Social Cognition, Vol. 34, No. 1, 2016, pp. 1–39

AN UNINTENTIONAL, ROBUST, AND REPLICABLE PRO-BLACK BIAS IN SOCIAL JUDGMENT

Jordan R. Axt, Charles R. Ebersole, and Brian A. Nosek
University of Virginia

Empirical evidence and social commentary demonstrate favoring of Whites over Blacks in attitudes, social judgment, and social behavior. In 6 studies ($N > 4{,}000$), we provide evidence for a pro-Black bias in academic decision making. When making multiple admissions decisions for an academic honor society, participants from undergraduate and online samples had a more relaxed acceptance criterion for Black than White candidates, even though participants possessed implicit and explicit preferences for Whites over Blacks. This pro-Black criterion bias persisted among subsamples that wanted to be unbiased and believed they were unbiased. It also persisted even when participants were given warning of the bias or incentives to perform accurately. These results suggest opportunity for theoretical and empirical innovation on the conditions under which biases in social judgment favor and disfavor different social groups, and how those biases manifest outside of awareness or control.

Keywords: prejudice, discrimination, race, attitudes, decision making, implicit, automaticity

If there is one conclusion to draw from decades of research on stereotyping, prejudice, and discrimination in the United States, it is that Whites are treated more favorably than Blacks (Greenwald & Pettigrew, 2014; List, 2004; Pager & Shepherd,

Declaration of Conflicting Interests. This research was partly supported by Project Implicit. B. A. Nosek is an officer and J. R. Axt and C. R. Ebersole are consultants of Project Implicit, Inc., a nonprofit organization with the mission of "develop[ing] and deliver[ing] methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender, or other factors." The authors declared that they had no other potential conflicts of interest with respect to their authorship or the publication of this article.

Author Contributions. All authors developed the concept of these studies. J. R. Axt programmed all studies and analyzed the data. J. R. Axt drafted the manuscript, and B. A. Nosek and C. R. Ebersole edited it. All authors approved the final version of the manuscript for submission.

Chronology of Studies. Studies are numbered 1–6 for narrative style. Chronologically, studies were run the numbered order except for Study 2, which was conducted last.

Address correspondence to Jordan Axt, Department of Psychology, University of Virginia, 485 McCormick Road, Box 400400, Charlottesville, VA 22904-4400; E-mail: jaxt@virginia.edu.

2008). Historically, this is understood in the context of slavery, lynchings, segregation, and state-sponsored discrimination. But, evidence from late 20th and early 21st century research suggests that the effects persist, if in subtler, even unintentional forms (Greenwald & Banaji, 1995; Nosek, Smyth, et al., 2007). Whites are advantaged or receive more benefits than Blacks in job openings (Bertrand & Mullainathan, 2004), housing opportunities (U.S. Department of Housing and Urban Development, 2013), offers on goods (Doleac & Stein, 2015), potential academic positions (Milkman, Akinola, & Chugh, 2015), and research funding (Ginther et al., 2011).

Controlled experimental paradigms demonstrate similar pro-White effects (e.g., Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Goff, Steele, & Davies, 2008; McConnell & Leibold, 2001). For instance, when evaluating hypothetical candidates with ambiguous qualifications for an academic position, participants were more likely to recommend a White over a Black applicant (Dovidio & Gaertner, 2000). Similarly, when reviewing fictional resumes of business leaders, profiles of Blacks were rated as having less leadership effectiveness than equivalent profiles of Whites (Rosette, Leonardelli, & Phillips, 2008). Or, when judging underperforming employees, participants were more likely to formally recommend firing a Black than White employee for committing the same actions (Biernat, Fuegen, & Kobrynowicz, 2010).

Evidence demonstrating that Whites are often favored over Blacks in judgments and behavior pervades the psychological literature. In a review of five popular introductory psychology textbooks (Cacioppo & Freberg, 2013; Feist & Rosenberg, 2012; King, 2011; Passer & Smith, 2011; Schacter, Gilbert, & Wegner, 2011), the sections discussing race referenced 14 instances of anti-Black associations, attitudes, or behavior, and zero instances of pro-Black associations, attitudes, or behavior. In the most recent *Handbook of Social Psychology* (Fiske, Gilbert, & Lindzey, 2010), the chapters on attitudes (Banaji & Heiphetz, 2010), intergroup relations (Yzerbyt & Demoulin, 2010), and intergroup bias (Dovidio & Gaertner, 2010) reviewed 20 studies in which Whites exhibited anti-Black attitudes, associations, or behaviors compared to one study where Whites exhibited a pro-Black attitude, association, or behavior (Kinzler, Shutts, DeJesus, & Spelke, 2009). Notably, in the latter study, White five-year-olds preferred a Black person with a native accent over a White person with a foreign accent. Finally, among articles published in four leading psychology journals in 2013 and 2014 (*Psychological Science*, *Journal of Personality and Social Psychology*, *Personality and Social Psychology Bulletin*, *Journal of Experimental Social Psychology*), there were 17 papers in which Whites exhibited associations, attitudes, or behaviors that favored Whites over Blacks, and two papers in which Whites had attitudes, associations, or behaviors that favored Blacks over Whites. In the first instance of pro-Black behavior (Unzueta, Everly, & Gutierrez, 2014), White participants reported greater liking for a Black than White person after each complained about experiencing discrimination. In the other instance (Galinsky, Hall, & Cuddy, 2013), a majority-White sample was more likely to choose a Black female than White female candidate for a leadership position within a company, though White males were preferred over Black males.

That such a majority of the research finds preferences for Whites over Blacks is perhaps not surprising given that, on average, Whites show explicit and implicit preferences for Whites over Blacks (e.g., Axt, Ebersole, & Nosek, 2014; Bar-Anan & Nosek, 2014; Nosek, 2007; Nosek, Smyth et al., 2007; Payne et al., 2010). Most perspectives on stereotyping and prejudice consider attitudes to play a key role in discriminatory judgments and behavior (e.g., Ajzen & Fishbein, 2005; Fazio & Towles-Schwen, 1999). Considering that explicit and implicit racial attitudes are so pervasively pro-White, it is easy to understand the emphasis in theory and research on why Whites are favored over Blacks. In particular, theoretical perspectives on the role of automatic racial evaluations influencing judgment and behavior have pro-White associations as a starting point for potential activation and influence on social judgment (Cunningham, Zelazo, Packer, & Van Bavel, 2007; Devine, 1989; Dovidio et al., 2002; Fazio, Jackson, Dunton, & Williams, 1995). Indeed, if attitudes are a determinant of discriminatory judgments, and both implicit and explicit racial attitudes favor Whites, then there is little reason that models would anticipate outcomes favoring Blacks, particularly when the influence of race occurs without awareness or control. We began the present research sharing this presumption and were surprised to find contrary evidence.

One of the challenges for investigating social judgment biases in the laboratory is that there are no established paradigms used for assessing such biases that (1) are reliable, (2) can distinguish clearly whether bias has occurred within an individual, and (3) are adaptable for a variety of research uses. We created a paradigm where participants make accept and reject decisions for applicants that are either qualified or unqualified (see Axt, Nguyen, & Nosek, 2015 for more information), and applied it to assess social judgment biases in an academic context toward Blacks compared to Whites. In Study 1, we unexpectedly observed a social judgment bias favoring Blacks over Whites. In 5 subsequent studies, we established that this effect is robust, replicable, and appears to occur partly outside of awareness and control. Such evidence presents an opportunity for advancing theory about stereotyping and prejudice to anticipate the conditions under which different groups will be favored or disfavored.

EXISTING EVIDENCE OF PRO-BLACK EFFECTS

Despite the prevailing evidence for pro-White biases, there are hints in the literature that Whites, and others, sometimes exhibit pro-Black judgments and behaviors. For instance, White participants have been shown to indicate greater self-reported liking (Mendes, Blascovich, Lickel, & Hunter, 2002; Vanman, Paul, Ito, & Miller, 1997) and more positive behaviors (laughing, smiling; Mendes & Koslov, 2013) toward a Black than White interaction partner. However, these pro-Black attitudes and behaviors have been interpreted as deliberate attempts to correct for automatic pro-White biases, as pro-Black attitudes were no longer present when executive resources were taxed via cognitive load (Mendes & Koslov, 2013), and occurred despite anti-Black physiological responses (Mendes et al., 2002; Vanman et al., 1997).

Perhaps the most striking example of pro-Black behavior comes from research on "casuistry," in which people engage in selective reasoning to justify their actions. In a hypothetical college admissions scenario, White participants chose between candidates that had equal qualifications. A Black candidate was roughly three times more likely to be selected than a White candidate (Norton, Vandello, & Darley, 2004). This behavior was partly driven by participants strategically altering what information they reported as being most important in their decisions. For example, when the Black applicant had a higher GPA, participants ranked GPA as more important than when the White applicant had a higher GPA (Norton, Vandello, Biga, & Darley, 2008). The preference for Black over White candidates persisted even after participants were told they would have to justify their decision to the experimenter and when participants had to report which criteria were most important before making their selection (Norton et al., 2004).

A comparable pro-Black bias in decision making was found in a recent natural experiment. When French companies were randomly assigned to receive resumes that either did or did not have the applicant's name, minority candidates (who were primarily African) were *less* likely to be interviewed or hired when their names and racial information were removed from the application (Behaghel, Crépon, & Le Barbanchon, 2015). Specifically, minority candidates received interviews at 9.3% of positions when names were attached to resumes but only 4.7% of positions when their resumes were anonymous.

In these cases, the only existing evidence suggests that preference for minority applicants is a function of deliberate processes (Mendes & Koslov, 2013)—such as affirmative action goals to correct perceived lack of equal opportunity or historical advantages. Given the existing literature's emphasis on unintended biases producing bias against Blacks compared to Whites that is corrected via deliberate adjustments (Cunningham et al., 2007; Devine, 1989; Greenwald & Banaji, 1995; Strack & Deutsch, 2004), that such an effect could occur automatically or outside of awareness would appear counter to existing models of automatic bias. The single exception is evidence that some people possess chronic egalitarian goals that can be activated automatically (Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). In this case, automatic activation of an egalitarian goal is anticipated to override automatic pro-White biases, but would not account for an automatic bias in behavior that favors Blacks over Whites.

OVERVIEW

In six studies, we observed replicable, robust evidence for a social judgment bias favoring Black over White candidates. Study 1 introduces our paradigm and finds evidence for a pro-Black bias in judgments for admission to an academic honor society. Studies 2–6 provide evidence that this behavior can occur without intention or awareness. All studies used a social judgment paradigm we developed to assess social judgment biases reliably and efficiently. Participants made accept and reject decisions for applicants to an academic honor society. Applicants were presented

with a photo and four pieces of relevant information (science GPA, humanities GPA, recommendation letter strength, interview score). We manipulated the race and qualifications of each applicant, such that half of the applicants were White and half were Black, and within each race, half of the applicants had qualifications that made them more qualified and half had qualifications that made them less qualified. The design therefore possessed objectively correct and incorrect answers, and recorded multiple judgments quickly to produce reliable estimates. Participants could show bias in two ways: differential ability by race in distinguishing less qualified and more qualified applicants, and differential criterion by race for selecting candidates (e.g., being more likely to select a White candidate with the same credentials as a Black candidate).

STUDY 1

METHOD

Participants

For all studies, we report how we determined our sample size, all data exclusions, all manipulations, and all measures.

We planned to collect sufficient data to achieve at least 80% power to detect a medium within-subjects effect size of Cohen's d = .50 (44 participants). Due to over-scheduling, our sample was slightly larger. Forty-seven White undergraduates (35 female; M_{age} = 18.57, SD = 1.14) participated in exchange for partial course credit. This sample provided 27% power at detecting d = .20, 92% power at detecting d = .50, and nearly 100% power for detecting d = .80.

Procedure

Participants completed the study at individual computer carrels with 0 to 3 other participants in the room at the same time. After providing consent, participants completed measures in the following order: academic decision-making task, measures of explicit and implicit racial attitudes in randomized order, and a demographics survey. Participants were then debriefed and given feedback on their implicit task performance (see https://osf.io/evzuh/ for materials, data, and analysis scripts from all studies).

Academic Decision-Making Task. Participants were instructed that they would first view all the applicants for an academic honor society, and then select or reject each applicant. In the viewing phase, each of 60 applications was shown one at a time for three seconds in a random order, and participants just observed passively. This provided participants with insight on the range of qualifications before making any accept or reject decisions. For the selection phase, participants saw the same applicants one at a time in randomized order, and were instructed to accept approximately half of them. Participants pressed the "I" key to accept and the "E" key to reject. There was no time limit for making the accept or reject decisions.

Each application included a picture of the applicant's face and four pieces of information: Science GPA (Range of 1–4); humanities GPA (1–4); recommendation letters (Poor, fair, good, excellent); and interview score (1–100). Participants were instructed to weigh each piece of information equally.

We used the four pieces of information to create 60 total applications, 30 that were *more qualified* and 30 that were *less qualified*. To do this, we standardized each piece of information to have a 1–4 range. The two GPAs already ranged from 1–4, and we converted the recommendation letters (poor = 1, fair = 2, good = 3, excellent = 4) and interview scores (dividing interview score by 25) to be on the same 1–4 scale as the GPAs. *Less qualified* applicants had information summing to 13 and *more qualified* applicants had information summing to 14.

For example, one *less qualified* applicant had the following qualifications: Science GPA = 3.6, Humanities GPA = 3.7, Recommendation Letters = Good, Interview Score = 67.5. When standardized, these pieces of information sum to 13 (3.7 + 3.6 + (Good = 3) + (67.5/25) = 13). One *more qualified* applicant had the following qualifications: Science GPA = 3.6, Humanities GPA = 3.4, Recommendation Letters = Excellent, Interview Score = 75. When standardized these pieces of information sum to 14 (3.6 + 3.4 + (Excellent = 4) + (75/25) = 14). See Appendix A for the applications used in all studies.

On the applications, 30 of the faces were Black males and 30 were White males. Faces and applications were randomly paired at the beginning of each study session with the restriction that Black and White faces were equally represented in the *more qualified* and *less qualified* groups (15 of each in each group).

Explicit Racial Attitudes. Participants completed the nine-item Symbolic Racism 2000 scale (Henry & Sears, 2002) followed by a single-item measure of preferences for Black compared to White people (Nosek, Smyth et al., 2007) that used a 7-point scale ranging from "I strongly prefer Black people to White people" (-3) to "I strongly prefer White people to Black people" (+3).

Implicit Racial Attitudes. Participants completed a seven-block Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2007) measuring the strength of the association between the concepts "Pleasant" and "Unpleasant" and the categories "White American" and "Black American." IAT responses were scored by the *D* algorithm (Greenwald, Nosek, & Banaji, 2003), such that more positive scores reflected a stronger association between White American and pleasant and Black American and unpleasant. The procedure followed the recommended design and exclusion criteria from Nosek, Greenwald, and Banaji (2005).

Demographics. Participants completed an 11-item demographics questionnaire. We only analyzed the items relating to race, gender, and age.

RESULTS

For all studies, we planned to exclude participants from analysis if they accepted less than 20% or more than 80% of the applicants on the decision-making task to

remove participants who likely disregarded the instructions to accept half of the applicants. Participants were also excluded if they accepted or rejected every applicant from either race. No participants were excluded by these criteria in Study 1. Two participants were excluded from IAT analyses for having more than 10% of IAT trial responses less than 300 ms following the Nosek, Greenwald, and Banaji (2005) guidelines.

Accuracy is defined as selecting *more qualified* candidates and rejecting *less qualified* candidates. Overall accuracy on the task was 73.4% (SD = 7.3), well in excess of chance, $t(46) = 21.97$, $p < .001$, $d = 3.20$, 95% CI [2.47, 3.91], but not so high that differences across conditions might be suppressed. The overall average acceptance rate was close to the recommended 50% (M = 50.3%, SD = 9.5).

Racial Bias in Selection for Honor Society. We used signal detection theory (SDT; Green & Swets, 1966/1974; MacMillan & de Creelman, 1991) to analyze the influence of qualifications on admissions judgments. This analysis assumes that on average, applicants with superior grades, recommendation letters, and interview scores (total score = 14) are more qualified for the honor society than applicants with lower values (total score = 13). The SDT framework assumes that the distributions of subjective perception of the quality of *more qualified* and *less qualified* applicants are normal and have equal variances.

SDT allows for two estimates of an individual's decision-making process: sensitivity (d') and criterion (c). *Sensitivity* concerns the extent to which participants can differentiate between the *more qualified* and *less qualified* distributions. Participants high in sensitivity are more effective at distinguishing these distributions than participants low in sensitivity. *Criterion* (c) refers to the decision threshold for accepting or rejecting a candidate. Above this threshold, participants accept the applicant; below the threshold, participants reject the applicant. Participants can have a more liberal threshold in which they are more likely to accept candidates regardless of qualifications, or a more conservative threshold in which they are less likely to accept candidates regardless of qualifications.

SDT analyses have been used frequently in social psychological research, most often for studying memory (e.g., Hugenberg, Miller, & Claypool, 2007), but also for studying decision making. For example, in the first-person shooter task (Correll, Park, Judd, & Wittenbrink, 2002), participants are presented with images of Black and White people, and they must quickly decide whether the person is holding a gun or harmless object. Typically, participants adopt a lower criterion for Black than White targets, meaning that the threshold to respond "gun" is lower when the person on screen is Black than White (e.g., Correll et al., 2002; Correll, Wittenbrink, Crawford, & Sadler, 2015). In that paradigm, there are rarely differences in sensitivity, meaning that participants are equally capable of distinguishing between a gun and a harmless object when held by either a Black or White person.

In the academic honor society paradigm, we investigated whether participants differed in sensitivity (the ability to differentiate between *more qualified* and *less qualified* applicants) and criterion (the threshold where participants are willing to accept or reject an applicant) for Black and White applicants. There was not a reliable difference in sensitivity (d') between Black applicants (M = 1.45, SD = .54) and

White applicants (M = 1.35, SD = .71) for distinguishing between more and less qualified applicants, $t(46)$ = .92, p = .361, d = .13, 95% CI [-.15, .42]. For criterion, Black applicants (M = -0.09, SD = .41) were held to a lower criterion than White applicants (M = 0.06, SD = .37), $t(46)$ = 2.33, p = .024, d = .34, 95% CI [.04, .63], meaning that Black applicants were more likely to be accepted than White applicants.

Racial Attitude Relations with Selection Decisions. IAT D scores indicated an implicit preference for Whites over Blacks (M = 0.46, SD = .37), $t(44)$ = 8.31, p < .001, d = 1.24, 95% CI [.85, 1.62]. The explicit preference item also indicated pro-White attitudes (M = 0.57, SD = .77), $t(46)$ = 5.10, p <.001, d = .74, 95% CI [.42, 1.06].

To assess the relationship between selection decisions and attitudes, we calculated the difference between White and Black criterion values, such that higher scores indicated a more relaxed criterion for Blacks than Whites. This criterion bias was not reliably correlated with pro-White implicit (r = -.10, p = .524, 95% CI [-.38, .20]) or explicit (r = -.24, p = .099, 95% CI [-.50, .05]) attitudes, as well as Symbolic Racism 2000 responses (r = -.18, p = .239, 95% CI [-.44, .12]). A reliable negative correlation would have indicated that weaker explicit and implicit preferences for Whites compared to Blacks was associated with a more relaxed criterion for accepting Blacks compared to Whites.

DISCUSSION

White participants adopted a lower threshold for admittance to an academic honor society when the applicant was Black compared to White. Simultaneously, White participants were equally able to distinguish between more and less qualified Black and White applicants. By constructing a stimulus set with objectively more and less qualified applicants that were somewhat difficult to distinguish, this decision-making paradigm enabled estimation of the direction and degree of decision-making bias. In this circumstance, we observed an affirmative action—a Black applicant with the same academic credentials as a White applicant was 8.7% more likely to be selected for the honor society.

Why did this occur? Participants may have consciously wanted to be more lenient on Black than White candidates. For example, in adopting this affirmative action criterion, participants may have been deliberately attempting to counter perceived historical disadvantages, perceived disadvantages in academic experience, perceived differences in diagnosticity of academic scores, or perceived cultural biases (e.g., Abrams, Bertrand, & Mullainathan, 2012; Tenenbaum & Ruck, 2007). Alternatively, they may have been trying to correct underrepresentation of Blacks in academic advancement, or was the result of attitudinal preferences for Blacks over Whites. Only the last of these is obviously implausible as participants showed both implicit and explicit attitudes favored Whites over Blacks.

It is also possible that this criterion bias was not intentional. Participants may have relaxed their criterion for Blacks without awareness or control, and believed that they were not letting race influence their judgments. In Study 2, we sought to replicate the criterion bias favoring Black over White applicants, and test whether it occurred unintentionally.

STUDY 2

In Study 2, we added measures of participants' perceptions of their performance as well as their desired performance in making selections for the honor society. If the criterion bias is deliberate, some participants may report a desire to favor Black over White applicants. Those participants who report that they treated, or wanted to treat, members of both racial groups equally should show no racial differences in criterion bias. However, if the bias is unintentional, then participants reporting a desire to be unbiased, and perceptions that they were so, may still show racial differences in criterion.

Finally, we included measures of explicit and implicit race attitudes, the internal and external motivations to control prejudice scale (Plant & Devine, 1998), and items measuring attitudes toward affirmative action to investigate how task performance relates to race-related attitudes and motivations.

METHOD

Participants

We sought to collect as many participants as possible during the Fall 2014 academic semester. One hundred and thirty-five White undergraduates (60 female; M_{age} = 19.18, SD = 1.32) participated in exchange for partial course credit. This sample size allowed for greater than 97% power to detect a criterion difference of the same size found in Study 1 (d = .34). The sample size allowed for 64% power at detecting an effect size d = .20 and nearly 100% power for detecting d = .50 and d = .80.

Procedure

Participants completed the study online. After providing consent, participants completed measures in the following order: academic decision-making task, a questionnaire regarding perceptions of performance on the task and a measure of explicit racial attitudes and motivations in randomized order, a demographics questionnaire, and a measure of implicit racial attitudes. Participants were then debriefed and given feedback on their implicit task performance.

Academic Decision-Making Task. Participants completed the same task as in Study 1 with two changes. First, applications were presented for one second (vs. three seconds) at a time during the passive-viewing phase. Second, participants were randomly assigned to 1 out of 12 task orders. Across the 12 orders, each face was equally likely to be assigned to either a *more qualified* or *less qualified* application.

Perceptions of Performance. Participants completed two items regarding their performance on the task. In the first item, participants rated their perceived performance on the task using a 7-point scale ranging from "I was extremely easier on Black applicants and tougher on White applicants" (-3) to "I was extremely easier on White applicants and tougher on Black applicants" (+3), and a neutral midpoint of "I treated both Black and White applicants equally" (0). Next, participants

reported how they wanted to perform on the task using a similar 7-point scale ranging from "I wanted to be extremely easier on Black applicants and tougher on White applicants" (-3) to "I wanted to be extremely easier on White applicants and tougher on Black applicants" (+3) and a neutral midpoint of "I wanted to treat both Black and White applicants equally" (0).

Explicit Racial Attitudes and Motivations. Participants completed the same single-item racial preference item as in Study 1, the 10-item Internal and External Motivation to Control Prejudice scale (Plant & Devine, 1998), and two items assessing attitudes toward affirmative action. In each of the affirmative action items, participants read a scenario where a White American and Black American candidate were equally qualified for a position, and participants made a "yes" or "no" decision on whether it was justifiable for the position to more frequently be awarded to the Black American candidate. We took the total number of "yes" responses to these two items as a measure of attitudes toward affirmative action, with higher values meaning greater endorsement of affirmative action policies. See Appendix B for the full text of both items.

Demographics. Participants completed a 6-item demographics questionnaire. We only analyzed the items relating to race, gender, and age.

Implicit Racial Attitudes. Participants completed a four-block, good-focal Brief Implicit Association Test (BIAT; Sriram & Greenwald, 2009) measuring the strength of the association between the concepts "Good" and "Bad" and the categories "White people" and "Black people." Due to the online sample in Studies 2–6, we used the BIAT in order to shorten the study time and minimize participant dropout.

In each block of the BIAT (20 trials), words or images were presented one at a time and participants categorized them as quickly as possible. Categorization errors had to be corrected before continuing to the next trial. In the first block, participants pressed the "I" key for all *Good* words (Love, Pleasant, Great, Wonderful) and for faces (2 male, 2 female) belonging to either Black or White faces and the "E" key for "any other images and words." The other items were *Bad* words (Hate, Unpleasant, Awful, and Terrible) and faces from the other racial group. In the second block, the structure was the same, except the racial groups paired with *Good* and "other" words were reversed. These same blocks were repeated through the critical third and fourth blocks. To reinforce each block's pairings, the first four trials in all blocks were limited to only Black and White faces, and were omitted from analyses (Sriram & Greenwald, 2009). Participants were randomly assigned to one out of two orders.

BIAT responses were scored by the *D* algorithm (Nosek, Bar-Anan, Sriram, Axt, & Greenwald, 2014), such that more positive scores reflected a stronger association between White people and good and Black people and bad. The procedure followed the recommended procedure and exclusion criteria from Nosek and colleagues (2014), except that a warm-up block of categorizing only good and bad words was not part of the procedure because of time constraints.

RESULTS

No participants accepted less than 20% or more than 80% of the applicants, or accepted or rejected all applicants from either race. So, no participants were excluded on these criteria. Seven participants did not complete the BIAT due to computer error, and two additional participants were removed from analyses involving the BIAT due to more than 10% of trial responses being less than 400 ms, as recommended in Nosek and colleagues (2014).

Racial Bias in Selection for Honor Society. Task accuracy (M = 68.9%, SD = 8.2) and acceptance rates (M = 49.8%, SD = 10.6) were similar to those observed in Study 1. As in Study 1, Black applicants (M = -0.04, SD = .39) received a lower criterion than White applicants (M = 0.08, SD = .42), $t(134) = 3.27$, $p = .001$, $d = .28$, 95% CI [.11, .45], indicating that a less qualified Black applicant was more likely to be accepted than a less qualified White applicant.[1] Again, there was not a reliable difference in sensitivity (d') between Black applicants (M = 1.11, SD = .67) and White applicants (M = 1.12, SD = .57), $t(134) = .12$, $p = .905$, $d = .01$, 95% CI [-.16, .18]. These results closely replicate the effects in Study 1.

Awareness of Selection Bias. Most participants (80.7%) indicated that they *had treated* both Black and White applicants equally. Among them, Black applicants (M = -0.04, SD = .38) received a lower criterion than White applicants (M = 0.06, SD = .41), $t(108) = 2.37$, $p = .020$, $d = .23$, 95% CI [.04, .42]. Likewise, most participants (85.9%) indicated a *desire* to treat Black and White applicants equally. Among them, Black applicants (M = -0.04, SD = .40) received a lower criterion than White applicants (M = 0.09, SD = .43), $t(115) = 2.90$, $p = .005$, $d = .27$, 95% CI [.08, .45]. Both effect sizes were only slightly smaller than observed with the whole sample.

Racial Attitude Relations with Selection Decisions. BIAT D scores indicated pro-White attitudes (M = 0.30, SD = .44), $t(124) = 7.56$, $p < .001$, $d = .68$, 95% CI [.48, .87]. The explicit preference item also indicated pro-White attitudes (M = 0.56, SD = .69), $t(134) = 9.39$, $p < .001$, $d = .81$, 95% CI [.61, 1.00]. Among participants who reported no explicit preference for White or Black people (51.9%), Black applicants (M = -0.10, SD = .38) received a lower criterion than White applicants (M = 0.04, SD = .42), $t(69) = 2.78$, $p = .007$, $d = .33$, 95% CI [.09, .57], an effect size nearly equivalent to that observed with the whole sample.

We computed the same difference score used in Study 1, subtracting White criterion from Black criterion values such that higher scores indicated a greater pro-Black bias in decision making. The criterion bias was negatively correlated with BIAT D scores (r = -.26, p = .003, 95% CI [-.42, -.09]), meaning that more positive associations with Whites were associated with a more stringent criterion for accepting Blacks compared to Whites. The criterion bias was weakly, but not reliably related to internal motivation to control prejudice (r = .15, p = .077, 95% CI [-.02, .31], external motivation to control prejudice (r = -.03, p = .757, 95% CI [-.20, .14]), explicit racial preferences (r = -.14, p = .113, 95% CI [-.30, .03]), attitudes toward af-

1. A mixed-model ANOVA including task order on criterion showed a significant main effect of race, $F(1, 123) = 11.36$, $p = .001$, $\eta^2_p = .085$, but no significant race by order interaction, $F(11, 123) = .92$, $p = .527$, $\eta^2_p = .076$.

firmative action (r = .10, p = .235, 95% CI [-.07, .27]), perceptions of performance (r = -.12, p = .155, 95% CI [-.29, .05]), and desired performance (r = -.04, p = .655, 95% CI [-.21, .13]).

DISCUSSION

We replicated the key result from Study 1 that White participants applied a more relaxed criterion to admit Black than White applicants to an academic honor society. A sizable majority of participants (greater than 80%) both desired to treat Black and White applicants equally and believed that they did so. Despite these intentions and beliefs, the racial criterion bias among those participants was still highly reliable. These results suggest that the bias occurred unintentionally and without awareness for many participants. In sum, a Black applicant with the same academic credentials as a White applicant was 8.9% more likely to be selected for the honor society, and participants were largely unaware of this bias.

Finally, in both Study 1 and Study 2, attitude and prejudice motivation measures were only weakly related to the criterion bias. In Study 3, we sought to test whether the relation with attitudes was weak but reliable by using a large sample to estimate the effect precisely. Also, we recruited a more diverse sample to test whether the racial criterion bias generalized beyond our relatively homogenous samples of undergraduate students in the lab.

STUDY 3

METHODS

Participants

We sought to collect a sample that would produce greater than 80% power at detecting a small correlation of r = .10 (779 participants) using Project Implicit (implicit.harvard.edu) as a data source. Studies at Project Implicit are posted and removed at a fixed time every week, which resulted in a larger than planned sample. One thousand and forty-four participants completed at least the academic decision-making task. Of that, 897 participants completed all study measures.[2] This sample size allowed for 85% power at detecting an effect size r = .10, and nearly 100% power for detecting r = .30 and r = .50.

The study was restricted to only White participants, who had completed demographic information when first registering for the Project Implicit research pool. Among those who provided data, 60.8% were female and the mean age was 33.1 (SD = 14.2). For political ideology, 18.2% were conservative and 48.4% liberal; 23.9% of the sample were non-U.S. residents. For Studies 3–6, sample sizes vary across tests due to missing data.

2. The study had 1,402 started sessions, with 1,044 providing data, and 897 completing the study (64% completion rate).

Procedure

After providing consent, participants completed the measures in the following order: academic decision-making task, a questionnaire assessing perceived and desired performance on the task, an item concerning racial preferences, and a BIAT. Participants were then debriefed and given feedback on their BIAT performance. All measures were the same as in Study 2, except for the racial attitudes questionnaire, which included only the single item regarding preferences for White relative to Black people.

RESULTS

Forty-four participants (4.2%) were excluded from the analyses for accepting less than 20% or more than 80% of the applicants, or for accepting or rejecting all applicants from either race. An additional 21 participants (2.3%) were excluded from analysis with the BIAT for having more than 10% of BIAT trials with response latencies below 400 ms.

Task accuracy (M = 67.9%, SD = 8.6) and acceptance rates (M = 50.3%, SD = 12.0) were very similar to those observed in the undergraduate samples.

Racial Bias in Selection for Honor Society. As in the previous samples, Black applicants (M = -0.11, SD = .46) received a lower criterion than White applicants (M = 0.11, SD = .42), $t(999) = 16.57$, $p < .001$, $d = .52$, 95% CI [.46, .59].[3] Unlike in previous samples, there were reliable differences in sensitivity (d') between Black applicants (M = 0.96, SD = .65) and White applicants (M = 1.18, SD = .67), $t(999) = 9.47$, $p < .001$, $d = .30$, 95% CI [.24, .36]. Participants were more capable of distinguishing more qualified from less qualified White applicants than Black applicants.

Awareness of Selection Bias. Again, most participants (74.8%) indicated that they *had treated* both Black and White applicants equally. Among them, Black applicants (M = -0.09, SD = .45) received a lower criterion than White applicants (M = 0.11, SD = .42), $t(661) = 13.06$, $p < .001$, $d = .51$, 95% CI [.43, .59] with an effect size very similar to the whole sample. Likewise, most participants (91.8%) indicated a *desire* to treat Black and White applicants equally. Among them, Black applicants (M = -0.12, SD = .45) received a lower criterion than White applicants (M = 0.10, SD = .42), $t(812) = 15.35$, $p < .001$, $d = .54$, 95% CI [.47, .61].

Racial Attitude Relations with Selection Decisions. BIAT D scores indicated pro-White attitudes (M = 0.22, SD = .47), $t(884) = 14.17$, $p < .001$, $d = .48$, 95% CI [.41, .55]. The explicit preference item also revealed pro-White attitudes (M = 0.37, SD = .68), $t(887) = 16.14$, $p < .001$, $d = .54$, 95% CI [.47, .61]. Among participants who reported no explicit preference for White or Black people (65%), Black applicants (M = -0.13, SD = .45) received a lower criterion than White applicants (M = 0.11, SD = .42), $t(576) = 14.65$, $p < .001$, $d = .61$, 95% CI [.52, .70].

3. A mixed-model ANOVA including task order on criterion, showed a main effect of race, $F(1, 988) = 295.63$, $p < .001$, $\eta^2_p = .23$, and a significant race by order interaction, $F(11, 988) = 6.88$, $p < .001$, $\eta^2_p = .071$. Though the race by order interaction was significant, Black applicants received lower criterion than White applicants within each of the 12 study orders.

Using the same criterion difference score as in previous studies, criterion bias was negatively correlated with explicit racial preferences (r = -.17, p < .001, 95% CI [-.24, -.11]) and BIAT D scores (r = -.18, p < .001, 95% CI [-.24, -.12]), indicating that weaker implicit and explicit preferences for Whites were associated with a more relaxed criterion for accepting Blacks compared to Whites. The criterion bias was also negatively correlated with perceptions of performance (r = -.20, p < .001, 95% CI [-.26, -.13]) and desired performance (r = -.09, p = .010, 95% CI [-.15, -.02]), indicating that a greater desire or perception for treating Whites more favorably was associated with a smaller pro-Black criterion bias.

A simultaneous linear regression with implicit and explicit attitudes, and perceived and desired performance, predicting race differences in criterion bias revealed that both explicit (β = -.14, p < .001) and implicit (β = -.14, p < .001) attitudes, and perceived (β = -.16, p < .001) and desired (β = -.10, p = .004) task performance contributed uniquely. Overall, those four variables accounted for 9.6% of the racial difference in criterion bias. Differences in sensitivity (White d'–Black d') were not correlated with racial differences in criterion, explicit or implicit attitudes, or perceived or desired performance, all rs < .04, all ps > .208.

DISCUSSION

We replicated the key effect from Studies 1 and 2 in a large, heterogeneous sample. In Study 3, a Black applicant with the same academic credentials as a White applicant was 14.2% more likely to be selected for the honor society. Moreover, this bias occurred at approximately the same magnitude among participants who wanted to evaluate both races equally, who believed they did evaluate both races equally, and who reported no explicit racial preference.

Part of the motivation for this study was to test whether the relationship between racial difference in criterion bias and perceptions of performance, desired performance, and racial attitudes was reliable but relatively small. This was the case. Both implicit and explicit racial attitudes, as well as perceived and desired performance, were reliable and independent predictors of racial differences performance, but accounted for only a small portion of the variance. That is, while a desire to treat all applicants equally, a perception of having done so, and equal preference for Whites and Blacks were all associated with smaller criterion biases, such goals, perceptions, and attitudes were not enough to account for race influencing decision making.

Unlike the first two studies, we also observed racial differences in sensitivity. White applicants received a higher sensitivity than Black applicants, indicating that participants were better able to distinguish *more qualified* from *less qualified* applicants when the candidates were White than Black. However, individual differences in this sensitivity bias were not reliably correlated with implicit and explicit racial attitudes, nor with perceived and desired task performance. Nonetheless, it would be interesting if racial differences in sensitivity occurred reliably, perhaps

suggesting a difference in attention depending on the race of the applicant or outgroup homogeneity. We examined this again in the subsequent studies.

While the Project Implicit sample is highly heterogeneous, participants may also be particularly attuned to issues of bias, as that is a feature of the site's educational mission. If anything, we would have expected reduced racial differences in criterion bias and sensitivity among this sample, but that is not what we observed. In any case, for Study 4, we compared another Project Implicit sample with a different heterogeneous sample source—Mechanical Turk (MTurk; mturk.com), to obtain more precise and generalizable evidence for the criterion bias effect, the newly observed sensitivity effect, and the relations with attitudes and performance expectations.

STUDY 4

METHOD

Participants

Participants were recruited from the Project Implicit research pool and MTurk during the same week. We sought to collect a sample of MTurk participants that would provide greater than 80% power for detecting a small within subjects effect size of $d = .20$ (199 participants). Since we could not recruit only White participants and had to estimate the percentage of White respondents from a larger sample, our final sample was slightly bigger. Two hundred and thirty-five White participants from MTurk (61.1% female, $M_{age} = 37.6$, $SD = 13.8$) completed at least the academic decision-making task in exchange for $0.60 (215 completed all study measures).

We sought twice the size of our planned MTurk sample in our Project Implicit sample (398 participants). Studies at Project Implicit are posted and removed at a fixed time every week, which resulted in a larger than planned sample. Four hundred and eighty-four White participants completed at least the academic decision-making task (411 completed all study measures).[4] Among those who provided data, 60.7% were female and the mean age was 33.6 ($SD = 14.8$). The Project Implicit sample was 20.4% politically conservative and 47.4% liberal; 16.9% of the sample were non-U.S. citizens. These sample sizes allow for 71% power at detecting a between-groups difference in criterion bias for an effect size of $d = .2$ and nearly 100% power for detecting $d = .5$ and $d = .8$.

Procedure

After providing consent, participants from both samples completed measures in the following fixed order: academic decision-making task, questionnaire assessing perceived and desired performance on the task, single-item measure of racial preferences, the affirmative action and motivation to control prejudice items used

4. On Project Implicit, the study had 619 started sessions, with 484 providing data and 411 completing the study (66.4% completion rate).

in Study 2, and the BIAT. MTurk participants completed a six-item demographics questionnaire before the BIAT, where we only analyzed race, gender, and age. Participants were then debriefed and given feedback on their BIAT performance.

RESULTS

Thirty-one participants (4.3%) were excluded from the analyses for accepting less than 20% or more than 80% of the applicants, or for accepting or rejecting all applicants from either race. BIAT data was excluded from an additional 17 (2.8%) participants for having more than 10% of BIAT trials with response latencies below 400 ms.

For both samples, average task accuracy (PI: M = 68.4%, SD = 8.2; MTurk: M = 67.8%, SD = 7.4) and acceptance rates (PI: M = 50.6%, SD = 12.2; MTurk: M = 50.3%, SD = 11.6) were similar to those observed in previous samples.

Racial Bias in Selection for Honor Society. In the Project Implicit sample, Black applicants (M = -0.15, SD = .50) received a lower criterion than White applicants (M = 0.14, SD = .46), $t(460) = 11.46$, $p < .001$, d = .53, 95% CI [.44, .63]. In the MTurk sample, Black applicants (M = -0.05, SD = .49) also received a lower criterion than White applicants (M = 0.05, SD = .49), $t(226) = 2.47$, $p = .013$, d = .16, 95% CI [.03, .29].[5] Project Implicit participants showed a larger gap between Black and White criterion than MTurk participants, $t(686) = 4.15$, $p < .001$, d = .34, 95% CI [.18, .50].

For sensitivity, there were no reliable differences between White (PI: M = 1.12, SD = .64; MTurk: M = 1.09, SD = .65) and Black (PI: M = 1.12, SD = .66; MTurk: M = 1.04, SD = .57) applicants in either the Project Implicit sample, $t(460) = .11$, $p = .915$, d = .01, 95% CI [-.09, .10], or the MTurk sample, $t(226) = 1.06$, $p = .291$, d = .07, 95% CI [-.06, .20]. This is a failure to replicate Study 3 and is consistent with what was observed in Studies 1 and 2.

Awareness of Selection Bias. Within each sample, most participants (PI: 72.6%; MTurk: 90.7%) indicated that they *had treated* both Black and White applicants equally, and this perception was more common in the MTurk than Project Implicit sample, $\chi^2(1, N = 676) = 29.61$, $p < .001$.[6] Among participants who indicated treating both races equally, Black applicants (PI: M = -0.15, SD = .49; MTurk: M = -0.04, SD = .48) received a lower criterion than White applicants (PI: M = 0.14, SD = .45; MTurk: M = 0.06, SD = .46), in both the Project Implicit sample, $t(325) = 11.18$, $p < .001$, d = .62, 95% CI [.50, .74] and MTurk sample, $t(205) = 2.62$, $p = .009$, d = .18, 95% CI [.05, .32]. The size of this criterion bias was larger in the Project Implicit than MTurk sample, $t(530) = 4.32$, $p < .001$, d = .38, 95% CI [.21, .56].

5. A mixed-model ANOVA including task order on criterion in both Project Implicit and Mechanical Turk samples showed a reliable main effect of race, Fs > 5.11, ps < .025, η^2_p > .023. Neither the Project Implicit or Mechanical Turk sample showed a reliable race by order interaction, Fs < .69, ps > .414.

6. Though MTurk participants were more likely to indicate having treated both races equally and a desire to do so, Project Implicit participants were more likely to indicate a perception of having favored Black applicants (PI = 21.2%; MTurk = 7.5%), $\chi^2(1, N = 676) = 20.38$, $p < .001$, and a greater desire to favor Black applicants (PI = 12%; MTurk = 4.8%), $\chi^2(1, N = 676) = 8.95$, $p = .003$.

Within each sample, most participants (PI: 85.3%; MTurk: 94.7%) also indicated that they *desired to treat* both Black and White applicants equally, and this desire was more common in the MTurk than Project Implicit sample, $\chi^2(1, N = 676) = 13.09$, $p < .001$. Among them, Black applicants (PI: $M = -0.15$, $SD = .49$; MTurk: $M = -0.05$, $SD = .47$) received a lower criterion than White applicants (PI: $M = 0.14$, $SD = .45$; MTurk: $M = 0.05$, $SD = .46$), in both the Project Implicit sample, $t(382) = 11.27$, $p < .001$, $d = .58$, 95% CI [.47, .68] and MTurk sample, $t(214) = 2.48$, $p = .014$, $d = .17$, 95% CI [.03, .30]. The size of this criterion bias was larger in the Project Implicit than MTurk sample, $t(596) = 4.48$, $p < .001$, $d = .38$, 95% CI [.21, .55].

Racial Attitude Relations with Selection Decisions. Though both samples showed pro-White BIAT scores on average, MTurk participants ($M = 0.34$, $SD = .43$) had stronger implicit preferences for Whites over Blacks than did Project Implicit participants ($M = 0.17$, $SD = .52$), $t(593) = 3.99$, $p < .001$, $d = .34$, 95% CI [.17, .51]. The two samples did not differ on the degree of pro-White explicit racial preference (Project Implicit: $M = 0.39$, $SD = .75$; MTurk: $M = 0.40$, $SD = .79$), $t(674) = .25$, $p = .802$.

Within each sample, most participants (PI: 70%; MTurk: 61.5%) reported no explicit preference for White or Black people, and this was more common in the Project Implicit than MTurk samples, $\chi^2(1, N = 676) = 4.83$, $p = .028$. Among participants with no reported racial preference, Black applicants (PI: $M = -0.20$, $SD = .50$; MTurk: $M = -0.07$, $SD = .47$) received a lower criterion than White applicants (PI: $M = 0.17$, $SD = .46$; MTurk: $M = 0.09$, $SD = .46$), in both the Project Implicit sample, $t(275) = 11.29$, $p < .001$, $d = .68$, 95% CI [.55, .81] and MTurk sample, $t(158) = 3.43$, $p = .001$, $d = .27$, 95% CI [.11, .43]. The size of this criterion bias was larger in the Project Implicit than MTurk sample, $t(433) = 3.88$, $p < .001$, $d = .37$, 95% CI [.18, .55].

Across both samples, the criterion racial difference score was negatively correlated with explicit racial preferences ($r = -.22$, $p < .001$, 95% CI [-.29, -.14]), BIAT D scores ($r = -.21$, $p < .001$, 95% CI [-.29, -.13]), perceptions of performance ($r = -.25$, $p < .001$, 95% CI [-.32, -.18]), and desired performance ($r = -.25$, $p < .001$, 95% CI [-.32, -.17]). Further, it was positively correlated with attitudes toward affirmative action ($r = .18$, $p < .001$, 95% CI [.11, .25]), and internal motivation to control prejudice ($r = .15$, $p < .001$, 95% CI [.08, .23]). The criterion bias racial difference was not reliably correlated with external motivation to control prejudice ($r = -.02$, $p = .701$, 95% CI [-.09, .06]).

A simultaneous linear regression with implicit and explicit attitudes, and perceived and desired performance, predicting race differences in criterion bias revealed that both explicit ($\beta = -.13$, $p = .002$) and implicit ($\beta = -.15$, $p < .001$) attitudes and perceived ($\beta = -.14$, $p = .001$) and desired ($\beta = -.13$, $p = .001$) task performance contributed uniquely. Overall, those four variables accounted for 10.9% of the racial difference in criterion bias. Adding attitudes toward affirmative action ($\beta = .12$, $p = .004$), internal motivation to control prejudice ($\beta = .04$, $p = .366$), and external motivation to control prejudice ($\beta = .02$, $p = .640$) increased the overall amount of variance explained to 12.9%.

DISCUSSION

In both Project Implicit and MTurk samples, we replicated the difference in criterion bias showing more relaxed standards for selecting Black than White applicants for an academic honor society. This criterion bias was present but significantly smaller among MTurk participants, at roughly one-third of the size of the effect among Project Implicit participants. The racial difference in sensitivity observed in Study 3 did not replicate in either sample.

Part of the explanation for the difference across samples may be that MTurk participants held a somewhat stronger belief that they had treated the racial groups equally. Also, Project Implicit participants held weaker pro-White implicit preferences. In exploratory analyses, both factors mediated the relationship between sample and racial difference in criterion bias.[7] Even so, the large majority of both samples believed that they showed no bias and nonetheless did so. This suggests that, not surprisingly, participants can deliberately favor one group over another, but that for the substantial portion of the sample who reported showing no favoritism, the criterion bias effect occurs unintentionally and without awareness.

Next, we sought to examine the possibility that this effect occurs unintentionally using a more stringent test. In Study 5, participants received instructions that either (1) stressed the importance of being fair but did not mention race, (2) warned participants about the possibility of favoring Black applicants, or (3) warned participants about the possibility of favoring White applicants.

STUDY 5

METHOD

Participants

We sought to collect enough participants to detect a small between-subjects effect size of $d = .2$ between any of the four experimental conditions on the academic decision-making task (1,576 participants; 394 participants per condition). Studies at Project Implicit are posted and removed at a fixed time every week, which resulted in a larger than planned sample: 1,825 White participants completed at least the academic decision-making task through the Project Implicit research pool (1,550 participants completed all study measures).[8] Among those who provided data, 60.1% were female and the mean age was 33.3 (SD = 13.5). This sample allowed

7. We tested whether sample differences (Project Implicit = 0, MTurk = 1) in racial differences in criterion bias were mediated by several outcome variables using bootstrapping procedures. Unstandardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. We found reliable mediation by both BIAT D scores (95% C.I. on indirect effect: -.06, -.02) and perceived performance (-.06, -.01).

8. The study had 2,475 started sessions, with 1,825 providing data, and 1,550 completing the study (62.6% completion rate).

for 85% power at detecting a between-subjects effect size of $d = .20$ between any experimental conditions, and nearly 100% power for $d = .50$ and $d = .80$.

Procedure

After providing consent, participants completed measures in the following fixed order: academic decision-making task, questionnaire assessing perceived and desired performance on the task, one-item measure of racial preferences, and the BIAT. Participants were then debriefed and given feedback on their BIAT performance.

All study elements were the same as those in Study 3. The only change came in the experimental manipulation. Participants were randomly assigned to one of four conditions: *Control*, *Be Fair*, *Don't Favor Whites*, or *Don't Favor Blacks*. Participants in the *Control* condition completed all measures in the order described above. Participants in the other three conditions were given additional instructions immediately before the testing phase of the decision-making task.

In the *Be Fair* condition, participants were told that decision makers are sometimes too easy on certain applicants and too tough on others, and were reminded to try to be as fair as possible when making their accept and reject decisions. In the *Don't Favor Blacks* condition, participants were told that prior research suggests that decision makers are much easier on the Black candidates and much tougher on the White candidates, and the researchers would like to see if people can be fair toward all applicants if told about this tendency beforehand. The *Don't Favor Whites* condition had the same wording but described how prior research suggests a tendency to be easier on White and tougher on Black candidates. See Appendix C for the full text from each manipulation.

RESULTS

Seventy participants (3.8%) were excluded from the analyses for accepting less than 20% or more than 80% of the applicants, or for accepting or rejecting all applicants from either race. BIAT data from an additional 14 participants were excluded for having more than 10% of BIAT trials with response latencies below 400 ms. Task accuracy ($M = 69.0\%$, $SD = 8.3$) and acceptance rates ($M = 49.9\%$, $SD = 11.3$) were similar to those observed in earlier studies.

Racial Bias in Selection for Honor Society. Within each condition, Black applicants received a lower criterion than White applicants, all ts > 6.06, all ps < .001, all ds > .29.[9] There were no reliable differences in sensitivity between Black and White

9. In each condition, a mixed-model ANOVA including task order on criterion showed a reliable main effect of race, Fs > 35.64, ps < .001, η^2_p > .074. The Favor Black, Favor White, and Be Fair conditions did not show a reliable race by order interaction, Fs < 1.50, ps > .129, whereas the Control condition did, $F(11, 424) = 1.83$, $p = .047$, $\eta^2_p = .045$. Within every order in the Control condition, Black applicants received a lower criterion than White applicants.

applicants in any condition, all *t*s < 1.58, all *p*s > .116, all *d*s < .07. See Table 1 for the criterion and sensitivity means and standard deviations for Black and White applicants in each condition as well as test statistics and effect sizes for comparisons between each race.

The *Don't Favor Blacks* condition had a reliably smaller pro-Black criterion bias than the *Control* condition, $t(890) = 2.61$, $p = .009$, $d = .17$, 95% CI [.04, .30]. However, relative to the *Control* condition, telling participants to *Be Fair* without mentioning race did not reduce the racial difference in criterion bias, $t(871) = .569$, $p = .570$, $d = .04$, 95% CI [-.09, .17], and telling participants *Don't Favor Whites* did not increase the bias, $t(860) = .81$, $p = .419$, $d = .06$, 95% CI [-.08, .19].

Awareness of Selection Bias. The experimental conditions differed on the frequency with which participants indicated that they *had treated* both Black and White applicants equally, $\chi^2(3, N = 1{,}692) = 12.07$, $p = .007$. Participants in the *Don't Favor Blacks* (80.8%) and *Be Fair* (80.9%) conditions were more likely to indicate having treated both races equally than participants in the *Don't Favor Whites* condition (72.5%), though no conditions showed reliable differences from the *Control* condition (75.7%). Across conditions, among participants who indicated that they *had treated* both Black and White applicants equally (77.5%), Black applicants ($M = -0.09$, $SD = .49$) received a lower criterion than White applicants ($M = 0.13$, $SD = .45$), $t(1{,}308) = 14.10$, $p < .001$, $d = .39$, 95% CI [.33, .45]. As in the full sample, participants who reported treating both races equally in the *Don't Favor Blacks* condition had a smaller criterion bias ($M = 0.13$, $SD = .55$) than similar participants in the *Don't Favor Whites* ($M = 0.27$, $SD = .59$), *Be Fair* ($M = 0.25$, $SD = .55$), or *Control* ($M = 0.23$, $SD = .56$) conditions, $F(3, 1{,}308) = 4.13$, $p = .006$, $\eta^2 = .01$, 95% C.I. [.001, .02].

There were no reliable differences across experimental conditions in reporting a *desire* to treat both Black and White applicants equally, $\chi^2(3, N = 1{,}692) = 4.83$, $p = .203$. Across conditions, among participants (92.1%) who indicated a *desire* to treat Black and White applicants equally, Black applicants ($M = -0.10$, $SD = .48$) received a lower criterion than White applicants ($M = 0.13$, $SD = .45$), $t(1{,}556) = 15.96$, $p < .001$, $d = .40$, 95% CI [.35, .46]. As in the full sample, participants who reported a desire to treat both races equally in the *Don't Favor Blacks* condition had a smaller criterion bias ($M = 0.15$, $SD = .54$) than participants in the *Don't Favor Whites* ($M = 0.28$, $SD = .58$), *Be Fair* ($M = 0.24$, $SD = .56$), or *Control* ($M = 0.24$, $SD = .54$), conditions $F(3, 1{,}553) = 3.76$, $p = .010$, $\eta^2 = .01$, 95% C.I. [.0004, .02].

Racial Attitude Relations with Selection Decisions. BIAT *D* scores indicated pro-White attitudes ($M = 0.19$, $SD = .49$), $t(1{,}529) = 15.09$, $p < .001$, $d = .39$, 95% CI [.33, .44]. The explicit preference item also revealed pro-White attitudes ($M = 0.34$, $SD = .65$), $t(1{,}684) = 21.36$, $p < .001$, $d = .52$, 95% CI [.47, .57]. Among participants who reported no explicit preference for White or Black people (66.6%), Black applicants ($M = -0.13$, $SD = .49$) received a lower criterion than White applicants ($M = 0.15$, $SD = .44$), $t(1{,}122) = 17.34$, $p < .001$, $d = .52$, 95% CI [.46, .58]. Curiously, participants who reported having no racial preference in the *Don't Favor Whites* condition had a larger criterion bias ($M = 0.37$, $SD = .56$) than participants in the *Don't Favor Blacks* ($M = 0.23$, $SD = .54$), *Be Fair* ($M = 0.29$, $SD = .56$), or *Control* ($M = 0.27$, $SD = .56$), conditions $F(3, 1{,}119) = 2.86$, $p = .036$, $\eta^2 = .01$, 95% C.I. [.0001, .02]. However, the effect was not large and the confidence interval was tight and close to 0. There

TABLE 1. Study 5 Sensitivity and Criterion by Race for Each Condition

Condition	*Black c (SD)*	*White c (SD)*	*t*	*p*	*d [95% CI]*
Control	-.11 (.46)	.14 (.44)	9.67	< .001	.46 [.36, .56]
Be Fair	-.08 (.49)	.16 (.42)	8.74	< .001	.42 [.32, .52]
Don't Favor Whites	-.16 (.49)	.13 (.48)	9.91	< .001	.48 [.38, .58]
Don't Favor Blacks	-.06 (.49)	.10 (.44)	6.18	< .001	.29 [.20, .38]
Condition	***Black d' (SD)***	***White d' (SD)***	***t***	***p***	***d [95% CI]***
Control	1.11 (.66)	1.14 (.64)	.95	.343	.05 [-.05, .14]
Be Fair	1.12 (.66)	1.17 (.64)	1.47	.142	.07 [-.02, .16]
Don't Favor Whites	1.10 (.65)	1.13 (.65)	.98	.327	.05 [-.05, .14]
Don't Favor Blacks	1.09 (.65)	1.14 (.66)	1.58	.116	.07 [-.02, .17]

Note. c = criterion, *d'* = sensitivity, *d* = Cohen's *d*.

were no reliable differences across experimental conditions in reporting no explicit preference for White or Black people, $\chi^2(3, N = 1{,}685) = 4.83$, $p = .203$.

Across all conditions, criterion bias was reliably and negatively correlated with explicit racial preferences ($r = -.17$, $p < .001$, 95% CI [-.22, -.13]), BIAT *D* scores ($r = -.12$, $p < .001$, 95% CI [-.17, -.07]), perceptions of performance ($r = -.16$, $p < .001$, 95% CI [-.21, -.11]), and desired performance ($r = -.07$, $p = .002$, 95% CI [-.12, -.03]).

A simultaneous linear regression with implicit and explicit attitudes, and perceived and desired performance, predicting race differences in criterion bias revealed that both explicit ($\beta = -.15$, $p < .001$) and implicit ($\beta = -.09$, $p = .001$) attitudes and perceived ($\beta = -.13$, $p < .001$) task performance contributed uniquely, while desired task performance ($\beta = -.02$, $p = .420$) did not. Overall, those four variables accounted for 5.7% of the racial difference in criterion bias.

DISCUSSION

An instruction to avoid favoring Black applicants reduced, but did not eliminate, the criterion bias difference favoring Black over White applicants. Instructions to be fair without mentioning race and to avoid favoring Whites had no impact on racial differences in criterion bias. Moreover, the instructions had little impact on participants' perceptions that they behaved in a biased manner. The only reliable result was ironic. Warning people to avoid favoring Whites made them less confident that they did so, even though they were actually favoring Blacks. Moreover, the instruction had no effect on actual performance, a finding that is consistent with previous work, which found that pro-Black biases in judgment were not affected by increased accountability (i.e., making participants believe they would need to justify their decisions afterwards) or through pre-commitment (i.e., asking participants to indicate which qualifications were most important beforehand; Norton et al., 2004). Study 5 results suggest that merely asking participants to adopt a certain strategy is insufficient to remove the pro-Black criterion bias, and reinforce the evidence that this bias occurs, in part, unintentionally and without awareness.

In Study 6, we tried a different approach to determine whether participants could act according to their intentions to be unbiased. Rather than instructions warning of possible bias, we offered a reward to those participants who could achieve high levels of task accuracy.

STUDY 6

METHOD

Participants

We sought to collect enough participants to have 80% power for detecting a small effect size of d = .20 between conditions on the academic decision-making task (788 participants; 394 per condition). Studies at Project Implicit are posted and removed at a fixed time every week, which resulted in a larger than planned sample. Nine hundred and twenty-two White participants completed at least the decision-making task through the Project Implicit research pool (781 participants completed all measures).[10] Among those who provided data, 65.3% were female and the mean age was 37.1 (SD = 14). This sample allowed for 85% power at detecting a between-subjects effect size of d = .20 between experimental conditions, and nearly 100% power for d = .50 and d = .80.

Procedure

After providing consent, participants completed the following measures in fixed order: academic decision-making task, questionnaire assessing perceived and desired performance on the task, single-item measure of racial preferences, and the BIAT. Participants were then debriefed and given feedback on their BIAT performance.

All study elements were the same as those in Study 3. The only change came in the experimental manipulation. Participants were randomly assigned to a *Control* or *Charity* condition. Participants in the *Control* condition completed all measures in the same order as Study 3. Participants in the *Charity* condition saw additional instructions immediately before the testing phase of the academic decision-making task. The instructions noted that participants in the top 10% of accuracy for selecting the most qualified and rejecting the least qualified applicants would earn a $15 donation to the charity of their choosing. Participants then selected a charity from a list of the 20 of the highest-rated charities according to the American Institute of Philanthropy. Participants were reminded to try their best to be accurate so that their charity would receive the $15.

In the debriefing of the *Charity* condition, participants were provided with a link that reported how much money was donated to each charity. Once data collection

10. The study had 1,204 started sessions, with 922 providing data, and 781 completing the study (64.9% completion rate).

was complete, we updated the link with the amount donated to each charity based on participant performance.

RESULTS

Forty-three participants (4.7%) were excluded from the analyses for accepting less than 20% or more than 80% of the applicants, or for accepting or rejecting all applicants from either race. BIAT data from an additional 12 participants were excluded for having more than 10% of BIAT trials with response latencies below 400 ms.

Task accuracy (M = 69.2%, SD = 8.2) and acceptance rates (M = 50.9%, SD = 12.4) were similar to those in earlier studies.

Racial Bias in Selection for Honor Society. In the *Control* condition, Black applicants (M = -0.15, SD = .50) received a lower criterion than White applicants (M = 0.15, SD = .49), $t(454) = 11.98$, $p < .001$, $d = .56$, 95% CI [.46, .66]. In the *Charity* condition, Black applicants (M = -0.18, SD = .49) also received a lower criterion than White applicants (M = 0.11, SD = .49), $t(423) = 10.62$, $p < .001$, $d = .52$, 95% CI [.41, .62]. The two conditions were not reliably different in racial difference in criterion bias, $t(877) = .36$, $p = .721$, $d = .02$, 95% CI [-.11, .16].[11]

For sensitivity, there were no reliable differences between White (Control: M = 1.21, SD = .62; Charity: M = 1.17, SD = .66) and Black (Control: M = 1.16, SD = .64; Charity: M = 1.14, SD = .62) applicants in either the *Control*, $t(454) = 1.28$, $p = .203$, $d = .06$, 95% CI [-.03, .15], or the *Charity* condition, $t(423) = .84$, $p = .403$, $d = .04$, 95% CI [-.05, .14].

Awareness of Selection Bias. There were no reliable differences between the *Control* (78.4%) and *Charity* (76.9%) conditions in reporting a desire to treat both Black and White applicants equally, $\chi^2(1, N = 861) = .28$, $p = .596$. Among participants who indicated that they *had treated* both Black and White applicants equally, Black applicants (M = -0.15, SD = .49) received a lower criterion than White applicants (M = 0.11, SD = .49), $t(668) = 12.57$, $p < .001$, $d = .49$, 95% CI [.41, .57], with no reliable differences in the size of this effect between conditions, $t(667) = .36$, $p = .716$, d = .03, 95% CI [-.12, .18].

There were no reliable differences between the *Control* (89.3%) and *Charity* (89.6%) conditions in reporting a *desire* to treat both Black and White applicants equally, $\chi^2(1, N = 860) = .02$, $p = .876$. Among participants who indicated a desire to treat both applicant races equally (89.4%), Black applicants (M = -0.16, SD = .49) received a lower criterion than White applicants (M = 0.12, SD = .48), $t(768) = 14.46$, $p < .001$, $d = .52$, 95% CI [.45, .60], with no reliable differences in the size of this effect between conditions, $t(767) = .04$, $p = .970$, $d = 0$, 95% CI [-.11, .12].

Racial Attitude Relations with Selection Decisions. BIAT D scores indicated pro-White attitudes (M = 0.20, SD = .48), $t(756) = 11.50$, $p < .001$, $d = .42$, 95% CI [.34,

11. In each condition, a mixed-model ANOVA including task order on criterion showed a reliable main effect of race, Fs > 125.42, ps < .001, $\eta^2_p > .233$. Both the control and experimental conditions also showed significant race by order interactions, Fs > 2.14, ps < .017, $\eta^2_p > .051$. In all orders within the control condition and 11 of the 12 orders in the experimental condition, Black applicants received lower criterion than White applicants.

TABLE 2. Racial Criterion for Each Level of Perceived Task Performance

Perceived Performance	*N*	*Black c (SD)*	*White c (SD)*	*t*	*p*	*d [95% CI]*
Extremely easier on Black applicants	41	-.20 (.63)	.20 (.64)	3.07	.004	.48 [.17, .80]
Moderately easier on Black applicants	102	-.20 (.53)	.26 (.46)	9.28	< .001	.92 [.69, 1.15]
Slightly easier on Black applicants	580	-.22 (.44)	.19 (.43)	20.70	< .001	.86 [.76, .95]
Treated both races equally	3281	-.11 (.48)	.12 (.45)	24.45	< .001	.43 [.39, .46]
Slightly easier on White applicants	187	-.02 (.51)	-.01 (.45)	.24	.813	.02 [-.13, .16]
Moderately easier on White applicants	41	.09 (.49)	.03 (.42)	-.61	.547	-.09 [-.40, .21]
Extremely easier on White applicants	13	.34 (.60)	-.29 (.68)	-2.27	.043	-.63 [-1.21, .02]

Note. c = criterion, *d* = Cohen's *d*.

.49]. The explicit preference item also revealed pro-White attitudes (M = 0.36, SD = .72), $t(857) = 14.62$, $p < .001$, d = .50, 95% CI [.43, .57]. There were no reliable differences between the *Control* (63.3%) and *Charity* (64.5%) conditions in reporting no preference between White and Black people, $\chi^2(1, N = 858) = .14$, $p = .714$. Among participants who reported no explicit preference for White or Black people, Black applicants (M = -0.18, SD = .48) received a lower criterion than White applicants (M = 0.11, SD = .48), $t(547) = 13.16$, $p < .001$, d = .56, 95% CI [.47, .65], with no reliable differences in the size of this effect between conditions, $t(546) = .06$, $p = .951$, d = .01, 95% CI [-.16, .17].

Across both conditions, criterion bias was reliably and negatively correlated with explicit racial preferences (r = -.11, p = .002, 95% CI [-.17, -.04]), BIAT D scores (r = -.14, $p < .001$, 95% CI [-.20, -.07]), perceptions of performance (r = -.17, $p < .001$, 95% CI [-.23, -.11]), and desired performance (r = -.15, $p < .001$, 95% CI [-.22, -.09]).

A simultaneous linear regression with implicit and explicit attitudes, and perceived and desired performance, predicting race differences in criterion bias revealed that implicit (β = -.12, p = .001) attitudes and perceived (β = -.11, p = .004) and desired (β = -.14, $p < .001$) task performance contributed uniquely, while explicit attitudes (β = -.06, p = .117) were not reliably related to differences in criterion. Overall, those four variables accounted for 6.5% of the racial difference in criterion bias.

DISCUSSION

Providing participants with an incentive to perform accurately on the task did not lessen the pro-Black criterion bias. Although our manipulation did not reward participants themselves with money for high task accuracy, we donated $15 to a charity of the participant's own choosing if task accuracy was in the top 10% of participants in the experimental condition. Despite this incentive to reduce bias on

TABLE 3. Racial Criterion for Each Level of Desired Task Performance

Desired Performance	*N*	*Black c (SD)*	*White c (SD)*	*t*	*p*	*d [95% CI]*
Extremely easier on Black applicants	21	-.23 (.71)	.14 (.82)	1.56	.135	.34 [-.10, .78]
Moderately easier on Black applicants	53	-.27 (.49)	.20 (.52)	5.84	< .001	.80 [.49, 1.11]
Slightly easier on Black applicants	280	-.19 (.48)	.23 (.42)	12.98	< .001	.78 [.64, .91]
Treat both races equally	3853	-.12 (.48)	.12 (.45)	27.70	< .001	.45 [.41, .48]
Slightly easier on White applicants	34	.30 (.51)	-.01 (.50)	-2.83	.008	-.49 [-.84,- .13]
Moderately easier on White applicants	5	.20 (.69)	-.30 (.45)	-3.04	.039	-1.36 [-2.58, -.06]
Extremely easier on White applicants	3	.36 (.87)	-.39 (1.0)	-.70	.557	-.40 [-1.55, .83]

Note. c = criterion, d = Cohen's d.

the task, participants in the experimental condition were no more accurate or less influenced by race than control participants. The criterion bias was again reliably correlated with implicit and explicit attitudes, as well as perceived performance and desired performance, though participants in both conditions who wanted to be unbiased, believed they were unbiased, and had no preference between Whites and Blacks still exhibited a pro-Black criterion bias.

META-ANALYSIS OF ALL STUDIES

Across all eligible participants from all studies, Black applicants (M = -0.12, SD = .48) receiver a lower criterion than White applicants (M = 0.12, SD = .45), $t(4,359) = 29.95$, $p < .001$, d = .45, 95% CI [.42, .48]. There was also a small but reliable effect on sensitivity; White applicants (M = 1.16, SD = .64) received a higher sensitivity than Black applicants (M = 1.09, SD = .65), $t(4,359) = 7.02$, $p < .001$, d = .11, 95% CI [.08, .14]

We also combined results from Studies 2–6 to examine how performance on the decision-making task was related to perceived and desired performance. See Table 2 for means and standard deviations of criterion for Black and White applicants for each level of perceived performance and Table 3 for each level of desired performance. While most participants (77.3%) indicated that they *had treated* Black and White applicants equally, these participants showed a significant pro-Black bias on the task, $t(3,280) = 24.49$, $p < .001$, d = .43, 95% CI [.39, .46]. Moreover, participants who indicated they were slightly easier on White and tougher on Black applicants (4.4%; n = 187) showed no reliable differences in criterion between Black (M = -0.02, SD = .51) and White applicants (M = -0.01, SD = .45), $t(186) = .24$, p = .813, d = .02, 95% CI [-.13, .16]. Even participants who indicated they were moderately easier on White and tougher on Black applicants (1.0%, n = 41) showed no reliable

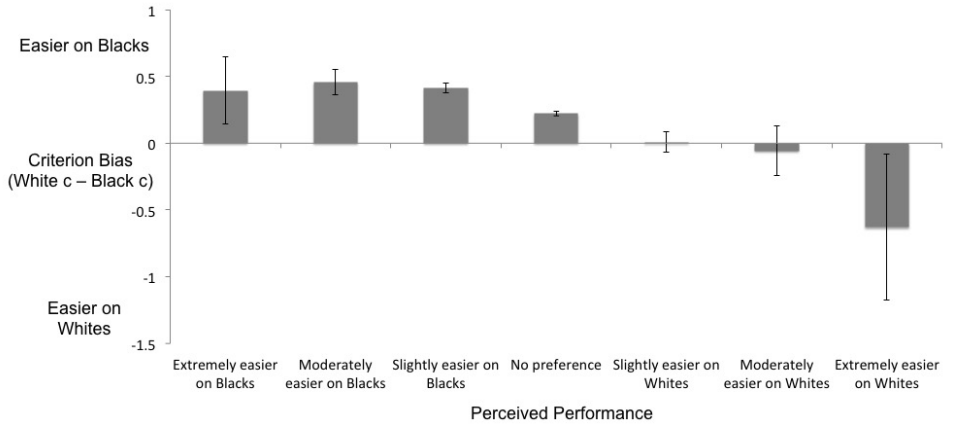

FIGURE 1. Criterion bias for each level of perceived performance across all eligible participants. Error bars denote 95% confidence intervals on the means.

difference in criterion between Black (M = 0.09, SD = .49) and White applicants (M = 0.03, SD = .42), $t(40)$ = -.61, p = .547, d = -.09, 95% CI [-.40, .21], though relatively few participants selected this option. See Figure 1 for a graphical display of the criterion bias for each level of perceived performance.

Similarly, while most (90.7%) participants indicated a *desire* to treat Black and White applicants equally, these participants had lower criteria for Black (M = -0.12, SD = .48) than White (M = 0.12, SD = .45) applicants, $t(3,852)$ = 27.70, $p < .001$, d = .45, 95% C.I. [.41, .48]. Unlike in perceived performance, participants who indicated a desire to slightly favor Whites over Blacks on the task (0.8%; n = 34) did show a lower criterion for White (M = -0.01, SD = .50) than Black (M = 0.30, SD = .51) applicants, $t(33)$ = -2.83, p = .008, d = -.49, 95% C.I. [-.84, -.13], but very few participants selected this response. See Figure 2 for a graphical display of criterion bias for each level of desired performance.

Finally, on suggestion of peer reviewers, we conducted an analysis of the relationship between the criterion bias and political orientation. We did not collect political orientation data in Studies 1 and 2 or in the Mechanical Turk sample in Study 4. All Project Implicit participants reported political orientation on a 7-point scale (-3 = Strongly conservative, 0 = Moderate/neutral, +3 = Strongly liberal) when they first registered at the site. We analyzed the relationship between the criterion bias and political orientation in a sample that included all PI participants in Studies 3–6. This analysis collapsed across experimental conditions in Studies 5 and 6, as we found no reliable interaction between political orientation and experimental condition in Study 5, $F(18, 1{,}639)$ = 1.00, p = .452, η^2p = .01, or Study 6, $F(6, 829)$ = 1.03, p = .402, η^2_p = .01 (see https://osf.io/tvm83 for full reporting of these analyses).

Criterion bias was positively and reliably correlated with political orientation (r = .12, $p < .001$, 95% C.I. [.09, .15]), such that higher levels of liberalism were associated with a larger pro-Black criterion bias. However, within every level of political orientation, there was a reliable pro-Black criterion bias, all ts > 3.89, all ps

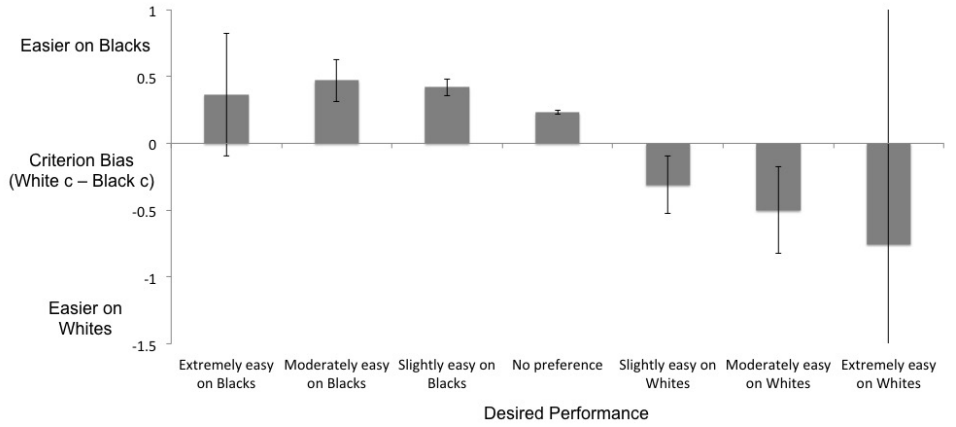

FIGURE 2. Criterion bias for each level of desired performance across all eligible participants. Error bars denote 95% confidence intervals on the means. The error bars in the "Extremely easy on Whites" are truncated due to a small sample size (n = 3).

< .001, all ds > 0.29. See Table 4 for means and standard deviations of criterion for Black and White applicants at each level of political orientation and Figure 3 for a graphical display of criterion bias for each level of political orientation.

A simultaneous linear regression with political orientation, implicit and explicit attitudes, and perceived and desired performance, predicting race differences in criterion bias revealed that political orientation (β = .07, p < .001), implicit attitudes (β = -.10, p < .001), explicit attitudes (β = -.11, p < .001), perceived performance (β = -.14, p < .001), and desired performance (β = -.07, p < .001) contributed uniquely. Overall, those five variables accounted for 7.2% of the racial difference in criterion bias.

GENERAL DISCUSSION

Across a variety of large, heterogeneous samples, White participants set a lower criterion for Black than White candidates when making accept and reject decisions for a hypothetical academic honor society. This criterion bias was present among participants who reported treating applicants from both races equally, who reported a desire to treat applicants from both races equally, and who reported having no explicit preferences between White and Black people. Moreover, the bias was lessened, but still present, even after warning participants about the likelihood of favoring Black candidates, and showed no changes after offering an incentive for participants to perform more accurately on the task.

To the extent that participants desire to show a racial bias on this decision-making task, they can do so. For example, if a participant adopted the criterion of accepting Black candidates and rejecting White candidates, then they could easily show a “perfect” bias in selection. Across studies, explicit attitudes, desires, and self-perceptions all suggest some intentionality influences in judgment by race. Larger pro-Black bias was related to: (1) weaker pro-White attitudes (Studies 3–6),

TABLE 4. Racial Criterion for Each Level of Political Orientation

Political orientation	*N*	*Black c (SD)*	*White c (SD)*	*t*	*p*	*d [95% CI]*
Strongly conservative	92	-.07 (.56)	.14 (.44)	3.89	< .001	.41 [.19, .62]
Moderately conservative	324	-.07 (.48)	.09 (.45)	5.23	< .001	.29 [.18, .40]
Slightly conservative	244	-.04 (.50)	.14 (.44)	5.34	< .001	.34 [.21, .47]
Moderate/neutral	1073	-.10 (.48)	.09 (.46)	12.08	< .001	.37 [.31, .43]
Slightly liberal	380	-.10 (.47)	.14 (.44)	9.69	< .001	.50 [.39, .60]
Moderately liberal	1049	-.15 (.48)	.15 (.44)	18.68	< .001	.58 [.51, .64]
Strongly liberal	738	-.18 (.48)	.18 (.46)	17.96	< .001	.66 [.58, .74]

Note. c = criterion, d = Cohen's d.

(2) stronger liberalism (see Meta-Analysis section), (3) a greater desire to favor and greater perception of having favored Black applicants (Studies 3–6), (4) higher levels of internal motivation to control prejudice (Studies 2 & 4), and (5) greater support of affirmative action (Study 4). The pro-Black bias was also reduced somewhat following instructions to not favor Black applicants (Study 5).

Our central question was whether the racial biases observed in this social judgment task were fully under conscious control. The evidence is clear that they were not. The bias was still present in the more than 75% of participants who reported a desire to show no racial bias and among the more than 90% who perceived that they showed no racial bias. Even instructing participants to not favor Black applicants did not eliminate the bias or shift the criterion to be pro-White. Further, on average, implicit and explicit attitudes revealed greater preference and more positive associations for Whites compared to Blacks. Despite these pro-White attitudes, behavior on the task suggested more favorable treatment of Blacks.

The decision-making paradigm used to measure the bias had a number of advantages. First, performance was an accumulation of many decisions instead of a single-shot assessment of one or two candidates, as is common in this literature. As such, the paradigm provided for a relatively reliable behavioral measure (across all eligible participants, $\alpha = .71$ for criterion on White candidates, $\alpha = .73$ for criterion on Black candidates, and $\alpha = .55$ for the criterion bias difference score). Moreover, the task elicited robust biases in social judgment (overall Cohen's $d = .45$), and in other research applications elicits social biases that are more consistent with attitudes and stereotypical expectations (Axt et al., 2015).

In addition, the task elicited a dissociation between bias awareness and intention; 23.7% of participants reported differing levels of perceived and desired task performance. Many participants perceived that they were not able to perform on the task in a manner that they desired to. This suggests that it is a relatively challenging task in productive ways for a variety of research uses. Finally, whereas previous studies used only a single decision and therefore could not indicate whether bias existed in any one participant (e.g., Norton, Sommers, Vandello, & Darley,

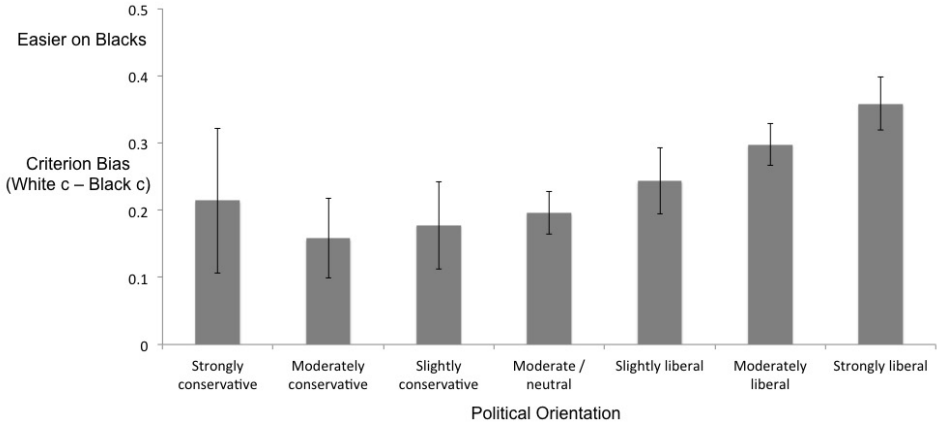

FIGURE 3. Criterion bias for each level of political orientation across all eligible participants. Error bars denote 95% confidence intervals on the means.

2006), the multiple decisions made among Black and White qualified and unqualified candidates in this paradigm allows for an estimate of racial bias in a single participant, and offered a clear means of identifying that no bias was present.

These results are not obviously anticipated by most models of unintentional or automatic effects on social judgment, where attitudes shape the direction of behavior, unless a corrective process intervenes (Cunningham et al., 2007; Devine, 1989; Dovidio et al., 2002; Fazio et al., 1995; Mendes et al., 2002). Here, the pro-Black behavior was in the opposite direction of participants' pro-White attitudes, and most participants (77.3%) did not report any conscious effort of altering their behavior to favor one race over another. From this perspective, the results are enigmatic. Based on existing theories of attitudes, it is difficult to explain how a pro-Black behavior emerged that ran contrary to explicit and implicit attitudes, and, for most participants, did not appear to be a result of consciously altering behavior so as to favor Blacks.[12]

DELIBERATE, JUST NOT REPORTED?

Other research shows that Blacks are disadvantaged compared to Whites in a variety of aspects of social life (e.g., Mitchell, Haw, Pfeifer, & Meissner, 2005; Stanley,

12. Blanton and colleagues (Blanton & Jaccard, 2006; see also Blanton, Jaccard, Strauts, Mitchell, & Tetlock, 2015) suggest that the racial attitude IAT does not have a rational zero point and overestimates pro-White bias (but see Greenwald, Nosek, & Sriram, 2006). From this perspective, there could actually be a strong pro-Black implicit bias that is directionally the same as the social judgment bias found here. However, explicit pro-White preferences are nonetheless pro-White and would be in contrast to both implicit preferences and the social judgment bias. Asserting that the explicit preference measure is also biased requires concluding that answering "I do not have a preference between Whites and Blacks" does not, in fact, indicate no explicit preference between Whites and Blacks. Regardless of one's position on this issue, implicit and explicit preferences cannot fully account for the observed racial bias on the social judgment task, for which the rational zero point of no bias is unambiguous.

Sokol-Hessner, Banaji, & Phelps, 2011). Given this, many participants may have believed that favoring Black over White applicants *is* the same as treating White and Black applicants equally. That is, participants may have consciously decided to lower their criterion for Black compared to White applicants, but reported showing no racial preference because they believed that strategy was most fair. This perspective would suggest that the pro-Black criterion bias did not necessarily occur outside of conscious awareness but simply appeared that way given the wording of our measures.

Another possibility is participants may have deliberately evaluated lower qualifications among Black applicants not as indicators of lesser academic ability, but as partly the byproduct of existing structural disadvantages against Blacks. For example, participants may have attributed lower interview scores among Black applicants not as the result of a poor interview but as the result of anti-Black bias by interviewers (e.g., Biernat & Eidelman, 2007). This account would mirror the results of Behaghel, Crépon, and Le Barbanchon (2015), who found that negative behaviors, such as long gaps in employment, were interpreted differently when attached to an anonymous resume (e.g., the result of the applicant's poor reliability) compared to a resume containing a minority name (e.g., the result of adverse economic conditions in disadvantaged neighborhoods).

Neither of these appears to be a viable account of our results as deliberate actions. For example, we would expect participants who deliberately favored Blacks but reported treating both races equally to support affirmative action in Study 2 and Study 4. However, among participants who indicated they had treated both applicants equally, wanted to treat both applicants equally, and opposed affirmative action policies (Study 2: $N = 67$; Study 4: $N = 328$), Black applicants still received a lower criterion (Study 2: $M = -0.02$, $SD = .40$; Study 4: $M = -0.10$, $SD = .48$) than White applicants (Study 2: $M = 0.10$, $SD = .38$; Study 4: $M = 0.07$, $SD = .46$), at about the same magnitude as the rest of the sample (Study 2: $t(66) = 2.10$, $p = .039$, $d = .26$, 95% CI [.01, .50]; Study 4: $t(327) = 5.72$, $p < .001$, $d = .32$, 95% CI [.20, .43]). The fact that the pro-Black bias occurred even among those participants who simultaneously opposed affirmative action, wanted to be unbiased, and believed they were so is further evidence that the racial difference in criterion bias can occur without intention or awareness.

RELEVANCE TO PROMINENT THEORETICAL ACCOUNTS OF RACIAL JUDGMENTS

The most obvious potential connection between our results and existing theoretical accounts is to the shifting standards model of stereotype-based judgments (Biernat, 2003; Biernat & Manis, 1994). Shifting standards refers to adjusting the relative meaning of criteria when assessing members of different groups. In the present case, participants may not have judged all applicants against a common standard, but instead created different standards for Blacks and Whites. Individual applicants would then be judged relative to an expectation for members of

that race rather than a shared expectation. This lower standard for Black applicants could have made their qualifications look more impressive than their White counterparts, leading to a lower acceptance criterion for Black applicants despite having equal qualifications as Whites (see Biernat & Fuegen, 2001, for a similar process occurring with gender).

This is an appealing explanation, and one that appears to align with our findings. Previous studies have also indicated that shifting standards may be an automatic process. For instance, gender-based shifting standards effects were shown to be stronger when under cognitive load, indicating that conscious processes only weakened rather than created the effect (Biernat, Kobrynowicz, & Weber, 2003). Our results further highlight how drastically such effects can occur without awareness or control.

The pro-Black criterion bias was related to but largely distinct from one's conscious goals or motivation. Even after making 60 judgments of Whites and Blacks in sequence, this race-based reweighting of applicants was not obvious to most participants. Furthermore, a pro-Black criterion bias was also observed among conservatives and those that opposed affirmative action policies, individuals likely to have the strongest motivation to not show a criterion bias. Participants, regardless of their explicit attitudes or political beliefs, appeared mostly unaware that they may have been comparing Black applicants to a subjectively different standard than White applicants. If shifting standards is the underlying mechanism, then the impact of shifting standards may be more pervasive than presently understood.

This decision-making paradigm may provide another measure of individual differences in shifting standards. Recent work has also attempted to measure a tendency to shift standards by comparing performance on objective versus subjective response scales (Biernat, Collins, Katzarska-Miller, & Thompson, 2009). For instance, participants predicted similar scores for White and Black applicants' ACT scores on a subjective scale ("Very poorly" to "Very well"), but the same participants predicted greater scores for Whites than Blacks on an objective scale (i.e., a numerical score). Similarly, our paradigm suggests that individual differences in shifting standards may also be assessed by first providing participants with relatively objective evaluation criteria (e.g., GPA), and seeing where they then create subjective admission standards for each race.

Finally, our results show that biases in criterion were weakly but reliably related to both implicit (r = -.15) and explicit (r = -.17) racial attitudes, such that greater pro-White implicit and explicit attitudes were associated with weaker pro-Black criterion judgments. Earlier research on shifting standards observed no reliable relationship with implicit or explicit race attitudes (Biernat et al., 2009). However, the present sample is over 20 times larger, perhaps providing a more precise estimate of a real, but relatively weak relationship.

Our findings also align with work on casuistry (Norton et al., 2004), wherein people engage in specious and misleading reasoning to justify difficult decisions. For example, individuals engage in casuistry when balancing the opposing goals of favoring members of stigmatized or under-represented groups while also maintaining a self-image as objective and unbiased. To resolve this conflict, people may

alter the relative importance of criteria to create a decision-making process that can both favor desired groups and preserve one's image as unbiased. In our studies, participants may have identified the particularly strong criterion that happened to be held by Black applicants, and then placed greater weight on that criterion for admission to the honor society. This way, participants might favor Blacks but also report having no racial preferences on the task, because they believed they would have likewise favored Whites if similar circumstances had emerged.

In many ways, for the present context, this casuistry explanation is similar to shifting standards. Pro-Black biases in judgment have been found in some casuistry research consistent with the present effects (e.g., Norton et al., 2006), and there is some evidence that casuistry can occur without intention or awareness (e.g., Lindner, Graser, & Nosek, 2014). However, earlier work found that when participants were allowed to make two consecutive decisions, each between a Black and a White candidate, a majority of participants (66%) showed no racial bias and elected to select one Black and one White candidate, regardless of qualifications (Norton et al., 2008). In our paradigm, the pro-Black bias in judgment was evident across many judgments of Black and White candidates. To clarify the viability of the casuistry explanation, future work could investigate whether this paradigm caused participants to abandon the "equal numbers" strategy used over two decisions, or if participants tried to accept the same amount of Black and White candidates but were unable to keep track given the larger number of trials.

Finally, the finding that this pro-Black bias occurred without conscious intention or awareness for most of our participants is related to previous work on implicit stereotype inhibition. Previous accounts of bias correction focus on effortful, conscious processes (e.g., Wegner & Petty, 1995, 1997), but more recent evidence has also suggested that bias correction can occur automatically. For instance, participants who wrote about a time where they failed to be egalitarian showed greater levels of stereotype inhibition on a reaction time measure than participants writing about a time where they were successfully egalitarian (Moskowitz & Li, 2011). Also, participants with a stronger implicit negative attitude toward prejudice, measured with an IAT pairing concepts of "prejudice" and "tolerant" with positive and negative words, were better able to inhibit stereotypes from influencing behavior (Glaser & Knowles, 2008). This literature suggests that bias correction can occur automatically and without awareness. In the present case, we cannot estimate whether the effects are an active corrective process. If they are, they provide a particularly strong example as the paradigm elicited strong, reliable outgroup favoritism that was opposing the participants' implicit and explicit racial attitudes.

DIRECTIONS FOR FUTURE RESEARCH

These data provide substantial evidence, and suggest multiple directions for additional research. First, though we used several heterogeneous samples, none of them are representative of any identifiable population. The strength or direction of

the criterion bias may differ among more representative samples or in non-White participants. However, because the effects were consistent among both conservatives and liberals in our samples, we predict that the bias will be quite pervasive. Second, the paradigm can be adapted to investigate the moderating conditions of eliciting pro-Black or pro-White biases in criterion judgments. For example greater pro-White biases may be observed by introducing more ambiguity into the materials (e.g., Dovidio & Gaertner, 2000), by informing participants that their decisions may have real-world impacts on scarce resources (e.g., Hodson, Dovidio, & Gaertner, 2002), or by converting the decision task to an assessment of potential dating partners (Axt et al., 2015).

WHAT THE DATA DO NOT SHOW

These results demonstrated a reliable and apparently automatic pro-Black bias in decision making for an academic honor society selection paradigm. At minimum, and possibly at maximum, they suggest that the prevailing emphasis on pro-White biases in judgment and behavior in the existing literature would improve by refining the theoretical understanding of under what conditions behavior favoring dominant or minority groups will occur. These results do not counter evidence for bias—in any direction—in other research applications. Further, because the effects are capable of occurring outside of awareness and independently of attitudes, the present results suggest opportunity for theoretical innovation in how automatic processes shape behavior. But, the results do not suggest that the existing models are wrong for characterizing other effects that have been observed. Rather, the models would appear to be incomplete in not anticipating how a social judgment that is counter to one's attitudes might occur automatically.

CONCLUSION

The present research suggests that pro-Black behavior can occur outside of awareness, apparently without a goal to favor Blacks, and that cannot be accounted for by attitudes that tend to be held in opposition to that behavior, both implicitly and explicitly. This suggests that rather than focusing on questions such as "Why are people biased against Blacks?"; "Why do people favor dominant groups?"; and "Why are minority groups discriminated against?"; theoretical models would better account for human behavior if they addressed questions such as "Why are people biased?"; "Under what conditions do people favor dominant or disadvantaged groups?"; "Under what conditions do people discriminate against majority or minority groups?"; and "How do these processes occur outside of awareness or control?"

APPENDIX A

QUALIFIED AND UNQUALIFIED APPLICATION INFORMATION

Unqualified Applicants

Science GPA	Humanities GPA	Rec. Letters	Interview Score
3.6	3.7	Good	67.5
3.6	3.3	Excellent	52.5
3.7	3.5	Good	70
3.2	3.7	Good	77.5
3.8	3.1	Good	77.5
3.5	3.6	Good	72.5
3.0	3.3	Excellent	67.5
3.8	3.4	Good	70
3.1	3.4	Excellent	62.5
3.2	3.1	Good	92.5
3.1	3.5	Excellent	60
3.8	3.3	Good	72.5
3.5	3.7	Good	70
3.5	3.4	Good	77.5
3.5	3.9	Good	65
3.1	3.7	Good	80
3.4	3.7	Good	72.5
3.1	3.4	Good	87.5
3.2	3.3	Excellent	62.5
3.5	3.2	Good	82.5
3.3	3.2	Good	87.5
3.2	3.1	Excellent	67.5
3.9	3.2	Good	72.5
3.2	3.4	Good	85
3.4	3.4	Good	80
3.5	3.0	Excellent	62.5
3.6	3.1	Good	82.5
3.3	3.4	Good	82.5
3.3	3.4	Excellent	57.5
3.7	3.2	Good	77.5

APPENDIX A (continued)

Qualified Applicants

Science GPA	Humanities GPA	Rec. Letters	Interview Score
3.8	3.3	Good	97.5
3.4	3.9	Good	92.5
3.7	3.2	Excellent	77.5
3.8	3.0	Excellent	80
3.4	3.5	Excellent	77.5
3.2	3.7	Excellent	77.5
3.6	3.7	Excellent	67.5
3.9	3.3	Good	95
3.6	3.7	Good	92.5
3.3	3.6	Excellent	77.5
3.7	3.6	Excellent	67.5
3.5	3.7	Excellent	70
3.2	3.4	Excellent	85
3.8	3.6	Good	90
3.8	3.8	Good	85
3.1	3.2	Excellent	92.5
2.9	3.4	Excellent	92.5
3.6	3.4	Excellent	75
3.5	3.4	Excellent	77.5
3.3	3.7	Excellent	75
3.7	3.9	Good	85
3.3	3.2	Excellent	87.5
3.8	3.4	Good	95
3.5	3.5	Excellent	75
3.1	3.4	Excellent	87.5
3.8	3.7	Good	87.5
3.4	3.6	Excellent	75
3.7	3.8	Good	87.5
3.9	3.7	Good	85
3.9	3.8	Good	82.5

APPENDIX B

AFFIRMATIVE ACTION ITEMS USED IN STUDIES 2 AND 4

1. A corporate personnel officer is evaluating an African American and a European American job applicant who are identically qualified except the European American has more prior experience in related work. Is there a reasonable justification for this personnel officer hiring the African American applicant rather than the European American?
2. A college admissions officer considers applications from African American and European American applicants with similar credentials and cannot accept all. Should the admissions officer more often accept African American than European American applicants?

APPENDIX C

TEXTS FROM MANIPULATIONS IN STUDY 5

Be Fair Condition

Decision makers are frequently too easy on some applicants, and too tough on others.
We would like to see if people can be fair toward all applicants if they are told about this tendency beforehand.
When you have to make your "Accept" and "Reject" decisions, try to be as fair as possible.

Don't Favor Blacks Condition

Decision makers are frequently too easy on some applicants, and too tough on others. Prior research suggests that decision makers are much easier on the Black candidates and much tougher on the White candidates.
We would like to see if people can be fair toward all applicants if they are told about this tendency beforehand.
When you have to make your "Accept" and "Reject" decisions, try to be as fair as possible.

Don't Favor Whites Condition

Decision makers are frequently too easy on some applicants, and too tough on others. Prior research suggests that decision makers are much easier on the White candidates and much tougher on the Black candidates.
We would like to see if people can be fair toward all applicants if they are told about this tendency beforehand.
When you have to make your "Accept" and "Reject" decisions, try to be as fair as possible.

REFERENCES

Abrams, D., Bertrand, M., & Mullainathan, S. (2012). Do judges vary in their treatment of race? *Journal of Legal Studies, 41*, 347-383.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173-221). Mahwah, NJ: Erlbaum.

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion and age. *Psychological Science, 25*, 1804-1815.

Axt, J. R., Nguyen, H., & Nosek, B. A. (2015). Decision criterion as a measure of bias in behavior. Manuscript in preparation.

Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In D. T. Gilbert & S. T. Fiske (Eds.), *Handbook of social psychology* (pp. 353-393). New York: Wiley.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*, 668-688.

Behaghel, L., Crépon, B., & Le Barbanchon, T. (2015). Unintended effects of anonymous resumes. *American Economic Journal: Applied Economics, 7*(3), 1-27.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*, 991-1013.

Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist, 58*(12), 1019-1027.

Biernat, M., Collins, E. C., Katzarska-Miller, I., & Thompson, E. R. (2009). Race-based shifting standards and racial discrimination. *Personality and Social Psychology Bulletin, 35*, 16-28.

Biernat, M., & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology, 37*(6), 1149-1175.

Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707-724.

Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin, 36*, 855-868.

Biernat, M., Kobrynowicz, D., & Weber, D. L. (2003). Stereotypes and shifting standards: Some paradoxical effects of cognitive load. *Journal of Applied Social Psychology, 33*(10), 2060-2079.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*, 5-20.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27-41.

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *Journal of Applied Psychology, 100*(5), 1468-1481.

Cacioppo, J. T., & Freberg, L. A. (2013). *Psychology: The science of the mind*. Boston: Houghton Mifflin.

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*, 1314-1329.

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of Personality and Social Psychology, 108*, 219-233.

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

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Doleac, J. L., & Stein, L. C. D. (2013). The visible hand: Race and online market outcomes. *The Economic Journal, 123*, F469-F492.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*, 319-323.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1084-1121). New York: Wiley.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

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*, 510-540.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013-1027.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 97-116). New York: Guilford.

Feist, G. J., & Rosenberg, E. L. (2012). *Psychology: Perspectives and connections*. New York: McGraw-Hill.

Fiske, S. T., Gilbert, D. T., & Lindzey, G. (Eds.). (2010). *Handbook of social psychology* (5th ed.). New York: Wiley.

Galinsky, A. D., Hall, E. V., & Cuddy, A. C. J. (2013). Gendered races: Implications for interracial dating, leadership selection, and athletic recruitment. *Psychological Science, 24*, 498-506.

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kinfton, R. (2011). Race, ethnicity and NIH research awards. *Science, 333*, 1015-1019.

Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology, 44*(1), 164-172.

Goff, P. A., Steele, C. M., & Davies, P. G. (2008). The space between us: Stereotype threat and distance in interracial contexts. *Journal of Personality and Social Psychology, 94*, 91-107.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psycho-physics.* New York: Wiley. (Reprinted 1974, Huntington, NY: Krieger).

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test (IAT). *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A. G, Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test (IAT): I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test (IAT): Comment on the article by Blanton and Jaccard. *American Psychologist, 61*, 56-61.

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist, 69*, 669-684.

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 scale. *Political Psychology, 23*, 253-283.

Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin, 28*(4), 460-471.

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology, 43*(2), 334-340.

King, L. A. (2011). *The science of psychology*: *An appreciative view*. New York: McGraw-Hill.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27*, 623-634.

Lindner, N. M., Graser, A., & Nosek, B. A. (2014). Age-based hiring discrimination as a function of equity norms and self-perceived objectivity. *PLOS ONE, 9*, e84752.

doi: 10.1371/journal.pone.0084752

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics, 119*, 49-89.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A users guide.* Cambridge, England: Cambridge University Press.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test (IAT), discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435-442.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin, 28*(7), 939-952.

Mendes, W. B., & Koslov, K. (2013). Brittle smiles: Positive biases toward stigmatized and outgroup targets. *Journal of Experimental Psychology: General, 142*(3), 923.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology, 100*(6), 1678-1712.

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior, 29*, 621-637.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*(1), 167.

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology, 47*(1), 103-116.

Norton, M. I., Sommers, S. R., Vandello, J. A., & Darley, J. M. (2006). Mixed motives and racial bias: The impact of legitimate and illegitimate criteria on decision-making. *Psychology, Public Policy, and Law, 12*, 36-55.

Norton, M. I., Vandello, J. A., Biga, A., & Darley, J. M. (2008). Colorblindness and diversity: Conflicting goals in decisions

influenced by race. *Social Cognition, 26*, 102-111.

Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87*, 817-831

Nosek, B. A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science, 16*, 65-69.

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J. R., & Greenwald, A. G. (2014). Understanding and using the Brief Implicit Association Test (BIAT): Recommended scoring procedures. *PLoS ONE*. http://dx.plos.org/10.1371/journal.pone.0110938.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test (IAT): II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test (IAT) at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265-292). New York: Psychology Press.

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*, 36-88.

Pager, D., & Shepherd, D. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology, 34*, 181-209.

Passer, M. W., & Smith, R. E. (2011). *Psychology*: *The science of mind and behavior*. New York: McGraw Hill.

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*, 367-374.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. (2008). The White standard: Racial bias in leader categorization. *Journal of Applied Psychology, 93*, 758-77.

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2011). *Psychology*. New York: Worth.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test (BIAT). *Experimental Psychology, 56*, 283-294.

Stanley, D., Sokol-Hessner, P., Banaji, M., & Phelps, E. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences, 108*, 7710-7715.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253-273.

U.S. Department of Housing and Urban Development. (2013). *Housing discrimination against racial and ethnic minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.

Unzueta, M. M., Everly, B. A., & Gutiérrez, A. S. (2014). Social dominance orientation moderates reactions to Black and White discrimination claimants. *Journal of Experimental Social Psychology, 54*, 81-88.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73*(5), 941-959.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology, 68*(1), 36-51.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29*, 142-208.

Yzerbyt, V. Y., & Demoulin, S. (2010). Intergroup relations. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1024-1083). New York: Wiley.