Perceived moral responsibility for attitude-based Discrimination

Liz Redford* and Kate A. Ratliff
University of Florida, Gainesville, Florida, USA

This research investigated judgements of moral responsibility for attitude-based discrimination, testing whether a wrongdoer's mental states – awareness and foresight – are central determinants of culpability. Participants read about and judged a target person who was described as consciously egalitarian, but harbouring negative attitudes that lead him to treat African Americans unfairly. Two studies showed that participants ascribed greater moral responsibility for discrimination when the target was aware of having negative attitudes than when he was unaware. Surprisingly, moral judgements were equally harsh towards a target who was explicitly aware that his bias could influence his behaviour as a target who was not. To explain this result, a second study showed that the path from awareness to moral responsibility was mediated by perceptions that the target had an obligation to foresee his discriminatory behaviour, but not by perceptions of the target's actual foresight. These results suggest that bias awareness influences moral judgements of those who engage in attitude-based discrimination because it obligates them to foresee harmful consequences. The current findings demonstrate that moral judges consider not just descriptive facts, but also normative standards regarding a wrongdoer's mental states.

Imagine a White person who considers her- or himself egalitarian, but actually harbours negative attitudes towards Black people. Now imagine this person is a company manager, making important hiring and promotion decisions. To what extent is the manager morally responsible if this negativity leads her or him to discriminate against Black people? Does it matter if the manager is aware or unaware of her or his negativity towards Black people and its influence on behaviour?

This scenario is typical of discrimination based on implicit biases. Although most White Americans want to view themselves as racially unbiased, they may harbor preferences for their own race relative to others that are activated automatically: Without awareness or intentional control (Dasgupta, 2013; O'Brien et al., 2010). For example, viewing pictures of Black people's faces generally facilitates White participants' responses to negative words and hinders their responses to positive words (Fazio, Jackson, Dunton, & Williams, 1995). Many White participants pair White people's faces with positive words and Black people's faces with negative words more quickly than the reverse; differences in response speeds from these pairings indicate greater associations of positivity with White people relative to Black people (Nosek et al., 2007). Despite conscious commitment to egalitarianism, such automatic racial preferences contribute to discriminatory behaviour in hiring decisions (e.g., Rooth, 2007), interpersonal behavior (e.g., Rudman & Ashmore, 2007), medical outcomes (e.g., Green et al. , 2007), and police shooting (e.g., Correll et al. , 2007) and may do so without the knowledge of the person who is discriminating.

The present research addresses the extent to which people hold others morally responsible for such attitude-based discriminatory behaviours. We answer two specific questions. First, when someone is not consciously aware of having negative attitudes towards a racial group, to what extent do others hold that person responsible for discriminatory behaviour resulting from that negative attitude? Second, when someone is aware of having negative attitudes, but is not consciously aware that those negative\

attitudes could influence their behaviour, to what extent do others hold that person responsible for subsequent discriminatory behaviour? To answer these questions, the current research draws on previous findings in moral judgement literature.

Moral responsibility for behaviour

Judgements about moral responsibility concern the extent to which someone is at fault for behaviour that violates moral norms (Eshleman, 2014). In general, people assign greater moral responsibility to wrongdoers who are aware that their behaviour could result in a negative outcome than to wrongdoers who are not aware that their behaviour could result in a negative outcome. For example, Nadler and McDonnell (2011) found that if a person's dog mauls a child, others blame her more if she was aware that her dog behaves badly than if shewas unaware of the dog's bad temperament. Similarly, Lagnado and Channon (2008) found that when a person is described as being aware that she made a chair poorly and that chair harms someone when it breaks, participants judge her as more blameworthy than if she believed she was unaware of her poor craftsmanship.

Because implicit racial biases are generally considered to operate automatically (Moors & De Houwer, 2006), they raise important questions concerning whether the concepts that guide moral responsibility for explicit attitudes also apply to implicit attitudes. For example, unlike explicit bias, implicit bias is not necessarily consciously endorsed (e.g., Gawronski & Bodenhausen, 2007; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Such endorsement (or 'identification with') is one widely accepted philosophical condition for moral responsibility (Frankfurt & Watson, 1982). Thus, it might be that discrimination based on implicit attitudes is bad, but that people are not willing to hold a person morally responsible for behaviour based on influences of which they are unaware (Kelly & Roedder, 2008).

Although many people can accurately predict their scores on tests of implicit attitudes (Hahn, Judd, Hirsh, & Blair, 2014), they also tend to disparage and disbelieve those scores if they indicate racial bias (Howell, Gaither, & Ratliff, 2015). In addition, many people simply never take a test of implicit attitudes. Moreover, many non-researchers learn about implicit attitudes via popular media, where implicit racial attitudes are often called 'unconscious' (e.g., Lyubansky, 2012). Therefore, popular understanding of implicit bias in the real world may or may not include its inaccessibility to consciousness.

Social cognition researchers widely accept that a person can be consciously aware of having an attitude without being consciously aware of that attitude's influence on other psychological processes or behaviours (Gawronski, Hofmann, & Wilbur, 2006). Thus, two types of awareness can be ascribed to a company manager who has negative attitudes towards Black people that lead him to discriminate against Black employees. The manager may be (1) completely unaware of his negative attitudes (and thus unaware that those attitudes could influence his behaviour), (2) aware of his negative attitudes but unaware that his attitudes have the potential to influence his behaviour, or (3) aware of both his negative attitudes and their behavioural influence.

Previous research on moral judgements for discrimination based on implicit attitudes, of which there is little, has neglected the distinction between these two types of awareness; it has instead focused on describing implicit bias as available to awareness, controllable, or neither. Such descriptions influence moral judgement: Participants ascribe

lower moral responsibility to a discriminator described as 'unaware of [this] subconscious dislike' than to one described as aware of but unable to control implicit attitudes, or one described as discriminating without mention of implicit attitudes (Cameron, Payne, & Knobe, 2010). This comparison of awareness to controllability is informative; however, the findings are mute to whether moral judgements are influenced by the two types of awareness identified by social cognition researchers.

The purpose of the current research was to test the causal influence of two types of awareness – awareness of bias and awareness of the behavioural influence of that bias – on moral judgements for discrimination based on implicit biases. In manipulating the two types of awareness identified by social psychologists, the current research comprehensively tests the effect of awareness on moral judgements. Because awareness has a clear, robust effect on judgements of moral responsibility, we predicted that moral judgements for attitude-based discriminatory behaviour would depend on both the target's awareness of his negative attitudes and on the target's awareness that those attitudes could lead to discriminatory behaviour.

Overview of the present research

Participants in two studies read vignettes about a target person who, despite being consciously egalitarian, has negative attitudes towards Black people that lead him to treat them unfairly. The target was either aware or unaware of his negative attitudes, and aware or unaware that those attitudes could influence his behaviour. After reading a vignette about the target, participants reported the extent to which they perceived him to be morally responsible for his discriminatory behaviour. Based on evidence that awareness increases perceived moral responsibility, we expected participants to perceive a target as more morally responsible when aware of his negative attitudes than when unaware. We also hypothesized that participants would perceive the target as more morally responsible when he was aware that his negativity could influence his behaviour than when he was unaware. We tested these hypotheses in both studies.

In the second study, we explored three additional variables expected to mediate the effect of the target's awareness perceptions of his moral responsibility: (1) perceptions of the target's foresight of his discriminatory behaviour, (2) perceptions of the target's obligation to foresee his discriminatory behaviour, and (3) perceptions that the target's negative attitudes reflect his true self. These hypotheses and relevant literature are discussed in more detail after the first study.

STUDY 1

Method

Participants

Participants were 386 U.S. citizen volunteers at the Project Implicit website (http://implicit.harvard.edu; Nosek, Banaji, & Greenwald, 2002) who completed all study materials. Total N was reduced to 286 after removal of those who answered the manipulation check incorrectly (Mage = 30.5 years, SD = 13.2, 66.4% women, 72.4% White people; for further discussion of manipulation, check below). This sample size was chosen based on an a priori decision to collect approximately 100 participants per awareness condition.

Materials and measures

Awareness manipulation

Participants read a vignette about a target person, John (adapted from Cameron et al., 2010). John was described as being consciously egalitarian, but having unconscious negative attitudes towards Black people that lead him to treat them unfairly. Participants were randomly assigned to one of three conditions: Unaware : The target was unaware of his racial negativity and unaware that it could influence his behaviour; Semi-Aware : The target was aware of his negativity, but unaware that it could influence his behaviour; and Fully-Aware : The target was aware of his negativity and aware of its behavioural influence. To strengthen generalizability to multiple domains of discrimination, we constructed several scenarios in which the target treated Black people unfairly: Hiring, promotions, university admissions, and health screenings. Participants were randomly assigned to one of these four different scenarios within one of the three awareness conditions. A sample vignette is:

John is in charge of promotions at a major company. He is supposed to decide between various candidates on the basis of merit. Consciously, John thinks people should be treated equally, regardless of race.

Despite this, John has an unconscious negativity towards African Americans. He is aware [unaware] of having this negativity, and disagrees [but if he knew, he would disagree] with this feeling because he sincerely believes in equality. When John decides whether or not to promote an employee, he tries to make the decision based only on merit. However, [because he is unaware of his unconscious negativity], he is not always successful at preventing it from influencing his judgment. As a result, John sometimes unfairly denies African Americans promotions.

After reading the vignette, participants could continue the experiment only after correctly answering two questions about the vignette.

Moral responsibility

Perceived moral responsibility was measured with a combination of the three-item Moral Responsibility Questionnaire (Cameron et al. , 2010) and five additional items, for a total of eight items (see Appendix). An example item is: John is at fault for treating African Americans unfairly . Participants responded on a 7-point scale ranging from 1 = Strongly disagree to 7 = Strongly agree . The items were combined into a single index, such that higher scores indicated greater perceived moral responsibility (a = .80).

Manipulation check

At the end of the study, participants identified the type of negativity that John displayed, choosing from the following options: *John has no unconscious negativity, John has no awareness of his unconscious negativity, John is aware of his negativity but not its influence on his behavior, and John is aware of his negativity and its possible influence on his behavior*. A total of 100 participants (26%) failed the manipulation check and were excluded from analysis.

Procedure

After random assignment to this study from the Project Implicit research pool, participants read the vignette about John and completed the measure of perceived moral responsibility. They then completed the manipulation check.

Results

To test the hypothesis that the target's awareness of his racial negativity and its potential influence on behaviour would affect moral judgements, we used a one-way ANOVA examining the effect of awareness condition (Unaware, Semi-Aware, Fully-Aware) on perceived moral responsibility, collapsing across the four types of scenarios (hiring, promotions, university admissions, and medical screenings).3 As expected, awareness condition had a significant effect on perceived moral responsibility, $F(2, 280) = 7.25$, $p = .001$, g2 = .05 (see Table 1 for means by condition).

Post-hoc tests indicated that participants saw the target as more morally responsible for his discriminatory behaviour in the condition where he was aware of both his negativity and its behavioural influence (Fully-Aware condition: $M = 4.59$, $SD = 1.14$) than when he was unaware of both (Unaware condition: $M = 4.01$, $SD = 0.98$), $t(193) = 3.81$, $p < .001$, Cohen's $d = 0.55$ (95% CI of the difference = 0.28, 0.88). In addition, the condition in which the target was aware of his negativity but unaware of its behavioural influence (Semi-Aware condition: $M = 4.42$, $SD = 1.19$) produced significantly higher moral responsibility scores than the Unaware condition, $t(191) = 2.64$, $p = .009$, $d = 0.38$ (95% CI of the difference = 0.10, 0.72). These results indicate that the target was perceived as less morally responsible when he was unaware of both his negative attitudes and their behavioural influence, than in either of the two conditions where he was aware of the attitudes.

Unexpectedly, moral responsibility scores in the Semi-Aware condition did not differ from those in the Fully-Aware condition, $t(176) = 0.96$, $p = .34$, $d = 0.15$ (95% CI of the difference = _0.18, 0.51). When the target was aware of his racial negativity, he was perceived as equally morally responsible for discriminatory behaviour whether he was aware or unaware that such attitudes could influence his behaviour.

Discussion

As expected, a target whose negative attitudes towards Black people led to discrimination was perceived as more morally responsible when he was aware of the negative attitudes that led to the behaviour than when he was unaware of them. This finding is consistent with other literature: Previous findings show that participants ascribe more blame to a target person who is aware of risk of harm to others than to one who is unaware (Lagnado & Channon, 2008; Nadler & McDonnell, 2011; Nelson-Le Gall, 1985).

When the target was aware of his negativity towards Black people, participants saw him as equally morally responsible whether or not he was aware that his attitudes had the potential to influence his behaviour. This finding is not consistent with previous research or with our hypothesis. Because foresight increases moral responsibility (e.g., Lagnado & Channon, 2008), we expected that awareness of behavioural influence would imply foresight and cause higher moral responsibility scores. So why did the conditions that

Table 1. Descriptive statistics for variables in Studies 1 and 2

	N	Mean	*SD*
Moral responsibility by scale (Study 1)			
Unaware	105	4.01	0.98
Semi-Aware	88	4.42	1.19
Fully-Aware	90	4.59	1.14
Moral responsibility to scale (Study 2)			
Unaware	177	4.23	1.10
Semi-Aware	156	4.60	0.99
Fully-Aware	153	4.68	1.06
Foresight			
Unaware	180	1.64	0.95
Semi-Aware	159	2.58	1.17
Fully-Aware	154	3.79	1.28
Obligation to foresee			
Unaware	178	2.96	1.15
Semi-Aware	159	3.62	0.88
Fully-Aware	153	3.78	1.01
True-self perceptions			
Unaware	178	2.91	0.97
Semi-Aware	158	3.01	0.87
Fully-Aware	154	3.18	0.94

varied awareness of behavioural influence produce similar moral judgements?

This unexpected effect could have a simple explanation: The target was described as aware of his attitudes' behavioural consequences in the Fully-Aware condition, and unaware in the Semi-Aware condition, but participants may yet have believed that the target actually did foresee his discriminatory behaviour in both conditions. Simply being aware of his negativity may have implied that the target foresaw his subsequent discriminatory behaviour, leading to equal moral responsibility judgements in the Semi-Aware and Fully-Aware conditions. Thus, in Study 2 we directly tested participants 'perceptions of the target's foresight.

Alternatively, moral judgement may depend not just on foresight, but on whether its target has a duty or obligation to foresee the harm caused. According to psychological (Alicke, Rose, &Bloom, 2011) and legal (Brady, 1996) theory, reasonable people who risk harm to others ought to foresee negative consequences; observers may perceive failure to meet this obligation as irresponsible and negligent. If so, then equivalent obligation to foresee between the Semi-Aware and Fully-Aware conditions may explain the equivalent moral judgements between them: People may think that in both conditions, the target person's awareness of his racial bias obligates him to foresee his discriminatory behaviour.

The argument that a person with implicit bias ought to foresee their potential to discriminate has also been considered in the philosophical literature. For example, a grader who anticipates the behavioural consequences of her or his implicit racial bias may correct for it by preferentially inflating Black students' grades (Kelly & Roedder, 2008).This anticipation, or foresight, allows the grader to meet an obligation to assign the most accurate grade possible.

Social psychologists have also considered moral obligation to be a necessary condition of moral responsibility. It has been theorized that people blame an agent for a

negative event only when the agent had both the ability to prevent the event and an obligation to prevent the event (Malle, Guglielmo, & Monroe, 2014). However, obligation outranks ability: People do not consider an agent's ability unless they first judge them to be obligated to prevent the event. For example, consider an adult who fails to provide food and shelter to a child. This adult would not be blamed for neglect unless she or he is obligated not to neglect – such as when the adult is the child's parent. Only then would other factors, such as whether the adult has the ability to care for the child, contribute to moral judgements.

In this study, the target person may have an obligation to foresee his discriminatory behaviour. In addition, it may be that such an obligation is a condition of moral responsibility for implicit bias: That failure to behaviourally correct for biases is blameworthy only when the need for correction is reasonably foreseeable. Thus, we designed Study 2 to test the possibility that moral judgements depend on obligation to foresee discriminatory behaviour rather than on actual foresight of that behaviour. We measured both perceptions of the target's foresight of his discriminatory behaviour and perceptions of his obligation to foresee that behaviour.

Moral responsibility for discrimination may also be explained by perceptions that the target's negative attitudes represent his true self . Moral traits outperform other traits as signals of one's essential self (Goodwin, Piazza, & Rozin, 2014; Strohminger & Nichols, 2014); if negative attitudes signal a wrongdoer's true self, they may seem morally relevant and contribute to moral judgements. In Study 2, we explored the possibility that beliefs about the target's true self would influence moral judgements.

Study 2 hypotheses and overview

In Study 2, we again hypothesized that participants would perceive a target as more morally responsible for discriminatory behaviour when aware of having racial negativity than when unaware. We also expected to again show equal moral responsibility judgements in the conditions where the target was aware or unaware that his negativity could influence his behaviour.

We tested three potential mediators of the effect of awareness on perceived moral responsibility: (1) foresight : Perceptions that the target foresaw his discriminatory behaviour, (2) obligation to foresee : Perceptions that the target ought to have foreseen his discriminatory behaviour, and (3) true self : Perceptions that the target's negative attitudes reflect his true self. We expected that perceived obligation to foresee would better predict moral responsibility judgements than beliefs that the target actually did foresee his discriminatory behaviour. We did not have a clear prediction about the direction in which participants' judgements about the target's true self would influence moral responsibility; this variable was included for exploratory purposes.

STUDY 2

Method

Participants

Participants were 657 U.S. citizen volunteers at the Project Implicit website (http:// implicit.harvard.edu; Nosek et al., 2002) who completed all study materials. Total N was reduced to 494 after removal of those who answered the manipulation check incorrectly

(Mage = 32.4 years, SD = 13.6, 62.6% women, 74.3% White people).

Materials and measures

Awareness manipulation
The awareness manipulation was identical to that of Study 1 except that the four vignette types were simplified to two (promotions and hiring).

Perceived responsibility for behaviour
Perceived moral responsibility was measured exactly as in Study 1 (a = .79).

Mediator variables Perceptions of target's foresight. Two items assessed participants' perceptions that the target foresaw the influence of his unconscious negativity on his discriminatory behaviour: *To what extent did John foresee the influence of his unconscious negativity on his unfair treatment of African Americans?* And *How likely is it that John foresaw the influence of his unconscious negativity on his unfair treatment of African Americans?* Participants responded on a scales ranging from 1 = Not at All to 7 = Completely for the first item and 1 = *Very Unlikely* to 7 = *Very Likely* for the second item. These two items were combined, such that higher scores indicated perception of greater likelihood and the extent to which the target foresaw the influence of his unconscious negativity on his discriminatory behaviour (a = .69).

Perceptions of target's obligation to foresee. One item assessed perceptions that *John should have foreseen the influence of his unconscious negativity on his unfair treatment of African Americans*. Participants responded on a scale ranging from 1 = Not at all to 7 = Completely.

Perceptions of whether target's unconscious negativity reflects a true self. Two items assessed perceptions that John's unconscious attitudes reflected his true self: *To what extent does John's unconscious negativity reflect his true self? and To what extent does John's unconscious negativity reflect his true attitudes toward African Americans?*

Table 2. Intercorrelations among variables in Study 2

	1	2	3	4
1. Moral responsibility	-			
2. Foresight	.118*	-		
3. Obligation to foresee	.356**	.319**	-	
4. True Self	.421**	.075	.272**	-

Note. *p < .05; **p < .01.

(a = .65). Participants responded on a scale ranging from 1 = *Not at all* to 7 = *Completely*. For intercorrelations among study variables, see Table 2.

Manipulation check
Participants were asked at the end of the study to identify the type of negativity that John displayed. A total of 163 participants (24.8%) failed or declined to respond to the manipulation check and were excluded from analysis.

Procedure

After random assignment to this study from the Project Implicit research pool, participants read a vignette about John and then completed the moral responsibility scale. They then completed the mediator measures in randomized order. Finally, they completed the manipulation check.2

Results

The influence of condition on moral responsibility

As in Study 1, collapsing across the two types of scenarios, awareness condition significantly influenced perceptions of moral responsibility, $F(2, 483) = 8.76$, $p < .001$, $g2 = .04$ (see Table 1 for means by condition).6 Post-hoc tests indicated that participants saw the target as more morally responsible for his discriminatory behavior when he was aware of both his racial negativity and its influence on his behaviour (Fully-Aware condition: M = 4.68, SD = 1.06) than when he was unaware of both (Unaware condition: M = 4.23, SD = 1.10), $t(328) = 3.77$, $p < .001$, $d = 0.42$ (95% CI of the difference = 0.22, 0.69). In addition, the Semi-Aware condition (M = 4.60, SD = 0.99) produced significantly higher moral responsibility scores than the Unaware condition, $t(331) = 3.21$, $p = .001$, $d = 0.35$ (95%CI of the difference = 0.14, 0.60). Again, the Fully-Aware condition did not differ from the Semi-Aware condition, $t(307) = 0.69$, $p = .49$, $d = 0.08$ (95% CI of the difference = _0.15, 0.31). When the target was aware of his negative attitudes, he was perceived as more morally responsible than if he was not aware of his negativity. Further, as in Study 1, he was seen as equally morally responsible whether he was aware or unaware that such attitudes could influence his behaviour. As in Study 1, moral responsibility scores for each condition fell between the scale midpoint *Neither disagree nor agree* and *Slightly agree.*

The influence of condition on mediator variables

Perceptions of target's foresight

Awareness condition significantly influenced perceptions of the target's foresight, $F (2,490) = 150.32$, $p < .001$, $g2 = .38$ (see Table 1 for means by condition). Post-hoc tests indicated that participants believed the target foresaw his discriminatory behavior more in the condition where he was aware of both his negativity and its influence on his behaviour (Fully-Aware condition: M = 3.79, SD = 1.28) than when he was unaware of both (Unaware condition: M = 1.64, SD = 0.95), $t (332) = 17.64$, $p < .001$, $d = 1.94$ (95% CI of the difference = 1.92, 2.40), and more than when he was aware of his negativity but unaware of its behavioural influence (Semi-Aware condition: M = 2.58, SD = 1.17), $t (311) = 8.76$, $p < .001$,$d = 0.99$ (95% CI of the difference = 0.94, 1.49). In addition, the Unaware condition produced significantly lower foresight scores than the Semi-Aware condition, $t (337) = _ 8.16$, $p < .001$, $d = _ 0.89$ (95% CI = _ 1.17, _ 0.71). When the target was aware of his racial negativity, he was perceived as having more foresight than when he was unaware. Further, he was perceived to have more foresight when he was aware that such attitudes could influence his behaviour than when he was unaware.

Perceptions of target's obligation to foresee

Condition significantly influenced perceptions that the target should have foreseen his discriminatory behaviour, F (2, 487) = 31.10, p < .001, g2 = .11 (see Table 1 for means by condition). Post-hoc tests indicated that participants thought the target was more obligated to foresee when he was aware of both his negative attitudes and their behavioural influence (Fully-Aware condition: M = 3.78, SD = 1.01) than when he was unaware of both (Unaware condition: M = 2.96, SD = 1.15), t (329) = 6.91, p < .001, d = 0.76 (95% CI of the difference = 0.59, 1.07). In addition, the Semi-Aware condition (M = 3.62, SD = 0.88) produced significantly higher obligation scores than the Unaware condition, t (335) = 5.94, p < .001, d = 0.65 (95% CI of the difference = 0.45, 0.89). As expected, scores in the Fully-Aware condition did not differ from scores in the Semi-Aware condition, t (310) = 1.51, p = .13, d = 0.17 (95% CI of the difference = _ 0.05, 0.37). Participants perceived the target to be more obligated to foresee his discriminatory behaviour when he was aware of his racial negativity than when unaware of it. Further, whether the target was aware or unaware that such attitudes could influence his behaviour, participants saw him as equally obligated to foresee resultant discrimination.

Perceptions of whether target's unconscious negativity reflects a true self

Condition significantly influenced perceptions that the target's negative attitudes represented his true self, F (2, 487) = 3.53, p = .03, g2 = .01 (see Table 1 for means by condition). Post-hoc tests indicated that participants thought the target's unconscious attitudes were more reflective of his true self when he was aware of both his negative attitudes and their behavioural influence (Fully-Aware condition: M = 3.18, SD = 0.94) than when he was unaware of both (Unaware condition: M = 2.91, SD = 0.97), t (330) = 2.56, p = .01, d = 0.28 (95% CI of the difference = 0.06, 0.48). The Semi-Aware condition (M = 3.01, SD = 0.87) did not produce significantly different obligation scores than the Unaware condition, t (334) = 0.92, p = .36, d = 0.10 (95% CI of the difference = _ 0.11, 0.29). In addition, scores in the Fully-Aware condition did not differ from scores in the Semi-Aware condition, t (310) = 1.72, p = .09, d = 0.20 (95% CI of the difference = _ 0.03, 0.38). Participants perceived the target's attitudes to be more reflective of his true self when he was aware of his racial negativity and its potential behavioural influence than when unaware of both, but true-self perceptions did not differ when he possessed only one type of awareness.

Mediation analyses

As described above, condition influenced three potential mediator variables: Perceptions of the target's foresight, perceptions of the target's obligation to foresee, and perceptions that the target's unconscious negative attitudes reflect his true self. Thus, we used a multiple-mediator model to test whether each of these three variables, entered simultaneously into the model, mediated the effect of awareness on perceived moral responsibility. All simple correlations between the predictor and mediator measures were significant (p < .05) except for that between true-self perceptions and foresight: Correlations ranged from r = .08 to r = .36 (see Table 2).

We followed the approach recommended by Hayes and Preacher (2014) for mediation analysis with ordered multicategorical independent variables. We used Hayes and Preacher's bootstrapping macro for SPSS. The macro uses thousands of random resamples of the data to generate an empirical sampling distribution, from which it

estimates effects. PROCESS uses listwise deletion based on all variables in the model, such that analyses for each path are based on the same subset of data (Hayes, n.d.). We created a system of contrast codes to examine the relative effects of being in one group relative to a reference group or groups. The first contrast coded for awareness of negative attitudes, such that the Semi-Aware and Fully-Aware conditions were coded as 0.5 and the Unaware condition was coded as _ 1. The second contrast coded for awareness of behaviour influence of negative attitudes, such that the Fully-Aware condition was coded as 2 and the Semi-Aware and Unaware conditions were coded as _ 1 (see Table 3).

Perceptions of target's unconscious attitudes as his true self did not mediate the effect of awareness condition on perceived moral responsibility

Attitude awareness, relative to unawareness of attitudes, did not significantly predict true self perceptions, b = .05 (95% CI = _ 0.08, 0.19), SE = 0.07, p = .56 (see Figure 1).

Table 3. Contrast coding of categorical predictor variable in Study 2

		Characteristic		
		Unaware	Semi-Aware	Fully- Aware
Contrast	C_1	-1	.5	.5
Code	C_2	-1	-1	2

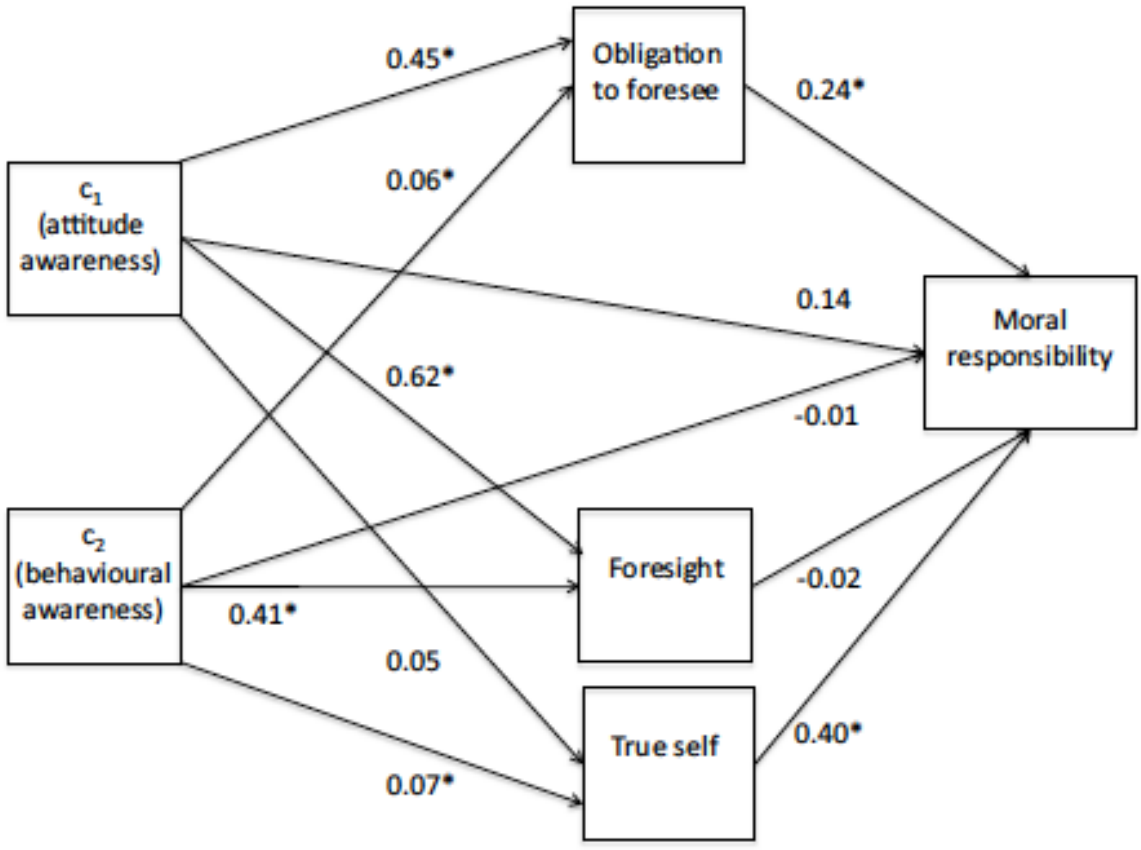

Figure 1. Results from Study 2: Effect of condition on moral responsibility as mediated by foresight, obligation, and perceptions that unconscious attitudes reflect the true self. Unstandardized coefficients

are shown (*p < .001).

Behavioural awareness, relative to behavioural unawareness, did not significantly predict true self perceptions, b = .07 (95% CI = 0.00, 0.14), SE = 0.04, p = .06. True-self perceptions significantly predicted perceived moral responsibility, b = .40 (95% CI = _ 0.31, 0.49), SE = 0.05, p < .0001.

True-self perceptions did not mediate between attitude awareness and perceived moral responsibility. The 10,000-sample bootstrapped estimate of the indirect effect was b = .02 (95% CI = _ 0.03, 0.08), SE = 0.03. Although true-self perceptions predicted perceived moral responsibility, they did not mediate the path from awareness to moral responsibility.

Perceptions of target's foresight did not mediate the effect of awareness condition on perceived moral responsibility

Attitude awareness, relative to unawareness of attitudes, significantly predicted perceptions of foresight, b = .57 (95% CI = 0.42, 0.73), SE = 0.08, p < .0001 (see Figure 1). Behavioural awareness, relative to behavioural unawareness, also significantly predicted perceptions of foresight, b = .30 (95% CI = 0.22, 0.38), SE = 0.04, p < .0001. However, perceptions of foresight did not significantly predict perceived moral responsibility, b = _ .01 (95% CI = _ 0.08, 0.05), SE = 0.03, p = .74. Although awareness influenced perceived foresight, perceived foresight was unrelated to perceived moral responsibility and did not mediate the path from awareness to moral responsibility.

Perceptions of target's obligation to foresee mediated the effect of awareness condition on perceived moral responsibility

Attitude awareness, relative to unawareness of attitudes, significantly predicted perceptions of obligation to foresee, b = .39 (95% CI = 0.25, 0.52), SE = 0.07, p < .0001. Behavioural awareness did not significantly predict perceptions of obligation to foresee, b = .03 (95% CI = _0.04, 0.10), SE = 0.03, p = .25. Perceptions of obligation significantly predicted perceived moral responsibility, b = .37 (95% CI = 0.30, 0.45), SE = 0.04, p < .0001.

Obligation mediated between attitude awareness and perceived moral responsibility. The 10,000-sample bootstrapped estimate of the indirect effect was b = .14 (95% CI = 0.09, 0.21), SE = 0.03. The significant effect of attitude awareness on moral responsibility scale scores, b = .17 (95% CI = 0.04, 0.30), SE = 0.07, p = .01, was no longer significant when perceived obligation to foresee was entered into the model, b = .03 (95% CI = _0.10, 0.16), SE = 0.07, p = .64. Perceptions that the target had an obligation to foresee his discriminatory behaviour mediated the effect of attitude awareness on moral judgements; attitude awareness led to greater obligation to foresee, which in turn contributed to greater perceived moral responsibility. Obligation did not mediate between behavioural awareness and perceived moral responsibility. The 10,000-sample bootstrapped estimate of the indirect effect was b = .01 (95% CI = _0.00, 0.03), SE = 0.01.

Discussion

Replicating the results of Study 1, participants judged a person whose discriminatory behavior was caused by negative attitudes as more morally responsible when he was aware of his negative attitudes than when he was unaware of them. Further, when he was aware of his negative attitudes, he was perceived as equally morally

responsible whether he was aware or unaware that those attitudes could influence his behaviour.

Study 2 also explained the pattern of moral judgements across the three awareness conditions. As expected, perceptions of the target's foresight did not predict moral judgements or mediate the effect of awareness on moral judgements. Also as expected, obligation to foresee did mediate the path from awareness to moral judgements. The more the target was perceived as having an obligation to foresee his discriminatory behaviour, the more participants held him morally responsible for that behaviour.

GENERAL DISCUSSION

The current research investigated perceptions of moral responsibility for attitude-based discrimination. Participants read about a target person whose negativity towards Black people (i.e., implicit attitudes) led him to discriminatory behaviour despite his conscious commitment to egalitarianism. Because moral judgements have been shown to depend on perceptions of the wrongdoer's awareness and foresight (Lagnado & Channon, 2008; Nadler & McDonnell, 2011; Nelson-Le Gall, 1985), we manipulated the manager's awareness of his racial negativity, and his awareness that his negativity could influence his behaviour.

The results of two studies showed that, as expected, participants perceived the target to be more morally responsible for his discriminatory behaviour when he was aware of having racial negativity than when unaware (i.e., his negativity was unconscious). Unexpectedly, participants saw the target as equally morally responsible whether he was aware or unaware that his negativity could influence his behaviour. Based on the previous experimental findings, we initially expected that lacking awareness of his attitudes' behavioural influence would partially excuse the target's discriminatory behaviour. However, unawareness that his negativity could influence his behaviour did not free the target from blame: Participants held him just as responsible as when he was aware of his attitudes' behavioural influence. These findings apply to Kelly and Roedder's (2008) hypothetical grader, about whom they ask: If she believes that she is racially biased, how could she justify continuing to give uncorrected grades? According to the current findings, the average person may answer that she could not justify it – or at least could not justify it as much as if she was unaware of her bias.

Study 2 was designed to explain why the two awareness types differently affected moral responsibility: Why the target's awareness of having negative implicit attitudes influenced his moral responsibility, but awareness of behavioural influence did not. We tested three potential mediators of the path from awareness to moral responsibility judgements: (1) perceptions that the target foresaw his discriminatory behaviour, (2) perceptions that the target had an obligation to foresee his discriminatory behaviour, and (3) perceptions that the target's negative attitudes reflected his true self.

We tested perceptions that the target foresaw his discriminatory behaviour to rule out the possibility that participants believed that the target actually did foresee his discriminatory behaviour in both the Semi-Aware and Fully-Aware conditions. If so, then simply being aware of his negativity may have implied that the target foresaw his subsequent discriminatory behaviour, leading to equal moral responsibility judgements in the Semi-Aware and Fully-Aware conditions. However, this was not the case; participants

did understand the distinction between the two conditions. They perceived the target person to have more foresight in the Fully-Aware condition than in the Semi-Aware condition. Thus, although participants distinguish between the level of awareness in the two conditions, that distinction does not matter for moral responsibility judgements.

What did matter for moral responsibility judgements was obligation to foresee. Of the three tested mediators, only obligation to foresee mediated the effect of awareness on moral responsibility. These results suggest that awareness of negative implicit attitudes brings an obligation to foresee the risk of discrimination posed by those attitudes. And the greater that obligation, the more one may be held morally responsible.

Although the influence of obligation on moral responsibility is intuitively compelling, there is sparse evidence for it (Malle et al. , 2014). The current research helps fill this gap in the literature, providing one of the first direct tests of the effects of obligation on blame. Moreover, it does so while also considering another central contributor to culpability: Capacity to prevent negative events, as represented by awareness.

Future research could further manipulate information about attitudes to investigate its influence on obligation, capacity, and moral responsibility. Several variations of information could work as manipulations of obligation and/or capacity, including descriptions of attitudes as weakly or strongly driving behaviour, as working with or against explicit attitudes, and as easy or difficult to recognize and control. Thus, flexibility in the public perception of attitudes is a strength of using them to investigate moral concepts.

The current results invite comparison with those of Cameron et al. (2010). Both our paper and theirs suggest that blame results when a target seems obligated to gain control over the behavioural effects of their negative attitudes, regardless of whether he actually can do so. In our Semi-Aware condition, the target person is aware of his negative attitudes, but not their potential to influence his behaviour. This differs somewhat from a conception of implicit bias as accessible to awareness but uncontrollable, as represented in Cameron et al. (2010) Automatic condition, where the target person 'has difficulty controlling' its influence on his behaviour. But in both our Semi-Aware and Cameron et al. (2010) Automatic conditions, the target is not freed from blame. Why might this be?

The current research suggests that bias awareness brings a responsibility to attempt to foresee (and thereby take control of) those biases' behavioural influences. So just as participants believed that the target in the Semi-Aware condition ought to try to foresee his attitudes' behavioural influence, Cameron et al. (2010) Automatic (aware but uncontrollable) condition may have made participants believe that the target should try harder to control his attitudes' behavioural influence. In both our paper and theirs, the control available to the target seems less important than the control participants think the target ought to have or ought to attempt to achieve.

Descriptive facts about a person's mental states (i.e., foresight) may offer only limited insight into moral psychology: People also form moral judgements based on normative standards about what one's mental states reasonably ought to be. When a target person fails to meet such normative standards – by neglecting to foresee his discriminatory behaviour – people might infer that he does not care, or does not care enough , to eliminate the risk posed by his implicit bias, or that he unreasonably fails to implement his intentions to prevent a harmful outcome (Brady, 1996).

Such normative standards also guide verdicts in criminal law. For example, legal culpability depends on a wrongdoer's obligation to manage their awareness and undertaking of risks and holds people to a 'reasonable' standard for the amount of care they must take to do so. Thus, the current findings suggest that 'reasonableness' standards in criminal law correspond with people's intuitive standards for culpability.

Because the current research shows the importance of obligation judgements for perceived moral responsibility, future research could focus on how people judge others' moral obligations. For example, in judging moral obligation, people may consider the probability and severity of the (im)moral outcome should the obligation not be met. People choose the best course of action (as in Expected Utility Theory; see Takemura, 2014) and courts impose legal obligations (U.S. v. Carroll Towing, 1947) by considering probability and severity of possible outcomes. A person may be expected to foresee even minimal risks when the possible outcome is severely harmful, such as death or serious injury. Thus, a wrongdoer who is ignorant of a possible harmful outcome may still be morally responsible when that outcome is severely harmful. In the current studies, some people may perceive the target's discriminatory behaviour as harmful enough to warrant an obligation to foresee the risk of its actualization. Thus, people who perceive discrimination as especially severe may more readily impose obligations and moral responsibility on others. On the other hand, perceiving discrimination as neutral or beneficial – for example discriminating against Nazis – may produce lower judgements of moral obligations.

Future research could also explore whether judgements of harmfulness, and subsequent obligation and moral judgements, differ depending on domain of discrimination and who is judging the harmfulness. Whereas, for example, religious people may attribute special harmfulness to religious discrimination, liberals may overtake conservatives in attributing harmfulness to racial discrimination. Conservatism is related to greater implicit and explicit racial bias than liberalism (Nosek et al. , 2007), and liberals value egalitarianism (e.g., Tetlock, Mitchell, & Anastasopoulos, 2013), so conservatism may negatively predict perceived harmfulness of racial discrimination. In turn, those who perceive discrimination as less harmful may attribute lower obligation and moral responsibility to a discriminator. Future studies could address this question by testing what factors (e.g., political ideology) influence perceived severity of attitude-based discrimination in different domains and whether perceived severity influences perceived obligations.

The participants in the present studies are perhaps more likely to care about intergroup biases than the general population as they volunteered for the study at the Project Implicit research website. It might be expected that these differences qualify the current findings. However, the participants are also somewhat more diverse and representative of the general population than a typical undergraduate sample (e.g., in these studies, the mean age was over 30, about 30% of the sample is non-White people, and the participants come from all over the United States). However, to test the generalizability of the current findings, we replicated Study 2 using a sample outside Project Implicit: Amazon Mechanical Turk, an online marketplace where individuals and organizations pay people for task completion. The mediation results from this sample, using a multiple-mediator model, replicate those of Study 2 (see full results at https://osf.io/2nqsu/).

Conclusion

The situation described in this study is unfortunately realistic – many people consider themselves egalitarian, but evidence shows that automatic racial bias is common and contributes to discriminatory behaviour (Jost et al. , 2009; O'Brien et al. , 2010). The current research suggests that awareness of attitudes influences moral judgements of those who engage in attitude-based discrimination because it obligates them to foresee their consequences. Ignorance about the behavioural influence of attitudes does not mitigate moral responsibility; simple awareness of attitudes is enough to obligate one to foresee and prevent harmful consequences. The current findings demonstrate that moral judges' consideration of mental states is not limited to descriptive facts, but also relies on normative standards.

References

Alicke, M., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *The Journal of Philosophy*, 108, 670–696.

Brady, J. B. (1996). Conscious negligence. *American Philosophical Quarterly*, 33, 325–335.

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, 23, 272–289.

Correll, J., Park, B., Judd, C., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, 92, 1006–1023.

Dasgupta, N. (2013). Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. In P. G. Devine & E. A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 233–279). Waltham, MA: Academic Press.

Eshleman, A. (2014). *Moral responsibility*. The Stanford Encyclopedia of philosophy. In E. N. Zalta (Ed.) Retrieved from http://plato.stanford.edu/archives/sum2014/entries/moralresponsibility/

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

Frankfurt, H. G.,&Watson, G. (1982). *Freedom of the will and the concept of a person* (pp. 81-95). Oxford, UK: Oxford University Press.

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, 25, 687–717.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and cognition*, 15, 485–499.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106, 148–168.

Green, A., Carney, D., Pallin, D., Ngo, L., Raymond, K., Iezzoni, L., . . . Banaji, M. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22, 1231–1238.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, 143, 1369.

Hayes, A. F. (n.d.). *Frequently asked question about my macros*. Retrieved from http://www.afhayes.com/macrofaq.html

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67, 451–470.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.

Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the middle defensive responses to IAT feedback among Whites, Blacks, and Biracial Black/Whites. *Social Psychological and Personality Science*, 6, 373–381.

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

Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass*, 3, 522–540.

Lagnado, D. A.,&Channon, S. (2008). Judgments of cause and blame: The influence of intentionality and foreseeability. *Cognition*, 108, 754–770.

Lyubansky, M. (2012). *Studies of unconscious bias: Racism not always by racists*. Psychology Today. Retrieved from https://www.psychologytoday.com/blog/between-thelines/201204/studies-unconscious-bias-racism-not-always-racists

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, 25, 147–186.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological bulletin*, 132, 297.

Nadler, J., & McDonnell, M. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review*, 97, 256–300. Northwestern Public Law Research Paper No. 11-43.

Nelson-Le Gall, S. A. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology*, 21, 323–337.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting intergroup attitudes and stereotypes from a demonstration website. *Group Dynamics*, 6, 101–115.

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

O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology*, 40, 917–946.

Rooth, D. (2007). *Implicit discrimination in hiring: Real world evidence* (IZA Discussion Paper No. 2764). Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor).

Rudman, L. A., & Ashmore, R. D. (2007). Discrimination and the implicit association test. *Group Processes and Intergroup Relations*, 10, 359–372.
Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131, 159–171.
Takemura, K. (2014). Expected utility theory and psychology. In *Behavioral decision theory* (pp. 49–61). Tokyo, Japan: Springer.
Tetlock, P. E., Mitchell, G., & Anastasopoulos, J. L. (2013). Detecting and punishing unconscious bias. *The Journal of Legal Studies*, 42(1), 83–110.
U.S. v. Carroll Towing, 159 F.2d 169 (2d Cir. 1947).

Appendix:
Perceived moral responsibility
1. John is morally responsible for treating African Americans unfairly.
2. John should be punished for treating African Americans unfairly.
3. John should not be blamed for treating African Americans unfairly (reversed).
4. John should not be held accountable for treating African Americans unfairly (reversed).
5. John should be excused for treating African Americans unfairly (reversed).
6. John should be considered a less moral person for treating African Americans unfairly.
7. John should be judged for treating African Americans unfairly.
8. John is at fault for treating African Americans unfairly.