Prejudiced and unaware of it: Evidence for the Dunning-Kruger model in the domains

of racism and sexism

Keon West [a]

Asia A. Eaton [b]

[a] Goldsmiths, University of London, London, UK

[b] Florida International University, Florida, USA

Word Count: 9977 (6933 without references, tables or figures)

Declarations of interest: none

Please address all correspondence to: Dr Keon West, Department of Psychology, Goldsmiths, University of London, UK, SE15 6NW, keon.west@gold.ac.uk, 0207 919 7171

Abstract

Prior research, and high-prolife contemporary examples, show that individuals tend to underestimate their own levels of bias. This underestimation is partially explained by motivational factors. However, (meta-) cognitive factors may also be involved. Conceptualising contemporary egalitarianism as type of skill or competence, this research proposed that egalitarianism should conform to the Dunning-Kruger model. That is, individuals should overestimate their own ability, and this overestimation should be strongest in the least competent individuals. Furthermore, training should improve metacognition and reduce this overestimation. Two studies on racism ($N = 148$), and sexism ($N = 159$) partially supported these hypotheses. In line with the Dunning-Kruger model, participants overestimated their levels of racial and gender-based egalitarianism, and this pattern was strongest among the most prejudiced participants. However, diversity training did not affect participants' overestimation of their egalitarianism. Implications for contemporary prejudice, and prejudice-reducing strategies are discussed.

"[President Trump is] the most racist, sexist, homophobic, bigoted president in history."

- Senator Bernie Sanders (Chasmar, 2018, p. 1)

"I am the least racist person that you have ever met."

- President Donald Trump (Scott, 2018, p. 16)

An overwhelming consensus of social-psychological research confirms that both race-based and gender-based bias pervade contemporary Western societies. Citing a few examples, even when all other differences are eliminated or accounted for, studies show that ethnic minorities (compared to White people) are treated with more suspicion in public places (Schreer, Smith, & Thomas, 2009), considered less desirable as romantic partners (Mendelsohn, Shaw Taylor, Fiore, & Cheshire, 2014; West, 2018; West, Lowe, & Marsden, 2017), less likely to receive offers of employment (Bertrand & Mullainathan, 2004; Booth, Leigh, & Varganova, 2012; Pager, 2003), judged more harshly for crimes they commit (Sommers & Ellsworth, 2000; West & Lloyd, 2017), more likely to be shot by police officers (Correll, Park, Judd, & Wittenbrink, 2007; Plant & Peruche, 2005), and less likely to receive adequate care from physicians (Green et al., 2007).

Similarly, even when other differences are eliminated or controlled, women (compared to men) are offered less pay for the same work (Auspurg, Hinz, & Sauer, 2017), offered less informal support when joining organisations (Milkman, Akinola, & Chugh, 2015), less likely to be offered high paying jobs and more likely to be offered low paying jobs (Neumark & Van Nort, Kyle, 1996), offered fewer leadership positions, disproportionately offered risky or precarious leadership positions

(Bruckmüller, Ryan, Rink, & Haslam, 2014; Ryan et al., 2016), and less likely to be greeted or approached by servers in stores (Bourabain & Verhaeghe, 2018).

However, despite the clear evidence for continued, contemporary discrimination, there is a contrasting tendency for people in privileged groups to deny the existence of bias (Augoustinos & Every, 2010; Bonam, Das, Coleman, & Salter, 2018; McIntosh, 1988; Nelson, Adams, & Salter, 2013; Swim, Aikin, Hall, & Hunter, 1995). Even among younger people (who tend to be more liberal) there is an impression that racism is not "that bad anymore" (Andreouli, Howarth, & Greenland, 2016, p. 171) or that it occurs only rarely, or in extreme circumstances. Indeed, an increasing number of White people in majority White countries believe that racism is a thing of the past, or that prejudice against White people is now a more serious concern than prejudice against Black people (Norton & Sommers, 2011). Similarly, men, compared to women, are less likely to notice or respond to sexism (Biggs, Hawley, & Biernat, 2018).

On a more individual level, people tend to underestimate *their own levels* of race- and gender-based bias, or the extent to which their behaviour is affected by another person's race or gender. For example, Norton, Sommers, Apfelbaum, Pura and Ariely (2006) found that White participants underestimated their own ability to categorise faces on the basis of race. Furthermore, though participants claimed to be "colourblind" (i.e., not to notice race), they nonetheless altered their speech about race depending on the race of their interaction partner (Black vs. White). A wealth of research on *implicit* biases (i.e., biases that can be detected or measured using methods that circumvent explicit control or self-presentation), consistently reveals that participants who express little or no explicit bias nonetheless consistently show

bias when tested with implicit measures (Nosek, Greenwald, & Banaji, 2007). This effect applies to sexism as well as racism (Rudman & Kilianski, 2000).

This tendency to underestimate one's own bias can (at least in part) be explained motivationally. Many individuals in Western societies report an internal desire to be egalitarian that persists even in the absence of external pressure (Crandall, Eshleman, & O'Brien, 2002; West & Hewstone, 2012). It is not unusual for individuals to explicitly explain their behaviour in intergroup contexts with such statements as "because I don't [or didn't] want to be racist" (Pauker, Apfelbaum, & Spitzer, 2015, p. 889; Watt, 2017, p. 410). This internalized motivation to act without prejudice has been shown to alter subsequent behaviour, including reducing expressions of implicit and explicit prejudice (Johns, Cullum, Smith, & Freng, 2008). Furthermore, much research on "aversive racism" supports the position that some individuals experience an internal struggle between their aversion to racial minorities and their genuinely-held principles of racial equality (Pearson, Dovidio, & Gaertner, 2009). These genuine desires to be (or think of oneself as) egalitarian could partially explain why individuals often present themselves as, or believe themselves to be, more egalitarian than they are.

However, the tendency to underestimate one's own bias can also be explained through cognitive variables. Large-scale cross-sectional studies in the UK and US have found that lower generalised intelligence and poorer abstract reasoning skills both predict more prejudice (Hodson & Busseri, 2012). Other studies have shown that individuals lacking in knowledge about historical racism are less able to recognise or acknowledge contemporary examples of racism (Nelson et al., 2013). Educating White individuals about historical racism can improve their recognition of contemporary racism (Bonam et al., 2018). Similarly, directing men's attention to

instances of sexism can improve their recognition of sexism and reduce modern sexist beliefs (Becker & Swim, 2011).

Interestingly, the two observations noted above – (a) that many individuals desire to be egalitarian, and (b) that being egalitarian appears to require a certain amount of both knowledge and cognitive ability – suggest that egalitarianism can be conceptualised as a skill, or type of competence. In other words, being egalitarian (i.e., non-racist or non-sexist) in contemporary society is not merely a matter of willingness, but also a matter of *ability*. If this is the case, egalitarianism should conform to the Dunning-Kruger model of incompetence and inflated self-assessments (Kruger & Dunning, 1999).

Kruger and Dunning (1999) argued that, across a range of fields, incompetent individuals suffered under a "dual burden" (p. 1121). The skills required to be competent in an activity are often the same as the skills required to recognise competence (or incompetence) in that activity. Therefore, those who are "unskilled" are also more likely to be "unaware" of how unskilled they are, and thus more likely than skilled individuals to severely overestimate their own competence. As Bertrand Russell observed, "the trouble with the world is that the stupid are cocksure and the intelligent are full of doubt" (Russell, 1979, p. 71).

The Dunning-Kruger model of incompetence and unawareness has been extremely influential throughout a variety of disciplines. The original paper (Kruger & Dunning, 1999) has been cited over 4,300 times (Google Scholar, 2018) and the model has been applied to a variety of domains including physicians decisions, health, education, the workplace, and driving (Anstey, Wood, Lord, & Walker, 2005; Davis et al., 2006; Dunning, Heath, & Suls, 2004). Alternative explanations for this specific pattern of overestimation have also been ruled out, including floor effects and a

negative skew in participant competence (i.e., a *backwards – J* distribution; Schlösser, Dunning, Johnson, & Kruger, 2013).

Still, despite the reach and influence of this model, no research to date has investigated whether it also applies to individuals' levels of egalitarianism or prejudice. Some research has tested related ideas. For example, de Keersmaecker, Onraet, Lepouttre, and Roets (2017) found that higher levels of externally measured intelligence predicted less prejudice, while higher levels of *self-perceived* intelligence predicted more prejudice. However, this current research is the first to investigate egalitarianism *itself* as a skill (rather than mediating the effect via another skill, such as intelligence). In this research, it is directly hypothesized that lower levels of egalitarianism (i.e., higher levels of prejudice) will predict greater overestimation of one's own egalitarianism.

Another hypothesis concerned a possible method of reducing this overestimation. Kruger and Dunning (1999) "predicted that training would provide incompetent individuals with the metacognitive skills needed to realize that they had performed poorly and thus would help them realize the limitations of their ability" (p. 1128). Testing this hypothesis with a logical reasoning task (Study 4), Kruger and Dunning showed that training in logical reasoning improved participants' (a) ability to recognise their prior mistakes (b) accuracy in their judgement of their performance and (c) perceptions of their performance relative to their peers. If this applies across domains, diversity training may have similar effects on participants' perceptions of their own egalitarianism.

That said, it is also possible that diversity training will fail to have the desired effect, largely due to the underdeveloped *training* aspect of much diversity training (Atewologun, Cornish, & Tresh, 2018; Noon, 2018). There are certainly teachable

strategies that have been empirically shown to cause significant, long-term reductions in bias, such as counter-stereotypic imaging and stereotype replacement (Devine, Forscher, Austin, & Cox, 2012). However, these techniques are rarely incorporated into diversity training. Rather, most diversity training courses merely include (a) an unconscious bias test, such as the implicit association test (Nosek et al., 2007) and an explanation of the results, (b) education on the psychological theory behind unconscious biases and (c) information on the impact of unconscious biases (Atewologun et al., 2018). There are also frequently suggestions for ways to mitigate the *impact* of unconscious biases, but no actual training in methods to reduce one's bias: either implicit or explicit (Atewologun et al., 2018; Noon, 2018). As such, an emerging consensus is that diversity training programmes do not typically reduce levels of bias, but merely raise awareness of the existence of bias – an important distinction (Dobbin & Kalev, 2016; Dobbin, Kalev, & Kelly, 2007).

Current Research and Hypotheses

In two studies this research investigated whether the Dunning-Kruger model could be applied to contemporary egalitarianism (specifically being non-racist and non-sexist). In line with this model there were three main hypotheses. First, individuals should tend to overestimate their level of egalitarianism. Second, more prejudiced individuals, compared to their more egalitarian counterparts, should more strongly overestimate their egalitarianism relative to external measures. Finally, we investigated whether giving participants feedback on their levels of bias, via diversity training, would reduce the tendency to overestimate one's egalitarianism. These hypotheses are investigated in two studies looking at racism (Study 1) and sexism (Study 2).

Study 1

Anti-Black racism in majority-White Western societies is a serious problem with a long, difficult history. In many ways, Black people continue to receive worse treatment than White people (Bertrand & Mullainathan, 2004; Milkman et al., 2015; Pager, Western, & Bonokowski, 2006), and racism against Black people has been shown to be stronger than racism against a variety of other ethnic groups, particularly in the UK (Leach, Peng, & Volckens, 2000; Rutland, Cameron, Milne, & Mcgeorge, 2005). Indeed, subtle measures of infrahumanization (Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014) show that many White people perceive Black people as subtly less than fully human. Nonetheless, contemporary social norms prohibit overt expressions of racism, and most White people report being opposed to racism (Crandall et al., 2002; Kawakami, Dunn, Karmali, & Dovidio, 2009; Pearson et al., 2009; West & Hewstone, 2012). This tension between egalitarian motives and continued prejudice makes anti-Black racism an ideal focus for the first test of the Dunning-Kruger model in this domain.

Method

To determine the sample sizes necessary for both studies a-priori power analyses were conducted using G*Power 3 (Faul, Erdfelder, Buchner, & Lang, 2009). Assuming a medium effect size for the hypothesised prediction of participants' overestimation of their egalitarianism by their externally measured egalitarianism - i.e., effect size (*slope H₁*) = .25, α = .05, power = .80 - it was found that 120 participants would be sufficient for adequate power. Similarly, assuming a medium, within-participants effect size for the effect of diversity training on self-perceived egalitarianism – i.e., effect size (*dz*) = .05, α = .05, power = .8 – it was found that 34 participants would be sufficient for adequate power.

Participants. One hundred and forty-eight participants (31 men, 115 women, and 2 who did not identify their gender, *mean age* = 23.19, *SD* = 5.10) were recruited through their involvement in a voluntary racism-related diversity training programme aimed at graduate students in London. Most of the participants were White (117, 79.1%). Smaller proportions of the participants were South Asian (13, 8.8%), East Asian (2, 1.4%), Middle Eastern (1, .7%) or ethnically mixed (15, 10.1%).

Procedure. At the start of the diversity training programme, participants indicated their perceptions of their own levels of racial egalitarianism using two items. First they were asked to indicate, on a scale from 0 to 99, "how egalitarian you are about race compared to the other people in this programme: 0 would indicate that you are at the very bottom of the programme or more racist than almost everyone else, 50 would indicate that you were 'exactly average', and a score of 99 would indicate that you were at the very top, or less racist than almost everyone else." In a similar fashion they were asked to indicate, on a scale from 0 to 99, "how egalitarian you are about race compared to other people in the UK: 0 would indicate that you were at the very bottom, or more racist than almost everyone else in the UK, 50 would indicate that you were 'exactly average', and a score of 99 would indicate that you were at the very top, or less racist than almost everyone else." These two items were strongly correlated ($r = .62, p < .001$), so the mean of the two items was used as the measure of participants' perception of their own racial egalitarianism. Higher values indicated a stronger perception of one's own racial egalitarianism.

Participants' actual levels of racial egalitarianism were also measured with two items: one explicit measure and one implicit measure. First, participants indicated "on a scale from 0 to 99, how you feel about Black people" (0 = *extremely unfavourable*, 50 = *neutral*, 99 = *extremely favourable*). Participants also completed

the free online version of the Black vs. White Implicit Associations Test (Project

Implicit, 2018). This free online version of the test automatically provided feedback

about participants' implicit associations that fell into one of seven categories.

For the purpose of this study, participants were asked to report the feedback

they received from the online task using the following coding: 1 = *strong automatic*

preference for White people compared to Black people; 2 = *moderate automatic*

preference for White people compared to Black people; 3 = *slight automatic*

preference for White people compared to Black people; 4 = *little to no preference*; 5 =

slight automatic preference for Black people compared to White people; 6 = *moderate*

automatic preference for Black people compared to White people; 7 = *strong*

automatic preference for Black people compared to White people. The implicit and

explicit prejudice scores were also significantly (though not strongly) correlated ($r =$

.24, $p =$ .003) and the pattern of results was the same whether the two items were used

together or used as individual items in separate analyses[i]. Thus, for simplicity of

presentation, the implicit scores were transformed such that they fell on the same

scale as the explicit scores (i.e., a minimum of 0 and a maximum of 99), and the mean

of the implicit and explicit scores was used as the external measure of participants'

racial egalitarianism. Higher values indicated more racial egalitarianism.

Participants then took part in a two-hour diversity training session that

followed the structure adopted by most diversity training sessions (Atewologun et al.,

2018). It included an explanation of participants' IAT results, education on the theory

behind unconscious bias, experimental results and statistics highlighting the

prevalence of contemporary racism, and suggestions for ways of reducing prejudice

or its effects. At the end of the session, participants once again completed the two

items measuring their perceptions of own levels of racial egalitarianism. These two

items were once again correlated ($r = .72, p < .001$), so the mean of these items was used as the measure of participants' self-perceptions after the diversity training.

Results

Descriptive statistics and correlations between relevant variables are included in Table 1. A multivariate analysis of variance with participant sex and ethnicity as independent variables revealed that neither participant sex, $F (6, 274) = .88, p = .51$, nor participant ethnicity, $F (12, 414) = 1.57, p = .10$, predicted any of the variables of interest, so neither was considered further. However, participant age was related to their externally measured egalitarianism scores ($r = -.25, p = .002$), and their self-perception scores at the second time point ($r = -.24, p = .008$), as well as marginally related to their self-perception scores at the first time-point ($r = -.16, p = .053$). Thus, wherever possible, age is controlled by including it as a covariate in the following analyses.

Applying the Dunning-Kruger model to (racial) egalitarianism. The first prediction was that participants would, overall, overestimate their racial egalitarianism. This was supported by the data. On average, participants rated their racial egalitarianism as being in the 75th percentile, which exceeded the actual mean percentile (50, by definition) by 25 percentile points, one-sample $t (147) = 21.32, p < .001, d = 1.75$. This overestimation occurred even though self-ratings of egalitarianism were positively correlated with the external measure ($r = .53, p < .001$).

The second prediction was that this overestimation would be highest among the least competent individuals. In the original paper on which this current research is based, Kruger and Dunning (1999) used an ANOVA-based analysis strategy in which participants were divided into quartiles based on their objective scores, and their

objective and self-perceived scores were compared by quartile to investigate patterns of overestimation.

In this current study, that ANOVA-based analysis also revealed the hypothesised pattern of increasing overestimation as competence decreased[ii]. However, to increase the clarity and simplicity of presentation, we used more straightforward regression-based analyses in which participants' overestimation of their egalitarianism was predicted by their externally measured egalitarianism (with age included as a predictor to control for its effects). To reduce the possibility of finding spurious relationships due to differences in units of measurement, all scores were standardized prior to the regression analyses. Participants' self-perceived and externally measured egalitarianism were derived as stated above. Participants' *overestimation* of their egalitarianism was derived by subtracting their standardized externally measured egalitarianism score from their standardized self-perceived egalitarianism score at Time 1. As hypothesised, higher levels of externally measured egalitarianism predicted less overestimation of one's egalitarianism ($ß = -.50$, $p <$ .001, $R^2 = .24$). Thus, as expected, the least egalitarian participants were also the ones who overestimated their egalitarianism the most (see Figure 1).

The effect of diversity training. Finally, to investigate whether the diversity training programme reduced participants' tendency to overestimate their levels of egalitarianism, we compared their overestimation of their egalitarianism before the diversity training to their overestimation of their egalitarianism after the diversity training (derived by subtracting their standardized externally measured egalitarianism score from their standardized self-perceived egalitarianism score at Time 2). This was done using a repeated-measures ANOVA with time (before vs. after) as the independent variable, overestimation as the dependent variable, and participant age

included as a covariate. However, we found no effect of diversity training on participants' overestimation of their egalitarianism $F(1, 145) = 2.42$, $p = .12$, $\eta_p^2 = .02$. Repeating the analyses using participants' unstandardized self-perception scores as the dependent variable, we similarly found no effect of the diversity training $F(1, 145) = .84$, $p = .36$, $\eta_p^2 = .01$, indicating that the diversity training did not reduce participants' self-perceptions or overestimation of their egalitarianism.

Supplementary analyses. A possible concern about the analyses above is that asking participants to assess themselves relative to the British population as a whole (rather than just the other participants for whom we had available data) is not a fair test of whether they overestimate their own egalitarianism. For this reason, the analyses were repeated using only participants' estimation of themselves "relative to the other people in this programme" as the measure of self-perceived egalitarianism. However, this did not alter the pattern of results. There was still a significant, negative relationship between participants' levels of externally measured egalitarianism and their overestimation of their egalitarianism ($\beta = -.49$, $p < .001$).

Study 2

Study 1 found support for the application of the Dunning-Kruger model in the domain of racism, but failed to find a significant effect of diversity training on participants' overestimation of their racial egalitarianism. Study 2 aimed to (a) replicate the Dunning-Kruger pattern of overestimation in a new domain, i.e., sexism and (b) perform a second test of the effect of diversity training on the tendency to overestimate one's egalitarianism.

Sexism remains a serious, widespread concern. In most places, even in contemporary times, women continue to hold less social, political and financial power than men do (Office for National Statistics, 2013; Zentner & Mitura, 2012). In some

populations the subordinate status of women is more explicit, but even in relatively

egalitarian societies, subtle and implicit sexism continue to play meaningful roles,

affecting women's abilities to achieve equal status (Glick & Fiske, 2001; Maltby,

Hall, Anderson, & Edwards, 2010; Ridgeway, 2001; Rudman & Kilianski, 2000;

Whyte, 1978). Sexism has been linked to a variety of negative outcomes, ranging

from resistance to female authority figures, to rape myth acceptance and sexual

assault (Abrams, Viki, Masser, & Bohner, 2003; Koss & Dinero, 1989; Rudman &

Kilianski, 2000; Taschler & West, 2017).

Method

Participants. One hundred and fifty-nine participants (34 men, 125 women,

mean age = 23.08, *SD* = 5.07) were recruited through their involvement in a sexism-

related diversity training programme aimed at graduate students in London. Most of

the participants were White (115, 72.3%). Smaller proportions of the participants

were Black (13, 8.2%) South Asian (16, 10.1%), East Asian (2, 1.3%), or other (15,

10.1%).

Procedure. The procedure was almost identical to that of Study 1, except that

all stimuli, diversity training, and measures related to sexism, rather than racism. This

required some minor modifications to the measures. For example, participants

indicated their perceptions of their own levels of gender-based (rather than race-

based) egalitarianism, indicated their responses to women in the workplace (rather

than to Black people), completed the Gender-Career Implicit Associations Test (rather

than the Black-White implicit associations test; Project Implicit, 2018), and took part

in diversity training related to sexism, rather than racism.

Apart from these minor alterations the procedure was the same. As in Study 1,

the two items measuring participants' self-perceptions of their egalitarianism were

strongly correlated both before ($r = .62, p < .001$) and after ($r = .63, p < .001$) the

diversity training, Thus the mean of the two items was used as the measure of

participants' perception of their own racial egalitarianism at both time points. Also, as

in Study 1, participants' explicit and implicit measures of bias were significantly

correlated ($r = .37, p < .001$), and the pattern of results was the same whether the two

items were used together or used as individual items in separate analyses[iii]. Thus, for

the sake of clarity and simplicity, the mean of these two items (after appropriate

transformation) was used as the external measure of participants' gender-based

egalitarianism.

Results

Descriptive statistics and correlations between relevant variables are included

in Table 2. A multivariate analysis of variance with participant sex and ethnicity as

independent variables revealed that neither participant sex, $F (3, 147) = .83, p = .48$,

nor participant ethnicity, $F (12, 447) = 1.33, p = .20$, predicted any of the variables of

interest, so neither was considered further. Participant age was also not related to any

of the variables of interest ($-.15 < r < -.05, .058 < p < .50$) and was not considered

further.

Applying the Dunning-Kruger model to (gender) egalitarianism. The first

prediction was that participants would, overall, overestimate their gender-based

egalitarianism. On average, participants rated their gender-based egalitarianism as

being in the 69th percentile, which exceeded the actual mean percentile (50, by

definition) by 19 percentile points, one-sample $t (158) = 17.17, p < .001, d = 1.36$.

This overestimation occurred even though self-ratings of egalitarianism were

positively correlated with the objective measure ($r = .27, p = .001$).

The second prediction was that this overestimation would be highest among the least competent individuals. As in Study 1, the original ANOVA-based analysis strategy used by Kruger and Dunning (1999) did reveal the hypothesised pattern of increasing overestimation as competence decreased[iv]. However, as in Study 1, we used a clearer, more straightforward regression-based analyses in which participants' overestimation of their egalitarianism was predicted by their externally measured egalitarianism. Again, all scores were standardized prior to the regression analyses, participants' overestimation of their egalitarianism was derived by subtracting their standardized externally measured egalitarianism score from their standardized self-perceived egalitarianism score at Time 1. As hypothesised, higher levels of externally measured egalitarianism predicted less overestimation of one's egalitarianism ($\beta = -.61$, $p < .001$, $R^2 = .37$). Thus, as expected, the least egalitarian participants were also the ones who overestimated their egalitarianism the most (see Figure 2).

The effect of diversity training. As in Study 1, we found no effect of diversity training on participants' overestimation of their egalitarianism $F(1, 157) < .001$, $p = .99$, $\eta_p^2 < .001$. Repeating the analyses using participants' unstandardized self-perception scores as the dependent variable, we similarly found no effect of the diversity training $F(1, 157) < .001$, $p = .99$, $\eta_p^2 < .001$, indicating that the diversity training did not reduce participants' self-perceptions or overestimation of their egalitarianism.

Supplementary analyses. As is Study 1, we repeated the analyses using only participants' estimation of themselves "relative to the other people in this programme" as the measure of self-perceived egalitarianism. However, this did not alter the pattern of results. There was still a significant, negative relationship between

participants' levels of externally measured egalitarianism and their overestimation of their egalitarianism ($\beta = -.61$, $p = .001$).

Discussion

In contemporary Western societies, the combination of ubiquitous prejudice and strong tendencies for individuals to deny or ignore their own levels of bias has spawned many areas of fruitful research in social psychology (Nosek et al., 2007; Pearson et al., 2009). While much prior research has investigated affective or motivational processes behind this apparent paradox, this current research investigated a potential meta-cognitive explanation.

Conceptualising egalitarianism as a skill, two studies found results in line with the Dunning-Kruger model of incompetence and overconfidence (Kruger & Dunning, 1999). Specifically, the least egalitarian individuals (i.e., the most racist in Study 1, and the most sexist in Study 2) were also those who most strongly overestimated their levels of egalitarianism. These findings suggested that highly prejudiced individuals deny their levels of bias (in part) because they lack the *meta-cognitive skill* necessary to recognise their own (lack of) egalitarianism. Interestingly, however, diversity training did *not* reduce the tendency to overestimate one's level of egalitarianism. Below, these findings are discussed with reference to study design, results, future research, and implications for contemporary bias and diversity training as a bias-managing strategy.

Implications

This current research can be added to the body of work showing that the unskilled are also unaware; i.e., that meta-cognitive deficits lead not only to incompetence, but also to an inability to recognise that incompetence (Anstey et al., 2005; Davis et al., 2006; Dunning et al., 2004; Kruger & Dunning, 1999). As such it

provides further support for the Dunning-Kruger model and strengthens the case for its general applicability. More specifically, concerning prejudice and egalitarianism, this research provides important evidence that being egalitarian may not simply be a matter of choice or motivation, but also a matter of skill or ability.

None of this undermines the wealth of research on motivational factors related to prejudice (e.g., Amodio, Harmon-Jones, & Devine, 2003; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Nosek et al., 2007; Pearson et al., 2009). However, it does suggest a variety of seldom explored avenues for research on prejudice-reduction strategies. Many strategies currently in use and under investigation seek to address the motivational side of egalitarianism, either by improving the participants' evaluation of the target of prejudice, or by making non-egalitarian behaviour seem less rewarding (Becker, Wright, Lubensky, & Zhou, 2013; Does, Derks, & Ellemers, 2011; Wright & Lubensky, 2008). However, if egalitarianism is also a skill, many strategies could be adapted from research on improving skills in other domains, such as reading, linguistic competence and academic achievement (Cejudo, Salido-López, & Rodrigo-Ruiz, 2017; Kleitman & Gibson, 2011; Sedaghat, Abedin, Hejazi, & Hassanabadi, 2011; Vázquez - Cano, 2017). Indeed, even if motivational concerns were fully addressed, this research suggests that some individuals would still lack the ability to act in highly egalitarian ways.

Research Design and Results

These studies have many strengths that should be acknowledged. First, the two studies investigated the applicability of the Dunning-Kruger model in two different domains of prejudice (racism and sexism) and nonetheless found remarkably similar results. This replication across the two studies should increase confidence in the reliability of the findings, as should the very large effect sizes that were found for the

hypothesised effects (Study 1; $d = 1.75$, $R^2 = .24$, Study 2; $d = 1.36$, $R^2 = .37$). Similarly, both studies used measures and procedures that were very similar to those of the original studies (Kruger & Dunning, 1999), and can thus be considered a successful replication of this research in a new domain.

That said, this research also has limitations. The participants for both studies were drawn from convenience samples from diversity training programmes aimed at graduate students. Thus, like many other studies in social psychology, this research used participants who were not likely to be representative of the broader population (Henrich, Heine, & Norenzayan, 2010; Sears, 1986). Given their relatively young age, student status, and voluntary involvement in a diversity training programme, these participants were probably more egalitarian, and more motivated to be egalitarian, than the broader population. Furthermore, as the measures were explicit and transparent, it is likely that social desirability led participants to present themselves as less biased. While these are genuine limitations that reduce the generalisability of these findings, it should be also noted that participants' higher levels of egalitarianism throughout the sample should have masked, rather than exaggerated, the effects found in these studies. Future research should replicate these findings with a more diverse sample of participants. However, it is likely that these replications would make the pattern of results more pronounced, not less so.

The measures of participants' self-perceptions were adapted from the original research (Kruger & Dunning, 1999) with very little alteration. However, the external measures of participants' egalitarianism were devised for this current research. While the combination of explicit and implicit measures of bias can be considered a strength, it is also noteworthy that both the explicit and implicit measures had limitations. Explicit egalitarianism was measured a single item, which could raise

concerns about the reliability of this measure, and how it would relate to other measures. However, contrary to some expectations, research in a variety of domains has found single-item measures to have a predictive validity similar to (or even, at times, equal to) that of multi-item measures (Bergkvist & Rossiter, 2007; Pettigrew & Tropp, 2006; Wanous, Reichers, & Hudy, 1997). Nonetheless, future research with multi-item measures would be useful for confirming and extending these findings.

It should also be noted that participants' levels of implicit bias were not measured directly. Rather, participants *reported* their levels of bias as indicated to them via the free online version of the Implicit Associations Test (IAT; Project Implicit, 2018). The free online version of the IAT conferred several advantages that made it easy and practical to use during the diversity training: speed, clarity of presentation, availability on multiple computers simultaneously, and instantaneous feedback. Due to ethical restrictions on participant information imposed during the diversity training, it was also necessary to allow participants to relay the results of their IAT's, rather than observe them directly. This raised a number of potential concerns. First, the free version of the IAT does not provide a very fine-grained analysis of participants' implicit biases, but merely one of seven options indicating (strong/moderate/slight/no) preference for one group over another. Second, allowing participants to report their own feedback allowed for the possibility of deception, error, or self-presentation biases, undermining some key advantages of implicit measures. Also, as the order of the measures was the same for all participants (with the implicit measures coming last), participants may have felt a desire for consistency, leading them to underplay their implicit biases. Again, this masking of implicit bias should have undermined, rather than exaggerated, the effects found here. Nonetheless,

future research should re-examine these hypotheses with more sensitive and more genuinely implicit measures.

On a more nuanced level, there is the question of whether the measures were appropriately matched. Specifically, can responses to Black people be used as an external measure of a concept as broad as participants' "racial egalitarianism"? Similarly, responses to women in a single domain (the workplace) are not the same as "gender egalitarianism", which contains many, sometimes conflicting, aspects (Glick & Fiske, 1996). While this potential mismatch is a limitation, it is a consequence of using participants' self-evaluations of a broad construct, while being practically restricted to using a measure of that construct that cannot capture every aspect of it. Notably, this limitation applies to all similar research on the Dunning-Kruger model. It is not the case, for example, that a single survey or written test could be a complete measure of competence in domains as broad as humour or logical reasoning (Kruger & Dunning, 1999). Indeed, it is unlikely that any single numerical score could be an ideal indicator of any of the skills, such as driving or health education, that have been explored in prior related research (Anstey et al., 2005; Davis et al., 2006; Dunning et al., 2004). The solution to this problem seems to be in the replication of these findings (as found in these two studies) across multiple domains with different target groups and measures. Future research could explore other aspects of both racism and sexism to investigate whether (or when) the pattern of results changes.

Also concerning the measures, neither of the external measures of egalitarianism focused on participants' *behaviours*. Both focused on participants thoughts or attitudes, whether explicit or implicit. While both explicit and implicit attitudinal measures are valid measures of egalitarianism (Nosek et al., 2007), much prior research has noted important disjunctions between thoughts, attitudes, and

behaviours such that one is not always strongly predictive of the others (Arcuri, Castelli, Galdi, Zogmaister, & Amadori, 2008; Bauer & Baltes, 2002; Crisp, Husnu, Meleady, Stathi, & Turner, 2010; Eastwick, Richeson, Son, & Finkel, 2009; Klein, Snyder, & Livingston, 2004; McConnell & Leibold, 2001). Behaviour is, at the very least, as important as thoughts or attitudes when considering an individual's egalitarianism. Thus, future research using behavioural measures would add meaningfully to this research.

Finally, though this research replicated the Dunning-Kruger pattern in the domain of egalitarianism, it did not investigate specific meta-cognitive deficits that could account for the findings. Kruger and Dunning (1999) alluded to at least two plausible mechanisms: simplistic or incomplete perceptions of the domain in question (e.g., an understanding of prejudice that only included overt prejudice), and an inaccurate estimation of the mean level of competence in the domain (e.g., a belief that the average person is more prejudiced than they really are). Future research could explore such explanatory meta-cognitive explanations to better understand why highly prejudiced people overestimate their egalitarianism.

The (lack of) effects of diversity training.

For the most part, these results closely replicated the original findings of Kruger and Dunning (1999). However, a noteworthy area of difference was the lack of effect of diversity training. These data cannot be used to explain that lack of effect, so any comment on it must be speculative. Nonetheless, as noted before, there are important differences between the training used in the original studies and the diversity training used in this current research.

First, while participants in this research received feedback concerning their *levels* of implicit bias, this is not the same thing as feedback on their levels of bias

overall, or on their levels of bias *compared to other people*. Thus, a participant who discovered that their levels of implicit bias were high, still received no information relevant to their *overestimation* of their levels of egalitarianism. It is possible that the overconfidence due to incompetence cannot be countered without this crucial aspect of the feedback. Alternatively, the difference may be due to the nature of the training offered. While diversity training is referred to as a type of "training", it has been noted that, unlike the training offered to participants by Kruger and Dunning (1999), most diversity training does not involve any actual training in techniques to reduce bias, but rather focuses on the delivery of information about bias: a crucial difference (Dobbin & Kalev, 2016; Kaiser et al., 2013). Without this important element of training, it is understandable that participants did not improve in the competence and meta-cognition necessary to alter their perceptions of their own egalitarianism. As such, this finding may do less to undermine the application of the Dunning-Kruger model in the domains of racism and sexism, and more to highlight a limitation of a wide-spread feature of contemporary diversity training: one that has also been highlighted in prior research (Atewologun et al., 2018; Dobbin & Kalev, 2016; Noon, 2018).

Conclusions

For many reasons, contemporary discussions of prejudice can be quite acrimonious. Members of socially advantaged groups may find such discussions difficult, unpleasant, or threatening (Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; Dover, Major, & Kaiser, 2016; Norton et al., 2006). Political divisions may lead members of both advantaged and disadvantaged groups to attribute overly negative motivations to the other group (Goff et al., 2014; Reeder, 2005; Taber, Brook, & Franklin, 2006). Motivation certainly forms an important part of the picture.

However, this research suggests that, even if such motivational considerations were accounted for, there may be important cognitive hindrances to understanding and reducing prejudice that would have to be addressed. In line with the Dunning-Kruger model, this research found that very prejudiced individuals (i.e., those low in egalitarianism) may be genuinely unaware of their shortcomings because they lack the meta-cognition necessary to perceive them. It is thus possible that some solutions to contemporary prejudice may rely less on motivation and more on education.

References

Abrams, D., Viki, G. T., Masser, B., & Bohner, G. (2003). Perceptions of stranger and acquaintance rape: Role of benevolent and hostile sexism in victim blame and rape proclivity. *Journal of Personality and Social Psychology*, *84*(1), 111–125. https://doi.org/10.1037/0022-3514.84.1.111

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, *84*(4), 738–753. https://doi.org/10.1037/0022-3514.84.4.738

Andreouli, E., Greenland, K., & Howarth, C. (2016). 'I don't think racism is that bad any more': Exploring the 'end of racism' discourse among students in English schools. *European Journal of Social Psychology*, *46*, 171–184.

Anstey, K. J., Wood, J., Lord, S., & Walker, J. G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review*, *25*(1), 45–65. https://doi.org/10.1016/j.cpr.2004.07.008

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: when older children underperform in social categorization. *Developmental Psychology*, *44*(5), 1513–1518. https://doi.org/10.1037/a0012835

Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., & Amadori, A. (2008). Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, *29*(3), 369–387. https://doi.org/10.1111/j.1467-9221.2008.00635.x

Atewologun, D., Cornish, T., & Tresh, F. (2018). *Unconscious bias training : An assessment of the evidence for effectiveness*. Manchester, UK. Retrieved from

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=

rja&uact=8&ved=2ahUKEwiws__lxLjgAhXOSRUIHQKRDa4QFjABegQIAhA

C&url=https%3A%2F%2Fwww.equalityhumanrights.com%2Fsites%2Fdefault

%2Ffiles%2Fresearch-report-113-unconcious-bais-training-an-assessment-of-

the-evidence-for-effectiveness-

pdf.pdf&usg=AOvVaw2kGtA5UC0i1LuhwSJC1rJG

Augoustinos, M., & Every, D. (2010). Accusations and denials of racism: Managing

moral accountability in public discourse. *Discourse & Society*, *21*, 251–256.

https://doi.org/10.1177/0957926509360650

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on

the gender pay gap from multifactorial survey experiments. *American

Sociological Review*, *82*(1), 179–210.

https://doi.org/10.1177/0003122416683393

Bauer, C. C., & Baltes, B. B. (2002). Reducing the effects of gender stereotypes on

performance evaluations. *Sex Roles*, *47*, 465–476.

https://doi.org/10.1023/A:1021652527696

Becker, J. C., & Swim, J. K. (2011). Seeing the unseen: attention to daily encounters

with sexism as way to reduce sexist beliefs. *Psychology of Women Quarterly*,

35(2), 227–242. https://doi.org/10.1177/0361684310397509

Becker, J. C., Wright, S. C., Lubensky, M. E., & Zhou, S. (2013). Friend or ally:

whether cross-group contact undermines collective action depends on what

advantaged group members say (or don't say). *Personality & Social Psychology

Bulletin*, *39*(4), 442–455. https://doi.org/10.1177/0146167213477155

Bergkvist, L. I., & Rossiter, J. (2007). The predictive validity of multiple-item versus

single-item measures of the same constructs. *Journal of Marketing Research*,

44(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than

Lakisha and Jamal? A field experiment on labour market discrimination.

American Economic Review, *94*, 991–1013.

Biggs, J., Hawley, P. H., & Biernat, M. (2018). The academic conference as a chilly

climate for women: Effects of gender representation on experiences of sexism,

coping responses, and career Intentions. *Sex Roles*, *78*(5–6), 394–408.

https://doi.org/10.1007/s11199-017-0800-9

Bonam, C. M., Das, V. N., Coleman, B. R., & Salter, P. (2018). Ignoring history,

denying racism: Mounting evidence for the Marley Hypothesis and

epistemologies of ignorance. *Social Psychology and Personality Science*.

https://doi.org/10.1177/1948550617751583

Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary

across minority groups? Evidence from a field experiment. *Oxford Bulletin of

Economics and Statistics*, *74*(4), 547–573. https://doi.org/10.1111/j.1468-

0084.2011.00664.x

Bourabain, D., & Verhaeghe, P. P. (2018). Could you help me, please? Intersectional

field experiments on everyday discrimination in clothing stores. *Journal of

Ethnic and Migration Studies*. https://doi.org/10.1080/1369183X.2018.1480360

Bruckmüller, S., Ryan, M. K., Rink, F., & Haslam, S. A. (2014). Beyond the glass

ceiling: The glass cliff and its lessons for organizational policy. *Social Issues and

Policy Review*, *8*(1), 202–232. https://doi.org/10.1111/sipr.12006

Cejudo, J., Salido-López, J. V., & Rodrigo-Ruiz, D. (2017). Effect of a programme to

enhance proficiency in linguistic competence in secondary education. *Revista de

Psicodidáctica (English Ed.)*, *22*(2), 135–141.

https://doi.org/10.1016/j.psicoe.2016.11.001

Chasmar, J. (2018, November 1). Bernie Sanders: Trump "most racist, sexist, homophobic, bigoted president in history." *The Washington Times*, pp. 1–2. Retrieved from https://www.washingtontimes.com/news/2018/nov/1/bernie-sanders-trump-most-racist-sexist-homophobic/

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*, 1102–1117. https://doi.org/10.1002/ejsp.450

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. https://doi.org/10.1037/0022-3514.82.3.359

Crisp, R. J., Husnu, S., Meleady, R., Stathi, S., & Turner, R. N. (2010). From imagery to intention: A dual route model of imagined contact effects. *European Review of Social Psychology*, *21*(1), 188–236. https://doi.org/10.1080/10463283.2010.543312

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R. T. K. E., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Jama*, *296*(9), 1094–1102.

de Keersmaecker, J., Onraet, E., Lepouttre, N., & Roets, A. (2017). The opposite effects of actual and self-perceived intelligence on racial prejudice. *Personality and Individual Differences*, *112*, 136–138. https://doi.org/10.1016/j.paid.2017.02.057

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term

reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). Exploring the relationship between implicit and explicit prejudice: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*(Journal Article), 835–848. https://doi.org/10.1037//0022-3514.82.5.835

Dobbin, F., & Kalev, A. (2016). Why diversity programmes fail. *Harvard Businnes Review*, *94*(7), 14–20. Retrieved from https://hbr.org/2016/07/why-diversity-programs-fail

Dobbin, F., Kalev, A., & Kelly, E. (2007). Diversity management in corporate america. *Contexts*, *6*(4), 21–27. https://doi.org/10.1525/ctx.2007.6.4.21

Does, S., Derks, B., & Ellemers, N. (2011). Thou shalt not discriminate: How emphasizing moral ideals rather than obligations increases Whites' support for social equality. *Journal of Experimental Social Psychology*, *47*(3), 562–571. https://doi.org/10.1016/j.jesp.2010.12.024

Dover, T. L., Major, B., & Kaiser, C. R. (2016). Members of high-status groups are threatened by pro-diversity organizational messages. *Journal of Experimental Social Psychology*, *62*, 58–67. https://doi.org/10.1016/j.jesp.2015.10.006

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Eastwick, P. W., Richeson, J. A., Son, D., & Finkel, E. J. (2009). Is love colorblind? Political orientation and interracial romantic desire. *Personality and Social*

Psychology Bulletin, *35*(9), 1258–1268.

https://doi.org/10.1177/0146167209338524

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses

using G * Power 3 . 1 : Tests for correlation and regression analyses. *Behavior*

Research Methods, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating

hostile and benevolent sexism. *Journal of Personality and Social Psychology*,

70(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent

sexism as complementary justifications for gender inequality. *American*

Psychologist, *56*(2), 109–118. https://doi.org/10.1037//O003-066X.56.2.1O9

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A.

(2014). The essence of innocence: Consequences of dehumanizing Black

children. *Journal of Personality and Social Psychology*, *106*(4), 526–545.

https://doi.org/10.1037/a0035663

Google Scholar. (2018). Items citing Kruger & Dunning, 1999. Retrieved November

19, 2018, from

https://scholar.google.co.uk/scholar?cites=1825888681341875116&as_sdt=2005

&sciodt=0,5&hl=en

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I.,

& Banaji, M. R. (2007). Implicit bias among physicians and its prediction of

thrombolysis decisions for black and white patients. *Journal of General Internal*

Medicine, *22*(9), 1231–1238. https://doi.org/10.1007/s11606-007-0258-5

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?

The Behavioral and Brain Sciences, *33*(2–3), 61–83.

https://doi.org/10.1017/S0140525X0999152X

Hodson, G., & Busseri, M. A. (2012). Bright minds and dark attitudes: Lower

cognitive ability predicts greater prejudice through right-wing ideology and low

intergroup contact. *Psychological Science*, *23*(2), 187–195.

https://doi.org/10.1177/0956797611421206

Johns, M., Cullum, J., Smith, T., & Freng, S. (2008). Internal motivation to respond

without prejudice and automatic egalitarian goal activation. *Journal of

Experimental Social Psychology*, *44*(6), 1514–1519.

https://doi.org/10.1016/j.jesp.2008.07.003

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R.

(2013). Presumed fair: Ironic effects of organizational diversity structures.

Journal of Personality and Social Psychology, *104*(3), 504–519.

https://doi.org/10.1037/a0030838

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting

affective and behavioral responses to racism. *Science*, *323*(5911), 276–278.

https://doi.org/10.1126/science.1164951

Klein, O., Snyder, M., & Livingston, R. W. (2004). Prejudice on the stage: self-

monitoring and the public expression of group attitudes. *The British Journal of

Social Psychology*, *43*, 299–314. https://doi.org/10.1348/0144666041501697

Kleitman, S., & Gibson, J. (2011). Metacognitive beliefs, self-confidence and primary

learning environment of sixth grade students. *Learning and Individual

Differences*, *21*(6), 728–735. https://doi.org/10.1016/j.lindif.2011.08.003

Koss, M. P., & Dinero, T. E. (1989). Discriminant analysis of risk factors for sexual

victimization among a national sample of college women. *Journal of Consulting

and Clinical Psychology*, *57*(2), 242–250. https://doi.org/10.1037/0022-

006X.57.2.242

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in

recognizing one's own incompetence lead to inflated self-assessments. *Journal*

of Personality and Social Psychology, *77*(6), 1121–1134.

https://doi.org/10.1037/0022-3514.77.6.1121

Leach, C. W., Peng, T. R., & Volckens, J. (2000). Is racism dead? Comparing

(expressive) means and (structural equation) models. *The British Journal of*

Social Psychology, *39*, 449–465. https://doi.org/10.1348/014466600164507

Maltby, L. E., Hall, M. E. L., Anderson, T. L., & Edwards, K. (2010). Religion and

sexism: The moderating role of participant gender. *Sex Roles*, *62*, 615–622.

https://doi.org/10.1007/s11199-010-9754-x

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association

Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes.

Journal of Experimental Social Psychology, *37*(5), 435–442.

https://doi.org/10.1006/jesp.2000.1470

McIntosh, P. (1988). White privilege and male privilege: a personal account of

coming to see correspondences through work in women's studies.

Understanding & Managing Diversity: Readings, Cases & Exercises, 1–20.

https://doi.org/10.1080/14725880701859969

Mendelsohn, G. A., Shaw Taylor, L., Fiore, A. T., & Cheshire, C. (2014).

Black/White dating online: Interracial courtship in the 21st century. *Psychology*

of Popular Media Culture, *3*(1), 2–18. https://doi.org/10.1037/a0035357

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.

https://doi.org/10.1140/epjd/e2007-00070-4

Nelson, J. C., Adams, G., & Salter, P. S. (2013). The Marley hypothesis: denial of

racism reflects ignorance of history. *Psychological Science*, *24*(2), 213–218.

https://doi.org/10.1177/0956797612451466

Neumark, R. J. B., & Van Nort, Kyle, D. (1996). Sex discrimination in restaurant

hiring: An audit study. *The Quarterly Journal of Economics*, *111*(3), 915–941.

Noon, M. (2018). Pointless diversity training: Unconscious bias, new racism and

agency. *Work, Employment and Society*, *32*(1), 198–209.

https://doi.org/10.1177/0950017017719841

Norton, M. I., & Sommers, S. R. (2011). Whites see racism as a zero-sum game that

they are now losing. *Perspectives on Psychological Science*, *6*(3), 215–218.

https://doi.org/10.1177/1745691611406922

Norton, M. I., Sommers, S. R., Apfelbaum, E. P., Pura, N., & Ariely, D. (2006). Color

blindness and interracial interaction. *Psychological Science*, *17*(11), 949–953.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association

Test at age 7: A methodological and conceptual review. *Automatic Processes in*

Social Thinking and Behavior, 265–292.

https://doi.org/10.1016/j.mrfmmm.2009.01.007

Office for National Statistics. (2013). Full report - Women in the labour market, 1–20.

Pager, D. (2003). The mark of a criminal record. *American Sociological Review*,

108(5), 937–975. https://doi.org/10.1086/374403

Pager, D., Western, B., & Bonokowski, B. (2006). Discrimination in a low wage labor

market: A field experiment. *American Sociological Review*, *74*(4469), 777–789.

Pauker, K., Apfelbaum, E. P., & Spitzer, B. (2015). When societal norms and social

identity collide: The race talk dilemma for racial minority children. *Social*

Psychological and Personality Science, *6*(8), 887–895.

https://doi.org/10.1177/1948550615598379

Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary

prejudice: Insights from aversive racism. *Social and Personality Psychology*

Compass, *3*, 1–25. https://doi.org/10.1111/j.1751-9004.2009.00183.x

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact

theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.

https://doi.org/10.1037/0022-3514.90.5.751

Plant, E. A., & Peruche, B. M. (2005). The consequences of race for police officers'

responses to criminal suspects. *Psychological Science*, *16*(3), 180–183.

https://doi.org/10.1111/j.0956-7976.2005.00800.x

Project Implicit. (2018). Project Implicit. Retrieved November 20, 2018, from

https://implicit.harvard.edu/implicit/

Reeder, G. D. (2005). On attributing negative motives to others who disagree with our

opinions. *Personality and Social Psychology Bulletin*, *31*(11), 1498–1510.

https://doi.org/10.1177/0146167205277093

Ridgeway, C. L. (2001). Gender, status , and leadership. *Journal of Social Issues*,

57(4), 637–655. https://doi.org/10.1111/0022-4537.00233

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward

female authority. *Personality and Social Psychology Bulletin*, *26*(11), 1315–

1328. https://doi.org/10.1177/0146167200263001

Russell, B. (1979). *A History of Western Philosophy (2nd Edition)*. London, UK:

Unwin Hyman.

Rutland, A., Cameron, L., Milne, A., & Mcgeorge, P. (2005). Social norms and self-

presentation: Children's implicit and explicit intergroup attitudes. *Child*

Development, *76*(2), 451–466. https://doi.org/10.1111/j.1467-8624.2005.00856.x

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016).

Getting on top of the glass cliff: Reviewing a decade of evidence, Explanations,

And impact. *Leadership Quarterly*, *27*(3), 446–455.

https://doi.org/10.1016/j.leaqua.2015.10.008

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the

unskilled? Empirical tests of the "signal extraction" counterexplanation for the

Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic*

Psychology, *39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Schreer, G. E., Smith, S., & Thomas, K. (2009). "Shopping while Black": Examining

racial discrimination in a retail setting. *Journal of Applied Social Psychology*,

39(6), 1432–1444.

Scott, E. (2018, January 17). Six times President Trump said he is the least racist

person. *The Washington Post*, pp. 15–18. Retrieved from

https://www.washingtonpost.com/news/the-fix/wp/2018/01/17/six-times-

president-trump-said-he-is-the-least-racist-

person/?noredirect=on&utm_term=.9e1ab53b958b

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow

data base on social psychology's view of human nature. *Journal of Personality*

and Social Psychology, *51*(3), 515–530. https://doi.org/10.1037/0022-

3514.51.3.515

Sedaghat, M., Abedin, A., Hejazi, E., & Hassanabadi, H. (2011). Motivation,

cognitive engagement, and academic achievement. *Procedia - Social and*

Behavioral Sciences, *15*, 2406–2410.

https://doi.org/10.1016/j.sbspro.2011.04.117

Sommers, S. R., & Ellsworth, P. C. (2000). Race in the courtroom: Perceptions of

guilt and dispositional attributions. *Personality and Social Psychology Bulletin*,

26(11), 1367–1379. https://doi.org/10.1177/0146167200263005

Swim, Aikin, Hall, & Hunter. (1995). Sexism and racism: Old fashioned and modern

prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199–214.

https://doi.org/0022-3514/95/S3.00

Taber, C. S., Brook, S., & Franklin, B. (2006). Motivated skepticism in the evaluation

of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Taschler, M., & West, K. (2017). Contact with counter-stereotypical women predicts

less sexism, less rape myth acceptance, less intention to rape (in men) and less

projected sexualisation of rape (in women). *Sex Roles*, *76*, 473–484.

https://doi.org/10.1007/s11199-016-0679-x

Vázquez - Cano, E. (2017). Analysis of difficulties of Spanish teachers to improve

students' digital reading competence. A case study within the PISA framework.

Pedagogika, *125*(1), 175–194. https://doi.org/10.15823/p.2017.13

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How

good are single-item measures ? *Journal of Applied Psychology*, *82*(2), 247–252.

https://doi.org/0021-90HV97/S3.00

Watt, D. (2017). Dealing with difficult conversations: anti-racism in youth &

community work training. *Race Ethnicity and Education*, *20*(3), 401–413.

https://doi.org/10.1080/13613324.2016.1260235

West, K. (2018). Interethnic bias in willingness to engage in casual sex versus

committed relationships. *Journal of Sex Research*.

https://doi.org/10.1080/00224499.2018.1546372

West, K., & Hewstone, M. (2012). Relatively socially acceptable prejudice within and

between societies. *Journal of Community and Applied Social Psychology*, *22*(3), 269–282. https://doi.org/10.1002/casp.1112

West, K., & Lloyd, J. (2017). The role of labeling and bias in the portrayals of acts of "terrorism": Media representations of Muslims vs. non-Muslims. *Journal of Muslim Minority Affairs*, *37*(2), 211–222. https://doi.org/10.1080/13602004.2017.1345103

West, K., Lowe, R., & Marsden, V. (2017). 'It don't matter if you're Black or White'?: Aversive racism and perceptions of interethnic romantic relationships. *Social Psychology Review*, *19*(1), 11–19.

Whyte, K. M. (1978). Cross-cultural codes dealing with the relative status of women. *Ethnology*, *17*(2), 211–237. https://doi.org/10.2307/3773145

Wright, S. C., & Lubensky, M. (2008). The struggle for social equality: Collective action vs prejudice reduction. In S. Demoulin, J. P. Leyens, & J. F. Dovidio (Eds.), *Intergroup misunderstandings: Impact of divergent social realities.* (pp. 291–310). New York: Psychology Press.

Zentner, M., & Mitura, K. (2012). Stepping out of the caveman's shadow: Nations' gender gap predicts degree of sex differentiation in mate preferences. *Psychological Science*, *23*(20), 1176–1185. https://doi.org/10.1177/0956797612441004

Tables

Table 1: Descriptive statistics and correlations between participants' externally-measured racial egalitarianism, their self-ratings of their egalitarianism, and their overestimation of their egalitarianism. The latter 2 measures come from 2 time-points – before and after diversity training. (Study 1).

		(1)	(2)	(3)	(4)	(5)
1. External measures						
2. Self-ratings (T1)		.53***				
3. Self-ratings (T2)		.59***	.83***			
4. Overestimation (T1)		-.49***	.49***	.25***		
5. Overestimation (T2)		-.46***	.34***	.46***	.81***	
	M	57.45	75.05	72.01	.00	-.01
	SD	16.31	14.29	15.33	.97	.91

Notes: 1) $* = p < .05$; $** = p < .01$; $*** = p < .001$

2) Both self-ratings and externally measured scores ranged from 0 to 99.

3) Participants' overestimation of their racial egalitarianism was derived by subtracting their standardized externally measured egalitarianism score from their standardized self-perceived egalitarianism score.

Table 2: Descriptive statistics and correlations between participants' externally-measured gender egalitarianism, their self-ratings of their egalitarianism, and their overestimation of their egalitarianism. The latter 2 measures come from 2 time-points – before and after diversity training. (Study 1).

		(1)	(2)	(3)	(4)	(5)
1. External measures						
2. Self-ratings (T1)		.27**				
3. Self-ratings (T2)		.26**	.75***			
4. Overestimation (T1)		-.61***	.61***	.41***		
5. Overestimation (T2)		-.61***	.40***	.61***	.83***	
	M	54.20	69.28	69.14	.00	-.01
	SD	15.86	14.16	14.90	1.21	1.22

Notes: 1) $* = p < .05$; $** = p < .01$; $*** = p < .001$

2) Both self-ratings and externally measured scores ranged from 0 to 99.

3) Participants' overestimation of their gender egalitarianism was derived by subtracting their standardized externally measured egalitarianism score from their standardized self-perceived egalitarianism score.

Figures

Figure 1: Participants' overestimation of their (racial) egalitarianism as predicted by

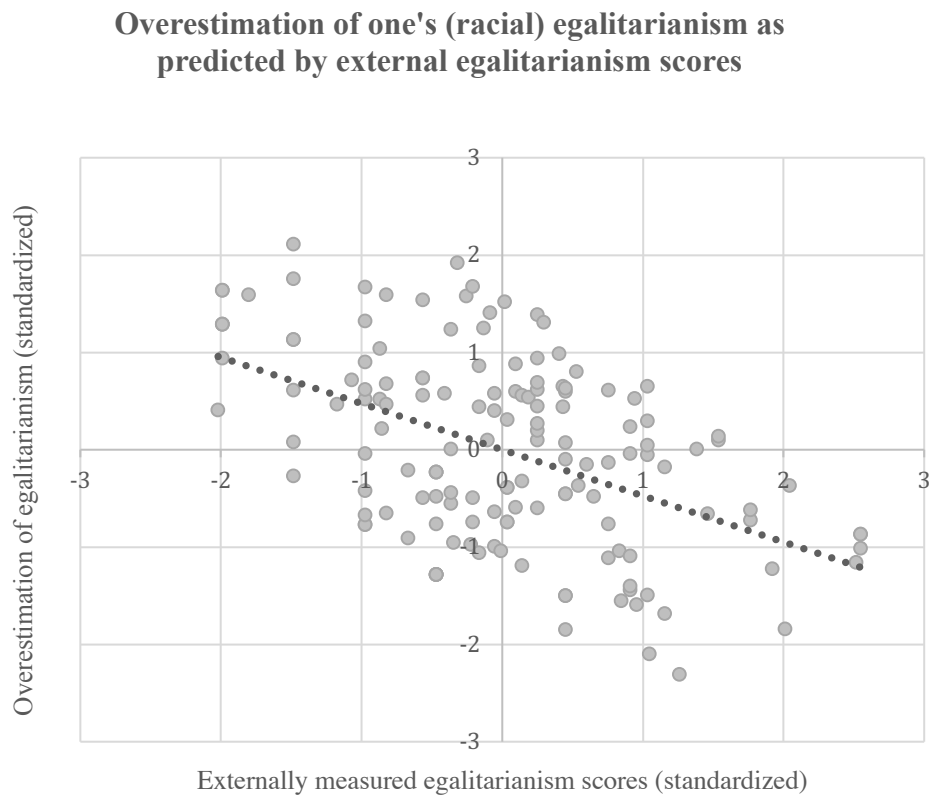their externally measured egalitarianism scores.

Overestimation of one's (racial) egalitarianism as predicted by external egalitarianism scores

Figure 2: Participants' overestimation of their (gender-based) egalitarianism as predicted by their externally measured egalitarianism scores.

Overestimation of one's (gender) egalitarianism as predicted by external egalitarianism scores

Externally measured egalitarianism scores (standardized)
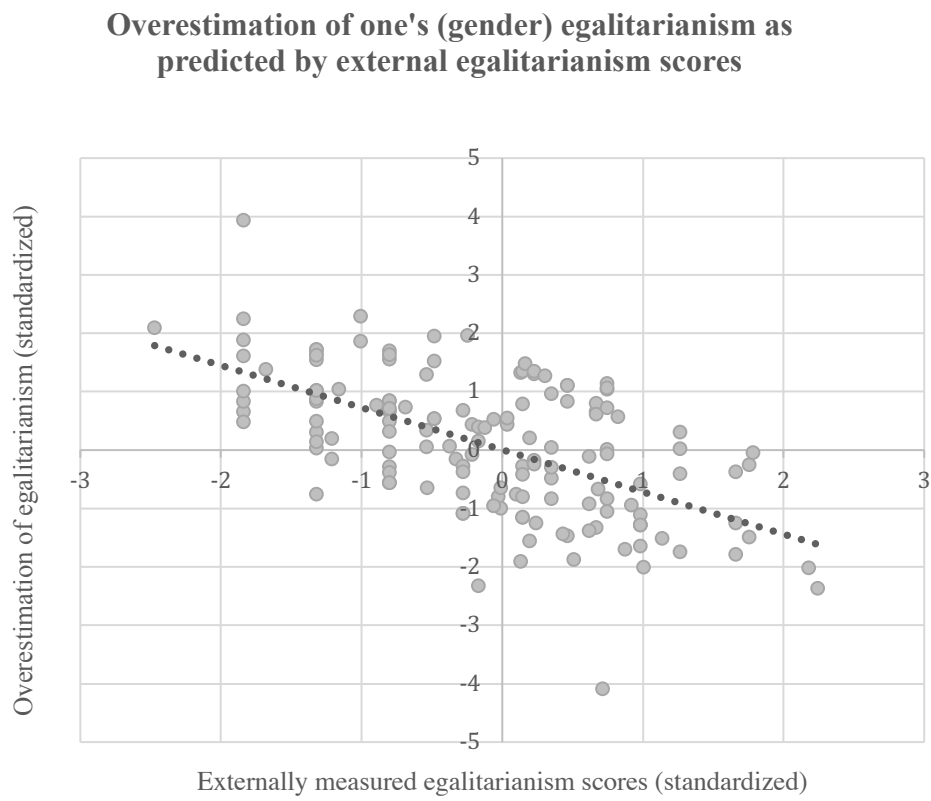
Notes

---

[i] In Study 1, the relationship between participants' *actual* levels of racial egalitarianism and their *overestimation* of their egalitarianism (the central analysis in question) was negative and significant whether analysed using combined implicit and explicit measures ($\beta = -.50$, $p < .001$), the implicit measure alone ($\beta = -.82$, $p < .001$), or the explicit measure alone ($\beta = -.64$, $p < .001$).

[ii] A 2 (Rating type: Self-perception vs. Objective score) x 4 (Quartile: 1st Quartile, i.e., bottom, 2nd Quartile, 3rd Quartile, 4th Quartile, i.e., top) analysis of variance with repeated measures on the first variable, age included as a covariate, and post-hoc Bonferroni-adjusted t-tests found the expected significant interaction of rating type and quartile $F(3, 143) = 23.14$, $p < .001$, $\eta_p^2 = .33$. This was in the expected direction. The least egalitarian participants most strongly overestimated their egalitarianism relative to their objective performance. While their actual performance fell in the 38th percentile, they rated themselves as being in the 67th percentile; $t(40) = 15.40$, $p < .001$. Contrastingly, the most egalitarian participants overestimated their performance the least. Their actual performance fell in the 78th percentile, and they rated themselves as being in the 83rd percentile; $t(36) = 2.77$, $p = .009$.

[iii] In Study 2, the relationship between participants' *actual* levels of racial egalitarianism and their *overestimation* of their egalitarianism (the central analysis in question) was negative and significant whether analysed using combined implicit and explicit measures ($\beta = -.61$, $p < .001$), the implicit measure alone ($\beta = -.76$, $p < .001$), or the explicit measure alone ($\beta = -.76$, $p < .001$).

[iv] A 2 (Rating type: Self-perception vs. Objective score) x 4 (Quartile: 1st Quartile, 2nd Quartile, 3rd Quartile, 4th Quartile) analysis of variance with repeated measures on the first variable and post-hoc Bonferroni-adjusted t-tests found the expected significant interaction of rating type and quartile $F(3, 155) = 33.15$, $p < .001$, $\eta_p^2 = .39$. This was in the expected direction. The least egalitarian participants most strongly overestimated their egalitarianism relative to their objective performance. While their actual performance fell in the 34th percentile, they rated themselves as being in the 64th percentile; $t(45) = 15.04$, $p < .001$. Contrastingly, the most egalitarian participants did not overestimate their performance at all. Their actual performance fell in the 73rd percentile, and they rated themselves as being in the 75th percentile; $t(33) = .97$, $p = .34$.