British Journal of Educational Psychology (2017), 87, 408–421

The British Psychological Society

www.wileyonlinelibrary.com

An unintentional pro-Black bias in judgement among educators

Jordan R. Axt*

Department of Psychology, University of Virginia, Charlottesville, Virginia, USA

Background. Previous work indicates widespread preference for White over Black people in attitudes and behaviour. However, there are instances where Black people receive preferential treatment over White people.

Aims. This study aimed to investigate whether a sample of education professionals would favour Black or White applicants to an academic honour society, and the extent to which any biases were related to conscious intentions.

Sample. Participants were education professionals (N = 618; 75.5% White) who completed an online study.

Methods. Participants completed a hypothetical admissions task where they evaluated more and less qualified applicants for an academic honour society, and applicants were either White or Black. Participants also completed measures of implicit and explicit racial attitudes.

Results. Educational professionals at all levels showed a pro-Black bias in judgement, adopting a lower acceptance criterion for Black compared to White applicants, replicating previous work using online and undergraduate samples. The bias was present among participants reporting they did not want to be biased or believed they were unbiased, suggesting that bias arose without conscious awareness or intention. Bias was also weakly but reliably related to racial attitudes.

Conclusions. These findings are consistent with the notion that educators automatically hold lower standards for Black versus White applicants. While education professionals likely have experience evaluating students from different racial and ethnic backgrounds, these professionals were, nevertheless, unable to eliminate the impact of race in their decision-making.

Research in prejudice, stereotyping, and discrimination has emphatically shown that White people are treated more favourably than Black people across a variety of domains and contexts (Bertrand & Duflo, 2016; Bertrand & Mullainathan, 2004; Greenwald & Pettigrew, 2014; List, 2004). These same anti-Black biases appear to exist among educators as well. For example, teachers in one study handed out more severe punishments to hypothetical students given a stereotypically Black (Darnell or Deshawn) versus White (Greg or Jake) name (Okonofua & Eberhardt, 2015). These experimental results compliment real-world data finding that Black students are disproportionately given infractions, suspended and expelled from school (Fabelo *et al.* 2011; Gregory & Weinstein, 2008; Kirwan Institute, 2014).

*Correspondence should be addressed to Jordan R. Axt, Department of Psychology, University of Virginia, Box 400400, Charlottesville, VA 22904-4400, USA (email: jra3ee@virginia.edu).

DOI:10.1111/bjep.12156

Behavioural biases against Black people in general and students in particular may not be surprising given evidence that a majority of people hold attitudes favouring White over Black people (e.g., Axt, Ebersole, & Nosek, 2014; Nosek *et al.*, 2007), and such attitudes are presumed to be an important determinant of behaviour (Ajzen & Fishbein, 2005). These anti-Black attitudes are present both explicitly, when responses are controlled and within conscious awareness, and implicitly, where responses may be automatic and reflect unconscious associations. An example of an explicit attitude measure would be a survey question asking the extent to which participants prefer White to Black people, while an example of implicit attitude measure would be the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). In an IAT, participants categorize both words (e.g., positive or negative words) and images (e.g., Black and White people) as quickly as possible using two computer keys as they appear one at a time on a computer screen. In some blocks, participants must categorize items from all four sets of stimuli (positive words, negative words, Black people, and White people) using just two keys (e.g., categorizing White people and positive words with one key, Black people and negative words with the other key). In other blocks, the pairing between words and images is switched (e.g., categorizing Black people and positive words with one key, White people and negative words with the other key). The difference in average reaction time between responses in the two types of blocks is used to infer the strength of associations between those categories and attributes in memory. Categorizing items faster when Black people and negative words (and White people and positive words) are paired together compared with when Black people and positive words (and White people and negative words) are paired together indicates more positive implicit associations with White compared to Black people. Such implicit measures assess attitudes without requiring introspection, and often differ in strength or direction compared to more controlled explicit attitudes (Nosek, Banaji, & Greenwald, 2002).

While implicit and explicit attitudes favouring White over Black people have been consistently found in large online, national samples (e.g., Rae, Newheiser, & Olson, 2015), they also exist among educators specifically. In a 2015 volunteer sample completing an IAT assessing evaluations of Black and White people, primary, secondary, and special education teachers (N = 2,796) had small pro-White explicit (Cohen's d = .28) and large pro-White implicit (d = 1.59) attitudes (data from Xu, Nosek, & Greenwald, 2014). These data support previous work suggesting that educators hold implicit and explicit attitudes that are much like those of the general public, such as in favouring thin over obese people (O'Brien, Hunter, & Banks, 2007) or native versus immigrant populations (Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010).

Although attitudes and behaviours favouring White over Black people is the norm in experimental research (e.g., Dovidio & Gaertner, 2000; Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Goff, Steele, & Davies, 2008), there exist several notable exceptions. For instance, White participants in one study reported greater liking for a Black than for a White target after each complained about experiencing discrimination (Unzueta, Everly, & Gutiérrez, 2014). In other work, White participants reported greater liking (Vanman, Paul, Ito, & Miller, 1997) and displayed more positive interpersonal behaviours (Mendes & Koslov, 2013) towards Black than White interaction partners. In judgement contexts, non-Black participants favoured Black over White applicants when making a hypothetical college admission decision (Norton, Vandello, Biga, & Darley, 2008), and this pro-Black admissions bias arose even after participants had to justify their decision to the experimenter or indicate which academic criteria were most important beforehand (Norton, Vandello, & Darley, 2004).

Moreover, past work has illustrated that White undergraduates and teachers give more positive feedback when evaluating work believed to come from a Black versus White student (Harber, 1998; Harber *et al.*, 2012), particularly among those high in motivation to avoid being prejudiced (Croft & Schmader, 2012).

While many of these pro-Black behaviours are interpreted as arising from deliberate attempts to correct for existing anti-Black attitudes (e.g., Harber, 1998; Mendes & Koslov, 2013), subsequent studies have shown that pro-Black behaviour can arise automatically. To date, the largest evidence of automatic and unintentional pro-Black behaviour comes from Axt, Ebersole, and Nosek (2016). Using a novel paradigm where participants made a series of accept or reject decisions towards Black and White candidates for a hypothetical academic honour society, White undergraduate and online samples (both paid and volunteer) had lower admissions criteria for Black than for White candidates (aggregate $d = .45$; $N > 4{,}000$). Favouritism towards Black candidates was still present even when participants were given a financial incentive to select the most qualified applicants, and despite the fact that participants held pro-White implicit and explicit attitudes on average.

Moreover, the pro-Black bias in behaviour persisted among participants who reported (1) showing no racial favouritism on the task (77% of the sample), (2) wanting to show no racial favouritism on the task (91% of the sample), and (3) having no explicit preferences between White and Black people (68% of the sample). While performance on the admissions task was reliably correlated with perceived performance, desired performance, and implicit and explicit attitudes, these results illustrate that for a majority of participants, the pro-Black bias in behaviour occurred outside of conscious awareness or intention. Finally, this pro-Black bias is more striking given that other uses of the same paradigm have elicited automatic biases in behaviour that align with participants' attitudes, such as favouritism towards more versus less physically attractive applicants, or for applicants from one's own versus a rival university (Axt, Nguyen & Nosek, 2017).

Given the growing literature of studies finding both pro-Black and anti-Black bias in behaviour, the present study tests whether the same pro-Black behavioural bias observed in Axt, Ebersole, *et al.* (2016) would also be found in a large sample likely holding real-world experience evaluating people from different racial and ethnic backgrounds: education professionals. The study also investigates how such biases relate to perceived and desired performance, as well as explicit and implicit racial attitudes.

This work furthers multiple areas of research. First, by using a sample of professionals who may make academic, administrative, and punitive decisions for students from various races, this study can strengthen the generalizability and potential real-world impact of earlier findings (Axt, Ebersole, *et al.* 2016), and can help identify contexts where behaviour may favour majority versus minority group members. Second, the sample adds to the relatively small amount of research concerning implicit attitudes among teachers (e.g., O'Brien *et al.*, 2007; Van den Bergh *et al.*, 2010). Finally, the large sample size can inform ongoing discussions regarding the extent to which implicit and explicit attitudes predict behaviour (Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013).

Method

Participants

I report how I determined the sample size, all data exclusions, as well as all manipulations and measures. Participants were recruited from a blog post from an

education website as part of a series on implicit bias. The post had general information about the issue of implicit bias and the logic behind the IAT. The post did not mention the academic decision-making task or reference the previous research that found a pro-Black bias in decision-making.[1] All participants provided consent before starting the study.

Data were collected between 15 September 2015 and 16 May 2016. Study end date was selected arbitrarily. The study had 1,641 started sessions, with 1,149 participants beginning the academic decision-making task and 839 participants completing all study measures. The completion rate was then 51.1% among participants who started the study and 73.0% among participants who started the academic decision-making task. This completion rate is relatively low, but is comparable to other online volunteer samples asked to do similar tasks (e.g., Study 3 in Axt, Ebersole, *et al.* 2016; had a completion rate of 64.0%, and Study 5 had a completion rate of 62.6%). The sample size allowed for greater than 99.9% power at detecting a Cohen's $d = .45$, which was the average effect size in differences between White and Black criterion in Axt, Ebersole, *et al.* (2016).

Among those reporting demographics, 75.8% were female, the mean age was 39.1 ($SD = 14.4$), and 96.2% were US citizens. By race, 71.8% were White, 1.4% were East Asian, 0.8% were South Asian, 11.1% were Black, 2.5% were biracial, and 12.4% indicated other or unknown racial membership. By ethnicity, 12.0% were Hispanic or Latino. Sample sizes vary across tests due to missing data.

Procedure

Participants completed measures in the following order: academic decision-making task, measures of task performance, explicit racial attitudes, implicit racial attitudes, and a demographics survey. Participants were then debriefed and given feedback on their implicit task performance. See https://osf.io/3bvct/ for materials and data.

Academic decision-making task

Participants completed the same academic decision-making task used in Studies 3–6 of Axt, Ebersole, *et al.* (2016). First, participants were instructed that they would view all the applicants for an academic honour society, which was now described as a high school honour society to create a context more familiar to the sample. After this viewing phase, participants would then go through and select or reject each applicant. In the viewing phase, participants observed passively as each of the 60 candidates was shown one at a time for one second in a random order. This viewing phase provided participants with insight into 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 a randomized order, and were instructed to accept approximately half of the applicants. Participants pressed the 'I' key to accept and the 'E' key to reject. There was no time limit for making these 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

[1] *The full post can be accessed at: http://www.edweek.org/ew/section/multimedia/test-yourself-a-survey-tool-for-gauging-bias.html.*

(poor, fair, good, excellent), and interview score (1–100). Participants were instructed to weigh each piece of information equally when making their evaluations.

These four pieces of information were used to create 60 unique applications, 30 that were more qualified and 30 that were less qualified. To do this, each piece of information was standardized to have a 1–4 range. The two GPAs already ranged from 1 to 4, and the recommendation letters (poor = 1, fair = 2, good = 3, excellent = 4) and interview scores (dividing interview score by 25) were converted to be on the same 1–4 scale. 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 = 3.5, humanities GPA = 3.6, recommendation letters = good, interview score = 72.5. When standardized, these pieces of information sum to 13 (3.5 + 3.6 + [good = 3] + [72.5/25] = 13). Conversely, one more qualified applicant had the following qualifications: science GPA = 3.7, humanities GPA = 3.9, recommendation letters = good, interview score = 85. When standardized, these pieces of information sum to 14 (3.7 + 3.9+ [good = 3] + [85/25] = 14).

In the applications, 30 of the faces were Black males and 30 were White males, with 15 faces in each race assigned to more qualified and 15 faces assigned to less qualified profiles. Participants were randomly assigned to one of 12 orders. Across the 12 orders, each face was equally likely to be assigned to either a more qualified or less qualified application, and each combination of qualifications was equally likely to be paired with a Black or White face.

Perceptions of performance

Participants completed two items regarding their task performance. First, 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 mid-point of 'I treated both Black and White applicants equally' (0). Second, participants rated how they wanted to perform on the task using a 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 mid-point of 'I wanted to treat both Black and White applicants equally' (0).

Explicit racial attitudes

Participants completed a single-item measure of preferences for Black relative to White people (Nosek *et al.*, 2007) using 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 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'. 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 *et al.* (2014), except that a warm-up block of categorizing only good and bad words was removed from the procedure.

Demographics

Participants completed a 7-item demographics questionnaire, reporting race, ethnicity, age, gender as well as country of citizenship and residence. Participants also reported their current job within education, with eight response options ('teacher', 'principal', 'superintendent', 'district staff', 'state or federal official', 'education researcher', 'other job in education', and 'I do not work in education').

Results

Analyses are limited to participants who reported having a job in education (N = 618). Analyses for the full sample as well as only among only White participants are available in the online supplement (https://osf.io/3bvct/). The pattern of results does not substantively change when including the full sample or only White participants.

As in Axt, Ebersole, *et al.* (2016), participants were excluded 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. Thirty-five participants (5.7%) were excluded by these criteria. Eleven additional participants were excluded from BIAT analyses for having more than 10% of BIAT trial responses less than 300 milliseconds, following the Nosek *et al.*'s (2014) guidelines.

Accuracy is defined as selecting more qualified candidates and rejecting less qualified candidates. Overall accuracy on the task was 69.9% (SD = 8.0), well above chance, $t(582) = 60.18, p < .001, d = 2.49$, 95% CI (2.33, 2.66), but not high enough for possible ceiling effects. The average acceptance rate was close to the recommended 50% (M = 53.4%, SD = 11.90).

Bias in selection for honour society

I used signal detection theory (SDT; Green & Swets, 1966/1974; MacMillan & Creelman, 1991) to analyse the influence of qualifications on admissions judgements. This analysis assumes that on average, applicants with superior grades, recommendation letters, and interview scores (those scoring 14) are more qualified for the honour society than applicants with lower values (those scoring 13), and also 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, and a sensitivity value of zero would indicate no ability to distinguish between more and less qualified candidates.

Criterion (c) refers to the decision threshold for accepting or rejecting a candidate. Participants can have a more liberal threshold in which they are more likely to falsely

accept unqualified candidates (a negative criterion value), or a more conservative threshold in which they are more likely to falsely reject qualified candidates (a positive criterion value).

SDT analyses have been used frequently in psychological research. 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 for White targets, meaning that the threshold to respond 'gun' is lower when the person on screen is Black than White (e.g., Correll, Wittenbrink, Crawford, & Sadler, 2015; Correll *et al.*, 2002). However, 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.

Using an SDT analysis, I compared criterion and sensitivity estimates for Black and White candidates. There were no reliable differences in sensitivity (d') between Black applicants ($M = 1.22$, $SD = 0.62$) and White applicants ($M = 1.25$, $SD = 0.64$), $t(582) = 0.84$, $p = .403$, $d = .03$, 95% CI (−.05, .12). Replicating the results found in Axt, Ebersole, *et al.* (2016), Black applicants ($M = -0.36$, $SD = 0.52$) received a lower criterion than White applicants ($M = 0.15$, $SD = 0.46$), $t(582) = 21.91$, $p < .001$, $d = .91$, 95% CI (.81, 1.00),[2] and this pattern held among White, Black, and Hispanic participants (see Table 1).

Within each education profession category, Black applicants received a lower criterion than White applicants. See Table 2 for sample sizes, criterion values, test statistics, and confidence intervals for each professional category as well as comparisons between professions. Comparing the size of the pro-Black criterion bias (the difference score between criterion for White and Black applicants), education professionals ($M = .51$, $SD = 0.56$) had a larger criterion bias than participants from other professions ($M = .38$, $SD = 0.51$), $t(769)$, $= 2.79$, $p = .005$, $d = .23$, 95% CI (.07, .40).

Awareness of selection bias

Most participants (69.8%) indicated that they *had treated* both Black and White applicants equally. Among them, Black applicants ($M = -0.30$, $SD = 0.53$) received a lower criterion than White applicants ($M = 0.13$, $SD = 0.46$), $t(401) = 15.50$, $p < .001$, $d = .77$, 95% CI (.66, .88). Likewise, most participants (80.3%) indicated a *desire* to treat

Table 1. Criterion values for White, Black, and Hispanic participants

Participant race or ethnicity	*N*	Black *c*	White *c*	*t*	*p*	*d*	95% CI
White	419	−.35 (.51)	.16 (.47)	18.63	<.001	.91	(.80, 1.02)
Black	62	−.39 (.54)	.06 (.36)	6.23	<.001	.79	(.50, 1.07)
Hispanic	44	−.33 (.55)	.10 (.45)	4.75	<.001	.72	(.38, 1.04)

Note. Black *c* = criterion mean (and *SD*) for Black applicants. White *c* = criterion mean (and *SD*) for White applicants. *d* = Cohen's *d* effect size for comparison of White *c* and Black *c*.

[2] *Lower criterion for Black vs. White applicants meant that when applicants were more qualified, accuracy rates were higher when evaluating Black (*M = *80.3%, SD = 15.4%) than White (*M = *66.3%, SD = 15.4%) applicants. Conversely, when applicants were less qualified, accuracy rates were higher when evaluating White (*M = *74.7%, SD = 18.3%) than Black (*M = *58.3%, SD = 20.5%) applicants.*

Table 2. Criterion values for educational professions

Profession	*N*	Black *c*	White *c*	*t*	*p*	*d*	95% CI
Teacher[a]	237	−.29 (.49)	.13 (.44)	11.08	<.001	.72	(.58, .86)
Principal[ac]	60	−.33 (.57)	.17 (.52)	7.05	<.001	.91	(.61, 1.21)
Superintendent[ac]	17	−.32 (.62)	.03 (.40)	2.08	.054	.50	(−.01, 1.00)
District Staff[bc]	74	−.42 (.55)	.17 (.45)	9.29	<.001	1.08	(.79, 1.36)
State or federal official[ab]	18	−.39 (.46)	.17 (.51)	5.94	<.001	1.40	(.73, 2.05)
Education researcher[bc]	31	−.53 (.54)	.12 (.44)	5.79	<.001	1.04	(.60, 1.47)
Other job in education[bc]	146	−.40 (.50)	.20 (.47)	14.45	<.001	1.20	(.98, 1.41)

Note. Black *c* = criterion mean (and *SD*) for Black applicants. White *c* = criterion mean (and *SD*) for White applicants. *d* = Cohen's *d* effect size for comparison of White *c* and Black *c*. Professions that do not share a superscript letter differ from each other in criterion bias effect size at $p < .05$.

Black and White applicants equally. Among them, Black applicants ($M = -0.34$, $SD = 0.52$) received a lower criterion than White applicants ($M = 0.14$, $SD = 0.46$), $t(464) = 18.36$, $p < .001$, $d = .85$, 95% CI (.75, .96).

Racial attitude relations with selection decisions

Surprisingly, BIAT *D* scores did not reveal pro-White attitudes ($M = 0.02$, $SD = 0.53$), $t(554) = 0.77$, $p = .439$, $d = .03$, 95% CI (−.05, .12), and neither did the explicit racial preference item ($M = 0.03$, $SD = 0.82$), $t(571) = 1.02$, $p = .307$, $d = .04$, 95% CI (−.04, .12).[3] However, among participants who reported no explicit preference for White or Black people (63.8%), Black applicants ($M = -0.32$, $SD = 0.53$) received a lower criterion than White applicants ($M = 0.15$, $SD = 0.48$), $t(364) = 15.60$, $p < .001$, $d = .82$, 95% CI (.70, .93).

Predictors of racial bias in criterion

To analyse the relationship between task performance and attitudes, criterion for Black applicants was subtracted from criterion for White applicants, creating a difference score such that higher values indicated a more relaxed criterion for Black relative to White applicants. This criterion bias was reliably and negatively correlated with implicit, $r = -.12$, $p = .004$, 95% CI (−.20, −.04), and explicit, $r = -.13$, $p = .002$, 95% CI (−.21, −.05), attitudes, as well as perceived performance, $r = -.22$, $p < .001$, 95% CI (−.30, −.14), and desired performance, $r = -.12$, $p = .005$, 95% CI [−.20, −.04). Here, a negative correlation indicates that stronger explicit and implicit preferences for Whites, as well as a greater perception of or desire to favour White over Black applicants, were associated with a smaller pro-Black criterion bias (these results replicate those reported in Axt, Ebersole, *et al.* 2016). See Table 3 for a correlation matrix between criterion bias, explicit attitudes, perceived performance, desired performance, and implicit attitudes.

A simultaneous linear regression predicting race differences in criterion bias from implicit attitudes, explicit attitudes, perceived performance, and desired performance

[3] *Among only White participants, BIAT D scores revealed weak pro-White attitudes (M = 0.08, SD = 0.53), t(496) = 3.20, p = .001, d = .14, 95% CI (.05, .23), as did the explicit racial preference item (M = 0.25, SD = 0.66), t(511) = 8.70, p <.001, d = .38, 95% CI (.29, .47).*

Table 3. Correlations between continuous study measures

	Criterion bias	Exp. attitudes	Perc. performance	Des. performance
Exp. attitudes	−.13**			
Perc. performance	−.22**	.14**		
Des. performance	−.12**	.17**	.41**	
Imp. attitudes	−.12**	.17**	.02	.04

Note. Exp. attitudes = explicit attitudes. Perc. performance = perceived performance. Des. performance = desired performance. Imp. attitudes = implicit attitudes (BIAT). Correlations with implicit attitudes exclude participants with more than 10% of trials faster than 300 milliseconds.
**$p < .01$.

revealed that explicit attitudes ($B = -.11$, $p = .048$), implicit attitudes ($B = -.09$, $p = .012$), and perceived performance ($B = -.20, p < .001$) contributed uniquely, while desired performance ($B = -.01$, $p = .820$) did not. Overall, those four variables accounted for 6.8% of the racial difference in criterion bias.

GENERAL DISCUSSION

A large sample of educational professionals showed a robust pro-Black bias in behaviour, setting a lower criterion towards Black than towards White candidates for admission into a hypothetical high school honour society. The bias was present among participants who reported showing no favouritism on the task (70% of respondents), not wanting to show favouritism on the task (80%), and having no explicit preferences between Black and White people (64%). However, biases in criterion were reliably but weakly related to implicit attitudes, explicit attitudes, perceived performance, and desired performance. These results suggest that performance on the task is partly under conscious control; participants wanting to show bias on the task by favouring White or Black applicants did so relative to participants wanting to be unbiased. Nevertheless, a perception of being fair, a desire to be fair, and no explicit racial preferences were not enough to eliminate racial information from impacting decision-making. For a majority of our sample, the pro-Black bias occurred outside of conscious awareness or intention, and countered explicit attitudes that suggested no racial preferences.

These results are most closely aligned with research on shifting standards (Biernat & Manis, 1994), where people adjust the value of criteria for members of different social groups. That is, the educational professionals in this sample may have used different standards for Black than for White applicants, meaning Black and White applicants were not being judged against each other but only against the standard for that racial group. Relative to expectations, the Black applicants may have seem more impressive, creating a lower admissions criterion and a higher acceptance rate to the academic honour society. While previous work illustrated a similar phenomenon in undergraduate and online samples (Axt, Ebersole, *et al.* 2016), the present study reveals automatic biases among a sample with possible professional experience evaluating people from different racial and ethnic backgrounds. The large difference in criterion bias found here (overall $d = .91$) suggests that even a sample likely containing relevant experience in academic evaluation still relied on shifting standards. In fact, the pro-Black criterion bias among education professionals was much larger than that found in past research ($d = .45$ in Axt, Ebersole,

et al. 2016) and results within this sample showed a reliably larger criterion bias for educational versus non-educational professionals. These results further indicate that the process of shifting racial standards occurs automatically and without conscious intent for most people.

Finally, whereas Axt, Ebersole, *et al.* (2016) was limited to all-White samples, this study found a pro-Black criterion bias among White, Black, and Hispanic participants. Such results increase the generalizability of the finding and improve understanding of the phenomenon. For one, the fact that Black participants also showed a pro-Black bias highlights that this behaviour is not merely outgroup racial favouritism but rather more specific to evaluations of Black versus White targets in an academic context. Moreover, while White participants may have favoured Black applicants due to processes like guilt or self-presentation, results from non-White participants indicate possible additional causes such as ingroup favouritism (for Black participants) or solidarity with other minority groups (for Hispanic participants). Subsequent work will need to examine whether this pro-Black behaviour is driven by shared or separate mechanisms among White and non-White samples.

Limitations and directions for future research

One clear weakness of this work is not only that participants came from a population interested in issues of implicit bias, but also that participants were aware their judgements were being studied. In addition, wording on measure anchors that required participants to admit 'strongly preferring' Black versus White people or going 'extremely easier' on Black versus White applicants may have introduced social desirability concerns, although prior work suggests a subset of people are willing to express their explicit prejudices or motivations to discriminate (e.g., Forscher, Cox, Graetz, & Devine, 2015). Moreover, there was a relatively high amount of dropout (51% of participants who started the study completed all measures).

These factors may have led to a population of participants that were particularly concerned with appearing racially prejudiced, creating motivated performance to favour Black over White applicants so as to appear unprejudiced. As a result, the participants who showed pro-Black biases in our study may have shown neutral or pro-White biases in contexts where they were unaware that their behaviour was being studied, and it will be necessary for future studies to obtain a more representative sample of educators (e.g., by collecting data in person or by offering a financial incentive to complete the study). The fact that participants on average showed no reliable pro-White implicit or explicit attitudes is a deviation from prior samples (Axt *et al.*, 2014; Bar-Anan & Nosek, 2014), and lends further support to the notion that our sample may have already been sensitive to issues of implicit bias and concerned about appearing racially unbiased.

However, given recent suggestions from the US Department of Education for implementing widespread implicit bias training (US Department of Education, 2014), or materials created by the National Education Association to help teachers deal with classroom diversity ('Diversity Toolkit', 2015), concerns over the presence and impact of implicit bias appear common among educators. While the educators in this study were very likely familiar with the concept of implicit bias before participating, many educators seem to be aware of the potential consequences of implicit bias.

Most studies of real-world behaviour often find that Black people receive worse treatment in employment (Bertrand & Mullainathan, 2004), economic (Doleac & Stein,

2013), academic (Milkman, Akinola, & Chugh, 2015), and medical (Freeman & Payne, 2000) contexts. However, similar to the results reported here, a recent field study also found an instance where minority groups received better treatment in contexts where their identity was salient. French companies randomly assigned to receive resumes that did not have applicants' names were less likely to interview minority candidates (who were primarily African) compared to companies who received resumes with names (Behaghel, Crépon, & Barbanchon, 2015). That is, minority applicants had better outcomes when employers had access to their ethnic information. Although this was not necessarily a pro-minority or pro-Black bias (majority applicants still received more interviews than minority applicants in both conditions), it does highlight a context where minority status helped rather than hurt individual outcomes. It will be crucial for future research to identify whether the same processes observed in this work are also present in other studies of real-world behaviour.

Notably, the pro-Black bias observed here has not arisen in other decision contexts using similar measures. In a study using a modified version of this decision-making paradigm, White participants evaluated more or less qualified potential dating partners on criteria such as intelligence and sense of humour. Unlike in an academic context, dating profiles labelled as White received a lower criterion to be accepted as a potential dating partner than profiles labelled as Black or Hispanic (Axt, Nguyen & Nosek, 2017). Future work should take the flexible paradigm used in these studies to investigate what contexts or conditions elicit pro-Black or anti-Black biases. For instance, previous work found that the positive feedback bias towards Black versus White students (Harber, 1998) was eliminated after participants could affirm their egalitarian self-images (Harber, Stafford, & Kennedy, 2010) or after receiving feedback that they were relatively low in levels of implicit bias (Ruscher, Wallace, Walker, & Bell, 2010). Similar affirmation manipulations may effectively reduce the pro-Black criterion bias found here. Furthermore, it is unclear how participants may perform when having to evaluate Black and White profiles for more informal (e.g., choosing friends) compared to more formal (e.g., employees to promote) decisions. Such work would illuminate whether the pro-Black bias observed here occurs more generally, or is specific to an academic context.

What the data do not show

These results found another instance of a robust and automatic pro-Black bias in judgement. Such results do not counter or invalidate results of bias found in previous research. Rather, these data suggest room for improvement in our theoretical understanding of the conditions that give rise to ingroup or outgroup favouritism. Moreover, while this work highlights the need for a better conception of how such behaviour can arise automatically while countering existing attitudes, it does not suggest that existing models of behaviour are wrong, but rather that they are incomplete.

Conclusion

Despite widespread evidence that Black people receive worse treatment than White people in a number of domains, a sample of professionals with possible experience evaluating people from multiple groups showed a robust, automatic, and unintentional bias favouring Black over White applicants in an academic context. These results also add to the growing literature concerning educators' implicit attitudes (O'Brien *et al.*, 2007; Van den Bergh *et al.*, 2010), and support the notion that implicit attitudes are weakly but

reliably predictive of behaviour (Greenwald *et al.*, 2015). Educators (and non-White participants) performed similarly as previous samples, strengthening the generalizability of this pro-Black bias and highlighting the need for better prediction of when people favour minority or majority group members.

Conflict of interests

This research was partly supported by Project Implicit. J. R. Axt is a consultant for Project Implicit, Inc., a non-profit organization with the mission to 'develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender, or other factors'.

References

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*(9), 1804–1815. doi:10.1177/0956797614543801

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

Axt, J. R., Nguyen, H., & Nosek, B. A. (2017). *The Judgment Bias Task: A reliable, flexible method for assessing individual differences in social judgment biases*. Manuscript submitted for publication. University of Virginia.

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

Behaghel, L., Crépon, B., & Barbanchon, T. L. (2015). Unintended effects of anonymous résumés. *American Economic Journal: Applied Economics*, 7, 1–27. doi:10.1257/app.20140185

Bertrand, M., & Duflo, E. (2016). *Field experiments on discrimination*. (Working paper No. 22014). National Bureau of Economic Research website: Retrieved from http://www.nber.org/papers/w22014

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. doi:10.1257/0002828042002561

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*, 5–20. doi:10.1037/0022-3514.66.1.5

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. doi:10.1037/0022-3514.83.6.1314

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. doi:10.1037/pspa0000015

Croft, A., & Schmader, T. (2012). The feedback withholding bias: Minority students do not receive critical feedback from evaluators concerned about appearing racist. *Journal of Experimental Social Psychology*, *48*, 1139–1144. doi:10.1016/j.jesp.2012.04.010

Diversity toolkit: Cultural competence for educators. (2015). Retrieved from http://www.nea.org/tools/30402.htm#S

Doleac, J. L., & Stein, L. C. D. (2013). The visible hand: Race and online market outcomes. *The Economic Journal*, *123*, F469–F492. doi:10.1111/ecoj.12082

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*, 319–323. doi:10.1111/1467-9280.00262

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. doi:10.1037/0022-3514.82.1.62

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

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks III, M. P. & Booth, E. A. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement: Council of State Governments Justice Center and The Public Policy Research Institute, Texas A&M University.

Forscher, P. S., Cox, W. T., Graetz, N., & Devine, P. G. (2015). The motivation to express prejudice. *Journal of Personality and Social Psychology*, *109*, 791–812. doi:10.1037/pspi0000030

Freeman, H. P., & Payne, R. (2000). Racial injustice in health care. *New England Journal of Medicine*, *342*(14), 1045–1047. doi:10.1056/NEJM200004063421411

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. doi:10.1037/0022-3514.94.1.91

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., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561. doi:10.1037/pspa0000016

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

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*, 669–684. doi:10.1037/a0036056

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology*, *46*(4), 455–475. doi:10.1016/j.jsp.2007.09.001

Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology*, *74*, 622–628. doi:10.1037/0022-3514.74.3.622

Harber, K. D., Gorman, J. L., Gengaro, F. P., Butisingh, S., Tsang, W., & Ouellette, R. (2012). Students' race and teachers' social support affect the positive feedback bias in public schools. *Journal of Educational Psychology*, *104*, 1149–1161. doi:10.1037/a0028110

Harber, K. D., Stafford, R., & Kennedy, K. A. (2010). The positive feedback bias as a response to self-image threat. *British Journal of Social Psychology*, *49*, 207–218. doi:10.1348/014466609X473956

Kirwan Institute (2014). *Implicit racial bias and school discipline disparities*. Retrieved from http://kirwaninstitute.osu.edu/wp-content/uploads/2014/05/ki-ib-argument-piece03.pdf

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics*, *119*, 49–89. doi:10.1162/003355304772839524

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

Mendes, W. B., & Koslov, K. (2013). Brittle smiles: Positive biases toward stigmatized and outgroup targets. *Journal of Experimental Psychology: General*, *142*, 923–933. doi:10.1037/a0029663

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*, 1678–1712. doi:10.1037/apl0000022

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. doi:10.1521/soco.2008.26.1.102

Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, *87*, 817–831. doi:10.1037/0022-3514.87.6.817

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115. doi:10.1037//1089-2699.6.1.101

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J. R., & Greenwald, A. G. (2014). Understanding and using the Brief Implicit Association Test: Recommended scoring procedures. *PLoS ONE*, *9*(12), e110938. doi:10.1371/journal.pone.0110938

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

O'Brien, K. S., Hunter, J. A., & Banks, M. (2007). Implicit anti-fat bias in physical educators: Physical attributes, ideology and socialization. *International Journal of Obesity*, *31*(2), 308–314. doi:10.1038/sj.ijo.0803398

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. doi:10.1177/0956797615570365

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. doi:10.1037/a0032734

Rae, J. R., Newheiser, A. K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the United States. *Social Psychological and Personality Science*, *6*(5), 535–543. doi:10.1177/1948550614567357

Ruscher, J. B., Wallace, D. L., Walker, K. M., & Bell, L. H. (2010). Constructive feedback in cross-race interactions. *Group Processes & Intergroup Relations*, *13*(5), 603–619. doi:10.1177/1368430210364629

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*, 283–294. doi:10.1027/1618-3169.56.4.283

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. doi:10.1016/j.jesp.2014.04.005

U.S. Department of Education (2014). *Guiding principles: A resource guide for improving school climate and discipline*. Washington, DC: Author.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*, 497–527. doi:10.3102/0002831209353594

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*, 941–959. doi:10.1037/0022-3514.73.5.941

Xu, K., Nosek, B., & Greenwald, A. (2014). Psychology data from the race Implicit Association Test on the Project Implicit demo website. *Journal of Open Psychology Data*, *2*(1). doi:10.5334/jopd.ac

Received 10 July 2016; revised version received 7 March 2017