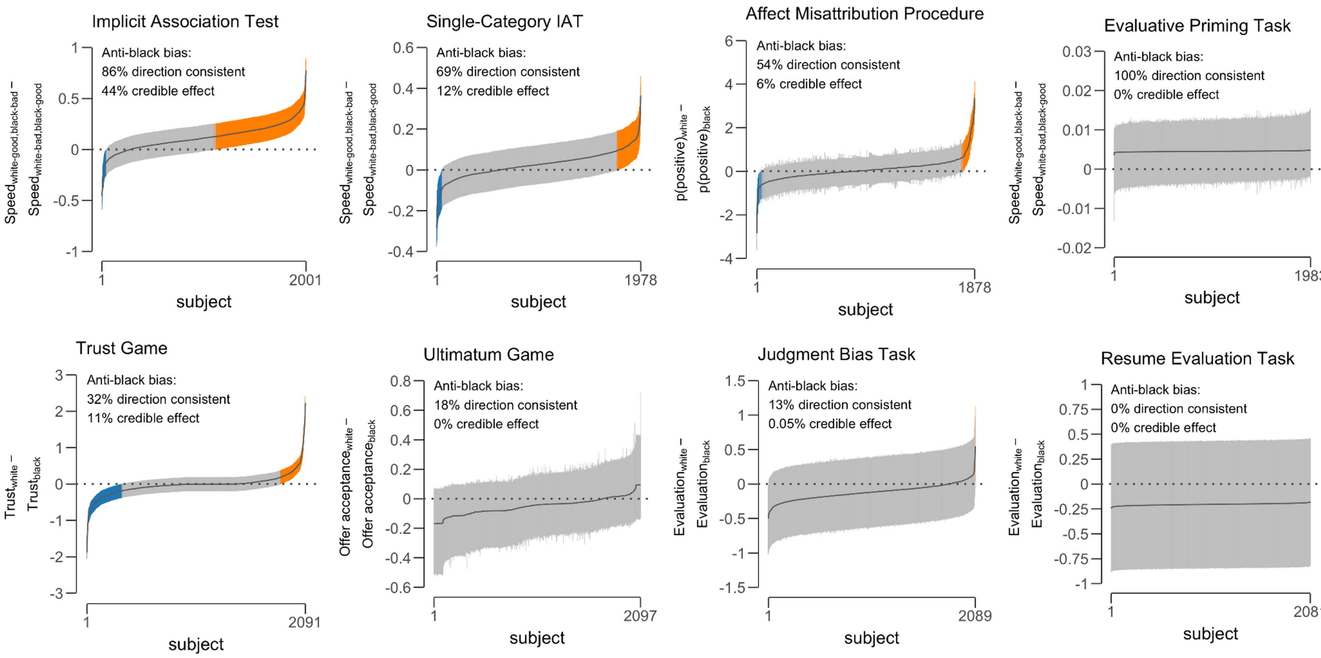
Note. The dashed line at zero reflects the neutral point, where there is no bias, with positive values reflecting pro-White/anti-Black bias and negative values reflecting pro-Black/anti-White bias. The solid line represents the regression line. The red squares highlight the correlations that are most relevant in this project. IAT = Implicit Association Test; SC-IAT = Single-Category IAT; AMP = affect misattribution procedure; EPT = evaluative priming task; ANES = American National Election Survey; DPM = direct preference measures; PI = Prejudice Index; SRS = Symbolic Racism Scale; UG = ultimatum game; TG = trust game; JBT = judgment bias task; RET = resume evaluation task. See the online article for the color version of this figure.

participants were aware of the study's purpose at the time of completing the behavioral measures.

Further measures of "study savviness" showed the following pattern: 12.1% of participants reported that they were influenced by the race of the target persons in the behavioral tasks when directly asked. In addition, 7.6% had participated in a similar study, 54.2% had taken a psychology course, and the median number of completed research studies was 100, range [1, 100,000].

Figure 2

Estimated Individual Racial Bias Effects on the Indirect Measures (Top Row) and the Behavioral Measures (Bottom Row)



Note. Gray points/lines reflect individuals for whom the estimated bias-effect includes zero, orange points/lines reflect individuals who showed a pro-White/anti-Black bias, and blue points reflect individuals who showed a pro-Black/anti-White bias. The plots are ordered by the size of the estimated effects. IAT = Implicit Association Test. See the online article for the color version of this figure.

Preregistered Frequentist Analysis

For each direct, indirect, and behavioral measure, we used previously established methods for assigning scores to participants and excluding low quality responses. These methods are detailed in the Supplemental Materials. For each of the behavioral and indirect measures, and for two of the direct measures (DPM and PI), scores above or below zero indicate anti-Black or pro-Black bias, respectively. The remaining direct measures (ANES and SRS) did not have meaningful zero points and were mean-centered. All scores were scaled to have a standard deviation of one.

Research Questions 1 and 2: Presence of Racial Bias

For research questions 1A and 2A, we computed participants' average scores on indirect (IAT, SC-IAT, EPT, AMP) and behavioral measures (UG, TG, JBT, RET). We then performed single-sample two-sided t tests on the average scores to test if means significantly differed from zero. These tests relied on observed scores due to their meaningful zero points (scores above/below zero indicating anti-Black/anti-White bias).

Indirect Measures. One-sample t tests for each indirect measure showed that mean scores for all measures were reliably greater than zero (see Table 3 for descriptive and test statistics), indicating anti-Black bias.

Behavioral Measures. One-sample t tests on the mean value for each behavioral measure were not reliably greater than zero; instead, scores were reliably below zero, indicating significant pro-Black bias for all but the UG task (see Table 4 for descriptive and test statistics).

As preregistered, equivalence tests were not conducted because results showed evidence of reliable effects when investigating RQs 1 and 2.

Research Questions 3 and 4: Predictive and Incremental Validity

Predictive Validity (RQ3A). To test for predictive validity, we first fit the measurement model including three latent factors (implicit attitudes, explicit attitudes, and discrimination), each measured by four observed indicators, with implicit and explicit latent factors

Table 3

Descriptive Statistics and One-Sample t Tests for Each Indirect Measure

Measure	M (SD)	t test against 0
IAT	0.47 (0.46)	$t(2063) = 46.29$, $p < .001$, $d = 1.02$, 95% CI [0.97, 1.07]
SC-IAT	0.19 (0.47)	$t(2088) = 18.49$, $p < .001$, $d = 0.40$, 95% CI [0.36, 0.45]
EPT	0.07 (0.51)	$t(2056) = 6.28$, $p < .001$, $d = 0.14$, 95% CI [0.09, 0.18]
AMP	0.02 (0.15)	$t(1954) = 5.49$, $p < .001$, $d = 0.12$, 95% CI [0.08, 0.17]

Note. IAT = Implicit Association Test; SC-IAT = Single-Category IAT; AMP = affect misattribution procedure; EPT = evaluative priming task.

Table 4
Descriptive Statistics and One-Sample t Tests for Each Behavioral Measure

Measure	*M* (*SD*)	*t* test against 0
UG	−0.09 (8.05)	$t(2{,}103) = -0.49$, $p = .621$, $d = -0.01$, 95% CI [−0.05, 0.03]
TG	−0.02 (0.33)	$t(2098) = -2.74$, $p = .006$, $d = -0.06$, 95% CI [−0.10, −0.02]
JBT	−0.03 (0.16)	$t(2094) = -7.76$, $p < .001$, $d = -0.17$, 95% CI [−0.21, −0.13]
RET	−0.23 (0.51)	$t(2032) = -20.15$, $p < .001$, $d = -0.45$, 95% CI [−0.49, −0.40]

Note. UG = ultimatum game; TG = trust game; JBT = judgment bias task; RET = resume evaluation task.

correlated, and each with causal paths to discrimination (see Figure 1). See Figure 2 for bivariate correlations between all direct, indirect, and behavioral measures.

Recent simulations demonstrated that common guidelines for assessing adequate fit in SEMs (e.g., Hu & Bentler, 1999) are biased when samples are large, models are complex, or measurement error is low (Shi et al., 2018, 2019). As a result, the same amount of misfit can lead to accepting a model in some conditions (e.g., small sample size), but rejecting it in others (e.g., high loadings). Alarmingly, these conditions vary across the many goodness-of-fit statistics that have been proposed, sometimes in opposite directions. Hence, in the preregistration we decided to accept adequate fit if either (a) the chi-square test of close fit was not significant or (b) the comparative fit index (CFI) was ≥0.95 and the lower bound of bootstrapped 95% confidence intervals around the unbiased standardized root-mean-square residuals (SRMRu) corrected for average communality (95% LL(SRMRu/$\_R2$)) was ≤0.10 (Ximénez et al., 2022).

Results for this model showed that the chi-square test was significant, $\chi^2(51) = 313.93$, $p < .001$, the CFI was 0.948, and the 95% confidence interval for the SRMRu included 0.1 (95% CI 0.10, 0.12). Since our thresholds were not met, we followed the preregistered analysis plan and inspected the modification indices, adjusting the model accordingly and achieving consensus among all authors on the changes. The modification indices suggested adding residual covariances between two pairs of explicit attitude scales, specifically between the ANES and SRS as well as between the DPM and PI. Adjusting and rerunning the model resulted in adequate model fit based on the preregistered criteria: while the chi-square test was still significant, $\chi^2(49) = 134.44$, $p < .001$, the CFI was 0.98, and the confidence interval for the SRMRu (95% CI 0.09, 0.13) included 0.1.

We then compared the SEM that constrained both the implicit attitude → discrimination path and the explicit attitude → discrimination path to zero (Model B) with the SEM that constrained only the explicit attitude → discrimination path to zero (Model C). The chi-square test showed that Model C significantly improved model fit over Model B, $\chi^2(1) = 181.85$, $p < .001$. The standardized path coefficient from the latent implicit attitude construct to the latent discriminatory behavior construct in Model C was $\beta = 0.90$, $SE = 0.05$, $p < .001$. These results indicate implicit racial attitudes predicted behavior in the expected direction, such that higher anti-Black implicit attitudes were associated with lower pro-Black discrimination.

Incremental Validity (RQ4A). To investigate whether latent implicit racial attitudes predicted latent tendencies on discrimination measures over and above latent explicit racial attitudes (i.e., incremental validity), we first verified that implicit and explicit attitudes are different constructs by fitting a model that constrained implicit and explicit attitudes to be the same construct (Model D), and comparing that to the full model where they are different constructs (Model A). The chi-square test showed that model A significantly improved model fit over Model D, $\chi^2(3) = 388.80$, $p < .001$.

Next, we fit a model that constrained the implicit attitudes → discrimination path to zero (Model E). The chi-square test showed that Model A significantly improved model fit over Model E, $\chi^2(1) = 8.42$, $p = .004$. The standardized path coefficient from the latent implicit attitudes construct to the latent discriminatory behavior construct in Model A was $\beta = 0.16$, $SE = 0.05$, $p = .004$. These results suggest that implicit racial attitudes continued to explain unique variance in discriminatory behavior after accounting for explicit racial attitudes (See Figure 3).

Bayesian Multiverse Analysis

We next conducted a Bayesian multiverse analysis that considered the relative strength of evidence produced for different theoretical perspectives given the specific treatment of the data (i.e., how variables were operationalized and whether they were included in the SEM). We constructed different SEMs in *lavaan* and used Bayesian Informative Hypothesis Testing, as implemented in the *bain* package (Version 353 0.2.11; Gu et al., 2024). The evidence is presented as posterior model probabilities of the considered perspectives (i.e., adding up to 1). Syntax for the multiverse analysis was developed using a blinded data set from which the crucial effects had been removed (Dutilh et al., 2019; MacCoun & Perlmutter, 2015).

Table 1 details the perspectives considered in the multiverse analysis and outlines the patterns of results that would provide the strongest support for each account.

Paths were largely determined by variations in (a) the scoring of indirect and behavioral measures, (b) inclusion of various measures in SEM (e.g., whether by using all measures or only those with factor loadings above 0.30), and (c) the inclusion of potential moderator variables (e.g., when including no moderators, when moderators are treated as continuous, or when moderators are dichotomized using median or theoretical splits). All told, we specified 36 different models based on these factors. The Supplemental Materials provides further details regarding how paths were determined and how the multiverse analysis was validated.

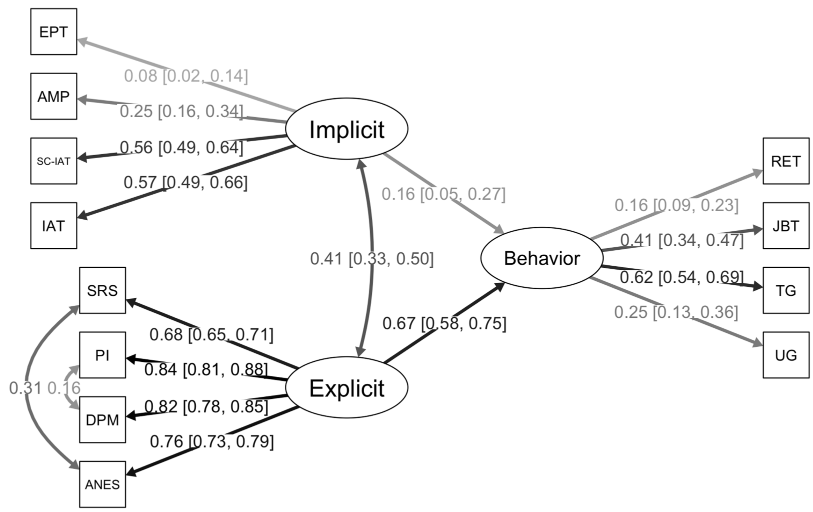
Figure 4 shows the results from the multiverse analysis. The *x*-axis specifies the different theoretical perspectives and the *y*-axis reflects the posterior model probabilities. The pink circles reflect the perspective that received the most evidence for each path (i.e., model) and the proportion of paths where each perspective was the winner is shown in the text labels at the top of the figure.

The incremental validity perspective received the most support across multiverse paths (63.9% of paths), followed by the motivated correction account (22.2% of the paths) and the executive control account (13.9% of the paths). The pervasive bias, false positives, limited value-add, situational fatigue, attitude strength, and study savviness accounts were never the best fit in any model.

This multiverse analysis thus provides the strongest relative support for the incremental validity perspective, where people who exhibit a stronger pro-White bias on indirect measures also show stronger pro-White behavior controlling for scores on direct attitude

Figure 3

Model A: Structural Equation Model With Latent Implicit and Explicit Attitudes Predicting Race-Related Behavior



Note. Parameter estimates are standardized, with 95% confidence interval in square brackets and transparency reflecting path strength. Implicit = implicit racial attitudes, Explicit = explicit racial attitudes, Behavior = relative treatment of Black versus White targets. Marker variables used for model identification were IAT, ANES, and UG. IAT = Implicit Association Test; SC-IAT = Single-Category IAT; AMP = affect misattribution procedure; EPT = evaluative priming task; ANES = American National Election Survey; DPM = direct preference measures; PI = Prejudice Index; SRS = Symbolic Racism Scale; UG = ultimatum game; TG = trust game; JBT = judgment bias task; RET = resume evaluation task.

measures. When comparing the evidence for the incremental validity and limited value-add accounts, we found Bayes factors ranging from 1.10 to 175 favoring the incremental validity perspective, with 100% of the paths preferring the incremental validity over the limited value-add perspective and 52.8% in which the incremental validity account strongly outperformed the limited value-add account (i.e., BF > 10).

At the same time, some paths suggested conditional effects where indirect measures were only predictive for people low in executive control or motivation to control prejudice. Specifically, in 14.8% of the paths, the data provide more evidence for the executive control over the incremental validity perspective. However, the Bayes factors range from 0 to 1.72, which is no or weak evidence for the moderation account over the general incremental validity account. The same holds for the motivated correction account; while it is preferred in 29.6% of paths, the Bayes factors range from 0 to 2.52, which is only weak evidence for the moderation account over the general incremental validity account.

Further Exploratory Analyses

We conducted various exploratory analyses to assess the robustness of the main findings, which are reported fully in the Supplemental Materials and briefly summarized here. Note that these analyses were neither preregistered nor prepared using blinded data and thus should be considered exploratory (Wagenmakers et al., 2012).

First, when predicting performance on individual behavioral tasks from both indirect and direct measures using model-averaged Bayesian linear regression, we found no consistent evidence of predictive validity for any indirect measure. The only consistent pattern across behavioral tasks was that direct measures of explicit attitudes consistently explained variance in behavior, although the specific measure varied across tasks. None of the moderators (e.g., motivation to control prejudice, executive function) showed a consistent pattern of substantial influence for any of the individual behavioral measures.

Second, we used Bayesian estimation to directly compare the predictive validity of direct versus indirect measures on our behavioral measures. By setting either the explicit-behavior or implicit-behavior path to zero, we found evidence that the explicit-bias-only model strongly outperformed the implicit-bias-only model. However, the model that set the implicit-behavior path to zero performed substantially worse than the model that forced the path to be above zero.

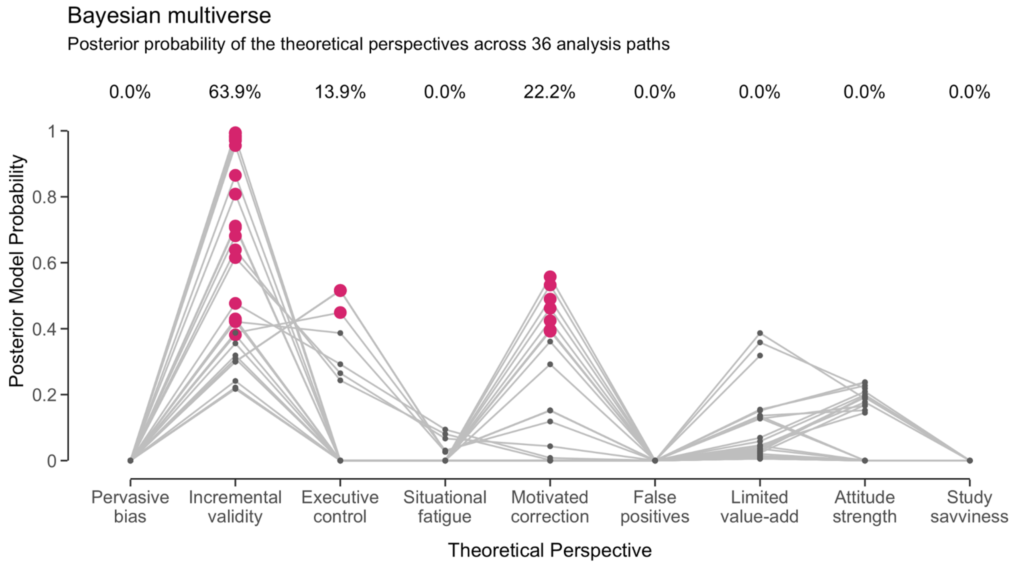
Third, we found that significant anti-Black behavioral bias only emerged when subsetting the sample to participants who displayed negative implicit racial attitudes at more than two standard deviations above the neutral zero-point ($n = 161$, 7.6% of the full sample). Fourth, conclusions and the estimate regarding the implicit-behavior association were unchanged following analyses that used alternative handling of missing data (i.e., imputation), excluded the EPT due to poor reliability, or weighted the data by participant gender and political affiliation.

General Discussion

In this adversarial collaboration, a large, stratified, and politically balanced sample of U.S. White adults completed four major indirect

Figure 4

Results From Bayesian Multiverse Analysis



Note. Posterior model probabilities are given for each theoretical perspective and each analysis path. The *x*-axis shows the different theoretical perspectives and the *y*-axis reflects the relative posterior probability. The lines connect the evidence for each path and the posterior probability of each path adds up to one. Pink points reflect the perspective that received most evidence for each path. The proportion of paths where each perspective is the winner is shown in the text labels at the top. See the online article for the color version of this figure.

measures of implicit racial attitudes and four behavioral measures of racial discrimination in separate sessions. Using multiple indicators of each construct allowed us to estimate latent variables and correct for the measurement error that may obscure the true relationship between implicit racial attitudes and behaviors. We further assessed several theoretically informed moderators: motivations to respond without prejudice, situational fatigue, executive control, study-savviness, and attitude strength. Collaborators—who included proponents of indirect measures, skeptics, and neutral third parties—all agree that the study provided a valid and fair test of our competing predictions.

Analyses confirmed that both indirect measures of implicit racial attitudes and direct, self-report measures of explicit racial attitudes correlated reliably in the expected direction with trust decisions, ultimatum game decisions, and hiring evaluations involving Black and White targets. Notably, direct measures were significantly better predictors of these behavioral outcomes than indirect measures.

And yet, both direct and indirect measures explained unique variance in race-related judgments and actions, though this pattern only emerged for indirect measures when using SEM analyses that created a larger implicit racial attitude construct. When exploratory analyses used linear regression models to predict performance on behavioral measures using a single indirect measure at a time, none showed consistent evidence of incremental predictive validity. These results provide further support for prior recommendations that researchers use multiple indirect measures when possible (e.g., Bar-Anan & Vianello, 2018) and use SEM in tests of incremental predictive validity to account for measurement error (Buttrick et al., 2020; Kurdi et al., 2019).

Providing some (limited) support for dual-process accounts, in a subset of multiverse branches indirect measures exhibited improved predictive and incremental validity among participants who scored lower on an objective test of executive control (Hofmann, Friese, et al., 2008; Unsworth et al., 2005) and self-report scale of motivations to control prejudice (Dunton & Fazio, 1997). This is consistent with models in which implicit processes shape judgments and actions unless deliberatively overridden (e.g., Fazio, 1990; Friese et al., 2008; Gawronski & Bodenhausen, 2006; Hofmann, Friese, et al., 2008; Perugini et al., 2010). At the same time, support for dual-process accounts was qualified by null results for the hypothesized moderating effects of self-reported situational fatigue, as well as the mixed support for moderation by executive control and motivation to respond without prejudice across alternative branches of the multiverse.

Although White Americans held more positive implicit racial attitudes toward their own group relative to Black Americans, their behavioral responses were generally characterized by pro-Black biases. At least among these measures of discrimination, participants who displayed higher levels of bias on indirect measures were only less pro-Black or race-neutral in their behavior, rather than engaging in outright anti-Black discrimination (see also Axt et al., 2016; Blanton et al., 2009). Only a relatively small subset of participants with extreme pro-White scores on indirect measures (7.61% of the sample) engaged in statistically significant discrimination against Black relative to White targets, and to only a small degree. Of course, if members of this subset of the population occupy positions of power, their discriminatory tendencies could prove consequential.

One Set of Findings, Multiple Interpretations

All collaborators agree that our study design was informative and that our reported results accurately characterize our data. However, our results and their implications can be interpreted in more than one way (as is perhaps true for most behavioral science findings; Martel et al., 2024; Mellers et al., 2001). To that end, below we present comparatively optimistic and pessimistic perspectives on the results.[4]

From a more optimistic lens, these results could be viewed as encouraging for the predictive validity and utility of indirect measures, which have taken a rhetorical beating in recent years (e.g., Blanton & Jaccard, 2024; Carlsson & Agerström, 2016; Cesario, 2022; Corneille & Gawronski, 2024; Cyrus-Lai et al., 2022; Forscher, 2023; Jussim et al., 2024; Machery, 2022; Meissner et al., 2019; Mitchell & Tetlock, 2017; Schimmack, 2021; Tetlock & Mitchell, 2024; Van Dessel et al., 2020). The present study represents an existence proof that implicit attitudes as assessed with commonly used indirect measures can reliably correlate with socially important behavioral outcomes (Kang, 2024) and explain unique variance beyond parallel self-report measures.

And yet, indirect measures were substantially outperformed by direct measures in both predicting and explaining unique variance in behavior. This is in sharp contrast to meta-analyses in which indirect measures performed as well or better than their self-report (direct) counterparts (Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013). Our divergent results could be attributable to the controlled nature of the decisions we examined using these behavioral measures (Dovidio et al., 2002; Fazio, 1990). Past meta-analyses may also have overestimated the predictive power of indirect measures by relying on mostly smaller samples that required large effect sizes to achieve statistical significance and pass through the publication filter. Indeed, though exceptions may be identified in future work, all authors in this collaboration agree that if researchers could only select between a single direct or indirect measure of racial attitudes to maximize prediction of a race-related behavior, then it makes sense to prioritize a direct measure (Corneille & Gawronski, 2024).

Even an optimistic interpretation of the present work suggests severe limits to the standard approach of deploying individual indirect measures to predict specific behaviors of interest. The zero-order correlations between each indirect measure and the behavioral measures were quite low, ranging from .00 to .19 (see Figure 1). The individual indirect measure-behavior correlations were statistically significant in only 10 of 16 cases, although four of the six nonsignificant correlations were for the EPT. By comparison, zero-order correlations between direct measures and behavioral measures were all statistically significant, ranging from .099 to .381. Moving forward, it will be productive to capture different race-related behaviors and create aggregated behavioral indices (Ajzen & Fishbein, 1977) as well as correct for measurement error and calculate latent implicit racial attitudes based on multiple indirect measures (Cunningham et al., 2001, 2004).

The 2.5% additional variance in behavior explained by latent implicit racial attitudes above and beyond latent explicit racial attitudes can be interpreted in different ways. Direct measures explained much more variance in behavior, raising the question of whether further administering time-intensive indirect measures is worth the return on investment. Ultimately, whether a further 2.5% variance explained is of practical and consequential significance is a subjective judgment that is likely to vary by context and researcher. A highly noisy measure could still be valuable if it shines light on a phenomenon that is otherwise not captured (e.g., attitudes the person does not deliberately endorse; Petty & Briñol, 2006) or is of great importance (racial discrimination in selection decisions). Researchers must decide if the variance potentially explained by indirect measures is of sufficient practical and theoretical interest.

Although there is reliable evidence of anti-Black racial discrimination in real-world settings—such as from field audits where fake applications are submitted to real employers (e.g., Nødtvedt et al., 2021; Quillian et al., 2017; Schwab & Singh, 2024)—none of the present behavioral measures produced anti-Black discrimination. One explanation for this result is a mix of impression management concerns and sincere egalitarian motives leading to pro-Black behavioral biases in the artificial tasks used here, but this explanation concedes that our methods were then not able to create appropriate analogs for the judgment contexts where anti-Black discrimination is most likely to occur outside of the laboratory. To that point, our results highlight the difficulties that researchers can expect to encounter when studying racial discrimination using lab-based methods that allow for effortful correction of responses, and it is possible that anti-Black discrimination (or race neutrality) would have emerged using higher-stakes decisions or more naturalistic contexts.

Additional research in laboratory contexts could use other measures of discriminatory behavior or manipulate aspects of the decision-making context to maximize the chances of finding anti-Black behavior. For instance, prior work suggests that socially desirable responding can be attenuated in contexts where judgments are made under time pressure (Ranganath et al., 2008) or when behavior occurs spontaneously (Dietrich & Sands, 2023). Using measures with speeded or more spontaneous behaviors may produce different results concerning evidence of racial bias and connections to implicit racial attitudes.

At the same time, focusing on spontaneous behaviors (e.g., seating distance) would undermine our goal of using outcomes that capture more consequential judgments (such as who to hire or trust). Similarly, placing participants under time pressure would reduce the ecological validity of the present study, since decisions about selection and resource allocation are often made with more deliberation. Identifying reliable measures of racial discrimination that provoke naturalistic, deliberate and suspicion-free judgments should be a clear priority for researchers.

A more pessimistic interpretation of results may follow once one considers that the incremental validity of the implicit racial attitudes construct was $\beta = 0.16$, which was very close to the collectively agreed threshold of $\beta < 0.15$ for declaring an effect that was small enough to be practically nonsignificant. Notably, the size of the implicit-behavioral estimate observed here was quite similar to the overall meta-analytic estimate from Kurdi et al. (2019). Moreover, 15 of the 16 zero-order correlations between indirect measures and behavior were below $r = .15$, with each behavior being more strongly predicted by direct than indirect measures of racial attitudes.

Direct measures of racial attitudes are easier to administer and showed consistently stronger correlations with behavior. To the extent that one's goal is the prediction of discriminatory behavior at the level of individuals, our results can be used to plausibly argue

[4] The original submission allowed authors to write their own subsections of the General Discussion section that provided interpretations of the results (e.g., Mellers et al., 2001). These are available in the Online Supplemental.

that there is little reason to continue administering indirect measures like those used in the present work. The incremental validity of the indirect measures barely passed the threshold for practical nonsignificance under what could be considered optimal conditions for finding an incremental gain, results that in turn stake the viability of indirect measures on the claim that they can explain an additional 2.5% in variance compared to the over 45% explained by direct measures. When the implicit social cognition revolution began, direct, self-report measures were framed as outdated (Banaji et al., 2004), with indirect measures like the IAT tapping into modern forms of bias that cause social group discrimination. Such claims are inconsistent with the current results.

Furthermore, indirect measures exhibited weak psychometrics and convergent validity (e.g., even the SC-IAT and IAT only correlated at $r = .35$), with all indirect measure intercorrelations lower than direct measure intercorrelations. When an individual responds to a direct measure of bias, we can fairly draw inferences about that individual's beliefs and attitudes (though labeling a respondent's view as biased may not be justified, at least not with responses to the symbolic racism measure, which is problematic for several reasons; see Sniderman & Tetlock, 1986). However, responses on indirect measures have a much more difficult time separating the "bias signal" from the noise, making it perilous and arguably unethical to label any individual's score as biased or unbiased. While the reliability of some indirect measures suggests that a signal (stable interindividual variability) can be separated from noise (random within-individual variability), it still remains difficult to know the degree to which that signal is influenced by factors like mental speed, cognitive flexibility, or some other artefact that contributes to scores on such tasks (Rothermund & Wentura, 2004).

Finally, that our behavioral measures generally produced effects favoring Black over White targets further complicates claims that anti-Black effects on indirect measures are a source of cumulative disadvantage. The argument that the negative effects of implicit racial attitudes accumulate across persons or time to produce sizable differences between groups (e.g., Greenwald et al., 2015; Kang, 2024) depends, of course, on such attitudes being associated with *negative* outcomes, not neutral or favorable ones (Cyrus-Lai et al., 2022).

Limitations and Future Directions

The present study is limited on many dimensions, such as its artificial laboratory setting, narrow tests of moderators, inability to address causality, and examination of simple direct effects. Additional limitations include claims that are restricted to a specific demographic (i.e., White American adults) and time period, as well as a lack of experimental manipulation for any potential moderator. Table 5 summarizes some of the limitations of this project.

Given these limitations and the many forms that discrimination can take, our results cannot be considered a definitive answer to the question of how indirect measures of racial attitudes predict discriminatory behavior. Our investigation used artificial tasks that had only surface-level connections to decisions completed outside of a laboratory context. Subsequent investigations that explore this question using other measures or settings may produce different conclusions than the present work. Indeed, this project leaves a number of open questions, ripe for adversarial collaboration, regarding indirect measures, and implicit social cognition.

Anti-Black Bias in the Laboratory

Despite using measures that in earlier studies produced anti-Black discrimination in an ultimatum game (Kubota et al., 2013) and hiring paradigms (Lahey & Oxley, 2021), analogous results did not emerge in the present investigation. That our behavioral measures consistently found pro-Black biases in judgment calls into question whether our results would generalize to other laboratory measures or contemporary real-world settings, or if rapid cultural change has already rendered prior studies unreplicable.

One reason for the lack of anti-Black discrimination in behavior could be that the original investigations using these measures included fewer race-related measures in the same study design, diminishing the chances that participants would realize the study was about discrimination. The population sampled in the preceding work were also from an earlier time period, prior to the George Floyd protests, where pro-Black sentiments and socially desirable responding were perhaps less prevalent (Barrie, 2020). Though overall levels of suspicion that the study was about race were low in absolute terms (9.1% of participants) and suspicion did not moderate the predictive validity of indirect measures, we cannot rule out the possibility that our awareness measures were insufficiently sensitive.

Several original studies recruited college students or local community members who participated in person (Kubota et al., 2013; Lahey & Oxley, 2021; Stanley et al., 2011), in contrast to our online sample of adults who likely had more prior experience taking part in research investigations. Our measures of study-savviness moderated the results in a prior study investigating group-based discrimination (Tierney et al., 2020); that similar moderation failed to emerge here suggests that the present results are not simply attributable to the use of an experienced online sample. However, our savviness measures may have been limited in ways specific to our population of relatively experienced study participants.

Moreover, our paradigmatic replication approach (Vohs et al., 2021) adapted and modified the original methods, which could have inadvertently affected the magnitude or direction of estimates of racial discrimination. For example, the RET (Lahey & Oxley, 2021) differed in this study compared with the original task in that the original resumes were completely randomized and unique for each participant, with less than 10% of resumes in each set having Black names to match the population of applicants. These and other changes in stimuli and procedure could potentially explain the departures in results from those reported in Lahey and Oxley (2021) and Kubota et al. (2013).

Finally, psychometric analyses revealed that three of the four behavioral measures suffered from very low internal reliability, with the lone exception being the TG. Indeed, the limited shared variance among both the implicit and behavioral indicators constrains the reliability of their latent constructs and should be considered when interpreting the structural relationships. While much prior work has explored the reliability and validity of indirect measures of implicit attitudes (Axt et al., 2024; Bar-Anan & Nosek, 2014), our findings suggest that additional effort is needed for developing reliable measures of discriminatory behavior (both in and outside of the laboratory). Such efforts will help identify whether the low intercorrelations among behavioral measures in the present work reflects

Table 5
Assessment of Limitations

Dimensions	Assessment
Internal validity	
Is the phenomenon diagnosed with experimental methods?	Although our behavioral and indirect measures use within-subject manipulations of target race, the primary research questions concerned the relationship between individual difference measures and did not experimentally manipulate relevant constructs.
Is the phenomenon diagnosed with longitudinal methods?	No.
Were the manipulations validated with manipulation checks, pretest data, or outcome data?	There were no between-subjects manipulations. Indirect measures were pretested, as were materials for behavioral measures (see Supplemental Materials for more information).
What possible artifacts were ruled out?	The use of multiple indirect, direct, and behavioral measures rules out the possibility that conclusions are driven by a single measure of each construct. Our SEM analyses also address the extent to which our conclusions are driven by measurement error.
Statistical validity	
Was the statistical power at least 80%?	Yes, for all primary analyses
Was the reliability of the dependent measure established in this publication or elsewhere in the literature?	Reliability of our individual behavioral measures is explored here and found to be weak for all tasks except the trust game (see page 21). This highlights the value of aggregating across behavioral measures.
If covariates are used, have the researchers ensured they are not affected by the experimental manipulation before including them in comparisons across experimental groups?	The study design had no between-subjects experimental manipulations.
Were the distributional properties of the variables examined and did the variables have sufficient variability to verify effects?	Distribution plots for all measures are available in Figure 1 (see page 23). The degree of true variability (relative to noise) is analyzed and reported for the main measures in the results section.
Generalizability to different methods	
Were different experimental manipulations used?	Though this work had no between-subjects manipulations, we used four measures for each of our implicit, explicit and behavioral constructs to ensure methodological diversity.
Generalizability to field settings	
Was the phenomenon assessed in a field setting?	No.
Are the methods artificial?	Yes, while behavioral measures were drawn from past work, they ultimately reflect artificial judgments completed in an online lab context.
Generalizability to times and populations	
Are the results generalizable to different years and historic periods?	This was not tested directly and as a result we cannot claim that results generalize to different years and historic periods.
Are the results generalizable across populations (e.g., different ages, cultures, or nationalities)?	Our sample is limited to White Americans who identified as democrat, republican, or independent, and as such we cannot claim that results generalize to other racial groups, nationalities, party affiliations, or political ideologies. We were also able to recruit fewer men relative to our target sample composition.
Theoretical limitations	
What are the main theoretical limitations?	The main theoretical limitation is the use of artificial behavioral measures that failed to produce anti-Black discrimination on average. This raises concerns about how well these results would generalize to more naturalistic settings or less controllable behavioral outcomes in the laboratory. An additional limitation concerns a lack of experimental manipulations of potential moderators, such as cognitive fatigue. Failing to use such manipulations provides a weaker and less conclusive test for the proposed moderators of the relationship between behavior and indirect measures.

Note. SEM = structural equation model.

a weakness in the measures or the targeted construct. That is, it is possible that intergroup behavior is such a complex, multifaceted construct that even seemingly similar decisions (like in selection vs. cooperation) will never be more than weakly associated with one another.

A future crowdsourced initiative might begin by first establishing a set of replicable anti-Black discriminatory biases in controlled laboratory settings in populations of interest and then in a second phase testing whether these behavioral effects can be predicted based on scores on indirect measures of racial attitudes or relevant stereotypes (e.g., criminality). Phase 1 might well reveal that many previously observed anti-Black biases fail to replicate, either due to noisy estimates in the original studies (Ioannidis, 2005), increased suspicion and sensitivity among research participants, or genuine cultural change in racial attitudes or discrimination (Barrie, 2020; Charlesworth & Banaji, 2022; Reny & Newman, 2021).

Field Investigations in Organizations

The pervasive bias thesis that racial discrimination remains robust and widespread even in ostensibly progressive societies is much better tested in ecologically valid field settings (Carlana, 2019; Régner et al., 2019; Rooth, 2010). Indeed, several members of the collaboration predicted a priori that our laboratory discrimination paradigms would fail to capture anti-Black bias in behavior, perhaps due to impression management concerns among research participants (see the "Perspectives and Predictions" section of the Supplemental Materials). To address this concern, a future adversarial collaboration could use a more inobtrusive but consequential outcome; for instance, managers in racially diverse workplaces could be administered direct and indirect measures of racial attitudes, examining how such measures predict their performance appraisals of subordinates or the employees' perceptions of the workplace climate (e.g., Glover et al., 2017).

Aside from capturing more naturalistic behavior, a longitudinal field setting would further afford empirical tests of whether small automatic biases either accumulate in their consequences over time (Greenwald et al., 2015; Hardy et al., 2022; Kang, 2024) or dissipate as more individuating information becomes available (Jussim, 2017; Jussim et al., 2024). A field study would also allow for an investigation into the competing predictions regarding whether the predictive validity of indirect measures is enhanced in more naturalistic settings due to lower impression management concerns or diminished due to increases in feelings of accountability.

Contextual Moderation

The primary focus of our design and analyses was on quantifying the strength of the relationship between indirect measures of racial attitudes and relevant judgments and behaviors. However, any such links are likely to be moderated by individual differences and situational factors. We did explore several variables anticipated to moderate the predictive validity of direct and indirect measures, finding some support for two (executive control and motivation to respond without prejudice).

The lack of support found for certain moderator variables could be for theoretical or psychometric reasons. For instance, while attitude strength has been shown to moderate the attitude–behavior relationship in other contexts (e.g., Holland et al., 2002), our own measures of attitude strength did not moderate the relationship between behavior and implicit racial attitudes. This pattern could be due to weaknesses in the reliability of our behavioral measures or a lack of sensitivity for the attitude strength measure, though unanticipated features of the context could also account for the lack of moderation. Additional confirmatory tests should be conducted capturing these and other candidate moderators under theoretically optimized conditions.

It would be especially informative to manipulate theorized moderators, rather than solely measuring them as we did. Experimentally varying factors like cognitive load (Gilbert & Hixon, 1991) and the controllability of behavior (Dovidio et al., 2002) could illuminate causal paths and provide more rigorous tests of dual-process theories (Fazio, 1990; Gawronski & Bodenhausen, 2006). Indeed, replicable evidence of anti-Black behavioral discrimination that emerges only under conditions of diminished executive control would support theories of automatic racial bias regardless of the results concerning indirect measures of individual differences. Conversely, if laboratory manipulations of process accountability break the link between measures of implicit attitudes and behavior—as suggested by our results regarding individual differences in motivations to respond without prejudice—the relevance of implicit attitudes to high stakes decision making in real organizations is cast into doubt (Tetlock & Mitchell, 2009b).

Correspondence

The theoretical question of interest here was how implicit and explicit racial attitudes concerning the overall categories of "Black Americans" and "White Americans" map onto decisions about individual members of those groups. Indeed, controversy over the idea of implicit attitudes has largely centered around the validity, meaning, and implications of the more positive attitudes that White Americans typically display for White versus Black people on the IAT and other indirect measures. However, classic research demonstrates that the (explicit) attitude—behavior correlation is strongest when the predictor and behavioral measures are matched in specificity (Ajzen & Fishbein, 1977). Future work should leverage the correspondence principle by tailoring the direct and indirect attitude measures to the targeted outcome (Irving & Smith, 2020)—for example, using race-crime stereotypes to predict racial bias in hypothetical shooting decisions (Correll et al., 2002).

Causality

A key follow-up question is whether implicit attitudes make a causal contribution to race-related judgments and behaviors (Wolsiefer & Blair, 2024). Even if correlated with relevant outcomes, implicit attitudes could be causally inert, constituting mere cognitive residue of past acts and shared environmental influences. Consistent with this account, a meta-analytic review of intervention studies found that laboratory manipulations that shifted performance on indirect measures did not exert parallel effects on behavioral measures (Forscher et al., 2019). Building on the present initiative and recent intervention tournaments (Lai et al., 2014, 2016; Roy et al., 2025), a future megastudy (Caruso et al., 2017; Milkman et al., 2021) could cross the replicable manipulations of performance on indirect measures (e.g., counter-stereotypical exemplars) with each of our indirect measure-DV pairs and test causal pathways from the interventions to behavior mediated by shifts in implicit racial attitudes.

Indirect Effects

Rather than pitting direct and indirect measures against one another, it is also possible to test interactive models that predict indirect effects. For example, negative implicit attitudes toward members of low-status outgroups could develop earlier in life than explicitly endorsed preferences (Baron & Banaji, 2006) and form the underpinnings of later ideologies. Thus, the shared variance between scores on direct and indirect measures could be of great interest, rather than something to be controlled away in tests of testing incremental validity (Greenwald et al., 2009). Indeed, scores on explicit attitudes questionnaires could partially or fully mediate the behavioral effects of implicit attitudes toward Black and White people on more controlled and nonhabitual behaviors (Ouellette & Wood, 1998).

At a macro level, cultural change could occur via shifts in implicit attitudes that gradually spill over to explicit attitudes (Charlesworth & Banaji, 2019) and then aggregate-level behaviors. Uniting the theory of reasoned action (Ajzen & Fishbein, 1977) with implicit social cognition suggests that the proximal causes of many more controlled choices could be explicit evaluations of the behavioral options and intentions to carry out the behavior, with implicit attitudes retaining a distal and indirect but still causal influence.

Cognitive Processing

Future work will also want to better understand the role of individual differences in cognitive processing on the relationship between implicit attitudes and discriminatory behavior. While a prior investigation (Ito et al., 2015) found no evidence that executive functioning moderated correlations between direct and indirect measures of racial attitudes, subsequent work could further examine the role of cognitive processes by including measures of task-switching ability (Klauer et al., 2010), adding conceptually unrelated indirect measures to account for shared method variance (Allidina et al., 2024), or using alternative forms of indirect measures that may limit the influence of cognitive ability on performance (Teige-Mocigemba et al., 2008).

Conclusion

For the first (but hopefully not last) time, we brought together proponents and critics of indirect measures of racial attitudes to design a study that all agreed represents an informative test of competing predictions. The empirical findings, based on a large, stratified sample of U.S. White adults, indicate that indirect measures of Black–White racial attitudes explain unique variance in basic social decisions regarding who to trust, allocate resources to, and select into organizations. And yet, as some critics have argued, relatively high levels of bias on indirect measures did not necessarily reflect anti-Black discriminatory tendencies on the four behavioral measures selected, and direct rather than indirect measures were superior in predictive and incremental validity for these outcomes. Additional research is needed to systematically replicate past laboratory findings, illuminate causal pathways, and identify contextual moderators. At the same time, it is in the field where competing narratives regarding contemporary racial prejudice can truly be put to the empirical fire. Regardless of the specific empirical approaches and settings chosen, we believe adversarial collaboration represents the most productive path forward.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–38). Springer-Verlag. https://doi.org/10.1007/978-3-642-69746-3_2

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Allidina, S., Long, E. U., Baoween, W., & Cunningham, W. A. (2024). Decoupling the conflicting evaluative meanings in automatically activated race-based associations. *Personality and Social Psychology Bulletin*, *50*, 987–1005. https://doi.org/10.1177/01461672231156029

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*(278), 200–203. https://doi.org/10.1080/01621459.1957.10501379

Arbuckle, J. L., Marcoulides, G. A., & Schumacker, R. E. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (p. 35). Taylor & Francis. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315827414-9/full-information-estimation-presence-incomplete-data-james-arbuckle

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would jesse jackson 'fail' the implicit association test?" *Psychological Inquiry*, *15*, 257–279. https://doi.org/10.1207/s15327965pli1504_01

Axt, J., Buttrick, N., & Feng, R. Y. (2024). A comparative investigation of the predictive validity of four indirect measures of bias and prejudice. *Personality and Social Psychology Bulletin*, *50*, 871–888. https://doi.org/10.1177/01461672221150229

Axt, J. R. (2018). The best way to measure explicit racial attitudes is to ask about them. *Social Psychological and Personality Science*, *9*(8), 896–906. https://doi.org/10.1177/1948550617728995

Axt, J. R., Bar-Anan, Y., & Vianello, M. (2020). The relation between evaluation and racial categorization of emotional faces. *Social Psychological and Personality Science*, *11*(2), 196–206. https://doi.org/10.1177/1948550619848000

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2016). An unintentional, robust, and replicable pro-Black bias in social judgment. *Social Cognition*, *34*(1), 1–39. https://doi.org/10.1521/soco.2016.34.1.1

Axt, J. R., Feng, T. Y., & Bar-Anan, Y. (2021). The good and the bad: Are some attribute words better than others in the implicit association test? *Behavior Research Methods*, *53*(6), 2512–2527. https://doi.org/10.3758/s13428-021-01592-8

Axt, J. R., Nguyen, H., & Nosek, B. A. (2018). The Judgment Bias Task: A flexible method for assessing individual differences in social judgment biases. *Journal of Experimental Social Psychology*, *76*, 337–355. https://doi.org/10.1016/j.jesp.2018.02.011

Axt, J. R., & Roy, E. (2025). Moderators of test–retest reliability in implicit and explicit attitudes. *Journal of Personality and Social Psychology*, *128*, 567–593. https://doi.org/10.1037/pspa0000419

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry*, *15*, 279–289. https://www.researchgate.net/publication/228926174_No_Place_for_Nostalgia_in_Science_A_Response_to_Arkes_and_Tetlock

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

Bar-Anan, Y., & Vianello, M. (2018). A multi-method multi-trait test of the dual-attitude perspective. *Journal of Experimental Psychology: General*, *147*(8), 1264–1272. https://doi.org/10.1037/xge0000383

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd Ed.). Cambridge University Press.

Barlow, M. R., & Lahey, J. N. (2018). What race is Lacey? Intersecting perceptions of racial minority status and social class. *Social Science Quarterly*, *99*(5), 1680–1698. https://doi.org/10.1111/ssqu.12529

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6, 10 & adulthood. *Psychological Science*, *17*(1), 53–58. https://doi.org/10.1111/j.1467-9280.2005.01664.x

Barrie, C. (2020). Searching racism after George Floyd. *Socius*, *6*, 1–3. https://doi.org/10.1177/2378023120971507

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

Blanton, H., & Jaccard, J. (2024). Listening to measurement error: Lessons from the IAT. Chapter in J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Cambridge handbook of implicit bias and racism* (pp. 337–373). Cambridge. https://doi.org/10.1017/9781108885492

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, *94*(3), 567–582. https://doi.org/10.1037/a0014665

Bobo, L., & Kluegel, J. R. (1993). Opposition to race-targeting: Self-interest, stratification ideology, or racial attitudes? *American Sociological Review*, *58*(4), 443–464. https://doi.org/10.2307/2096070

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631–643. https://doi.org/10.1037/0022-3514.79.4.631

Buttrick, N., Axt, J., Ebersole, C. R., & Huband, J. (2020). Re-assessing the incremental predictive validity of implicit association tests. *Journal of Experimental Social Psychology*, *88*, Article 103941. https://doi.org/10.1016/j.jesp.2019.103941

Caldwell, A. R. (2022). *Exploring equivalence testing with the updated TOSTER R package*. PsyArXiv.

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*, *4*, 330–350. https://doi.org/10.1177/1088868312440047

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, *134*(3), 1163–1224. https://doi.org/10.1093/qje/qjz008

Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, *57*, 278–287. https://doi.org/10.1111/sjop.12288

Carpenter, T. P., Goedderz, A., & Lai, C. K. (2023). Individual differences in implicit bias can be measured reliably by administering the same implicit association test multiple times. *Personality and Social Psychology Bulletin*, *49*(9), 1363–1378. https://doi.org/10.1177/01461672221099372

Caruso, E. M., Shapira, O., & Landy, J. F. (2017). Show me the money: A systematic exploration of manipulations, moderators, and mechanisms of priming effects. *Psychological Science*, *28*, 1148–1159. https://doi.org/10.1177/0956797617706161

Casad, B. J., Flores, A. J., & Didway, J. D. (2013). Using the implicit association test as an unconsciousness raising tool in psychology. *Teaching of Psychology*, *40*, 118–123. https://doi.org/10.1177/0098628312475031

Cesario, J. (2022). So close, yet so far: Stopping short of killing implicit bias. *Psychological Inquiry*, *33*(3), 162–166. https://doi.org/10.1080/1047840X.2022.2106753

Charlesworth, T. E., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Charlesworth, T. E., & Banaji, M. R. (2022). Patterns of implicit and explicit attitudes: IV. Change and stability from 2007 to 2020. *Psychological Science*, *33*(9), 1347–1371. https://doi.org/10.1177/09567976221084257

Clark, C. J., & Tetlock, P. E. (2023). Adversarial collaboration: The next science reform. In C. Frisby, R. Redding, W. O'Donohue, & S. Lilienfeld (Eds.), *Political bias in psychology: Nature, scope, and solutions*. Springer. https://doi.org/10.1007/978-3-031-29148-7_32

Connor, P., Clark, C. J., Axt, J. R., & Hoogeveen, S. (2025, November 10). *On the relationship between implicit prejudice measures and discrimination: An adversarial collaboration*. https://osf.io/wsu4q.

Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology*, *3*, 835–846. https://doi.org/10.1038/s44159-024-00376-z

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

Crabtree, C., Kim, J. Y., Gaddis, S. M., Holbein, J. B., Guage, C., & Marx, W. W. (2023). Validated names for experimental studies on race and ethnicity. *Scientific Data*, *10*(1), Article 130. https://doi.org/10.1038/s41597-023-01947-0

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. https://doi.org/10.1177/0146167204264654

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measurement: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170. https://doi.org/10.1111/1467-9280.00328

Cyrus-Lai, W., Tierney, W., du Plessis, C., Nguyen, M., Schaerer, M., Clemente, E., & Uhlmann, E. L. (2022). Avoiding bias in the search for implicit bias. *Psychological Inquiry*, *33*(3), 203–212. https://doi.org/10.1080/1047840X.2022.2106762

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353. https://doi.org/10.1111/spc3.12111

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science*, *14*(5), 835–840. https://doi.org/10.1177/17456916198556

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 176–196). Guilford Press.

DeBruine, L. (2023). *faux: Simulation for Factorial Designs*. R package Version 1.2.1. Zenodo. https://doi.org/10.5281/zenodo.2669586

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Diekmann, K. A., Samuels, S. M., Ross, L., & Bazerman, M. H. (1997). Self-interest and fairness in problems of resource allocation: Allocators versus recipients. *Journal of Personality and Social Psychology*, *72*, 1061–1074. https://doi.org/10.1037/0022-3514.72.5.1061

Dietrich, B. J., & Sands, M. L. (2023). Seeing racial avoidance on New York City streets. *Nature Human Behavior*, *7*, 1275–1281. https://doi.org/10.1038/s41562-023-01589-7

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*, 315–319. https://doi.org/10.1111/1467-9280.00262

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, *36*, 4–56. https://doi.org/10.1016/S0065-2601(04)36001-6

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326. https://doi.org/10.1177/0146167297233009

Dutilh, G., Sarafoglou, A., & Wagenmakers, E.-J. (2019). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese*, *198*, 5745–5772. https://doi.org/10.1007/s11229-019-02456-7

Ekstrøm, C. T. (2022). *MESS: Miscellaneous esoteric statistical scripts* [R package Version 0.5.9]. https://CRAN.R-project.org/package=MESS

Fanelli, D. (2010). Positive results increase down the hierarchy of the sciences. *PLOS ONE*, *5*(4), Article e10068. https://doi.org/10.1371/journal.pone.0010068

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, *23*, 75–109. https://doi.org/10.1016/S0065-2601(08)60318-4

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*(6), 1013–1027.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Forscher, P. S. (2023). *A brief intellectual history of implicit bias*. https://persistentastonishment.blogspot.com/2023/01/a-brief-intellectual-history-of.html

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*, 522–559. https://doi.org/10.1037/pspa0000160

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behaviour. *British Journal of Social Psychology*, *47*(3), 397–419. https://doi.org/10.1348/014466607X241540

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Academic Press.

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. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., De Houwer, J., & Sherman. J. W. (2020). Twenty-five years of research using implicit measures. *Social Cognition*, *38*, s1–s25. https://doi.org/10.1521/soco.2020.38.supp.s1

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. https://doi.org/10.1037/0022-3514.60.4.509

Glover, D., Pallais, A., & Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*, *132*, 1219–1260. https://doi.org/10.1093/qje/qjx006

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

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561. https://doi.org/10.1037/pspa0000016

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*, 945–967. https://doi.org/10.2307/20439056

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Oakes, M. A., & Hoffman, H. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, *39*, 399–405. https://doi.org/10.1016/S0022-1031(03)00020-9

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Gu, X., Hoijtink, H., Mulder, J., & van Lissa, C. (2024). *bain: Bayes factors for informative hypotheses*. (Version 353 0.2.11, R package). https://CRAN.R-project.org/package=bain

Hagiwara, N., Kron, F. W., Scerbo, M. W., & Watson, G. S. (2020). A call for grounding implicit bias training in clinical and translational frameworks. *The Lancet*, *395*(10234), 1457–1460. https://doi.org/10.1016/S0140-6736(20)30846-1

Hardy, J. H., III, Tey, K. S., Cyrus-Lai, W., Martell, R. F., Olstad, A., & Uhlmann, E. L. (2022). Bias in context: Small biases in hiring evaluations have big consequences. *Journal of Management*, *48*(3), 657–692. https://doi.org/10.1177/0149206320982654

Hekman, D. R., Aquino, K., Owens, B. P., Mitchell, T. R., Schilpzand, P., & Leavitt, K. (2010). An examination of whether and how racial and gender biases influence customer satisfaction. *Academy of Management Journal*, *53*(2), 238–264. https://doi.org/10.5465/amj.2010.49388763

Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, *23*(2), 253–283. https://doi.org/10.1111/0162-895X.00281

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes & Intergroup Relations*, *11*(1), 69–87. https://doi.org/10.1177/1368430207084847

Hofmann, W. G., Friese, M., Wiers, R. W., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*(4), 962–977. https://doi.org/10.1037/a0012705

Holland, R. W., Verplanken, B., & Van Knippenberg, A. (2002). On the nature of attitude–behavior relations: The strong guide, the weak follow. *European Journal of Social Psychology*, *32*, 869–876. https://doi.org/10.1002/ejsp.135

Hoogeveen, S., Berkhout, S. W., Gronau, Q. F., Wagenmakers, E.-J. & Haaf, J. M. (2023). Improving statistical analysis in team science: The case of a Bayesian multiverse of Many Labs 4. *Advances in Methods and Practices in Psychological Science*, *6*(3), Article 25152459231182318. https://doi.org/10.1177/25152459231182318

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*, 696–701. https://doi.org/10.1371/journal.pmed.0020124

Irving, L. H., & Smith, C. T. (2020). Measure what you are trying to predict: Applying the correspondence principle to the implicit association test. *Journal of Experimental Social*, *86*, Article 103898. https://doi.org/10.1016/j.jesp.2019.103898

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*, 187–218. https://doi.org/10.1037/a0038557

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340. https://doi.org/10.1037/0096-3445.110.3.306

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, *47*, 802–809. https://doi.org/10.1037/0003-066X.47.6.802

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69. https://doi.org/10.1016/j.riob.2009.10.001

Jussim, L. (2017). Accuracy, bias, self-fulfilling prophecies, and scientific self-correction. *Behavioral and Brain Sciences*, *40*, e18. https://doi.org/10.1017/S0140525X16000339

Jussim, L., Careem, A., Goldberg, Z., Honeycutt, N., & Stevens. S. (2024). IAT scores, racial gaps, and scientific gaps. In J. A. Krosnick, T. H Stark & A. L. Scott (Eds.), *The Cambridge handbook of implicit bias and racism* (pp. 374–411). Cambridge University Press. https://doi.org/10.1017/9781108885492.021

Kang, J. (2024). Little things matter a lot: The significance of implicit bias, practically and legally. *Daedalus*, *153*, 193–212. https://doi.org/10.1162/daed_a_02055

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32. https://doi.org/10.1037/0022-3514.91.1.16

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the implicit association test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology*, *63*, 595–619. https://doi.org/10.1080/17470210903076826

Koppehele-Gossel, J., Hoffmann, L., Banse, R., & Gawronski, B. (2020). Evaluative priming as an implicit measure of evaluation: An examination of outlier-treatments for evaluative priming scores. *Journal of Experimental Social Psychology*, *87*, Article 103905. https://doi.org/10.1016/j.jesp.2019.103905

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the ultimatum game. *Psychological Science*, *24*(12), 2498–2504. https://doi.org/10.1177/0956797613496435

Kurdi, B., Ratliff, K. A., & Cunningham, W. A. (2021). Can the Implicit Association Test serve as a valid measure of automatic cognition? A response to Schimmack (2021). *Perspectives on Psychological Science*, *16*(2), 422–434. https://doi.org/10.1177/1745691620904080

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*(5), 569–586. https://doi.org/10.1037/amp0000364

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*, 423–434. https://doi.org/10.1038/s41562-019-0787-z

Lahey, J. N., & Oxley, D. R. (2021). Discrimination at the intersection of age, race, and gender: Evidence from a lab-in-the-field experiment. *Journal of Policy Analysis and Management*, *40*(4), 1083–1119. https://doi.org/10.1002/pam.22281

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765–1785. https://doi.org/10.1037/a0036260

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., … Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*, 1001–1016. https://doi.org/10.1037/xge0000179

Lakens, D. (2017). Equivalence tests: A practical primer for *t*-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*(1), 53–76. https://doi.org/10.1207/s15327906mbr3201_3

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature*, *526*(7572), 187–189. https://doi.org/10.1038/526187a

Machery, E. (2022). Anomalies in implicit attitudes research. *WIREs Cognitive Science*, *13*(1), Article e1569. https://doi.org/10.1002/wcs.1569

Martel, C., Rathje, S., Clark, C. J., Pennycook, G., Van Bavel, J. J., Rand, D. G., & van der Linden, S. (2024). On the efficacy of accuracy prompts across partisan lines: An adversarial collaboration. *Psychological Science*, *35*, 435–450. https://doi.org/10.1177/09567976241232905

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., & Rothermund, K. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology*, *10*, Article 2483. https://doi.org/10.3389/fpsyg.2019.02483

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*(4), 269–275. https://doi.org/10.1111/1467-9280.00350

Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., Beshears, J., Bonacorsi, L., Camerer, C., Chang, E., Chapman, G., Cialdini, R., Dai, H., Eskreis- Winkler, L., Fishbach, A., Gross, J. J., … Duckworth, A. L. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, *600*(7889), 478–483. https://doi.org/10.1038/s41586-021-04128-4

Mitchell, G., & Tetlock, P. (2006). Antidiscrimination law and the perils of mindreading. *Ohio State Law Journal*, *67*, 1023–1121.

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 164–195). Wiley Blackwell. https://doi.org/10.1002/9781119095910.ch10

Nødtvedt, K. B., Sjåstad, H., Skard, S. R., Thorbjørnsen, H., & Van Bavel, J. J. (2021). Racial bias in the sharing economy and the role of trust and self-congruence. *Journal of Experimental Psychology: Applied*, *27*(3), 508–528. https://doi.org/10.1037/xap0000355

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental psychology. General*, *134*, 565–584. https://doi.org/10.1037/0096-3445.134.4.565

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*(6), 625–666. https://doi.org/10.1521/soco.19.6.625.20886

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of racial prejudice: What are we measuring? *Psychological Science*, *14*, 636–639. https://doi.org/10.1046/j.0956-7976.2003.psci_1477.x

Onyeador, I. N., Hudson, S. T. J., & Lewis, N. A., Jr. (2021). Moving beyond implicit bias training: Policy insights for increasing organizational diversity. *Policy Insights From the Behavioral and Brain Sciences*, *8*(1), 19–26. https://doi.org/10.1177/2372732220983840

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. https://doi.org/10.1037/a0032734

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*(1), 54–74. https://doi.org/10.1037/0033-2909.124.1.54

Park, B., Petty, R. E., Correll, J., Feldman Barret, L., Hutchings, V., Langer, G., Otten, S., Parker, C., & von Hippel, W. (2025). Report from the NSF conference on implicit bias. In J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Cambridge handbook of implicit bias and racism* (pp. 43–69). Cambridge University Press. https://doi.org/10.1017/9781108885492.003

Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology*, *46*(2), 367–374. https://doi.org/10.1016/j.jesp.2009.11.001

Payne, B. K., & Lundberg, K. B. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, *8*, 672–686. https://doi.org/10.1111/spc3.12148

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 255–278). Guilford Press.

Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: Attitude clarity and attitude correctness. *Journal of Personality and Social Psychology*, *92*(1), 30–41. https://doi.org/10.1037/0022-3514.92.1.30

Petty, R. E., & Briñol, P. (2006). A meta-cognitive approach to “implicit” and “explicit” evaluations: Comment on Gawronski and Bodenhausen (2006). *Psychological Bulletin*, *132*(5), 740–744. https://doi.org/10.1037/0033-2909.132.5.740

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. William C. Brown.

Petty, R. E., & Krosnick, J. A. (Eds.), (1995). *Attitude strength: Antecedents and consequences*. Erlbaum Associates.

Peyton, K., & Huber, G. A. (2021). Racial resentment, prejudice, and discrimination. *The Journal of Politics*, *83*(4), 1829–1836. https://doi.org/10.1086/711558

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology*, *71*(3), 431–449. https://doi.org/10.1037/0022-3514.71.3.431

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, *114*(41), 10870–10875. https://doi.org/10.1073/pnas.1706255114

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*(2), 386–396. https://doi.org/10.1016/j.jesp.2006.12.008

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019). Committees with implicit biases pro mote fewer women when they do not believe gender biases exist. *Nature Human Behavior*, *3*, 1171–1179. https://doi.org/10.1038/s41562-019-0686-3

Reny, T. T., & Newman, B. J. (2021). The opinion-mobilizing effect of social protest against police violence: Evidence from the 2020 George Floyd protests. *The American Political Science Review*, *115*(4), 1499–1507. https://doi.org/10.1017/S0003055421000460

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

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139–165. https://doi.org/10.1037/0096-3445.133.2.139

Rouder, J. N., & Mehrvarz, M. (2024). Hierarchical-model insights for planning and interpreting individual-difference studies of cognitive abilities. *Current Directions in Psychological Science*, *33*(2), 128–135. https://doi.org/10.1177/09637214231220923

Roy, E., Jaeger, B., Evans, T., Lai, C. K., & Axt, J. R. (2025). A contest study to reduce attractiveness-based discrimination in social judgment. *Journal of Personality and Social Psychology*, *128*, 508–535. https://doi.org/10.1037/pspa0000414

Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes & Intergroup Relations*, *5*, 133–150. https://doi.org/10.1177/1368430202005002541

Schimmack, U. (2021). The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science*, *16*(2), 396–414. https://doi.org/10.1177/1745691619863798

Schwab, S. D., & Singh, M. (2024). How power shapes behavior: Evidence from physicians. *Science*, *384*, 802–808. https://doi.org/10.1126/science.adl3835

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, *79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, *53*, 676–694. https://doi.org/10.1080/00273171.2018.1476221

Shirom, A., & Melamed, S. (2006). A comparison of the construct validity of two burnout measures in two groups of professionals. *International Journal of Stress Management*, *13*(2), 176–200. https://doi.org/10.1037/1072-5245.13.2.176

Sniderman, P. M., & Tetlock, P. E. (1986). Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues*, *42*, 129–150. https://doi.org/10.1111/j.1540-4560.1986.tb00229.x

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, *108*(19), 7710–7715. https://doi.org/10.1073/pnas.1014345108

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A single block IAT. *European Journal of Psychological Assessment*, *24*, 237–245. https://doi.org/10.1027/1015-5759.24.4.237

Tetlock, P. E., & Mitchell, G. (2009a). A renewed plea for adversarial collaboration. *Research in Organizational Behavior*, *29*, 71–72. https://www.researchgate.net/publication/238382693_A_renewed_appeal_for_adversarial_collaboration

Tetlock, P. E., & Mitchell, G. (2009b). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior*, *29*, 3–38. https://doi.org/10.1016/j.riob.2009.10.002

Tetlock, P. E., & Mitchell, G. (2024). Stretching the limits of science: Was the implicit-racism debate a “bridge too far” for social psychology? In J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Cambridge handbook of implicit bias and racism* (pp. 149–168). Cambridge University Press. https://doi.org/10.1017/9781108885492.009

Tierney, W., Hardy, J. H., III, Ebersole, C., Leavitt, K., Viganola, D., Clemente, E., Gordon, M., Dreber, A. A., Johannesson, M., Pfeiffer, T., Hiring Decisions Forecasting Collaboration, & Uhlmann, E. (2020). Creative destruction in science. *Organizational Behavior and Human Decision Processes*, *161*, 291–309. https://doi.org/10.1016/j.obhdp.2020.07.002

Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M. D., & Johnson, R. E. (2012). Getting explicit about the implicit: A taxonomy of implicit measures and guide for their use in organizational research. *Organizational Research Methods*, *15*, 553–601. https://doi.org/10.1177/1094428112442750

Uhlmann, E. L., Poehlman, T. A., Tannenbaum, D., & Bargh, J. A. (2011). Implicit puritanism in American moral cognition. *Journal of Experimental Social Psychology*, *47*, 312–320. https://doi.org/10.1016/j.jesp.2010.10.013

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (2020). Reflecting on 25 years of research using implicit measures: Recommendations for their future use. *Social Cognition*, *38*(Suppl.), S223–S242. https://doi.org/10.1521/soco.2020.38.supp.s223

Vargas, P. T., von Hippel, W., & Petty, R. E. (2004). Using partially structured attitude measures to enhance the attitude–behavior relationship. *Personality and Social Psychology Bulletin*, *30*(2), 197–211. https://doi.org/10.1177/0146167203259931

Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., … Albarracín, D. (2021). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science*, *32*(10), 1566–1581. https://doi.org/10.1177/0956797621989733

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wolsiefer, K. J., & Blair, I. V. (2024). Implicit bias and discrimination: Evidence on causality. In J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Cambridge handbook of implicit bias and racism* (pp. 246–261). Cambridge University Press. https://doi.org/10.1017/9781108885492.014

Ximénez, C., Maydeu-Olivares, A., Shi, D., & Revuelta, J. (2022). Assessing cutoff values of SEM fit indices: Advantages of the unbiased SRMR index and its cutoff criterion based on communality. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(3), 368–380. https://doi.org/10.1080/10705511.2021.1992596

Yuan, K. H., & Bentler, P. M. (2000). 5. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, *90*, 554–562. https://doi.org/10.1037/0021-9010.90.3.553

Received June 18, 2025
Revision received November 10, 2025
Accepted November 13, 2025 ■