Punitive and prosocial reactions to discrimination attributed to implicit bias

by

Stuart Miller

B.S., University of Iowa, 1995
M.S., Kansas State University, 2014

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

Theories of moral judgment suggest that people may respond less punitively and more prosocially to discrimination attributed to implicit, compared to explicit, bias. Additionally, what people believe about the nature of implicit bias may affect their perceptions of the blameworthiness of discrimination caused by implicit bias. The two studies presented here were designed to test these assumptions. In Study 1, participants read a scenario in which a case of racial discrimination was caused by either implicit or explicit bias, and the effects on mental state attributions involved in blame (e.g., awareness, control, intent), as well as how these attributions relate to support for punishment and forgiveness were examined. Study 2 examined how scientific communications about implicit bias affected these judgments by framing implicit bias as something that people are unconscious of and cannot control, or something that people can be aware of and control with effort. Additionally, both studies examined the role of individual differences related to perceptions of bias and discrimination in moderating these effects. Results of these studies were consistent with theories of blame that emphasize the importance of perceptions of intent in attributions of blame. When discrimination was attributed to implicit, compared to explicit bias (Study 1), and when implicit bias was framed as unconscious and uncontrollable, compared to more conscious and controllable (Study 2), the perpetrators' behavior was perceived as less intentional, blameworthy, and deserving of punishment. The results of Study 1 also suggest that people are more sympathetic toward, and forgiving of, perpetrators who unintentionally discriminate. These findings contribute to our theoretical understanding of moral judgments and have practical implications for the consequences of how scientific knowledge of implicit bias is communicated to the public.

Punitive and prosocial reactions to discrimination attributed to implicit bias

by

Stuart Miller

B.S., University of Iowa, 1995
M.S., Kansas State University, 2014

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Dr. Donald A. Saucier

Copyright

Abstract

Theories of moral judgment suggest that people may respond less punitively and more prosocially to discrimination attributed to implicit, compared to explicit, bias. Additionally, what people believe about the nature of implicit bias may affect their perceptions of the blameworthiness of discrimination caused by implicit bias. The two studies presented here were designed to test these assumptions. In Study 1, participants read a scenario in which a case of racial discrimination was caused by either implicit or explicit bias, and the effects on mental state attributions involved in blame (e.g., awareness, control, intent), as well as how these attributions relate to support for punishment and forgiveness were examined. Study 2 examined how scientific communications about implicit bias affected these judgments by framing implicit bias as something that people are unconscious of and cannot control, or something that people can be aware of and control with effort. Additionally, both studies examined the role of individual differences related to perceptions of bias and discrimination in moderating these effects. Results of these studies were consistent with theories of blame that emphasize the importance of perceptions of intent in attributions of blame. When discrimination was attributed to implicit, compared to explicit bias (Study 1), and when implicit bias was framed as unconscious and uncontrollable, compared to more conscious and controllable (Study 2), the perpetrators' behavior was perceived as less intentional, blameworthy, and deserving of punishment. The results of Study 1 also suggest that people are more sympathetic toward, and forgiving of, perpetrators who unintentionally discriminate. These findings contribute to our theoretical understanding of moral judgments and have practical implications for the consequences of how scientific knowledge of implicit bias is communicated to the public.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I have a tremendous amount of gratitude for my major professor, Don Saucier, for his leadership, support, patience, and mentoring. I wish to convey thanks as well to my committee members, Laura Brannon, Gary Brase, and Amelia Hicks for their feedback on this project. I would also like to thank Tianjun Sun and Jin Lee for their help with the path model analyses in the present studies.

My graduate student lab mates over the years, Conor O'Dea, Tiffany Lawless, Amanda Martens, Evelyn Stratmoen, Svyatoslav Prokhorets, Tucker Jones, Megan Strain, Jericho Hockett, Russ Webster, Jessica McManus, Ashley Schiffer, Noah Renken, Andrew Olah, and Ted Wheeler, as well as the many undergraduates I have worked with, have been a wonderful source of inspiration and support during my graduate studies. I would also like to thank the people at Kansas State University for the nurturing culture they have created here.

An especially heart-felt thank you to my wife, Kim Kirkpatrick, daughter, Sadie Miller, parents, Bob & Sue Miller, and Blue Breese, siblings, Laura Hammes, and Jeff Miller for their love, understanding, and support through the ups and downs of a long graduate career.

Chapter 1 - Punitive and Prosocial Reactions to Discrimination Attributed to Implicit Bias

Racial disparities exist in important areas, such as income inequality and in the criminal justice system (Hetey & Eberhardt, 2018; Kraus et al., 2019; Sørensen et al., 2018). These disparities have often been attributed to racial bias. Although in recent decades prejudice has evolved, in response to stronger social norms against blatant racial bigotry (Crandall & Eshleman, 2003), additional forms of expressing racial animosity through more socially acceptable and socially defensible attitudes have persisted (McConahay, 1983; Sears & Henry, 2003). Even more liberally minded individuals who more actively suppress personal prejudice may discriminate against racial outgroups at a more unconscious or uncontrollable level (Dovidio & Gaertner, 2000). These subtle forms of prejudice may partially explain the types of discriminatory behaviors that contribute to the persistence of racial inequalities (Daumeyer et al., 2017; Fiske, 2004; Payne et al., 2017; Spencer et al., 2016).

Understanding how people perceive subtle forms of bias and subsequently make judgments of moral responsibility (e.g., blame, punishment) may help to inform approaches to reducing discrimination. One such subtle form of bias is when implicit bias causes discriminatory behavior. Because people may perceive less intent in these cases, people may perceive discrimination attributed to implicit bias as less blameworthy, and potentially less problematic, than cases in which people perceive greater intent to discriminate. The present research examined how discrimination caused by implicit biases may be perceived as less intentional and blameworthy than discrimination caused by explicit biases, and how these

judgments are associated with subsequent punitive and prosocial[1] responses. The following sections discuss how implicit bias has been conceptualized and how beliefs about implicit bias may affect moral judgments, how theories of moral responsibility provide models of the social judgment processes involved in blame, and how the competing hypotheses derived from these models predict punitive and prosocial responses to discrimination attributed to implicit bias.

Implicit Bias

Implicit bias has frequently been conceptualized by researchers in psychology as a form of bias that is predominantly unintentional, automatic, or outside of people's conscious awareness. Implicit bias is thought to originate from implicit stereotypes that are conditioned through exposure to cultural representations of social groups (Greenwald & Banaji, 1995; Greenwald & Lai, 2020). For example, exposure to representations of Black people as being dangerous, or in other ways inferior, may be encoded into memory and associations may form between these stereotypes and Black people as a social group. Although people are generally aware of such socially shared stereotypes, they may not consciously endorse them because of social norms about their unacceptability (Devine, 1989). In contrast, explicit bias describes a form of bias that includes acceptance or endorsement of group-based stereotypes, consciously held negative attitudes about a social group, and intentional discrimination against members of that social group. However, even for people who endorse egalitarian values and genuinely believe that they are unbiased, implicit stereotypes may still exist at a level that is much less accessible to conscious awareness (Devine, 1989; Fiske, 2004; Perry et al., 2015; Wilson et al., 2000).

[1] The term, prosocial, is used throughout to capture a broad range of helpful, sympathetic, or otherwise nonpunitive responses, such as forgiveness.

Implicit associations have the potential to affect behavior in ways that are unintentional and beyond awareness. For example, implicit bias tends to predict automatic or spontaneous behaviors, whereas explicit bias tends to predict deliberative behaviors (Amodio & Devine, 2006; Dovidio et al., 2002; Fazio, 1990; Wilson et al., 2000). Although the associations between implicit bias and behavior are relatively weak in studies examining person-level correlations between implicit bias and discriminatory behavior, aggregate levels of implicit bias (e.g., nations, states) show much stronger associations with disparate outcomes (Payne et al., 2017) which may explain the persistence of group-based disparities despite apparent reductions in explicit prejudice and the illegality of discriminating against some social groups (Jost et al., 2004).

There is some debate about the unconscious nature and explanatory power of implicit bias in the psychological literature, with some researchers putting forward alternative conceptualizations of implicit bias (or even eschewing the term "implicit" altogether) that avoid troubling issues with defining it as something that exists in the unconscious mind (Corneille & Hütter, 2020; de Houwer, 2019; Fazio & Olson, 2003; Gawronski, 2019; Gawronski et al., 2020; Greenwald & Banaji, 2017; Greenwald & Lai, 2020; Newell & Shanks, 2014). Furthermore, there is convincing evidence that people can spontaneously become aware of their implicit biases (Hahn et al., 2014; Hahn & Gawronski, 2019). Despite these new developments in how implicit bias is conceptualized by scientists, implicit bias is still most commonly portrayed to the public as unconscious associations that people lack introspective awareness of in themselves that result in biased behaviors that are automatic, unintentional, or uncontrollable (Daumeyer et al., 2019; K. Payne et al., 2018). A recent Google search of the term "implicit bias" in recent news publications confirms that when implicit bias is defined at all, the definition contains one or more

of these features. Furthermore, public references to this conceptualization of implicit bias are pervasive, suggesting that many people have high level of awareness of the phenomenon of implicit bias and its usage in explaining concepts like systemic racism and disparate outcomes (Gawronski, 2019).

The goal of the present research was not to resolve the extant debates about what implicit bias is or how it affects behavior or contributes to existing disparities, but rather to examine how people react to instances of discrimination attributed to implicit bias and how different conceptualizations of implicit bias affect these judgments. If people conceptualize implicit bias as unconscious negative attitudes toward a group of people, psychological theories of responsibility and blame discussed in the following section suggest that people will judge others who discriminate because of their implicit biases (or otherwise express their implicit biases) less harshly than when discrimination is due to explicit biases. Central to these judgments may be the extent to which perceivers blame others for behaviors that result in discrimination against members of a social group.

Theories of Responsibility and Blame

Attributions of blame are both explanatory and evaluative (Tooby & Cosmides, 2010). When learning about a harmful outcome, people want to know what actions caused the outcome and who is responsible—they seek to explain the event. Blame attributions are also evaluative in the sense that people also judge the wrongness of the action and form an impression of the actor's moral character to anticipate and safeguard against future harmful behavior. Additionally, when blame precedes attempts to regulate others' behaviors (e.g., through shaming or other forms of punishment), the ascription of blame is fundamentally a social act that often must be justified to others (Malle et al., 2014; Voiklis & Malle, 2017). Because the expression of blame

is a social behavior that can have severe consequences for the accused, blame is regulated by social norms. Individuals who unjustifiably blame others for discriminatory outcomes may face social consequences. Thus, blame must be warranted by acceptable evidence. The constraint of warrant motivates people to systematically evaluate information to arrive at defensible conclusions (Malle et al., 2014).

Theories of moral responsibility and blame suggest that mental state attributions about awareness, control, and intent play a significant role in punitive and prosocial responses to others' actions (Alicke, 2000; Cushman, 2008; Gray et al., 2012; Guglielmo, 2015; Guglielmo et al., 2009; Heider, 1958; Malle et al., 2014; Monroe & Malle, 2017, 2019; Shaver, 1985; Weiner, 1995). Assuming discrimination against a social group is believed to be wrong (and there are groups for which prejudice is seen as more socially acceptable, e.g., Crandall et al., 2002), when implicit bias is understood by perceivers as something that is unconscious or uncontrollable, they may blame others less when implicit bias is perceived to be the reason behind discriminatory behaviors.

According to theories of responsibility and blame, perceptions of intent are a crucial factor in moral judgments about the wrongness of others' actions, and people condemn and punish intentional norm violations more severely than unintentional norm violations (Alicke, 1992; Cushman, 2008; Darley & Shultz, 1990; Gray et al., 2012; Gray & Wegner, 2008; Young & Saxe, 2009). Malle and colleagues' (2014) Path Model of Blame additionally contends that observers' perceptions of others' awareness and control over their behaviors are important precursors to judgments of intent. There is some initial evidence to support this hypothesized process, i.e., that perceivers first consider awareness, then control, and then intent before arriving at judgments of blame (Monroe & Malle, 2017, 2019). In the context of moral evaluations of

implicit bias, when people think of implicit bias as unconscious, automatic, or uncontrollable, they may perceive implicit bias that results in discrimination as unintentional, and thus have less reason to blame, and subsequently punish, others who discriminate because of implicit bias. This hypothesis is also consistent with lay intuitions of moral responsibility (Pizarro et al., 2003) as well as legal definitions of conditions of moral responsibility that require evidence of intent, awareness of the implications of one's behavior, and control over the actions' implementation (Kelly & Roedder, 2008).

However, alternative theories of moral responsibility suggest that blame may not be mitigated when implicit bias results in discrimination. Some have argued that the unconsciousness of biases does not excuse people from moral responsibility (Nosek & Hansen, 2008; Sher, 2015; Smith, 2015). Others have argued that people can control behaviors influenced by implicit biases (Amodio & Swencionis, 2018; Correll et al., 2014; Devine et al., 2002, 2012; Nosek et al., 2011; Suhler & Churchland, 2009), which implies that implicit bias may be morally blameworthy. Regardless of whether it is right to blame others for discrimination caused by implicit bias, people may be motivated to blame and punish others for these behaviors.

Alicke (1992, 2000) convincingly argued that previous theories failed to fully consider how cognitive limitations and motivational biases (Kunda, 1990; Nickerson, 1998) affect the blame attribution process. Alicke's Culpable Control Model of Blame acknowledges that people deliberately consider information pertaining to intent in making judgments of blame, but that these judgments are also influenced by relatively automatic processes Alicke referred to as spontaneous evaluations. These spontaneous evaluations arise from extraevidential factors, such as emotional reactions arising from witnessing harm caused to others (e.g., anger, outrage, disgust) or social biases (e.g., stereotypes, assumptions about the perpetrator's character, simple

dislike of the perpetrator). According to Alike, spontaneous evaluations create a motivational state of "blame-validation," in which people blame first and then seek evidence to support that judgment. Specifically, people's desire to find blame biases perceptions of a perpetrator's awareness, control, and intent ways that confirm their initial judgments of blame.

The Culpable Control Model of Blame is consistent with broad theories of social cognition (Crandall et al., 2007; Heider, 1958), as well as theories that argue moral judgments are more intuitive than rational (Haidt, 2001, 2007). Directly supporting the model are several empirical findings, such as incidental anger (i.e., anger not related to evaluations of the event being judged) increases perceptions of criminal intent (Ask & Pina, 2011); unlikeable targets are blamed more than likeable targets (Alicke & Zell, 2009); and people tend to perceive that a side-effect of an action is more intentional when it results in a negative, compared to a positive, outcome (Cova et al., 2016; Leslie et al., 2006).

This motivated social-cognitive theory of blame suggests that when perceivers are outraged by discriminatory behavior they may blame first and then perceive greater awareness, control, or intent than may be warranted when discrimination is due to implicit bias. Thus, according to this theory, blame may not be reduced when discrimination is attributed to implicit compared to explicit bias. The theory also suggests factors that may bias blame judgments. For example, discrimination that results in greater levels of harm could increase negative affect and the motivation to blame. Additionally, there are individual differences (discussed in more details below) in how people react to instances of prejudicial behavior (Miller, Peacock, et al., 2021; Miller & Saucier, 2018; Perry et al., 2015; Pinel, 1999) that are associated with greater tendencies to perceive others' behaviors as prejudiced. Theses individual differences may also be

associated with greater motivation to blame others who discriminate because of their implicit biases.

Thus, there are different theories about how judgments of blame are made. The Path Model in which the process begins with considering the available evidence of awareness, control, and intent leading to degrees of blame. And the Culpable Control Model, based on theories of motivated cognition, which argues that this process may sometimes happen in reverse: beginning with blame and then proceeding to attributions of awareness, control, and intent, with the consideration that motivation to blame may bias these attributions. The former model predicts that blame will be mitigated when discrimination is due to implicit bias, while the latter suggests there are conditions in which it may not.

Empirical Tests of Blame for Implicit Biases

Earlier studies provide evidence that information about intent and harm affect attributions to prejudice (Swim et al., 2003), and a small body of recent experimental findings supports the hypothesis that blame related judgments are less severe when discrimination is attributed to implicit, compared to explicit bias (Cameron et al., 2010; Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016). In one of the first set of studies to test this hypothesis, the researchers presented participants with a scenario in which a corporate manager discriminated against Black employees when making decisions about promotions despite his commitment to treating others equally, regardless of race (Cameron et al., 2010). Participants were either told that the manager's discriminatory behavior was the result of an unconscious dislike of Black people or that the manager was aware of gut feelings of dislike of Black people, but he was unable to control this from affecting his decisions. Participants attributed less moral responsibility to the manager when discrimination was due to unconscious bias compared to conscious but automatic bias.

Additionally, mediation models in Study 2 provided evidence for a psychological process of blame beginning with attributions of intent and control leading to judgments of moral responsibility, supporting the Path Model of Blame (Malle et al., 2014). Yet equally plausible was the reverse causation mediation model beginning with attributions of moral responsibility leading to perceptions of intent and control, supporting motivated theories of blame that argue that intent may be inferred from harmful actions (Alicke, 2000).

These findings were replicated and extended by experiments in which participants made judgments about a company manager who was described as endorsing egalitarian values but nonetheless held negative attitudes about Black people that caused him to discriminate against Black candidates for promotions (Redford & Ratliff, 2016). Participants were led to believe that the target was either fully aware of his negative attitudes and how they affected his behavior, unaware of his negative attitudes but aware of how they affected his behavior, or unaware of both his attitudes and how they affected his behavior. Participants blamed the target more for his biased behavior in the fully aware than the fully unaware condition. However, blame did not differ based on whether the target was aware or unaware of how his negative attitudes affected his behavior. Furthermore, blame was mediated by the degree to which participants believed that the target had a moral obligation to prevent his prejudice from treating his Black employees unfairly (Study 2).

These studies suggest that perpetrators' awareness of their negative attitudes, and the control they have over these attitudes affecting their discriminatory behaviors, influences perceivers' judgments of moral responsibility. Both sets of studies show that people blame others less for fully unconscious and uncontrollable biases. However, for consciously held negative attitudes that are difficult to control, people are still held morally accountable, potentially

because people believe that others have an obligation to prevent their conscious attitudes from resulting in discrimination. These studies described the targets as attempting to prevent their attitudes from negatively affecting their behavior, yet participants only blamed the targets less when they were unaware of their bias—they did not blame less when targets were aware of their bias but unable to control it. Thus, when discrimination is attributed to implicit bias, it matters how people conceptualize implicit bias. If people believe implicit bias is unconscious, they may be more forgiving of discriminatory actions than if they believe that implicit bias is something that people can be aware of (even if they find it difficult to control), because they may believe that people *ought* to prevent their biases from causing harm.

More recent evidence that people blame others less for discrimination attributed to implicit, compared to explicit bias, further examined how these different forms of bias affected people's concerns about discrimination and the extent to which they supported efforts to reduce discrimination (Daumeyer et al., 2019). Across four studies, participants read ostensible news articles describing recent research finding evidence of discrimination in different fields (e.g., healthcare, policing) which was either attributed to implicit or explicit bias. Implicit bias in these studies was framed in terms of a lack of awareness of the negative attitudes and stereotypes affecting perpetrators' discriminatory behaviors. Across the four studies, participants held perpetrators less morally accountable and supported punishment less when discrimination was attributed to implicit compared to explicit bias, even when discrimination resulted in severe harms (e.g., premature death via differential healthcare decisions). Furthermore, when discrimination was attributed to implicit bias, participants showed less concern about discrimination and less support for reform efforts, suggesting that, in addition to holding individuals less morally responsible, people may also hold institutions less accountable for

addressing the negative effects of implicit bias. The practical implications of this finding are troubling for contexts in which implicit bias is pervasive, assuming people genuinely desire to create a more egalitarian society.

The results of the body of research reviewed here are consistent with theories of blame that suggest people will blame less when discrimination is attributed to implicit, compared to explicit, bias. However, only one of these studies assessed how components of blame (perceived control, intent, anger) mediate attributions of moral responsibility (Cameron et al., 2010, Study 2). In the remaining studies, the relative strengths of the paths from perceptions of intent, awareness, and control leading to blame and punishment were not assessed. Nor have researchers examined how these components of moral judgment are associated with prosocial responses. This limitation prevents drawing conclusions about the relative weight these components may have in judgments of discrimination. For example, it may be the case that perceptions of awareness and intent play a greater role than perceptions of control when blame for discrimination is mitigated, as suggested by some of the findings reviewed here (Cameron et al., 2010; Redford & Ratliff, 2016). The present studies were designed to contribute to our understanding of these processes.

Furthermore, a plausible alternative interpretation for why blame is reduced for unconscious bias, but not uncontrollable bias, is that people blame others for having consciously held biases regardless of whether they intentionally try to prevent their biases from causing harm. Simply harboring negative attitudes about a social group (whether they are aware of these attitudes or not) may go against normative standards about the acceptability of prejudice, and the resulting negative evaluations of the targets may override any consideration of whether discriminatory behaviors are intentional or not. Additionally, it would be informative to examine

how individual differences in concerns about prejudice are associated with degrees of blame and how they are associated with perceptions of the components of blame. As already discussed, motivated blame theories (e.g., Alicke, 2000) suggest that greater perceptions of harm may bias judgments of moral responsibility by increasing perceptions of intent, awareness, and control. Therefore, the present research examined potential moderators and mediators of blame for discrimination resulting from implicit bias (discussed in greater detail below).

Potential Prosocial Reactions to Implicit Bias

So far, research on the effects of attributing discrimination to implicit bias has only assessed punitive reactions (e.g., blame, punishment). This small but reliable body of existing research has been limited in scope by only examining how discrimination attributed to implicit bias affects judgments of accountability and responsibility (Cameron et al., 2010; Redford & Ratliff, 2016), as well as support for punitive consequences (Daumeyer et al., 2019, 2021). Although these attitudes and judgments are important antecedents and consequences of blame, when norm-violating behavior is attributed to factors beyond an agent's awareness or ability to control, Weiner's (1995, 1996) theory of responsibility proposes that, in addition to mitigating blame and the resulting motivations to punish, perceivers may feel sympathy for norm-violators and react to them in prosocial ways (e.g., sympathy, non-judgmental/non-punitive help in managing bias). Like the Path Model, Weiner's (1995, 1996) theory of responsibility emphasizes the role of intent. However, when behavior is unintentional, this theory argues that people consider whether the behavior is the result of a lack of effort or a lack of ability. When there is a lack of effort, blame ensues. When there is a lack of ability, people respond with sympathy, forgiveness, and help. Therefore, if people attribute discrimination to implicit bias, they may perceive a lack of ability to prevent discrimination, (especially if they believe implicit bias is

unconscious and uncontrollable), and thus respond more prosocially compared to when discrimination is attributed to more conscious and controllable forms of bias.

Although previous research implies that people may partially forgive others for discriminatory behaviors resulting from implicit biases, this inference can only be made indirectly given that participants punished perpetrators of discrimination less when their behaviors were attributed to implicit compared to explicit bias (prosocial reactions, such as forgiveness, were not directly measured). However, the reduction in punishment for discrimination attributed to implicit bias was small compared to the reduction in blame: Cohen's d = 0.24 and 0.41 respectively (Daumeyer et al., 2019, internal meta-analyses). It is possible, given what people may believe about implicit bias, that people may desire that discriminatory actions caused by implicit bias have punitive consequences, and still show sympathy and forgiveness for perpetrators. While punitive and prosocial responses are likely to be negatively correlated, these responses may not be entirely redundant. For example, someone could be punished for an act of discrimination because punishment is expected or required in that situation, but still be understood to have unintentionally caused harm, and then be treated in a way that does not defame their moral character but supports their commitment to being more egalitarian. Therefore, both punitive and prosocial responses should be assessed independently to examine a more complete range of possible responses to discrimination attributed to implicit bias.

Furthermore, people often respond defensively to accusations of prejudice and these kinds of defensive responses may be especially likely when people who are consciously egalitarian learn about their implicit biases (Howell et al., 2013, 2015, 2017). Additionally, people with strong egalitarian values are likely to experience guilt when faced with evidence that

they have violated their egalitarian values (Devine et al., 1991; Monteith et al., 1993; Monteith & Walters, 1998). Sometimes the experience of guilt leads to efforts to control bias (Czopp & Monteith, 2003; Devine et al., 1991, 2002; Mallett & Monteith, 2019; Monteith, 1993), but guilt could also lead to avoidance (Amodio et al., 2007). Perhaps a more prosocial approach to others who discriminate due to implicit biases would be beneficial in helping people avoid defensive reactions to the perception that they are being accused of bias. Research on confrontations of prejudice shows that confrontations are more likely to be effective in reducing bias if they avoid creating hostility or threat, and validate positive, egalitarian self-images (for a recent review, see Monteith et al., 2019). Furthermore, interventions that decrease perceived moral blameworthiness can reduce defensive reactions to feedback that one possesses implicit biases (Vitriol & Moskowitz, 2021). If people respond with kindness to others who unintentionally discriminate by helping them understand that their moral character is not being called into question (i.e., that they are not being judged as bad people), this may reduce the perceived threat and hostility of being confronted. Thus, more prosocial approaches to confronting unintentional discrimination may be more effective by validating perpetrators' egalitarian beliefs and reinforcing their commitment to reducing and controlling their bias. The present research aims to increase the understanding of factors that lead to more prosocial reactions to bias.

Individual Differences and Motivated Blame

Some people believe prejudice is more of a social problem than others, which may motivate perceptions of discrimination in ways consistent with the motivated blame theories discussed above. Specifically, the degree to which perceiving an act of discrimination as morally wrong may motivate people to perceive higher levels of perpetrator awareness, control, and intent leading to greater blame and punishment for discrimination attributed to implicit bias.

Similarly, individual differences in beliefs about how morally wrong prejudice and discrimination are may affect perceptions of the degree of harm caused by acts of discrimination, which in turn may provide greater motivation for blame and greater perceptions of intent. Some research suggests people often infer intent from harmful outcomes, rather than using intent to inform blame (Knobe, 2006).

In popular culture, the idea that some individuals are motivated to punish injustices caused by prejudice has been referred to as cancel, woke, or call-out culture (Bouvier & Machin, 2021; Clark, 2020; Norris, 2021; Romano, 2020). Cancel culture appears to be construed as unreasonably punitive responses especially when the intentions behind what are perceived as expressions of prejudice are ambiguous. Some have argued that in cancel culture intent is irrelevant when prejudice is expressed and what matters is only whether people perceive harm (G. Greenwald, 2021). The idea of cancel culture implies that individual differences in concerns about the harmful consequences of prejudice play a role in moral evaluations of discrimination: individuals who more strongly believe that prejudice is wrong or who perceive discrimination as more harmful may be more likely to blame and punish those who discriminate regardless of how intentional the behaviors appear to be.

There is some initial support for the idea that individual differences in perceivers' motivations may shape their judgments about discrimination resulting from implicit bias. For example, White and Black perceivers take different perspectives when evaluating instances of potential racial discrimination. Black perceivers tend to focus on both intent and harm, whereas White perceivers primarily focus on intent and downplay harm in evaluating the degree to which an action qualifies as discrimination (Simon et al., 2019). In one study, higher levels of internal motivations to respond without prejudice (Plant & Devine, 1998) were associated with greater

blame for discrimination regardless of whether perpetrators' behavior was due to implicit or explicit bias (Daumeyer et al., 2019, Study 3). Few other studies have examined how individual differences are associated with blame-related judgments, with the notable exception of one set of studies that examined how shared group membership shapes perceptions of discrimination (Daumeyer et al., 2021). In these studies, women blamed male perpetrators of gender discrimination more than did men, yet shared group membership did not moderate the effects of attributing less blame for discrimination caused by implicit, compared to explicit, bias. These findings suggest motivational factors may affect degrees of blame (and associated attributions) but may not alter the effects of attributing discrimination to implicit, compared to explicit, bias.

Additionally, when other individual difference variables were assessed in the existing published research they have been found to be associated with blame, but evidence of consistent moderation of the effects of attributing discrimination to implicit or explicit bias effects were unreliable. To date, researchers have tested for the moderating role of internal motivations to respond without prejudice (Plant & Devine, 1998), bias awareness (Perry et al., 2015), and victim and perpetrator perspective-taking (Daumeyer et al., 2021). The present studies attempted to replicate previous findings and examine additional individual differences that may be associated with blame-related judgments (described in the sections that follow), as well as their potential to moderate the effects of attributing discrimination to implicit, compared to explicit, bias.

Perspective-Taking

The degree to which perceivers take the perspective of the victim or the perpetrator could affect their moral evaluations of discrimination. Previous research has found victim and perpetrator perspective-taking to mediate the relationship between gender identity and judgments

of blame in responding to gender discrimination (Daumeyer et al., 2021). Specifically, when evaluating instances of men discriminating against women, women more strongly took the perspective of the victim which was associated with higher degrees of blaming the perpetrator, whereas men more strongly took the perspective of the perpetrator which was associated with lower degrees of blaming the perpetrator. In another relevant study, when asked to take the perspective of the Black victim in cases of anti-Black discrimination, White observers increased their perceptions of intent, harm, and wrongful discrimination (Simon et al., 2019). Thus, taking the perspective of the victim may increase blame (and related judgments), whereas taking the perspective of the perpetrator may decrease blame (or increase prosocial responses) in situations where discrimination is attributed to implicit bias.

Bias Awareness

Individuals differ in the degree to which they believe that they may be unconsciously racially biased and in the degree to which they are concerned about unintentionally behaving in prejudiced ways (Perry et al., 2015). Higher levels of bias awareness have been associated with perceiving greater racism in subtle incidents of discrimination (Perry et al., 2015), suggesting that higher bias awareness may also be associated with perceiving discrimination caused by implicit bias as more harmful, blameworthy, and deserving of punishment. Furthermore, believing oneself to be biased in subtle ways may lead to the belief that one has the ability to control one's biases and the expectation that others should also be aware of and control their biases. Conversely, self-awareness of one's subtle, but unintentional, biases might be related to having greater empathy for individuals who are perceived to unintentionally discriminate due to their implicit biases because perceivers may be more aware that they themselves risk similar mistakes for which they would not like to be judged unfavorably (e.g., perceived as racist).

Therefore, it is equally plausible that bias awareness may be related to less punitive and more prosocial responses to discrimination attributed to implicit bias.

The Propensity to Make Attributions to Prejudice

Individuals differ to the extent to which they believe prejudice is a pervasive problem and in their vigilance for detecting cues for prejudice. The propensity for making attributions to prejudice scale (PMAPS), which measures these attitudes and tendencies has been associated with perceiving ambiguously discriminatory behaviors as racially prejudiced (Miller & Saucier, 2018). Greater tendencies to make attributions to prejudice are also related to stronger beliefs that racism is a systemic problem, and more positive attitudes and support for social protests and movements such as Black Lives Matter (Miller, O'Dea, et al., 2021). More relevant to the present studies, higher levels of PMAPS are associated with higher levels of perceived intent and harm when intent is ambiguous (Miller, 2014). Therefore, PMAPS may be related to perceiving intent and harm in discrimination attributed to implicit bias and may even reduce or eliminate the differences between blame for implicit, compared to explicit, bias. Conversely, higher levels of PMAPS are associated with being more influenced by the amount of evidence pertaining to potential racial discrimination (Miller, Peacock, et al., 2021; Miller & Saucier, manuscript in preparation), and better judgment about when a joke is intended to disparage people of color or to confront prejudice as unacceptable (Miller et al., 2019; Saucier et al., 2018). Therefore, it is equally plausible that PMAPS would not moderate the effects of attributing discrimination to implicit, compared to explicit, bias because they perceive implicit bias as unintentional.

Lay Conceptualizations of Racism

People also differ in the degree to which they believe racial disparities are the result of systemic racism (e.g., a history of structural, cultural, and institutionalized racism) as opposed to

conceptualizing racism as primarily a problem of a relatively few individual bigots (Bonam et al., 2019; Daumeyer et al., 2017; Miller, O'Dea, et al., 2021; Nelson et al., 2013; Rucker & Richeson, 2021a; Salter & Adams, 2013; Salter et al., 2018). People who conceptualize racism as more of a systemic problem show greater support for social protests and criminal justice reform (Miller, O'Dea, et al., 2021; Rucker & Richeson, 2021b) and perceive more racism in events such as in the aftermath of Hurricane Katrina (O'Brien et al., 2009). It is possible people who more strongly believe systemic racism is a problem are more likely to believe that implicit bias is a consequence of, and a significant contribution to, racial disparities. They may also more strongly believe that implicit bias is a social problem and be more motivated to address it. Thus, greater acknowledgement of the role that implicit bias may play in causing social harms may influence blame and punishment judgments in response to cases of discrimination caused by implicit bias.

How Implicit Bias is Conceptualized

Individual differences may be related to how people respond to discrimination attributed to implicit bias, but as already suggested, differences in how people conceptualize implicit bias may also affect blame-related judgments. If people think of implicit bias as unconscious and uncontrollable, then they may be less likely to perceive discrimination attributed to implicit bias as intentional because awareness and control are key components of intent (Cameron et al., 2010; Malle et al., 2014). Subsequently, if people perceive discriminatory behavior as unintentional, they may have less reason to blame and punish, and more reason to respond prosocially to, perpetrators when discrimination is caused by implicit bias. Alternatively, if people think of implicit bias as something that people can become consciously aware of in themselves, they may then think that they have a moral obligation to control their implicit bias from leading to

discriminatory behavior (Malle et al., 2014; Redford & Ratliff, 2016). Therefore, if implicit bias is conceptualized as something that people have the potential to be self-aware of and control, blame and punishment may be stronger, and prosocial responses weaker, when implicit bias leads to discrimination.

Despite implicit bias being often conceptualized as unconscious, some evidence suggests that people can become spontaneously aware of their implicit biases (Corneille & Hütter, 2020; de Houwer, 2019; Gawronski, 2019; Hahn et al., 2014; Hahn & Gawronski, 2019). Whether the behaviors that results from implicit biases are controllable (Amodio & Swencionis, 2018; Correll et al., 2014; Devine et al., 2002, 2012; Nosek et al., 2011; Suhler & Churchland, 2009) is another consideration that has important implications for antidiscrimination policies and bias interventions (Gawronski et al., 2020). As a scientific study, answering questions about awareness and control of implicit bias will help to better define and understand this type of bias, but it would also be informative to know how lay perceivers' understanding of implicit bias affects their judgments of discrimination caused by implicit bias. Such information could help shape public discourse and lead to a better understanding of the implications of communicating facts about implicit bias to the public (Daumeyer et al., 2019). The present research tested how different lay conceptualizations of implicit bias as unconscious and uncontrollable, or potentially conscious and controllable, affect punitive and prosocial responses to discrimination caused by implicit bias.

Acknowledging Implicit Bias

A related question is how perpetrators' prior acknowledgement[2] of their implicit bias might affect attributions of awareness, control, and blame. If people understand that to some degree most people have implicit biases, then they may be willing to acknowledge their own. One of the first steps in many bias interventions is to get the participants to acknowledge this possibility, for example, by using implicit association tests to reveal participants' biases (Fitzgerald et al., 2019; Greenwald et al., 2022; Lai et al., 2014). Acknowledgement suggests awareness and may therefore be perceived as an obligation to prevent known biases from causing harmful behavior, thus making discrimination attributed to implicit bias more blameworthy. However, people may also believe that implicit biases are difficult to control. If a perpetrator of discrimination has made a prior commitment to controlling their implicit biases and to being vigilant in preventing their biases from having unfair or harmful consequences, others may give them the benefit of the doubt when making judgments about how to respond to unintentional discrimination. To date, no studies have systematically manipulated prior acknowledgement of, and commitment to control, implicit bias to examine how this shapes perceivers moral judgments of discrimination. The present studies tested this extension of the research.

Hypotheses

The current research was designed to test the following hypotheses. First, based on the Path Model and Weiner's (1995, 1996) theory of responsibility, the intent-dominant hypothesis predicts that if people's reactions to discrimination are primarily affected by their perceptions of perpetrators' intent which are based on perceptions of perpetrators' mental states (e.g., awareness and control) rather than perceptions of harm, then people will respond more prosocially and less

[2] The word "acknowledge" is used here because it implies an admission or a declaration of the truth or existence of some reality. "Recognize" may also be appropriate, and has a similar meaning, but for consistency, "acknowledge" is used throughout this document.

punitively when discrimination is attributed to implicit, compared to explicit, bias. The following pattern was defined as more prosocial responses toward perpetrators:

- Attributing less intent, awareness, and control to perpetrators' mental states.
- Having more prosocial intentions (e.g., forgiving, consoling, expressing sympathy) toward perpetrators.
- Making less severe judgments of blame, responsibility, and accountability.
- Feeling lower levels of anger at perpetrators and higher levels of sympathy for perpetrators.
- Experiencing lower levels of desires to punish perpetrators.
- Perceiving higher levels of shame and guilt experienced by perpetrators.
- Perceiving perpetrators more positively in terms of their moral character.

More punitive responses toward perpetrators were defined as the reverse of this pattern (e.g., attributing more intent, making more severe judgments of blame, experiencing higher levels of desires to punish).

The current research also tested an alternative second hypothesis, the harm-dominant hypothesis, derived from the Culpable Control motivated blame theory that if people's judgments are not influenced by perpetrators' awareness of their biases, but are instead primarily focused on the harmful consequences of an act of discrimination, then we would *not* expect to find differences in people's reactions to discrimination attributed to implicit, compared to explicit, bias. Under the harm-dominant hypothesis, perceived harm should result in judgments of intent and blame, as well as punitive and prosocial reactions. In the present studies, the harmful outcomes to the victim were the same and only the type of bias differed. Therefore, the harm-dominant hypothesis predicts that judgments of intent and blame would not differ between

discrimination attributed to implicit, compared to explicit, bias. Note that although the Culpable Control theory does *not* argue that intent is irrelevant for blame, the harm-dominant hypothesis is framed as a stronger prediction based on the theory so that it is a testable hypothesis.

Study 2 additionally tested the hypothesis that people will respond more punitively and less prosocially toward perpetrators of discrimination attributed to implicit bias when they are primed with a conceptualization of implicit bias framed as a form of bias that people can be aware of and control with effort, compared to a form of bias that people are primarily unconscious of and cannot control. This hypothesis was based on the intent-dominant hypothesis because it predicts that people will perceive implicit bias that is potentially conscious and controllable as more blameworthy than implicit bias that is entirely unconscious and uncontrollable, perhaps because greater awareness and control imply an obligation to try to prevent discriminatory behavior (Malle et al., 2014; Redford & Ratliff, 2016). Study 2 was also designed to test the hypothesis that people will respond more punitively and less prosocially toward perpetrators of discrimination attributed to implicit bias when perpetrators acknowledge and express a commitment to controlling their implicit biases, compared to when they deny they have implicit biases, against the alternative hypothesis that people will respond less punitively and more prosocially because they are empathetic to someone who is aware of, and may be trying to control, their bias.

Additionally, the set of individual differences related to attitudes about bias and discrimination described above may account for levels of blame and related judgments (e.g., awareness, intent, harm) in addition to the type of bias attributed to discrimination (i.e., significant main effects of individual differences). These individual differences could also potentially moderate the effects of attributing discrimination to implicit, compared to explicit,

bias such that the differences in punitive and prosocial responses to these types of bias are minimized for people who are more likely to believe that racial discrimination is wrong (e.g., higher levels of PMAPS). It is likely that, in the present studies, individuals higher (but not lower) in these individual differences will perceive similar levels of harm when the consequences to the victims of discrimination are the same because the outcomes for the victim will be the same across the experimental manipulations in the present studies. Consistent with the harm-dominant hypothesis, for people who are likely to be more angry about racial discrimination (e.g., at higher levels of PMAPS or systemic conceptualizations of racism), perceiving similar levels of harm between more conscious (e.g., explicit) and less conscious (e.g., implicit) forms of bias should result in a similar pattern of moderation for perceptions of intent, attributions of blame, and punitive and prosocial reactions. However, if harm is moderated by these individual differences in the way just described, but intent, blame, punishment, and forgiveness are not (i.e., the differences between forms of bias are not minimized at higher, compared to lower, levels of the individual differences), then this pattern of moderation for harm and no moderation of blame would support a rejection of the harm-dominant hypothesis. In other words, perceived harm would not appear to be driving perceptions of intent and blameworthiness.

Chapter 2 - Study 1

Study 1 tested the intent-dominant hypothesis against the harm-dominant hypothesis by manipulating whether an act of discrimination was attributed to implicit or explicit bias and measuring participants' punitive and prosocial responses. Specifically, participants read scenarios in which an academic advisor was found to discriminate against Black students by advising them to choose majors in less intellectually demanding fields because of conscious or unconscious attitudes and stereotypes. Participants then provided their perceptions of perpetrator awareness, control, intent, and blame, as well as their anger and sympathy for the perpetrator and their support for punishment and prosocial responses.

Method

Participants

Participants were recruited using CloudResearch. Although these samples are not truly representative of the population, they are more diverse than college student samples with respect to age, education, and political views. Recruitment was limited to residents of the United States who were 18 years or older. Furthermore, eligible participants needed to have at least 100 approved hits with a 95% approval rating. The top 5% of workers—those who participate in 56% of the studies on CloudResearch—were also excluded from eligibility to improve the naivete of the participants.

A target sample size of 260 participants was determined using power analyses based on 80% power to detect the small to moderate effect sizes, $d = .35$, $f = .15$, found in previous studies (Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016). Participants were compensated $1.00 for their time (approximately 15 minutes to complete the materials). Although this amount is less than minimum wage, participants were informed prior to signing up for the study that "We are a

non-profit research group at a public university without grant funding. We are offering what we can as a token of our appreciation for your time and help in our research. We understand that the amount we are offering is not a fair hourly wage, and we ask that you not participate if you are discouraged by the amount we are able to offer you at this time."

Based on expected loss of participants due careless responding, 312 participants were initially recruited. Participants' data were removed for the following reasons: not fully completing the study (n = 9), completing the study too quickly (under 270 seconds, n = 8), not indicating that they read the vignette (n = 2), failing the manipulation attention check (n = 34). The result was a final sample size of 261 participants (ages 19 to 68, M = 37.54, SD = 11.55; 71% White; 66% Female; 95% United States nationality; 90% had at least some college-level education).

Bias Manipulation

Study 1 was a between-groups experimental design manipulating whether an act of racial discrimination was attributed to explicit bias or implicit bias. Participants were asked to read a short passage describing a case in which a college advisor racially discriminated against a Black student by advising the student to choose a less challenging academic major (see Appendix A for the full wording of the vignettes that was used). In the vignette the student revealed that she believed that she was discriminated against, which was followed up with an investigation by university officials. In the explicit bias condition, the vignette stated that the investigatory team concluded that the advisor's behavior was due to conscious attitudes and stereotypes. In the implicit bias condition, the vignette stated that the investigatory team concluded that the advisor's behavior was due to unconscious attitudes and stereotypes. Following the vignette, participants completed a manipulation check pertaining to the type of bias

(conscious/unconscious) attributed to the discriminatory behavior (see Appendix A for manipulation check materials).

Dependent Variables

The following outcome measures were based on materials used in prior research examining reactions to discrimination attributed to implicit bias (Daumeyer et al., 2019, 2021). Measures of the dependent variables were modified to correspond to the scenario described in the vignettes and additional items were created to measure emotional and prosocial reactions (see Appendix A for a complete description of the materials). Unless otherwise stated, participants were given the instructions "Please indicate your agreement with the following statements from 1 (Strongly Disagree) to 7 (Strongly Agree)" for each scale. The order of each dependent variable scale, as well as all the items within a scale, were randomized. Each of the scales used as dependent variables were analyzed to achieve acceptable levels of reliability ($\alpha > .70$) and reliable sets of items were averaged together to create composite variables prior to analyzing the results of the study.

Mental State Attributions of the Perpetrator. Several scales were created to measure participants' inferences about the mental states of the perpetrator.

Awareness. Inferences about whether the perpetrator was aware of her bias and its effect on her behavior was measured using three items (e.g., *Melissa knew she was discriminating against Brianna, Melissa was aware of how her bias was affecting her behavior*), $\alpha = .97$.

Control. Inferences about how much control the perpetrator had over her behavior was measured using four items (e.g., *Melissa could have chosen to not discriminate against Brianna, There is no way that Melissa could have stopped herself from discriminating against Brianna*—reverse scored), $\alpha = .84$.

Intent. Perceptions of the perpetrator's intent to racially discriminate were measured using five items (e.g., *Melissa intended to discriminate against Brianna, Melissa discriminated against Brianna on purpose*), $\alpha = .97$.

Foreseeability. The extent to which the perpetrator should have been able to foresee the potential for her bias to result in discrimination was measured with two items (e.g., *There is no way Melissa could have known she would discriminate against Brianna*—reverse scored), $\alpha = .72$.

Blame Judgments (Responsibility/Accountability). The extent to which participants blamed the perpetrator for racially discriminating was measured using three items (e.g., *I blame Melissa for discriminating against Brianna, Melissa should be held accountable for discriminating against Brianna*), $\alpha = .94$.

Emotional Reactions. Participants' level of anger at the perpetrator was measured using four items (e.g., *I am outraged at Melissa's discriminatory behavior*), $\alpha = .95$. Participants' level of sympathy for what the perpetrator was going through was measured using four items (e.g., *I sympathize with what Melissa must be going through*), $\alpha = .91$.

Punishment. Participants' support for punishing the perpetrator was measured using nine items describing different severities of punishment (see Appendix A). Because these items reflected different levels of punishment severity, exploratory factor analysis was used to identify the latent variables in the set. A principal components analysis with varimax rotation extracted two factors using parallel analysis. These factors were interpreted as mild punishment (three items, e.g., *Melissa should be forced to undertake bias sensitivity training*) and severe punishment (four items, e.g., *Melissa should be fired*). Two additional items were also included in the materials, *Melissa should be punished*, and *Melissa should be put on probation*, but these

items cross-loaded onto both mild and severe punishment factors in the principal components analysis, so these two items were not included in the composite variables and were not analyzed further. Two composite variables were created by averaging the relevant items together: mild punishment $\alpha = .81$, and severe punishment $\alpha = .90$.

Prosocial Reactions. Prosocial reactions to the perpetrator were measured using nine items. Because these items reflected different kinds of prosocial behaviors, exploratory factor analysis was used to identify the latent variables in the set. A principal components analysis with varimax rotation extracted two factors using parallel analysis. These factors were interpreted as consoling/forgiving (five items, e.g., *Someone should console Melissa*, *Melissa should be forgiven for her behavior*) and helping to correct the perpetrator's bias (three items, e.g., *Someone should help Melissa learn from her mistake*). One additional item was included in the materials, *I think Melissa will become a better person from this experience*, but was omitted because it did not reliably load onto either of the two factors and was not analyzed further. Two composite variables were created by averaging the relevant items together: consoling/forgiving $\alpha = .87$, and helping to correct $\alpha = .76$.

Perceptions of the Perpetrator's Moral Character. Perceptions of the perpetrator's moral character were measured using three items (e.g., *Melissa is morally flawed*, *Melissa is a good person that made an unfortunate mistake*), $\alpha = .83$.

Institutional Reform. The extent to which the institution should take action to reduce discrimination on campus was measured using three items (e.g., *The university should implement bias awareness training for all of its staff, The university should implement policies that reduce the likelihood that discrimination will occur anywhere on campus*), $\alpha = .96$.

Harm. Perceptions of how much harm was caused to the victim was measured using five items (e.g., *Hurtful for Brianna, Potentially limiting Brianna's opportunities* on a scale from 1 (*Not at all*) to 7 (*Very Much*), $\alpha = .91$.

Redress. Support for redress to the victim was measured using four items (e.g., *Brianna should be awarded a large sum of money in a legal settlement, Brianna should receive an apology*), $\alpha = .75$.

Sympathy for the Victim. Participants' sympathy for the victim was measured using four items (e.g., *I sympathize with what Brianna was going through, I don't feel bad for Brianna*—reverse scored), $\alpha = .90$.

Evaluations of Bias. Three separate items were used to measure participants' general evaluations of the wrongness of the act of discrimination (*The discrimination that Brianna experienced was morally wrong*), and the pervasiveness (*The kind of bias described in the passage is a widespread problem*) and acceptability (*Some level of bias in situations like the one described should just be expected*) of the kind of bias described in the vignette.

Individual Differences

Perspective-Taking. The extent to which the participants took the perspective of the victim was measured using three items (e.g., *When I read the passage, I imagined myself being in Melissa's situation*), $\alpha = .83$. The extent to which the participants took the perspective of the perpetrator was measured using three items (e.g., *When I read the passage, I imagined myself in Brianna's situation*), $\alpha = .87$.

Propensity to Make Attributions to Prejudice. (Miller & Saucier, 2018)**.** General tendencies to attribute potentially biased behavior to the racial prejudice of others were measured using the propensity to make attributions to prejudice scale (PMAPS). PMAPS is a 15-item

measure comprised of four factors: pervasiveness (four items, e.g., *Racist behavior is more widespread than people think it is*), trivialization (four items, e.g., *Minorities today are overly worried about being victims of racism*), vigilance (four items, e.g., *I am on the lookout for instances of prejudice or discrimination*), and confidence (three items, e.g., *I am quick to recognize prejudice*). In the current study, the overall composite of the 15 items was used, α = .90.

Lay Conceptualizations of Racism. (Miller, O'Dea, et al., 2021)**.** Participants' beliefs about racism were measured using four items measuring beliefs that racism is due to systematic oppression (e.g., *A history of policies that systematically disadvantaged People of Color*), α = .94, and four items measuring beliefs that racism is due to individual acts of prejudice (e.g., *Individuals' beliefs about White racial superiority*), α = .89. Participants were given the following instructions: "Please rate each of the following statements in terms of how much each contributes to the problem of racism in the United States today on a scale from 1 (*Not at all*) to 7 (*Very much*)."

Bias Awareness. (Perry et al., 2015)**.** Participants' awareness of their implicit bias was measured using four items (e.g., *When talking to people, I sometimes worry that I am unintentionally acting in a prejudiced way*; *I worry that I have unconscious biases toward some social groups*), α = .83.

Procedure

After providing informed consent, participants were randomly assigned to one of the two experimental conditions and were asked to read the vignette corresponding to their assigned condition. After completing the manipulation checks, participants were asked to respond to measures of the dependent variables, followed by the measures of the potential moderating

variables (presented in a randomized order). Following these measures, participants completed relevant demographic information. Lastly, participants were debriefed about the nature of the study, and informed that the scenario they read about was a fictional story that was created for the purpose of this study. Participants were provided with contact information if they had questions or concerns, thanked for their participation, and given instructions for how to receive payment.

Results and Discussion

Bivariate Correlations

Correlations Between Dependent Variables. As expected, the bivariate correlations (see Table 2.1) between the dependent variables measuring the perceived mental states of the perpetrator (awareness, control, and intent) were all moderately to strongly correlated. Notably, perceptions of awareness were more strongly[3] correlated with attributions of intent than were perceptions of control. In turn, these mental state attributions were moderately to strongly correlated with blame, a finding that supports models claiming mental state attributions are a significant precursor to (Malle et al., 2014), or consequence of (Alicke, 2000), blame. Blame was strongly and positively correlated with anger at the perpetrator, and moderately and negatively correlated with sympathy for the perpetrator. Mild and severe punishment were moderately correlated, suggesting a moderate degree of overlap between decisions about relatively moderate and severe forms of punishment. Both forms of punishments were more strongly correlated with blame and anger than with sympathy for the perpetrator, suggesting that blame and negative emotions may play more of a role in decisions to punish than are more prosocial emotions. Notably, severe punishment was more strongly correlated with perceptions of intent than was

[3] Effect size comparisons discussed throughout this manuscript are descriptive, not inferential, comparisons.

mild punishment, suggesting that attributions of intent may play more of a role in decisions about relatively more severe forms of punishment. More severe forms of punishment may require greater evidence that warrants their use, and perceptions of intent may work to justify severe punishment. This interpretation is consistent with the idea that intent provides warrant for blame and punishment (Malle et al., 2014).

Conversely, intentions to forgive and console the perpetrator were more strongly correlated with sympathy for the perpetrator than with blame and anger. Intentions to help the perpetrator correct her bias were weakly, but positively correlated with blame and anger, and moderately correlated with support for mild punishment, but not significantly correlated with sympathy for the perpetrator. This suggests that the items measuring this kind of help may be construed as somewhat punitive. However, helping to correct bias was weakly, and positively correlated with intentions to forgive and console the perpetrator. Together, this pattern of correlations suggests that considerations about helping a person to correct their bias may not be entirely prosocial, but rather a mix of prosocial and punitive attitudes. Further analyses and conclusions will take this into account.

Perceptions of the moral character of the perpetrator were correlated with the mental state attributions, blame, anger, sympathy, punishment, and prosocial responses (except for intentions to help correct bias). This suggests that people who perceived the behavior as more blameworthy and punishment-worthy, were also more likely to perceive the perpetrator as being more morally flawed. It also suggests that people may be more likely to attribute instances of discrimination that are perceived to be more blameworthy to the moral character of the perpetrators.

Perceptions of harm to the victim were positively correlated with perceptions of the perpetrator's awareness, control, and intent—a pattern consistent with previous findings that

intentional harms are more painful (Ames & Fiske, 2013; Gray & Wegner, 2008). Blame, anger, and punishment were even more strongly correlated with perceptions of harm. Additionally, perceived harm was strongly correlated with sympathy for the victim, support for institutional reform, and support for compensating the victim (redress).

Overall, these patterns of correlations were expected and are consistent with theories of blame and moral responsibility. Mental state attributions (awareness, control, intent) were positively correlated with blame, anger, and punishment, and negatively correlated with sympathy and forgiveness. Further tests of the potential psychological processes involved in these judgments are examined in the Competing Models of the Blame Process section below. Additionally, these patterns of correlations support the convergent validity of the scales that were used to measure these constructs because the nomological network of these variables was theoretically consistent.

Table 2.1. Bivariate Correlations Between Dependent Variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Awareness	(.97)															
2. Control	.59***	(.84)														
3. Intent	.89***	.59***	(.97)													
4. Foresee	.62***	.73***	.63***	(.72)												
5. Blame	.69***	.65***	.72***	.69***	(.94)											
6. Anger	.62***	.59***	.59***	.61***	.76***	(.95)										
7. Sympathy - Perpetrator	-.52***	-.45***	-.55***	-.50***	-.61***	-.62***	(.91)									
8. Mild Punishment	.45***	.44***	.40***	.59***	.66***	.73***	-.52***	(.81)								
9. Severe Punishment	.64***	.37***	.67***	.50***	.64***	.69***	-.53***	.51***	(.90)							
10. Help to Correct	.03	.14*	-.09	.21***	.20***	.34***	-.09	.59***	.08	(.76)						
11. Console/Forgive	-.46***	-.40***	-.56***	-.39***	-.47***	-.41***	.59***	-.24***	-.43***	.27***	(.87)					
12. Moral Character	-.65***	-.43***	-.71***	-.56***	-.66***	-.67***	.63***	-.54***	-.75***	-.03	.62***	(.83)				
13. Institutional Reform	.29***	.32***	.24***	.46***	.48***	.60***	-.37***	.79***	.39***	.67***	-.12*	-.41***	(.96)			
14. Harm	.35***	.34***	.30***	.47***	.52***	.64***	-.43***	.69***	.46***	.51***	-.15*	-.47***	.69***	(.91)		
15. Redress	.39***	.37***	.38***	.46***	.55***	.63***	-.43***	.65***	.67***	.37***	-.23***	-.52***	.58***	.67***	(.75)	
16. Sympathy - Victim	.32***	.36***	.27***	.49***	.52***	.70***	-.38***	.73***	.42***	.56***	-.15*	-.46***	.75***	.73***	.64***	(.90)

Note. Cronbach's alphas are on the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$

Effects of Discrimination Attributed to Explicit or Implicit Bias

To test the main hypothesis about the effects of discrimination attributed to explicit or implicit bias, a multivariate analysis of variance (MANOVA) on the dependent variables was conducted to control for type I error rates. The omnibus MANOVA was significant, $F(19, 241) = 27.07$, $p < .001$. Overall, the pattern of results was consistent with the intent-dominant hypothesis (see Table 2.3, Figure 2.1). Large effects of bias attribution were observed for perceptions of the perpetrator's mental states, such that participants perceived less awareness, control, intent, and foreseeability when discrimination was attributed to implicit, compared to explicit, bias. It was not surprising that that the effects on awareness and control were large given that the vignettes described the perpetrator as having either a conscious or unconscious bias against Black people, and thus may be considered checks on the effectiveness of the manipulation. More importantly, the large effects of the bias attribution manipulation on perceptions of intent and degrees of blame replicates prior research on the consequences of attributing discrimination to implicit bias (Cameron et al., 2010; Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016).

Also consistent with the intent-dominant hypothesis, participants reported feeling less anger and more sympathy for the perpetrator when discrimination was attributed to implicit, compared to explicit, bias. Perhaps a combination of less blame and anger explains participants having less desire to punish the perpetrator when discrimination was attributed to implicit, compared to explicit, bias (this was examined further in the path analyses described below).

Interestingly, although both effects were significant, the size of the effect of the bias attribution manipulation on more mild forms of punishment was descriptively weaker than the effect on more severe forms of punishment. Perhaps this makes sense in the context of the fact that discrimination occurred in both conditions in the study. For weaker forms of punishment,

such as being forced to apologize or complete a bias sensitivity course, when discrimination is caused by implicit bias, people may still support these corrective actions (as suggested by the observed mean of 5.26 on a seven-point scale for mild punishment in the implicit bias condition) and they may be only slightly more supportive of these actions when discrimination is caused by explicit bias. For more severe forms of punishment, however, people may feel as though they need more justification of wrongdoing (e.g., intentional discrimination) and the rather large effect size for severe punishment supports this interpretation.

Intentions to help the perpetrator correct her bias were not affected by the bias attribution manipulation. Like support for mild forms of punishment, perhaps this kind of corrective action is something that people would support when discrimination occurs, regardless of whether it was intentional or not. However, participants reported a greater willingness to console and forgive the perpetrator when discrimination was attributed to implicit, compared to explicit, bias. This was predicted by the intent-dominant hypothesis, and in conjunction with the higher levels of sympathy for the perpetrator in the implicit, compared to explicit, bias condition, it is also consistent with Weiner's (1995, 1996) theory of responsibility. Additionally, attributing discrimination to implicit, rather than explicit, bias had a large effect on perceptions of the perpetrator's moral character.

Because the consequences for the victim were constant between the two bias attribution conditions in this experiment, it might be reasonable to assume that perceptions of harm would not be affected by whether discrimination was caused by implicit or explicit bias. However, there was a moderate effect of the bias attribution manipulation, such that there was greater perceived harm and greater sympathy for the victim when discrimination was attributed to explicit, compared to implicit, bias. This finding makes sense, however, considering findings that

intentional harms are more psychologically painful than unintentional harms (Gray & Wegner, 2008), and in the present study, discrimination attributed to explicit bias was perceived as much more intentional than discrimination attributed to implicit bias. Additionally, support for institutional reform and compensation to the victim (redress) were lower in the implicit, compared to explicit, bias condition. Perhaps this was because of the lesser harm perceived in the implicit, compared to explicit, bias condition.

Finally, discrimination attributed to explicit bias was perceived as more wrong and less acceptable than discrimination attributed to implicit bias. These overall evaluations of the appropriateness of the perpetrator's actions suggest that people are making moral judgments about discrimination based on the perpetrator's mental states. Overall, these findings support the intent-dominant hypothesis and fail to support the harm-dominant hypothesis that people's judgments are not influenced by perpetrators' intent and are instead primarily affected by the harmful consequences of an act of discrimination. Although the bias attribution manipulation had a moderate effect on perceived harm, this effect was not proportionate to rather large effects on intent, blame, and punishment. Additionally, a follow-up multivariate analysis of covariance controlling for harm did not significantly alter the effects already described. Thus, the less punitive and more prosocial responses to discrimination attributed to implicit, compared to explicit, bias appear to be driven by participants' consideration for whether the perpetrator's bias was conscious and intentional, or unconscious and unintentional.

Table 2.2. Effects of Attributing Discrimination to Implicit, Compared to Explicit, Bias

Dependent Variable	Explicit Bias *M* (*SD*)	Implicit Bias *M* (*SD*)	*F*	*p*	*d*
Awareness	5.80 (1.17)	2.62 (1.47)	367.96	< .001	2.38
Control	6.08 (1.12)	4.89 (1.39)	57.11	< .001	0.94
Intent	5.54 (1.38)	2.64 (1.44)	257.35	< .001	1.99
Foresee	6.13 (1.01)	4.71 (1.39)	83.91	< .001	1.14
Blame	6.12 (1.14)	4.36 (1.68)	95.15	< .001	1.21
Anger	5.69 (1.53)	4.30 (1.63)	50.40	< .001	0.88
Sympathy - Perpetrator	2.39 (1.37)	3.54 (1.57)	39.19	< .001	-0.78
Mild Punishment	6.08 (1.20)	5.23 (1.61)	23.25	< .001	0.60
Severe Punishment	4.10 (1.74)	2.31 (1.28)	90.54	< .001	1.18
Help to Correct	5.69 (1.16)	5.80 (1.33)	0.49	.485	-0.09
Console/Forgive	3.57 (1.40)	4.72 (1.22)	50.20	< .001	-0.88
Moral Character	3.39 (1.43)	4.84 (1.25)	76.16	< .001	-1.08
Institutional Reform	6.26 (1.20)	5.88 (1.67)	4.54	.034	0.26
Harm	6.13 (1.00)	5.55 (1.42)	14.19	< .001	0.47
Redress	5.27 (1.21)	4.55 (1.22)	23.25	< .001	0.60
Sympathy - Victim	6.10 (1.15)	5.58 (1.50)	9.92	.002	0.39
Wrongness	6.30 (1.12)	5.46 (1.84)	19.82	< .001	0.55
Pervasiveness	5.90 (1.46)	5.49 (1.84)	3.96	.048	0.25
Acceptability	2.50 (1.73)	3.03 (1.88)	5.65	.018	-0.29

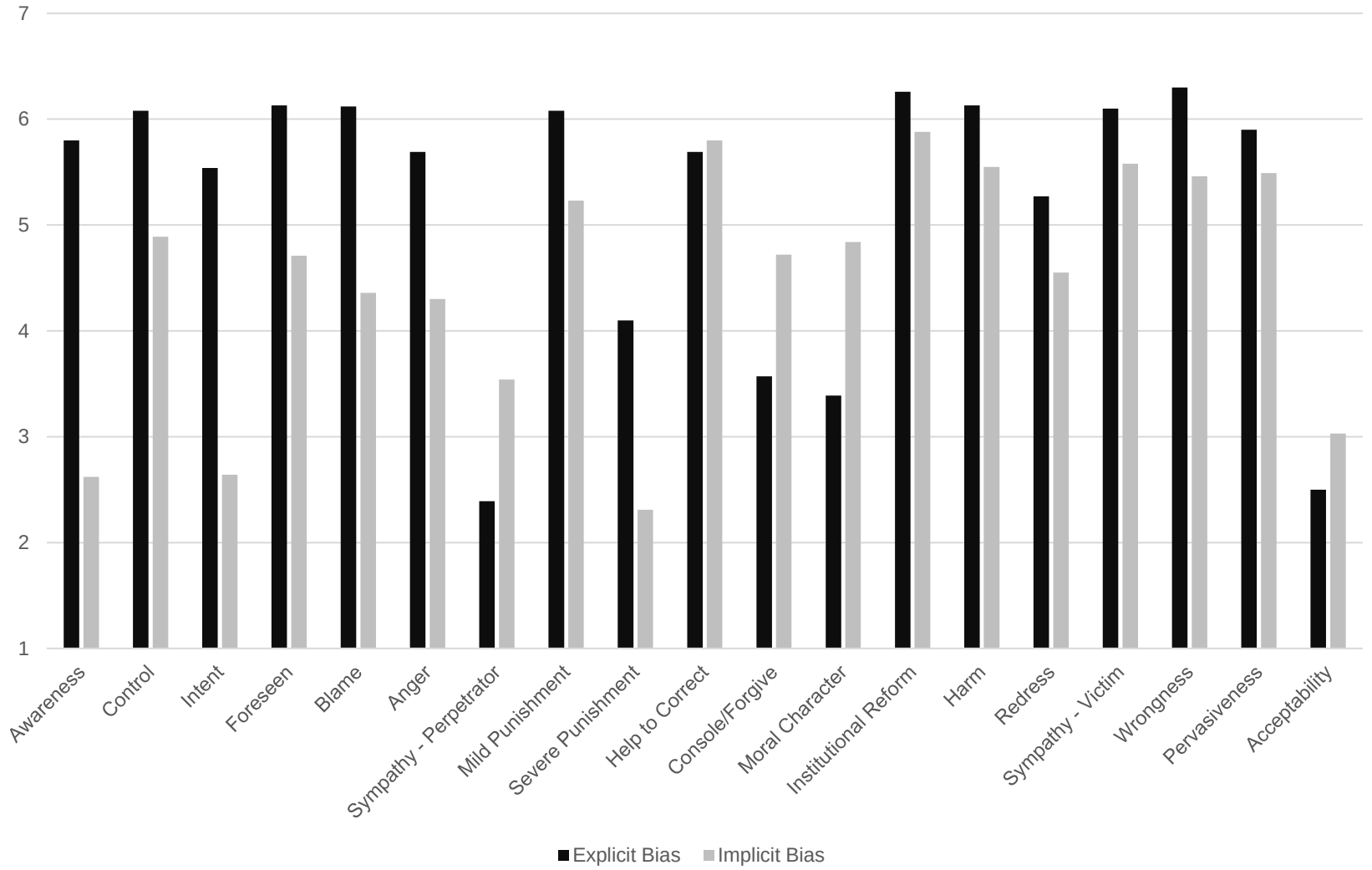


Figure 2.1. Effects of attributing discrimination to implicit, compared to explicit, bias.

Individual Differences

To examine how individual differences in perspective-taking, bias awareness, PMAPS, and lay conceptualizations of racism are associated with punitive and prosocial responses to discrimination, and to test whether the effects of attributing discrimination to implicit, compared to explicit, bias are moderated by these individual differences, a series of separate regression models were conducted on each of the dependent variables. Each regression model contained the experimentally manipulated independent variable, the continuous individual difference predictor variable, and their two-way interaction. In the sections that follow, the bivariate correlations between individual differences are examined first, followed by the bivariate correlations between the individual differences and the dependent variables (Table 2.4). The following sections describe the main effects (i.e., bivariate correlations) of the individual differences along with the significant interactions that were found, but for the sake of brevity, they do not report the main effects of the bias attribution manipulation because these were already discussed in the previous section. The final part of this section interprets the general trends that were consistently found across the individual differences to draw conclusions from these patterns.

Correlations Between Individual Differences. Table 2.3 contains the bivariate correlations between the individual difference measures. More strongly taking the perspective of the perpetrator was related to lower levels of PMAPS, lower levels of perceiving racism as a problem associated with individual bigots, and higher levels bias awareness. Conversely, more strongly taking the perspective of the victim was related to higher levels of PMAPS, higher levels of perceiving racism as a systemic problem, and higher levels of perceiving racism as an individual problem. Higher levels of PMAPS in turn were related to higher levels of perceiving

racism as a systemic problem and higher levels of perceiving racism as an individual problem. Interestingly, PMAPS was more strongly related to perceptions that racism is a systemic problem than with perceptions that racism is the result of individual bigots, a pattern consistent previous research (Miller, O'Dea, et al., 2021). Furthermore, higher levels of bias awareness were also more strongly related to higher levels of perceiving racism as a systemic problem than with perceiving racism as an individual problem. Perhaps this demonstrates an awareness that one's own subtle biases may be a consequence of systems of oppression that perpetuate inequities between groups.

Overall, these patterns of correlations were expected given the nature of the constructs. However, that bias awareness was positively correlated with perpetrator perspective-taking is a unique finding in the research to date. One interpretation of this relationship is that people who are more aware of their own subtle biases (i.e., higher bias awareness) may be more likely to understand how such bias in others may result in discriminatory behavior and may therefore more easily take the perspective of people who unintentionally discriminate. Future research should test the validity of this explanation.

Table 2.3. Bivariate Correlations Between Individual Differences

	1.	2.	3.	4.	5.	6.
1. Perpetrator Perspective-Taking	(.87)					
2. Victim Perspective-Taking	-.02	(.83)				
3. PMAPS	-.16*	.28***	(.90)			
4. Systemic Racism	-.10	.24***	.78***	(.94)		
5. Individual Racism	-.13*	.28***	.61***	.67***	(.89)	
6. Bias Awareness	.48***	-.06	.20**	.23***	.12*	(.83)

Note. PMAPS = propensity to make attributions to prejudice scale; Cronbach's alphas are on the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$

Table 2.4. Bivariate Correlations Between Individual Differences and Dependent Variables

	PT Perpetrator	PT Victim	PMAPS	Systemic Racism	Individual Racism	Bias Awareness
Awareness	-.31***	.16**	.22***	.20**	.17**	-.06
Control	-.22***	.05	.29***	.30***	.33***	-.02
Intent	-.37***	.14*	.22***	.18**	.16*	-.12
Foresee	-.30***	.15*	.43***	.42***	.39***	.02
Blame	-.38***	.24***	.37***	.36***	.38***	-.06
Anger	-.40***	.30***	.54***	.47***	.44***	-.07
Sympathy - Perpetrator	.55***	-.21***	-.35***	-.34***	-.33***	.23***
Mild Punishment	-.29***	.25***	.54***	.54***	.50***	.06
Severe Punishment	-.33***	.29***	.40***	.35***	.25***	-.07
Help to Correct	.11	.13*	.43***	.41***	.47***	.28***
Console/Forgive	.45***	-.03	-.18**	-.15*	-.14*	.19**
Moral Character	.40***	-.32***	-.39***	-.38***	-.33***	.07
Institutional Reform	-.13*	.25***	.65***	.66***	.54***	.18**
Harm	-.16**	.32***	.58***	.55***	.53***	.10
Redress	-.26***	.31***	.53***	.51***	.45***	.04
Sympathy - Victim	-.20**	.36***	.62***	.59***	.57***	.06
Wrongness	-.22***	.28***	.55***	.50***	.52***	.09
Pervasiveness	-.05	.33***	.72***	.78***	.61***	.26***
Acceptability	.45***	.01	-.23***	-.16**	-.21***	.23***

Note. PT = Perspective-Taking; PMAPS = propensity to make attributions to prejudice scale.
* $p < .05$; ** $p < .01$; *** $p < .001$

Table 2.5. Regression Coefficients Testing Individual Difference x Bias Attribution Interactions

	PT Perpetrator	PT Victim	PMAPS	Systemic Racism	Individual Racism	Bias Awareness
Awareness	-0.14	< 0.01	-0.07	-0.29*	0.27	-0.19
Control	-0.06	-0.15	-0.08	-0.22	0.24	-0.11
Intent	-0.16	0.12	0.30	-0.26	0.56**	-0.15
Foresee	-0.18*	-0.03	-0.19	-0.35***	0.35*	-0.26**
Blame	-0.11	-0.12	-0.23	-0.25*	-0.07	-0.22*
Anger	-0.18	-0.23*	-0.31*	-0.24	-0.18	-0.28*
Sympathy - Perpetrator	-0.08	0.16	0.20	0.39**	-0.35	0.06
Mild Punishment	-0.23*	-0.26**	-0.59***	-0.41***	-0.13	-0.32**
Severe Punishment	-0.15	0.13	0.34*	0.07	0.03	-0.17
Help to Correct	-0.13	-0.30***	-0.44***	-0.24*	-0.12	-0.19*
Console/Forgive	0.05	< 0.01	-0.33*	0.13	-0.46**	-0.02
Moral Character	0.12	-0.06	-0.07	0.19	-0.26	0.19
Institutional Reform	-0.36***	-0.27**	-0.47***	-0.24*	-0.12	-0.46***
Harm	-0.18	-0.19*	-0.32**	-0.18	-0.19	-0.34***
Redress	-0.14	-0.18*	0.02	0.04	-0.16	-0.23*
Sympathy - Victim	-0.35***	-0.23**	-0.36**	-0.21*	-0.08	-0.42***
Wrongness	-0.30**	-0.29**	-0.62***	-0.35**	-0.06	-0.42***
Pervasiveness	-0.18	-0.32**	-0.31*	-0.20*	0.09	-0.40**
Acceptability	0.16	0.05	0.09	0.19	-0.18	0.01

Note. Cells in this table show the regression coefficients testing the interaction between the individual difference and the bias attribution manipulation; PT = Perspective-Taking; PMAPS = propensity to make attributions to prejudice scale.
* $p < .05$; ** $p < .01$; *** $p < .001$

Perpetrator and Victim Perspective-Taking. Generally, the main effects of perpetrator perspective-taking showed that higher levels of perpetrator perspective-taking were associated with less punitive and more prosocial responses to discrimination. Perpetrator perspective-taking

was also negatively correlated with perceived harm and overall wrongness, and positively correlated with the acceptability of the behavior. A few two-way interactions were found between perpetrator perspective-taking and the bias attribution manipulation (see Table 2.5). The overall pattern of these interactions (see Figure 2.2) revealed larger differences in reactions to discrimination attributed to implicit, compared to explicit, bias at lower levels of perpetrator perspective taking. For these measures, the slopes for perpetrator perspective taking were larger in the explicit, compared to the implicit, bias condition.

For victim perspective-taking, the bivariate correlations revealed that higher levels of victim perspective-taking were associated with more punitive responses to discrimination, but victim perspective-taking was not significantly correlated with intentions to console and forgive the perpetrator. Several two-way interactions were found between victim perspective-taking and the bias attribution manipulation (see Table 2.5). The overall pattern of these interactions (see Figure 2.3) showed greater differences in reactions to discrimination attributed to implicit, compared to explicit, bias at lower levels of victim perspective-taking. In contrast to the patterns observed for perpetrator perspective-taking, there were stronger relationships between victim perspective-taking and these measures in the implicit, compared to explicit, bias condition. At higher levels of victim perspective-taking, perceptions of harm, wrongness, and pervasiveness, and support for mild punishment and institutional reform were similar for discrimination attributed to implicit and explicit bias.

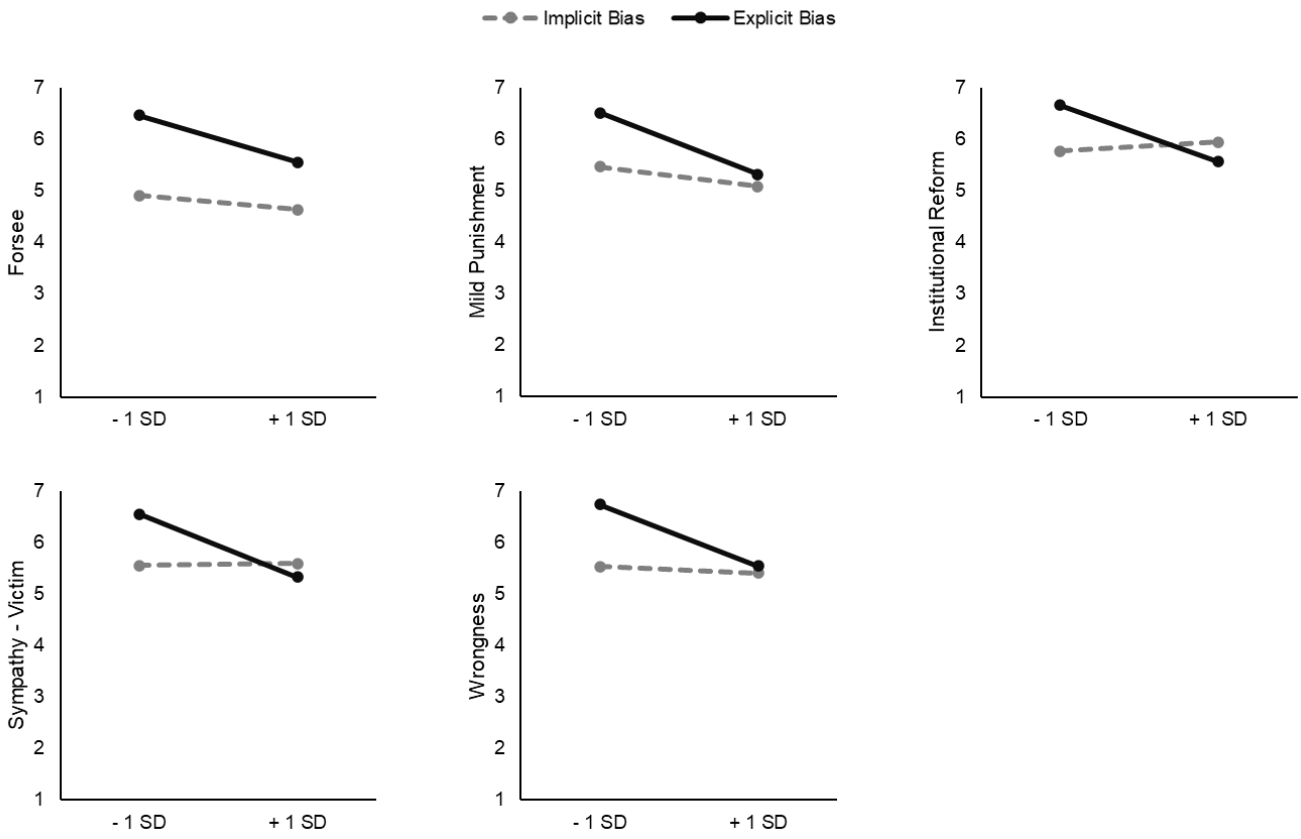


Figure 2.2. Perpetrator perspective-taking moderating bias attribution effects. Data are plotted at one standard deviation below and above the sample mean for perpetrator perspective-taking.

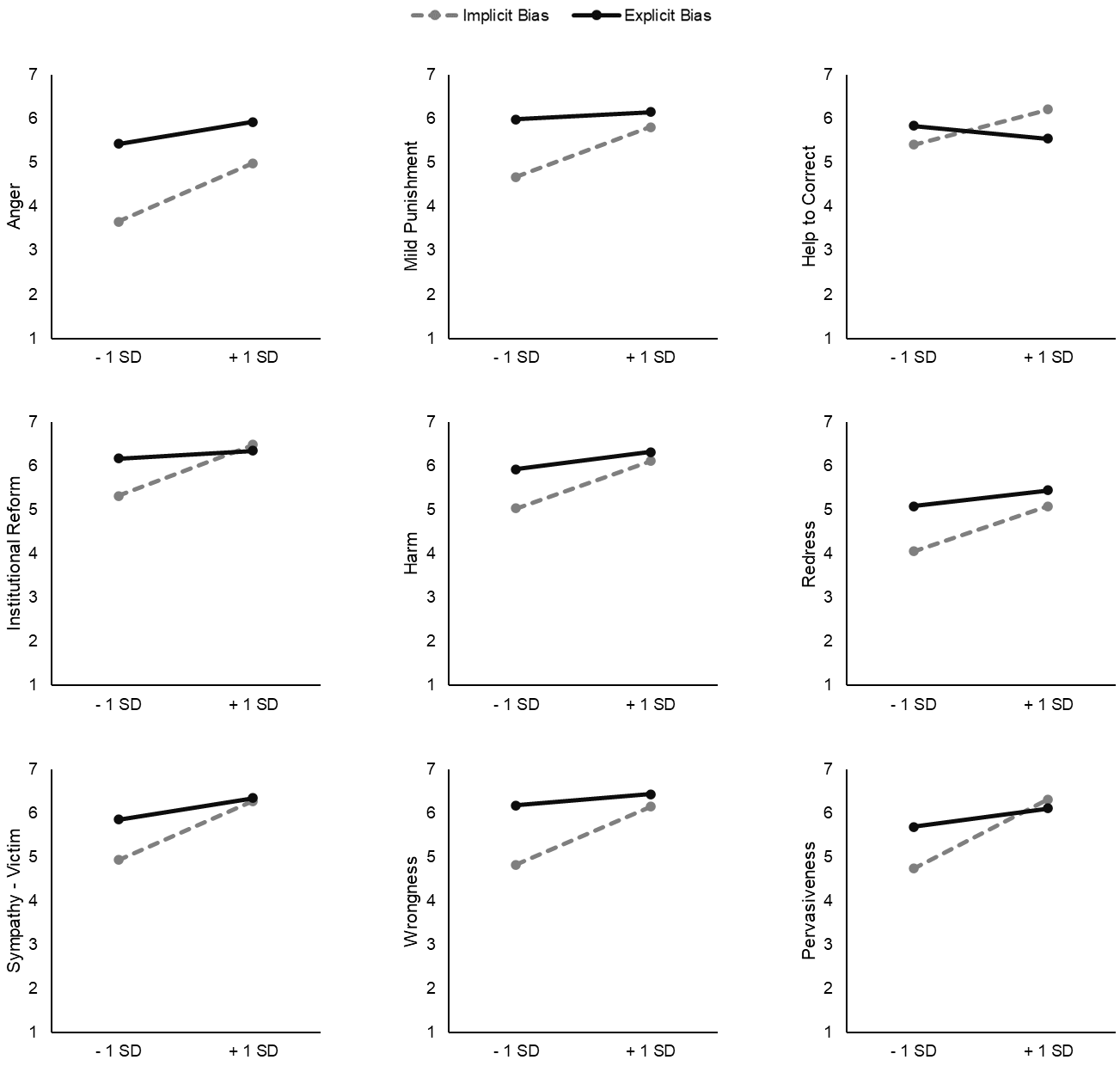


Figure 2.3. Victim perspective-taking moderating bias attribution effects. Data are plotted at one standard deviation below and above the sample mean for victim perspective-taking.

Propensity to Make Attributions to Prejudice. Generally, the bivariate correlations revealed that higher levels of PMAPS were associated with more punitive and less prosocial reactions to discrimination (see Table 2.4). The significant two-way interactions between PMAPS and the bias attribution manipulation (see Table 2.5) showed greater differences in reactions to discrimination attributed to implicit, compared to explicit, bias at lower levels of

PMAPS for anger, mild punishment, institutional reform, harm, sympathy for the victim, wrongness, and pervasiveness (see Figure 2.4). For these variables, the slopes for PMAPS were larger in the implicit, compared to explicit, bias condition. For support for severe punishment, intentions to help the perpetrator correct her bias, and intentions to console and forgive, larger differences in reactions to discrimination attributed to implicit, compared to explicit, bias were seen at higher levels of PMAPS. For these variables, the slopes for PMAPS were larger in the explicit, compared to implicit, bias condition, with the exception of intentions to help the perpetrator correct her bias where the slope was larger in the implicit bias condition.

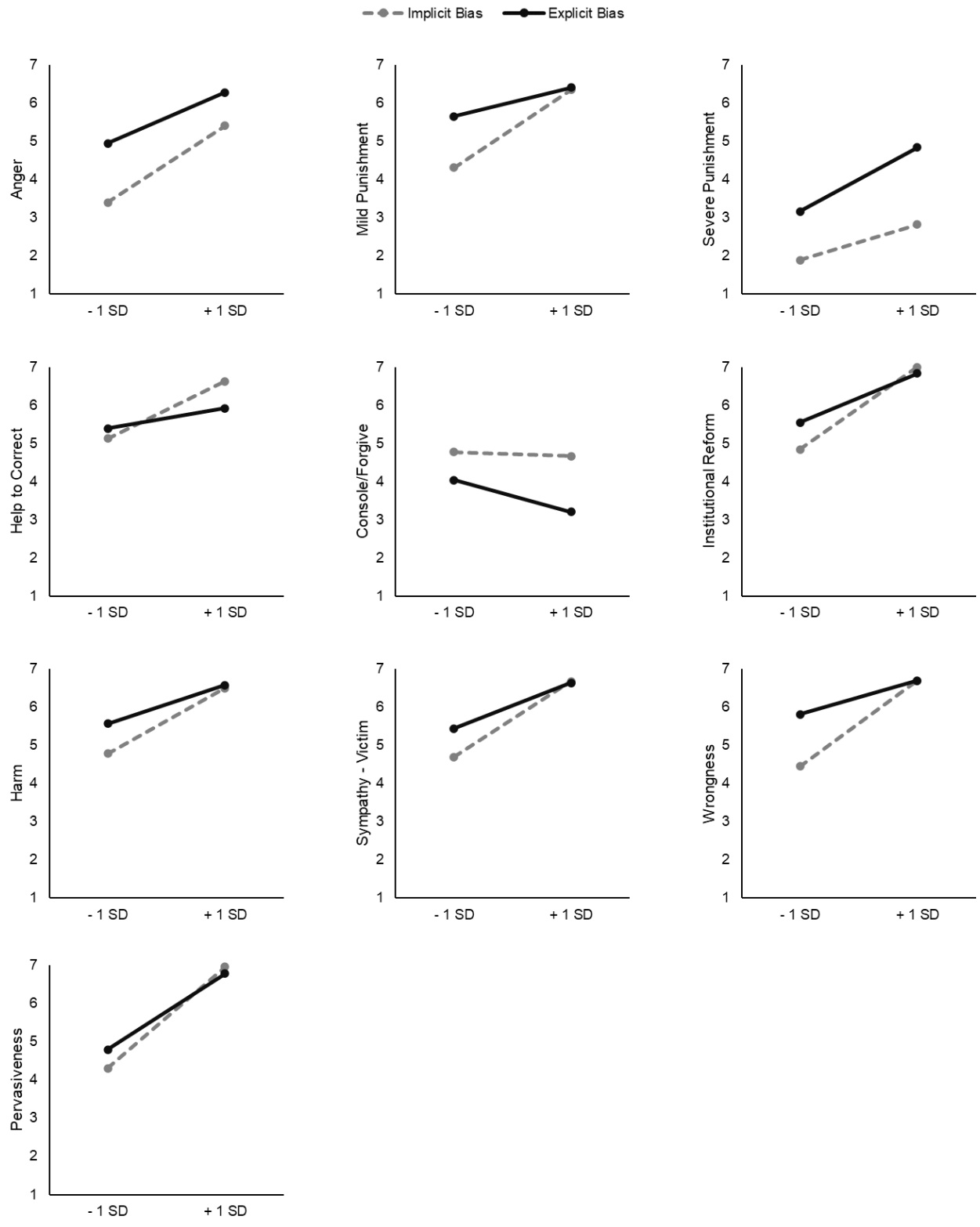


Figure 2.4. PMAPS moderating bias attribution effects. Data are plotted at one standard deviation below and above the sample mean for PMAPS.

Lay Conceptualizations of Racism. The bivariate correlations revealed that higher levels of both systemic and individual conceptualizations of racism were associated with more punitive and less prosocial reactions to discrimination (see Table 2.4). Both systemic and individual conceptualizations of racism variables were entered along with the bias attribution manipulation into the regression models to test the two-way and three-way interactions (see Appendix C – Exploratory Analyses for the results of the interactions between systemic and individual conceptualizations of racism). For the variables that had significant two-way interactions between systemic conceptualizations of racism and the bias attribution manipulation (see Table 2.5) the pattern of interactions showed greater differences in reactions to discrimination attributed to implicit, compared to explicit, bias at lower levels of systemic conceptualizations of racism (see Figure 2.5). For these measures, the slopes for systemic conceptualizations of racism were larger in the implicit, compared to explicit, bias condition.

For the variables that had significant two-way interactions between individual conceptualizations of racism and the bias attribution manipulation (see Table 2.5), there were greater differences in reactions to discrimination attributed to implicit, compared to explicit, bias at higher levels of individual conceptualizations of racism (see Figure 2.5). This pattern was due to larger slopes for individual conceptualizations of racism in the explicit, compared to implicit, bias condition.

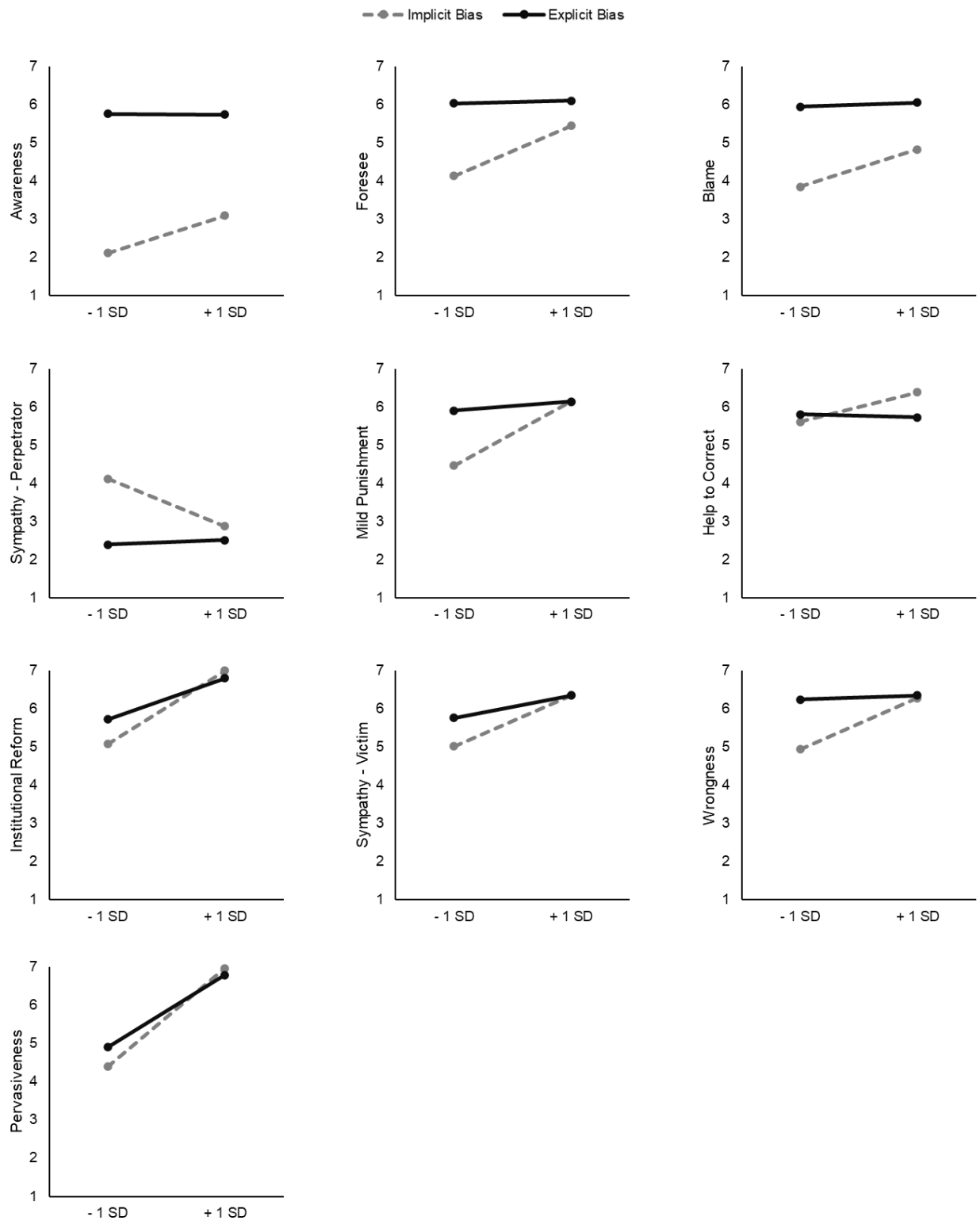


Figure 2.5. Systemic racism conceptualization moderating bias attributions effects. Data are plotted at one standard deviation below and above the sample mean for systemic racism.

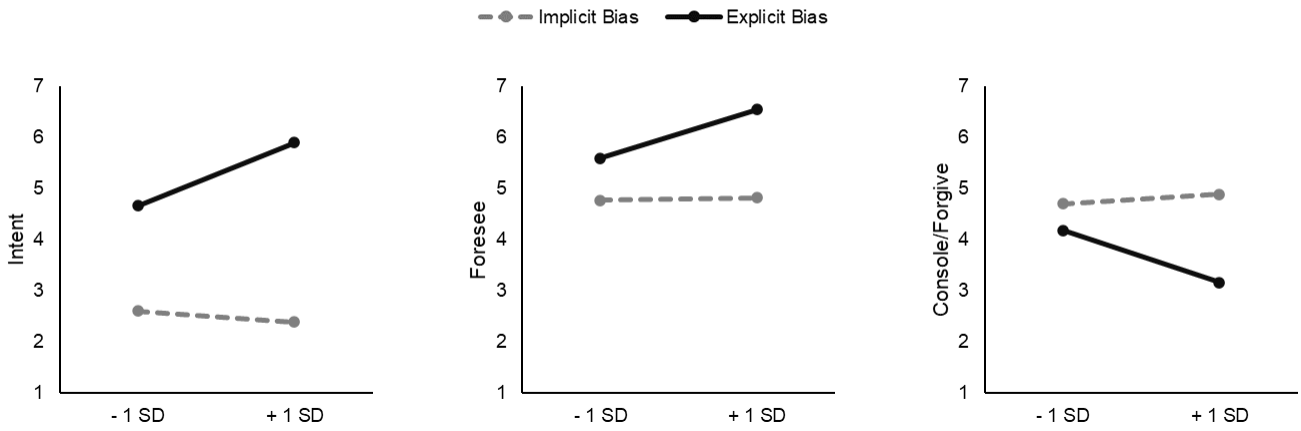


Figure 2.6. Individual racism conceptualization moderating bias attributions effects. Data are plotted at one standard deviation below and above the sample mean for individual racism.

Bias Awareness. Significant bivariate correlations between bias awareness and the dependent variables were notably limited to the measures related to prosocial responses (sympathy for the perpetrator, help to correct, and console/forgive), and bias awareness was not significantly associated with punitive responses (see Table 2.4). Additionally at the bivariate level, higher levels of bias awareness were associated with greater support for institutional reform, and greater perceptions of the pervasiveness and acceptability of discrimination. The significant two-way interactions between bias awareness and the bias attribution manipulation (see Table 2.5) showed greater differences in reactions to discrimination attributed to implicit, compared to explicit, bias at lower levels of bias awareness for all variables except for intentions to help the perpetrator correct her bias where the difference was larger at higher levels of bias awareness (see Figure 2.7). For each of these variables, the slopes for bias awareness had a more positive slope in the implicit, compared to explicit, bias condition.

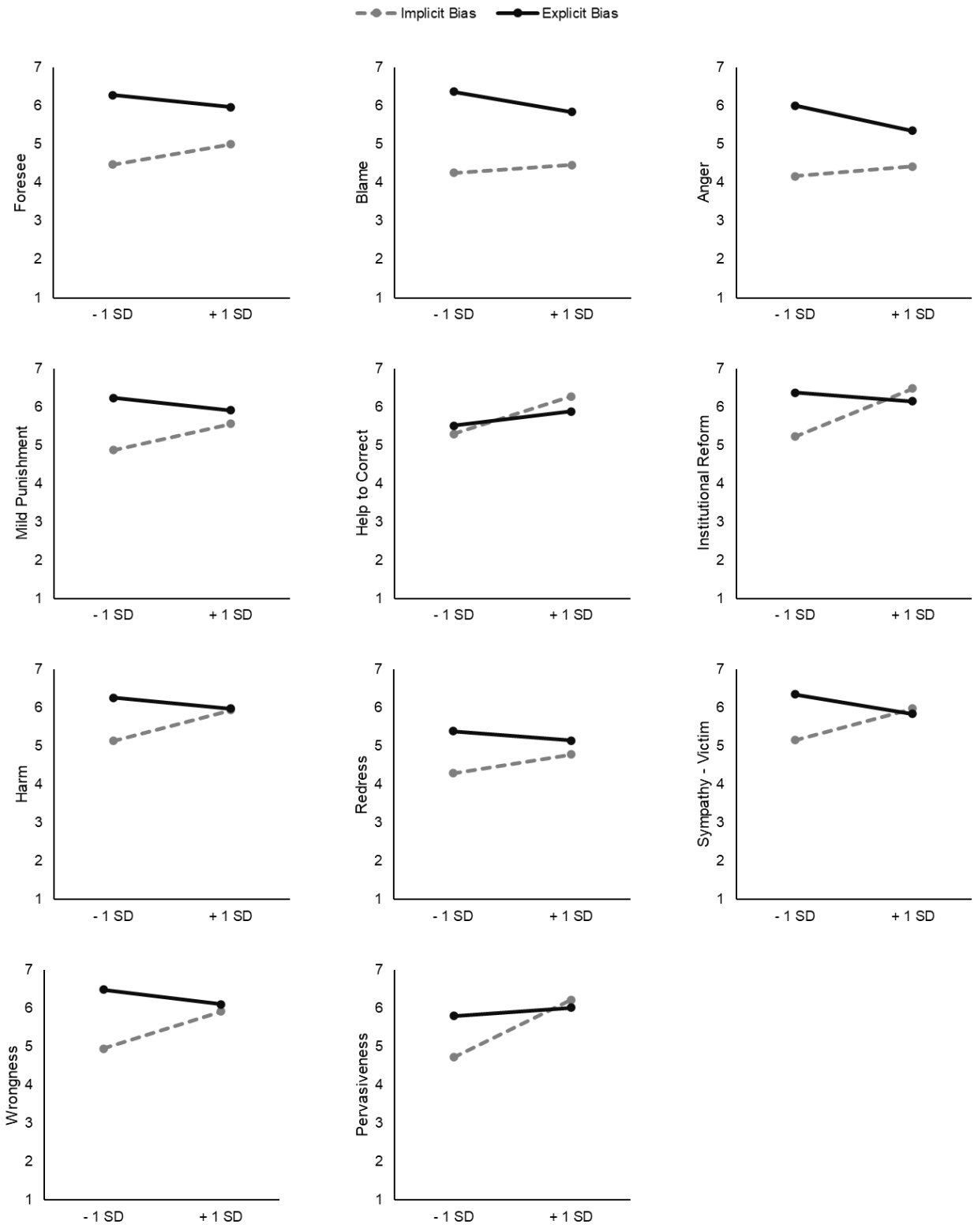


Figure 2.7. Bias awareness moderating bias attribution effects. Data are plotted at one standard deviation below and above the sample mean for bias awareness.

Individual differences Interpretations and General Conclusions. The main effects showed, as expected, that individual differences related to attitudes about bias and discrimination are associated with more punitive and less prosocial responses to perpetrators of racial discrimination. The evidence for the moderating effects of these individual differences also has implications for the hypotheses derived from theories of blame. The general pattern across the interactions showed that at higher, compared to lower, levels of victim perspective-taking, PMAPS, and bias awareness, judgments of the harm caused to the victim and the wrongness of the perpetrator's behavior, as well as sympathy for the victim, and support for mild punishment and institutional reform were very similar for discrimination attributed to implicit and explicit bias (see the figures plotting the significant interactions). However, there were no similar patterns of interactions for judgments of intent or blame. This suggests that although individuals higher in these individual differences perceive discrimination caused by implicit bias to be as harmful and wrong as that caused by explicit bias, judgments of intent and blame are not being affected to the same degree. In other words, individuals higher in these individual differences may judge discrimination caused by implicit bias as more unintentional and less blameworthy, despite being just as harmful, wrong, and deserving of mild punishment, than that caused by explicit bias. These results seem to contradict the harm-dominant hypothesis (which predicts that reactions to harm are what drive blame)[4].

Instead, the results of the moderation tests suggest that individual differences related to general perceptions of prejudice and attitudes about discrimination show greater support for the

[4] One exception to this overall pattern was for systemic conceptualizations of racism which moderated the bias attribution effects for perceptions of awareness and blameworthiness, such that the difference between implicit and explicit bias was minimized (but not eliminated) at higher levels of this individual difference. However, systemic conceptualizations of racism did not also moderate perceptions of harm, as would be expected by the harm-dominant hypothesis if harm was causing perceptions of intent and blame.

intent-dominant hypothesis. Discrimination attributed to implicit, compared to explicit, bias was perceived as less intentional and blameworthy, and the size of this difference was maintained (rather than minimized) at increasingly higher levels of the individual difference variables. This absence of an interaction suggests that although people at higher levels of, for example PMAPS, may generally perceive more intent than people at lower levels, they do not tend to equate implicit bias with explicit bias when it comes to judgments of intent and blame. Additionally, the differences in support for severe punishment, and in intentions to console and forgive, between discrimination attributed to implicit, compared to explicit, bias were greater at higher, compared to lower, levels of PMAPS. This finding suggests that higher levels of PMAPS are associated with seeing a greater difference between implicit and explicit bias when it comes to punitive and prosocial reactions presumably because they perceive implicit bias as less blameworthy than explicit bias, even though they perceive similar levels of harm to the victim across the two forms of bias.

Competing Models of the Blame Process

As a further test of the intent-dominant and harm-dominant hypotheses, two different path models of the psychological process involved in making judgments about discrimination were derived from the competing theories about the blame process that formed the bases of these hypotheses. The Path Model of Blame (Malle et al., 2014) describes a sequence of perceptions and judgments starting with attributions about mental states, leading to degrees of blame, and resulting in motivations to respond punitively or prosocially (see Figure 2.10). In contrast, the Culpable Control Model of Blame (Alicke, 2000), consistent with theories of motivated cognition, argues that the process often begins with perception of harm, leading to anger and

punishment, which create the motivation to blame, followed by post-hoc mental state attributions that are made to justify blame (see Figure 2.11).

To compare these process models, ordinary least squares path analyses were used to compare model fit indices and the strength of the indirect effects. To simplify the number of variables and paths in these models, a few variables were combined into single composites (i.e., observed variables) after using structural equation modeling to assess model fit and determine the appropriateness of combining these variables. Awareness, control, and intent were entered as latent variables along with their measured items to form a second-order latent variable representing mental state attributions. The model had an acceptable fit (SRMR = 0.03, CFI = 0.99, RMSEA = 0.07 [95% CI lower = 0.05, upper = 0.09]). Therefore, a composite variable for mental state attributions was created by averaging together the items measuring awareness, control, and intent (α = .96). Similarly, mild punishment and severe punishment were entered as latent variables along with their measured items to form a second-order latent variable representing punishment. This model also had an acceptable fit (SRMR = 0.03, CFI = 0.99, RMSEA = 0.07 [95% CI lower = 0.04, upper = 0.11]). Therefore, a composite variable for punishment was created by averaging together the items measuring mild and severe punishment (α = .88). Additionally, only the variable measuring intentions to console and forgive the perpetrator were used to represent prosocial responses in these models because of the ambiguity in the interpretation of the variable measuring intentions to help the perpetrator correct her bias discussed above (see the Correlations Between Dependent Variables section).

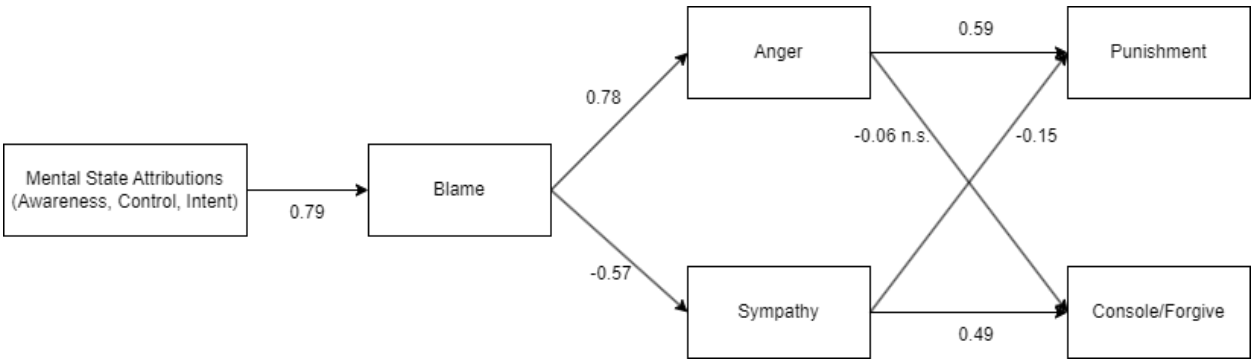


Figure 2.8. The path diagram for the Path Model of Blame.
All paths were significant at $p < .001$ except for the path from anger to console/forgive which was not significant. The harm variable was also entered into the model as a covariate to allow for fit indices comparisons between this model and the Culpable Control model.

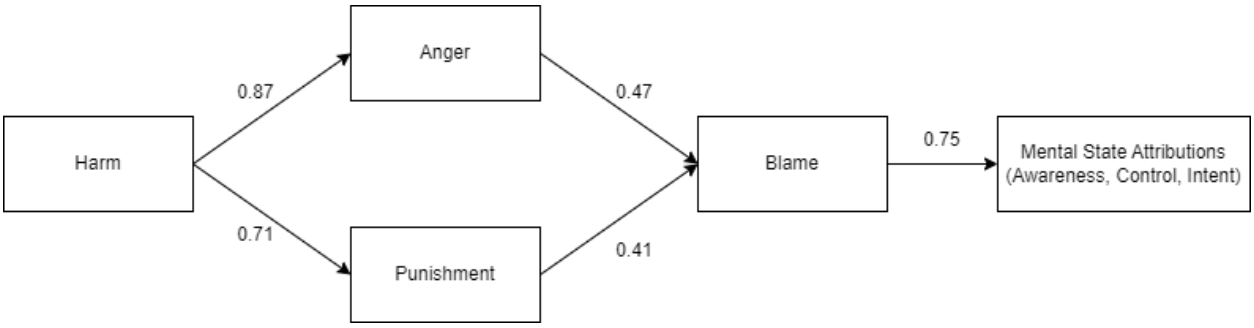


Figure 2.9. The path diagram for the Culpable Control Model of Blame.
All paths were significant at $p < .001$. Perpetrator sympathy and console/forgive variables were also entered as covariates in the model to allow for fit indices comparisons between this model and the Path Model.

Both models had significant paths (Figures 2.10 and 2.11) and indirect effects (Tables 2.6 and 2.7), and although these effects support the potential for the causal process to have happened in the order depicted in these models, the Path Model of Blame path model was the better-fitting model, $\chi^2 = 91.58$, df = 7; Comparative Fit Index (CFI) = 0.92; Standardized Root Mean Squared Residual (SRMR) = 0.08, than the Culpable Control Model of Blame path model, $\chi^2 = 185.27$, df = 5; CFI = 0.81; SRMR = 0.13; χ^2 difference test $p < .001$. Thus, the model fit comparison favors the intent-dominant over the harm-dominant hypothesis. It should be noted though that the cross-

sectional nature of the methods used in the current study do not allow for causal conclusions about the direction of the hypothesized paths in these models.

Table 2.6. Indirect Effects of the Path Model of Blame Path Model

		95% CI *b*	
Path	*b*	Lower	Upper
Mental States → Blame → Anger → Punishment	0.36***	0.29	0.45
Mental States → Blame → Anger → Console/Forgive	-0.04	-0.12	0.03
Mental States → Blame → Sympathy → Punishment	0.07***	0.03	0.11
Mental States → Blame → Sympathy → Console/Forgive	-0.22***	-0.30	-0.15
Blame → Anger → Punishment	0.46***	0.38	0.55
Blame → Anger → Console/Forgive	-0.05	-0.15	0.04
Blame → Sympathy → Punishment	0.08***	0.04	0.13
Blame → Sympathy → Console/Forgive	-0.28***	-0.37	-0.19

Note. CI = confidence interval; indirect effects were calculated using adjusted bias corrected 1,000 bootstrapped samples.
*** $p < .001$

Table 2.7. Indirect Effects of the Culpable Control Model of Blame Path Model

		95% CI *b*	
Path	*b*	Lower	Upper
Harm → Punishment → Blame → Mental States	0.22***	0.13	0.32
Harm → Anger → Blame → Mental States	0.31***	0.20	0.42
Punishment → Blame → Mental States	0.31***	0.19	0.43
Anger → Blame → Mental States	0.35***	0.24	0.47

Note. CI = confidence interval; indirect effects were calculated using adjusted bias corrected 1,000 bootstrapped samples.
*** $p < .001$

Another interesting finding from the path analyses of the Path Model was that the indirect effect of blame on punishment was significantly mediated by both anger and sympathy for the perpetrator, but the indirect effect of blame on intentions to console and forgive the perpetrator was only significantly mediated by sympathy for the perpetrator and not anger (see Table 2.6). This significant role of sympathy in prosocial behaviors is consistent with Weiner's (1995, 1996) theory of responsibility.

Conclusions

Study 1 extends past research on moral evaluations of discrimination attributed to implicit, compared to explicit, bias by examining prosocial responses and more thoroughly examining the mental state attributions and emotional reactions involved in punitive and prosocial responses. Consistent with past research (Cameron et al., 2010; Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016), the results of Study 1 suggest that people make less punitive and more prosocial responses to discrimination when it is attributed to implicit, compared to explicit bias. These results supported the intent-dominant hypothesis and failed to support the harm-dominant hypothesis. The greater intentions to console and forgive the perpetrator when discrimination was attributed to implicit, compared to explicit, bias, as well as the finding that sympathy for the perpetrator mediated the path from blame to intentions to console and forgive the perpetrator, also support Weiner's (1995, 1996) theory of responsibility.

The moderating effects of the individual differences provided additional support for the intent-dominant hypothesis and failed to support the harm-dominant hypothesis. For perceived harm, higher levels of victim perspective-taking, PMAPS, and bias awareness were related to a reduction in the differences between discrimination attributed to implicit, compared to explicit, bias for perceived harm and several related variables (e.g., sympathy for the victim, support for mild punishment)—virtually eliminating the difference at one standard deviation above the mean on these individual differences. However, these same individual differences did not consistently moderate perceptions of intent or blameworthiness as would be expected by the harm-dominant hypothesis which argues that perceptions of harm lead to blame and the supporting attribution of intent. Study 1 was also the first study (as far as the published research shows) to find evidence

of individual differences moderating the effects of attributing discrimination to implicit, compared to explicit, bias.

Study 1 additionally modeled the processes described by two theories of blame: the Path Model of Blame in which the process of blame begins with mental state attributions of awareness, control, and intent (Malle et al., 2014), and the Culpable Control Model of Blame (i.e., motivated blame model) in which the process begins with perceived harm, anger, and desires to punish. The Path Model was a significantly better fitting model than the Culpable Control Model. In combination with the results of the effects of the bias attribution manipulation and the moderating effects of the individual differences, these findings consistently supported the intent-dominant hypothesis over the harm-dominant hypothesis. Study 2 provides further tests of the intent-dominant hypothesis as well as a replication of the path analyses from Study 1.

Chapter 3 - Study 2

Study 2 tested the intent-dominant hypothesis that framing implicit bias as something that people can potentially be aware of and control will result in more punitive and less prosocial responses to discrimination caused by implicit bias. Study 2 additionally examined how this framing affected beliefs and concerns about implicit bias. Conceptualizations of implicit bias were manipulated using a fabricated scientific press release that framed implicit bias either as something people are unaware of and cannot control, or something that people can be made aware of and control with effort (see Appendix A). Study 2 additionally manipulated whether the perpetrator of discrimination acknowledged their implicit bias or not in a 2 (implicit bias framing: no awareness and control/awareness and control) x 2 (perpetrator acknowledgement: yes/no) between-groups factorial design. Participants responded to these scenarios on similar measures of perceptions of the perpetrator and support for punitive and prosocial responses used in Study 1. The same individual differences used in Study 1 were also included to replicate the relationships between the individual differences and reactions to discrimination, as well as test their moderating effects in exaggerating or minimizing the effects implicit bias framing and perpetrator acknowledgement.

Method

Participants

Participants were recruited using CloudResearch. Recruitment was limited to residents of the United States who were 18 years or older. Furthermore, eligible participants needed to have at least 100 approved hits with a 95% approval rating. The top 5% of workers—those who participate in 56% of the studies on CloudResearch—were also excluded from eligibility to improve the naivete of the participants.

A target sample size of 500 participants was determined using power analyses based on 80% power to detect the small to moderate effect sizes, $d = .35$, $f = .15$, found in previous studies (Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016). Participants were compensated $1.00 for their time (approximately 15 minutes to complete the materials). Although this amount is less than minimum wage, participants were informed prior to signing up for the study that "We are a non-profit research group at a public university without grant funding. We are offering what we can as a token of our appreciation for your time and help in our research. We understand that the amount we are offering is not a fair hourly wage, and we ask that you not participate if you are discouraged by the amount we are able to offer you at this time."

Based on expected loss of participants due careless responding, 638 participants were initially recruited, but due to a larger than expected loss of data because of failed attention checks, 225 additional participants were recruited. Participants' data were removed for the following reasons: not fully completing the study ($n = 75$), completed the study more than once ($n = 3$), completing the study too quickly (under 310 seconds, $n = 12$), failing the manipulation attention checks ($n = 302$)[5], reporting that they did not carefully or honestly respond to the items in the study ($n = 3$). The result was a final sample size of 468 participants (ages 19 to 80, $M = 38.74$, $SD = 12.37$; 72% White; 65% Female; 96% United States nationality; 89% had at least some college-level education).

Experimental Manipulations

[5] The primary reason for the large number of attention check failures was due to a total of 226 participants in the unconscious/uncontrollable framing condition failing the manipulation attention check by indicating a different response to the correct option, *People cannot become aware of or control their implicit biases*, in response to the prompt to *Please select the option that best summarizes the conclusions from the article*. Perhaps the attention check was more ambiguous for participants in this condition, or perhaps participants were answering the question based on what they believed about implicit bias and not what the article concluded.

Study 2 was a 2 (implicit bias framing) x 2 (acknowledgement of bias) between-groups design manipulating the framing of implicit bias and whether the perpetrator acknowledges his implicit bias.

Framing of Implicit Bias. Participants were asked to read a mock article titled "Understanding Implicit Bias" that reported that some research has concluded that implicit bias is something that people can be spontaneously aware of and control with some effort, or something that is unconscious and uncontrollable (see Appendix B for the complete wording of the article). In the aware and controllable framing condition, the article concluded with the following paragraph:

However, some research suggests that, although implicit bias works subconsciously, people can become conscious of these biases with a little effort. By drawing our attention to instances when stereotypes pop into mind, or by receiving feedback from others about our behaviors, we can become aware of biases that would normally remain hidden to us. Further research suggests that by gaining awareness of our implicit biases, with some effort we can control how they affect our behavior. In other words, if we put in the work, we can change our implicit biases.

In the unaware/uncontrollable framing condition, the article concluded with the following paragraph:

Some research suggests that implicit bias works subconsciously, and that people are not aware of these biases. In other words, we are unable to simply self-examine our conscious attitudes and behaviors to learn whether we have implicit biases. Further research suggests that we can only know about our implicit biases through feedback from scientific instruments that

measure our implicit associations, but that we cannot control how they affect our behavior. In other words, our implicit biases are not something we can typically be aware of or change.

Acknowledgement of Bias. Participants were asked to read about a case of racial discrimination attributed to implicit bias in which a bank manager denies a Black couple a loan to buy a new home (see Appendix B for a complete wording of the vignette). In the acknowledged bias condition, the vignette concluded with the following passage:

The leader of the training program that Hillsborough completed earlier that year also noted that, at the time, Hillsborough appeared to acknowledge the possibility that he may have implicit biases, and that he seemed committed to trying to prevent those biases from affecting his decisions.

In the unacknowledged bias condition, the vignette concluded with the following passage:

The leader of the training program that Hillsborough completed earlier that year also noted that, at the time, Hillsborough did not appear to acknowledge the possibility that he may have implicit biases, nor did he seem committed to trying to prevent those biases from affecting his decisions.

Manipulation Checks. Following the article and vignette, participants were asked which option best summarizes the conclusions from the article: a) "With effort, people can become aware of and control their implicit biases" or b) "People cannot become aware of or control their implicit biases," and asked to rate their agreement with the statement "With effort, people can become aware of and control their implicit biases." Participants also responded to a rating scale measuring the degree to which participants agree that the victim's experience of discrimination was caused by the perpetrator's implicit bias, and a rating scale measuring participants level of

agreement that the perpetrator discriminated against the victims because of their race (see Appendix B for manipulation check materials).

Measures

Participants were asked to complete the same dependent measures and individual difference measures used in Study 1 which were modified to correspond to the institution and names of the perpetrator and victims in the vignette (see Appendix B). Reliability alphas for the composite variables were all acceptable and similar in size the alphas observed in Study 1 (see Table 3.1 and 3.2 for Cronbach's alphas). One additional dependent variable was included in Study 2 to examine whether attitudes about implicit bias change in response to the different framings of implicit bias. Participants were asked to complete a 24-item implicit bias attitudes scale (Miller & Saucier, unpublished data) that measures concern (e.g., *I am concerned about the effects of implicit biases*), $\alpha = .92$, nihilistic attitudes (e.g., *There is nothing we can do to prevent the negative consequences of implicit biases*), $\alpha = .73$, and perceptions of the normality of implicit bias (e.g., *Almost everyone holds some degree of implicit bias*), $\alpha = .86$.

Procedure

After providing informed consent, participants were randomly assigned to one of the four experimental conditions. Participants were first asked to read the article "Understanding Implicit Bias" containing the implicit bias framing manipulation, and then asked to read the passage describing the case of discrimination attributed to implicit bias. After completing the manipulation checks, participants were asked to respond to the dependent variables, followed by the measures of the potential moderating variables (presented in a randomized order). Following these measures, participants completed relevant demographic information. Lastly, participants were debriefed about the nature of the study, informed that the article explaining implicit bias

was created for the purposes of the study, provided with additional information about implicit bias, and informed that scenario they read about was a fictional story that was created for the purpose of this study. Participants were provided with contact information if they have questions or concerns, thanked for their participation, and given instructions for how to receive payment.

Results and Discussion

Bivariate Correlations

Consistent with theories of blame and moral responsibility, the bivariate correlations between the dependent variables (see Table 3.1) replicated the results from Study 1. Blame and anger were significantly correlated with mental state attributions (awareness, control, intent). In turn, blame was positively correlated with punishment and negatively correlated with intentions to console and forgive the perpetrator. However, blame was not significantly correlated with intentions to help the perpetrator correct his bias (unlike in Study 1 where there was a weak, but significantly positive correlation). This may be additional evidence that participants construed the items measuring this variable as a combination of punitive and prosocial reactions. Again, the help to correct variable appears to be more ambiguous than the other outcome measures and the pattern of correlations with the punishment variables and the console/forgive variable also replicated the general patterns found in Study 1—suggesting that this variable may represent a combination of punitive and prosocial reactions. Overall, the pattern of correlations in Study 2 was nearly identical to those found in Study 1, and further supports the convergent validity of the items used to measure these constructs.

Table 3.1. Bivariate Correlations Between Dependent Variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Awareness	(.94)															
2. Control	.50***	(.84)														
3. Intent	.70***	.53***	(.96)													
4. Foreseen	.41***	.64***	.50***	(.68)												
5. Blame	.49***	.67***	.62***	.64***	(.91)											
6. Anger	.37***	.53***	.48***	.52***	.74***	(.93)										
7. Sympathy - Perpetrator	-.43***	-.45***	-.61***	-.43***	-.60***	-.52***	(.87)									
8. Mild Punishment	.25***	.43***	.27***	.47***	.63***	.64***	-.43***	(.76)								
9. Severe Punishment	.45***	.43***	.59***	.46***	.61***	.60***	-.52***	.45***	(.89)							
10. Help to Correct	-.15**	.08	-.25***	.09	.08	.12**	.08	.41***	-.18***	(.70)						
11. Console/Forgive	-.31***	-.25***	-.44***	-.30***	-.32***	-.30***	.53***	-.16***	-.49***	.47***	(.87)					
12. Moral Character	-.45***	-.43***	-.64***	-.46***	-.58***	-.53***	.55***	-.37***	-.69***	.26***	.57***	(.74)				
13. Institutional Reform	.19***	.37***	.18***	.41***	.45***	.53***	-.37***	.68***	.28***	.46***	-.13**	-.26***	(.94)			
14. Harm	.20***	.42***	.23***	.45***	.53***	.62***	-.36***	.64***	.37***	.33***	-.13**	-.31***	.67***	(.93)		
15. Redress	.28***	.46***	.35***	.47***	.61***	.64***	-.42***	.62***	.63***	.14**	-.28***	-.49***	.56***	.68***	(.78)	
16. Sympathy - Victim	.23***	.44***	.25***	.41***	.54***	.62***	-.29***	.58***	.34***	.26***	-.11*	-.28***	.61***	.73***	.63***	(.86)

Note. Cronbach's alphas are on the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$

Effects of Implicit Bias Framing and Acknowledgement of Bias

To test the effects of implicit bias framing and the perpetrator's acknowledgement of his implicit biases, a MANOVA on the dependent variables was conducted to control for type I error rates. The omnibus MANOVA revealed significant main effects for framing, $F(19, 446) = 4.22$, $p < .001$, and acknowledgement, $F(19, 446) = 6.08$, $p < .001$, but no significant two-way interactions, $F(19, 446) = 1.07$, $p = .377$. Because none of the two-way interactions were significant, the main effects are discussed separately in the sections below.

Implicit Bias Framing. Overall, the effects of implicit bias framing were consistent with the intent-dominant hypothesis (see Table 3.2) and conceptually replicated findings from Study 1 where discrimination was attributed to either implicit or explicit bias. In Study 2, participants responded less punitively when implicit bias was framed as unconscious and uncontrollable, compared to when it was framed as potentially conscious and controllable. The effects of framing on perceptions of awareness and control provide evidence that the framing manipulation was successful. More importantly, perceptions of intent, anger at the perpetrator, attributions of blame, and support for mild and severe punishment were significantly lower in the unaware/uncontrollable, compared to the aware/controllable condition. Additionally, sympathy for the perpetrator was significantly higher when implicit bias was framed as unaware/uncontrollable, compared to aware/controllable. However, no significant effects were found for intentions to console and forgive or intentions to help the perpetrator correct his bias. Perhaps the gender of the perpetrator (a man) and act of discrimination described in the current study (denying a mortgage application) were less forgivable compared to the scenario described in Study 1 (a woman giving improper advice to a student). Alternatively, although sympathy was affected, the framing manipulation may have been too subtle to affect prosocial intentions.

Still, the significant effects observed in Study 2 are consistent with theories of blame that emphasize how perceptions of intent (e.g., Malle et al., 2014) and ability (e.g., Weiner, 1996) influence moral judgments. The subtle suggestion in the aware/controllable condition that implicit bias can be managed with effort increased blame and punitive responses, and decreased sympathy compared to framing implicit bias as entirely unconscious and uncontrollable. It is possible that perceivers thought that if people can become aware of their implicit biases and that these biases can be controlled with effort, then people have an obligation to prevent their biases from resulting in discriminatory judgments and behaviors. Although these perceived obligations were not directly measured in the current study, this interpretation is consistent with theories of blame (Malle et al., 2014) and previous empirical findings (Redford & Ratliff, 2016).

Another similarity to Study 1 where explicit, compared to implicit, bias resulted in greater perceived harm, was that framing implicit bias as potentially conscious and controllable with effort, as opposed to unconscious and uncontrollable, resulted in a small (d = 0.18), but significant (p = .046) increase in perceived harm. Again, perhaps this was because greater intent was perceived in the aware/controllable condition and intentional wrongdoing is perceived to be more harmful (Gray & Wegner, 2008). In contrast to Study 1, implicit bias framing had no significant effects on support for victim redress, sympathy for the victim, or global evaluations of the wrongness and pervasiveness of this type of discrimination. However, framing implicit bias as potentially conscious and controllable with effort did result in lower levels of perceived acceptability compared to the unaware/uncontrollable framing.

Table 3.2. Effects of Implicit Bias Framing on Moral Judgments

Dependent Variable	Aware/Controllable *M* (*SD*)	Unaware/Uncontrollable *M* (*SD*)	*F*	*p*	*d*
Awareness	4.27 (1.75)	3.67 (1.70)	15.05	< .001	0.35
Control	5.83 (1.06)	4.98 (1.38)	54.81	< .001	0.69
Intent	3.98 (1.73)	3.38 (1.49)	16.33	< .001	0.37
Foresee	5.54 (1.29)	5.07 (1.40)	14.57	< .001	0.35
Blame	5.43 (1.42)	4.96 (1.59)	11.32	< .001	0.31
Anger	5.39 (1.37)	4.90 (1.59)	12.63	< .001	0.33
Sympathy - Perpetrator	2.85 (1.46)	3.27 (1.54)	9.52	.002	-0.28
Mild Punishment	5.92 (1.20)	5.55 (1.46)	9.23	.003	0.28
Severe Punishment	3.76 (1.74)	3.24 (1.55)	11.67	< .001	0.32
Help to Correct	5.64 (1.21)	5.46 (1.28)	2.47	.117	0.15
Console/Forgive	3.97 (1.50)	4.07 (1.29)	0.60	.440	-0.07
Moral Character	4.00 (1.34)	4.36 (1.17)	9.43	.002	-0.28
Institutional Reform	6.45 (1.04)	6.19 (1.25)	6.16	.013	0.23
Harm	6.29 (0.98)	6.09 (1.20)	3.99	.046	0.18
Redress	5.54 (1.14)	5.35 (1.30)	2.63	.106	0.15
Sympathy - Victim	6.14 (1.01)	5.94 (1.19)	3.64	.057	0.18
Wrongness	6.15 (1.25)	5.98 (1.36)	2.12	.146	0.13
Pervasiveness	5.92 (1.35)	5.71 (1.65)	2.14	.144	0.14
Acceptability	2.73 (1.70)	3.12 (1.71)	6.08	.014	-0.23

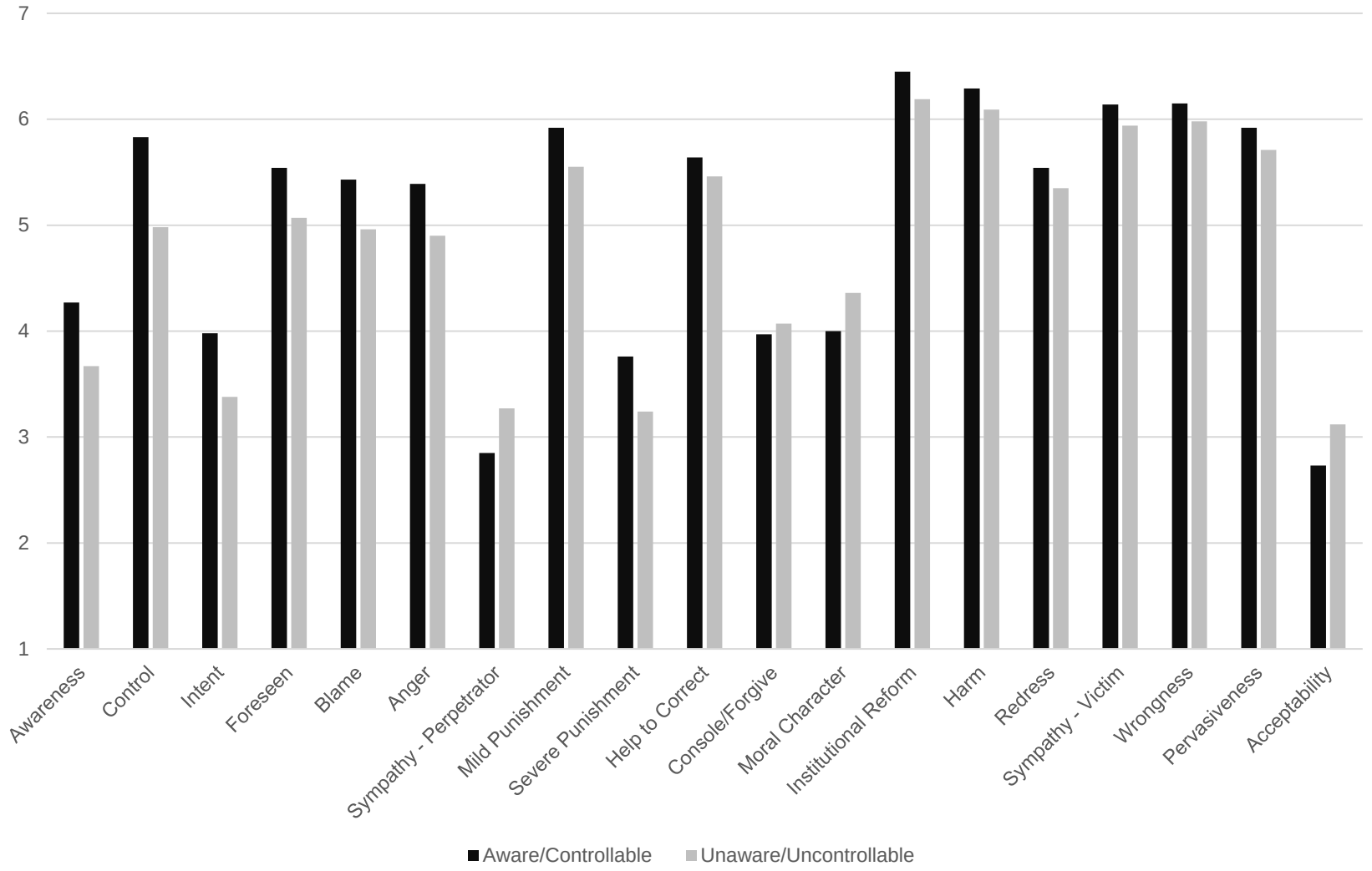


Figure 3.1. Effects of bias framing on moral judgments.

Perpetrator Acknowledgement. The manipulation of perpetrator acknowledgement was designed to test if people respond more punitively and less prosocially when perpetrators acknowledge, compared to when they do not acknowledge, the potential that they have implicit biases. The rationale for this hypothesis was that because acknowledgement may indicate awareness, people may expect those who acknowledge their bias to be vigilant in taking care not to let their biases affect their behavior in negative ways. The alternative hypothesis was that people respond less punitively and more prosocially because they are more empathetic to those who acknowledge their implicit biases and therefore may be assumed to be trying to prevent their biases from causing harm. Neither of these possibilities were strongly supported, however. Only perceptions of awareness, foreseeability, and sympathy for the perpetrator were significantly affected by the manipulation of the perpetrator's prior acknowledgement of his implicit biases. Thus, prior acknowledgment may indicate that a perpetrator is more aware of their biases and may be able to foresee how that may affect their behavior. However, the current study does not provide evidence that prior acknowledgement affects judgments of intent and blame. Furthermore, although participants showed greater sympathy for the perpetrator when he acknowledged his implicit bias, there was no evidence that acknowledgment affects intentions to punish or forgive others for discriminating because of their implicit biases. It is possible that participants did not believe or trust that the acknowledgement was genuine. Future research would be needed to examine this possibility.

Table 3.3. Effects of Perpetrator Acknowledgement on Moral Judgments

Dependent Variable	Acknowledged *M* (*SD*)	Unacknowledged *M* (*SD*)	*F*	*p*	*d*
Awareness	4.41 (1.60)	3.59 (1.79)	29.36	< .001	0.48
Control	5.45 (1.31)	5.39 (1.29)	0.54	.461	0.05
Intent	3.60 (1.62)	3.77 (1.66)	1.05	.305	-0.10
Foresee	5.47 (1.25)	5.16 (1.44)	6.74	.010	0.23
Blame	5.23 (1.46)	5.17 (1.58)	0.30	.581	0.04
Anger	5.18 (1.49)	5.12 (1.51)	0.27	.603	0.04
Sympathy - Perpetrator	3.20 (1.53)	2.91 (1.49)	4.17	.042	0.19
Mild Punishment	5.85 (1.32)	5.64 (1.36)	3.25	.072	0.16
Severe Punishment	3.56 (1.67)	3.47 (1.68)	0.42	.515	0.05
Help to Correct	5.64 (1.14)	5.47 (1.33)	2.26	.134	0.14
Console/Forgive	4.01 (1.35)	4.03 (1.45)	0.03	.868	-0.01
Moral Character	4.18 (1.23)	4.16 (1.32)	0.01	.922	0.02
Institutional Reform	6.32 (1.13)	6.33 (1.17)	< 0.01	.956	-0.01
Harm	6.23 (1.05)	6.16 (1.13)	0.49	.485	0.06
Redress	5.50 (1.19)	5.40 (1.25)	0.74	.391	0.08
Sympathy - Victim	6.06 (1.11)	6.03 (1.11)	0.16	.685	0.03
Wrongness	6.15 (1.22)	6.00 (1.38)	1.66	.199	0.12
Pervasiveness	5.84 (1.44)	5.79 (1.56)	0.16	.692	0.03
Acceptability	2.97 (1.74)	2.88 (1.69)	0.31	.581	0.06

Attitudes About Implicit Bias. Additionally, the effects of implicit bias framing on participants' general attitudes about implicit bias were examined in the current study. Independent-samples *t*-tests revealed that the unaware/uncontrollable framing made participants significantly less concerned about implicit bias ($t(466) = 2.50$, $p = .013$, $d = 0.23$), more nihilistic

about the existence of implicit bias ($t(466) = -3.60$, $p < .001$, $d = -0.33$), and believe that implicit bias is more normal ($t(466) = -.2.40$, $p = .017$, $d = -0.22$), compared to the aware/controllable framing. These findings have potential implications for how psychologists communicate information about implicit bias to the public. If implicit bias is something that people can be spontaneously self-aware of or at least be inferred from behavior, and if implicit bias can to some extent be controlled, as recent evidence suggests (Amodio & Swencionis, 2018; Corneille & Hütter, 2020; Correll et al., 2014; Devine et al., 2002, 2012; Hahn et al., 2014; Hahn & Gawronski, 2019; Nosek et al., 2011; Suhler & Churchland, 2009), then public awareness of these possibilities may make people more concerned about implicit bias, and potentially make them less nihilistic and more confident that efforts to reduce implicit bias are worthwhile.

Individual Differences

To examine how individual differences in perspective-taking, bias awareness, PMAPS, and lay conceptualizations of racism are associated with punitive and prosocial responses to discrimination, and to test whether the effects of attributing discrimination to implicit, compared to explicit, bias are moderated by these individual differences, a series of regression models were performed on the dependent measures. Each regression model contained the experimentally manipulated independent variables, the continuous individual difference predictor variable, and their two-way and three-way interactions. In the sections that follow, the bivariate correlations between individual differences are examined first, followed by the bivariate correlations between the individual differences and the dependent variables (Table 3.5). The following sections describe the main effects (i.e., bivariate correlations) of the individual differences along with the significant interactions that were found, but for the sake of brevity, they do not report the main effects of implicit bias framing or acknowledgement because these were already discussed in the

previous section. The final part of this section interprets the general trends that were consistently found across the individual differences to draw conclusions from these patterns.

Correlations Between Individual Differences. Table 3.4 contains the bivariate correlations between the individual difference variables. The pattern of correlations generally replicated the patterns observed in Study 1. Descriptively, the strength of the correlations observed in the current study were similar to Study 1. Again, the unique, and perhaps surprising, finding across both studies was that bias awareness was positively correlated with perpetrator perspective-taking (see Study 1 for a discussion of a potential explanation for this).

Table 3.4. Bivariate Correlations Between Individual Differences

	1.	2.	3.	4.	5.	6.
1. Perpetrator Perspective-Taking	(.86)					
2. Victim Perspective-Taking	-.12**	(.84)				
3. PMAPS	-.12**	.27***	(.89)			
4. Systemic Racism	-.11*	.19***	.77***	(.94)		
5. Individual Racism	-.15**	.20***	.62***	.69***	(.89)	
6. Bias Awareness	.39***	-.05	.28***	.32***	.13**	(.83)

Note. PMAPS = propensity to make attributions to prejudice scale; Cronbach's alphas are on the diagonal.
* $p < .05$; ** $p < .01$; *** $p < .001$

Table 3.5. Bivariate Correlations Between Individual Differences and Dependent Variables

	PT Perpetrator	PT Victim	PMAPS	Systemic Racism	Individual Racism	Bias Awareness
Awareness	-.31***	.20***	.19***	.20***	.19***	-.15**
Control	-.30***	.13**	.37***	.33***	.35***	-.02
Intent	-.41***	.22***	.27***	.28***	.25***	-.13**
Foresee	-.18***	.15***	.41***	.36***	.35***	.10*
Blame	-.34***	.21***	.48***	.46***	.48***	-.01
Anger	-.28***	.26***	.60***	.52***	.53***	.13**
Sympathy - Perpetrator	.52***	-.13**	-.35***	-.37***	-.34***	.09*
Mild Punishment	-.16***	.09	.52***	.51***	.50***	.15**
Severe Punishment	-.31***	.25***	.42***	.44***	.36***	-.01
Help to Correct	.22***	-.07	.16***	.11*	.15***	.25***
Console/Forgive	.41***	-.11*	-.27***	-.28***	-.18***	.09*
Moral Character	.35***	-.20***	-.39***	-.40***	-.29***	.04
Institutional Reform	-.11*	.17***	.51***	.49***	.47***	.21***
Harm	-.11*	.18***	.54***	.54***	.56***	.19***
Redress	-.16***	.29***	.54***	.63***	.52***	.11*
Sympathy - Victim	-.16***	.28***	.64***	.55***	.57***	.15**
Wrongness	-.25***	.17***	.49***	.51***	.48***	.11*
Pervasiveness	-.10*	.17***	.65***	.74***	.58***	.32***
Acceptability	.28***	< .01	-.28***	-.23***	-.29***	.09*

Note. PT = Perspective-Taking; PMAPS = propensity to make attributions to prejudice scale.
* $p < .05$; ** $p < .01$; *** $p < .001$

Perpetrator and Victim Perspective-Taking. Generally, the main effects of perpetrator perspective-taking (see Table 3.5) showed that higher levels of perpetrator perspective-taking were associated with less punitive and more prosocial responses to discrimination. Perpetrator perspective-taking was also negatively correlated with perceived harm and overall wrongness,

and positively correlated with the acceptability of the behavior. These patterns replicated the correlations observed in Study 1 with the exception that in the current study perpetrator perspective-taking was also significantly positively correlated with intentions to help the perpetrator correct his bias.

The regression coefficients testing the interactions between perpetrator perspective-taking and the manipulated variables are shown in Table 3.6. There was a single three-way interaction between perpetrator perspective-taking, framing, and acknowledgment in predicting anger at the perpetrator. This pattern showed that in the aware/controllable framing condition there were slightly higher levels (descriptively) of anger at lower levels of perpetrator perspective-taking when the perpetrator acknowledged, compared to when he did not acknowledge, his bias, and at higher levels of perpetrator perspective-taking there were slightly lower levels of anger when the perpetrator acknowledged, compared to when he did not acknowledge, his bias. This pattern was reversed in the unaware/uncontrollable framing condition (see Figure 3.2).

Several two-way interactions between perpetrator perspective-taking and framing were also found. The overall pattern of these interactions (see Figure 3.3) revealed that the slopes for perpetrator perspective-taking were stronger in the aware/controllable, compared to the unaware/uncontrollable condition. The result was that, descriptively speaking, framing had less of an effect on perceptions and reactions to discrimination at higher, compared to lower, levels of perpetrator perspective-taking, except for the console/forgive variable where the effect of framing was similar in size but reversed when moving from lower to higher levels of perpetrator perspective-taking. The pattern of interactions here suggests that higher levels of perpetrator perspective-taking minimize the effects of framing implicit bias as more conscious and

controllable. Perhaps this is because individuals who more strongly take the perspective of the perpetrator are more lenient in their judgments of the perpetrator's behavior.

Additionally, one two-way interaction between perpetrator perspective-taking and acknowledgement was found, such that there were slightly greater intentions to console and forgive the perpetrator at higher levels of perpetrator perspective-taking when the perpetrator acknowledged, compared to when he did not acknowledge, his bias and this was reversed at lower levels of perpetrator perspective-taking (see Figure 3.4).

Table 3.6. Regression Coefficients Testing the Moderating Effects of Perpetrator Perspective-Taking

	Framing X Perpetrator PT	Acknowledgement X Perpetrator PT	Framing X Acknowledgement X Perpetrator PT
Awareness	0.08	0.02	-0.25
Control	0.07	-0.06	-0.13
Intent	0.14	0.11	-0.14
Foresee	0.15*	-0.02	-0.17
Blame	0.09	0.05	0.01
Anger	0.19*	-0.01	-0.30*
Sympathy - Perpetrator	-0.05	-0.06	-0.07
Mild Punishment	0.16*	0.05	-0.07
Severe Punishment	0.21*	-0.12	0.002
Help to Correct	-0.08	0.06	0.03
Console/Forgive	-0.33***	0.14*	-0.07
Moral Character	-0.21***	-0.04	-0.01
Institutional Reform	0.07	0.03	-0.14
Harm	0.10	-0.02	-0.12
Redress	0.06	-0.02	-0.13
Sympathy - Victim	0.11	-0.02	-0.21
Wrongness	0.12	-0.01	0.06
Pervasiveness	0.13	0.01	-0.06
Acceptability	-0.05	0.01	0.24

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between perpetrator perspective-taking and the manipulated variables; PT = Perspective-Taking.
* $p < .05$; ** $p < .01$; *** $p < .001$

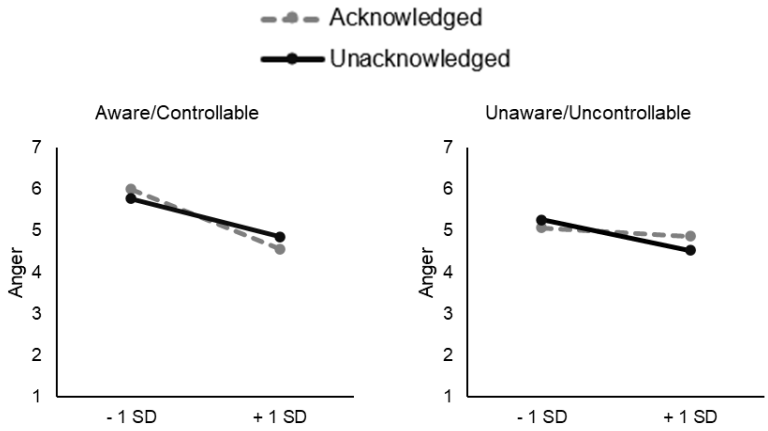


Figure 3.2. Perpetrator perspective-taking interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for perpetrator perspective-taking.

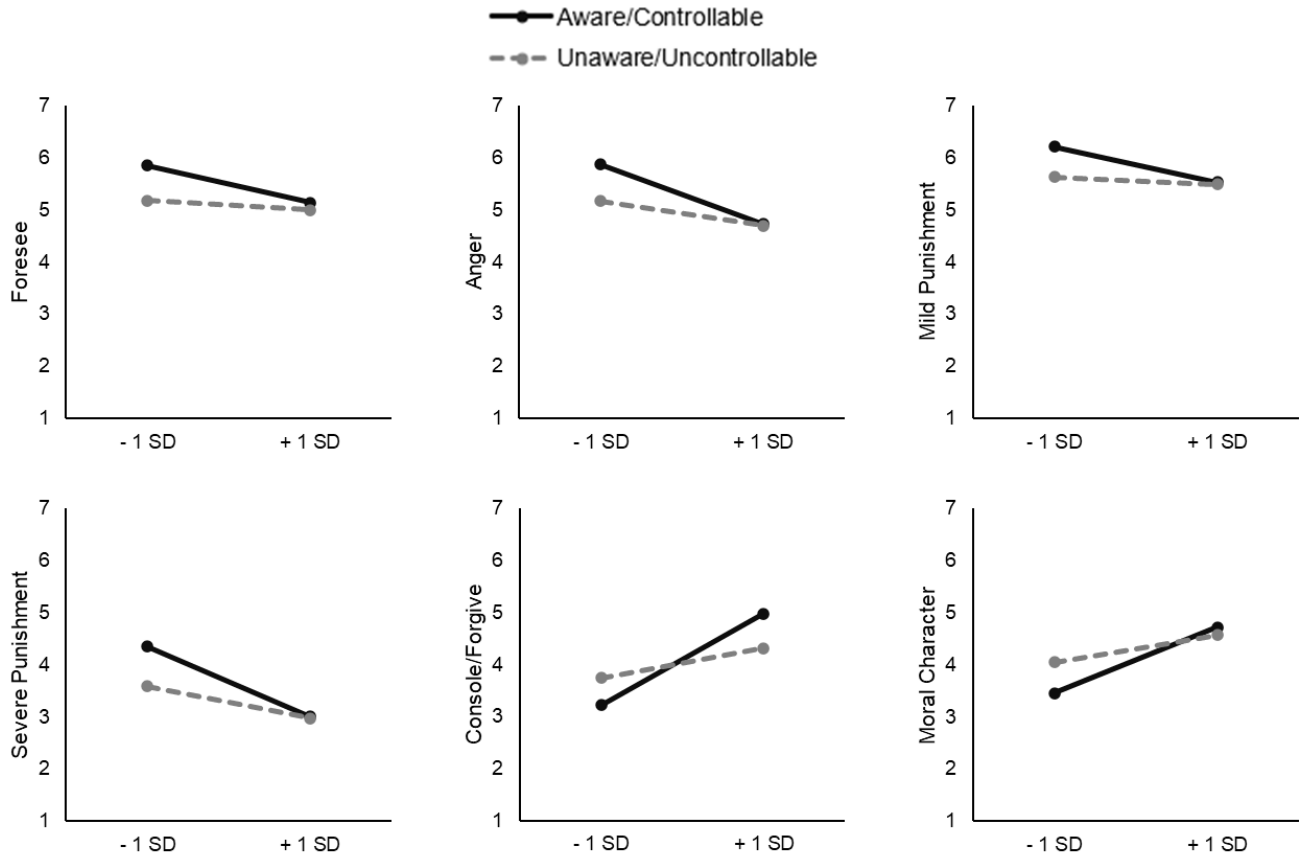


Figure 3.3. Perpetrator perspective-taking interacting with implicit bias framing. Data are plotted at one standard deviation below and above the sample mean for perpetrator perspective-taking.

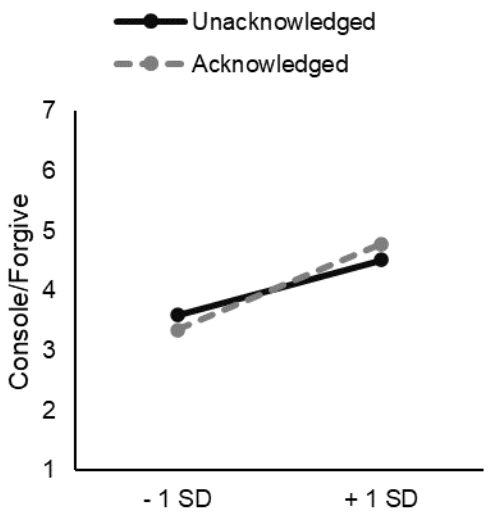


Figure 3.4. Perpetrator perspective-taking interacting with perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for perpetrator perspective-taking.

For victim perspective-taking, the bivariate correlations revealed that higher levels of victim perspective-taking were associated with higher levels of punitive responses and lower levels of prosocial responses to discrimination, generally replicating the findings from Study 1. The only differences were that in the current study there was a significant correlation between victim perspective-taking and perceived control, which was not significant in Study 1, and nonsignificant correlations for mild punishment and help to correct which were significant in Study 1. The only significant interaction was a three-way interaction between victim perspective-taking, framing, and acknowledgement for perceptions of awareness (see Table 3.7). The pattern of this interaction (see Figure 3.5) showed that victim perspective-taking more strongly moderated the effects of acknowledgement when implicit bias was framed as unconscious and uncontrollable than when it was framed as conscious and controllable. In the unaware/uncontrollable condition, acknowledgement had a stronger effect on perceptions of awareness at lower, compared to higher, levels of victim perspective-taking. None of the two-way interactions were significant.

Table 3.7. Regression Coefficients Testing the Moderating Effects of Victim Perspective-Taking

	Framing X Victim PT	Acknowledgement X Victim PT	Framing X Acknowledgement X Victim PT
Awareness	-0.09	0.13	0.50*
Control	-0.02	0.05	0.16
Intent	-0.08	0.06	0.26
Foresee	0.02	0.03	0.09
Blame	-0.09	0.05	0.04
Anger	-0.004	0.007	-0.09
Sympathy - Perpetrator	0.01	-0.06	-0.16
Mild Punishment	-0.04	-0.05	-0.08
Severe Punishment	-0.08	0.05	0.17
Help to Correct	-0.04	-0.08	0.02
Console/Forgive	0.05	-0.07	0.10
Moral Character	0.05	0.002	-0.13
Institutional Reform	-0.001	-0.02	-0.01
Harm	0.06	0.03	-0.06
Redress	0.06	-0.01	0.03
Sympathy - Victim	0.03	0.10	0.03
Wrongness	-0.06	0.02	0.05
Pervasiveness	0.03	0.001	0.20
Acceptability	-0.04	-0.05	0.22

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between victim perspective-taking and the manipulated variables; PT = Perspective-Taking.

* $p < .05$; ** $p < .01$; *** $p < .001$

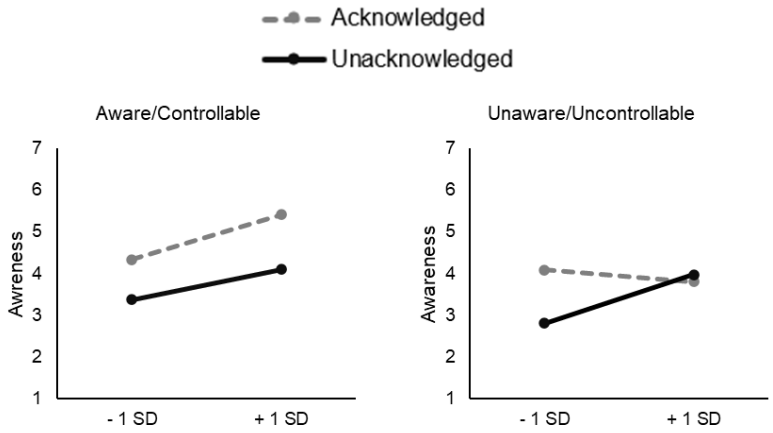


Figure 3.5. Victim perspective-taking interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for victim perspective-taking.

Propensity to Make Attributions to Prejudice. Replicating the results from Study 1, PMAPS was significantly correlated with all of the outcome measures (see Table 3.5), such that levels of PMAPS were associated with more punitive and less prosocial reactions to discrimination. Several significant three-way interactions were found between PMAPS, framing, and acknowledgement (see Table 3.8). Generally, the slopes for PMAPS were larger when the perpetrator acknowledged his bias in the aware/controllable, compared to the unaware/uncontrollable, condition (see Figure 3.6). When the perpetrator did not acknowledge his bias, the slopes for PMAPS were generally larger in the unaware/uncontrollable, compared to the aware/controllable, condition. Many of these interactions were noticeably small. However, the interaction for perceived awareness stands out from the rest. Here, the pattern shows the highest levels of perceived awareness were for individuals higher in PMAPS when implicit bias was framed as more conscious and controllable, and the perpetrator acknowledged his bias.

Additionally, a few two-way interactions between PMAPS and implicit bias framing were found (see Table 3.8). The slope for PMAPS in predicting intentions to console and forgive the perpetrator was more negative in the aware/controllable, compared to the

unaware/uncontrollable, condition (see Figure 3.7). For overall perceptions of the wrongness and pervasiveness of the kind of discrimination described in the scenario, the slopes for PMAPS were descriptively more positive in the unaware/uncontrollable, compared to aware/controllable, condition. But again, the patterns of these interactions were noticeably small. The two-way interactions between PMAPS and acknowledgement were all non-significant.

Table 3.8. Regression Coefficients Testing the Moderating Effects of PMAPS

	Framing X PMAPS	Acknowledgement X PMAPS	Framing X Acknowledgement X PMAPS
Awareness	-0.04	-0.11	0.96***
Control	0.03	-0.05	0.31
Intent	-0.05	0.08	0.60*
Foresee	0.09	0.06	0.44*
Blame	0.04	0.08	0.32
Anger	0.04	-0.01	-0.04
Sympathy - Perpetrator	0.10	-0.03	-0.43
Mild Punishment	0.08	0.06	0.44*
Severe Punishment	-0.12	0.15	0.44
Help to Correct	0.16	-0.10	0.46*
Console/Forgive	0.29*	-0.04	-0.05
Moral Character	0.08	0.05	-0.23
Institutional Reform	0.09	-0.01	0.35*
Harm	0.12	0.11	0.25
Redress	0.15	0.11	0.23
Sympathy - Victim	0.10	-0.02	0.003
Wrongness	0.21*	0.13	0.03
Pervasiveness	0.25*	0.09	0.42*
Acceptability	0.06	0.28	-0.20

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between PMAPS and the manipulated variables.
* $p < .05$; ** $p < .01$; *** $p < .001$

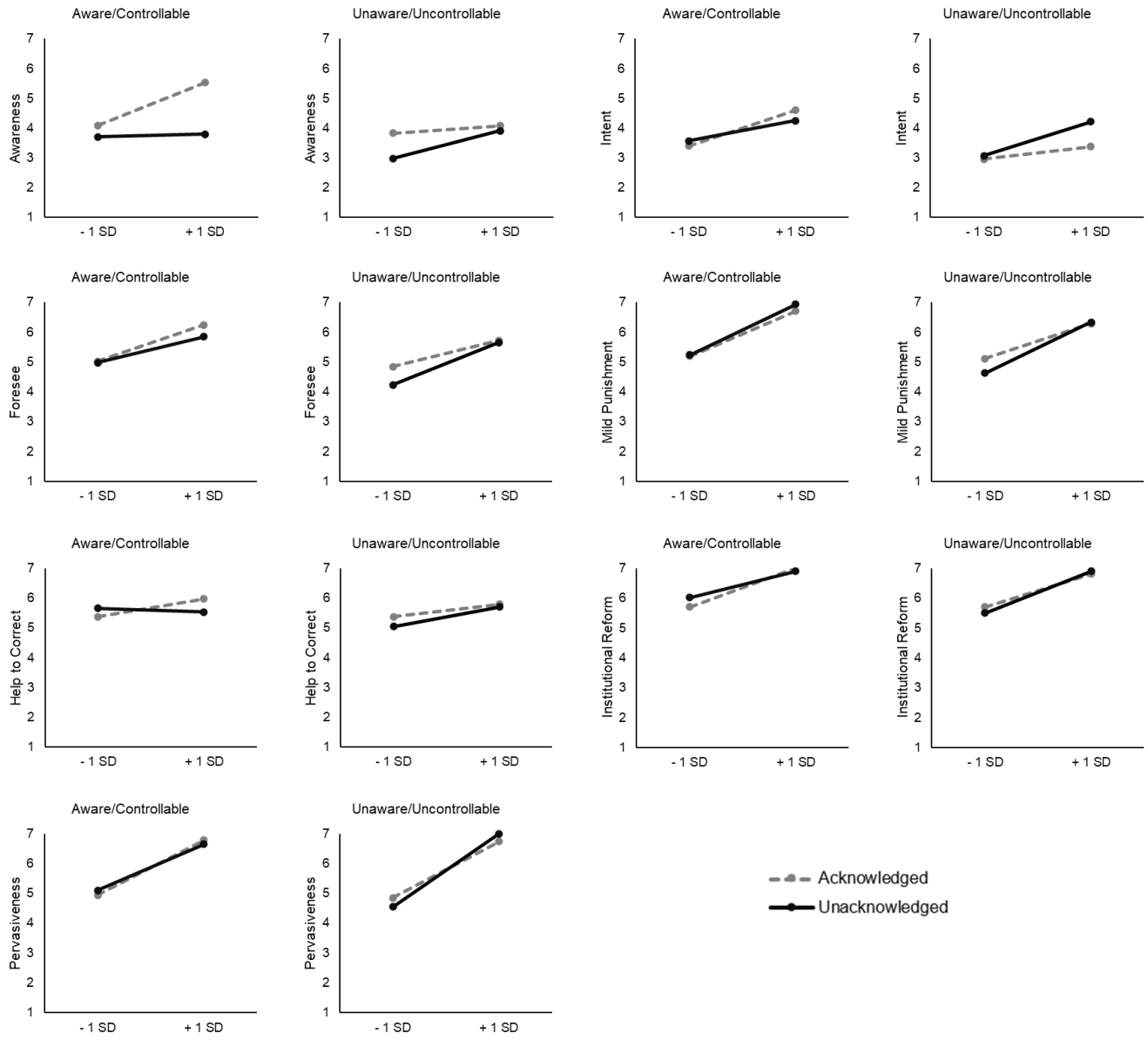


Figure 3.6. PMAPS interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for PMAPS.

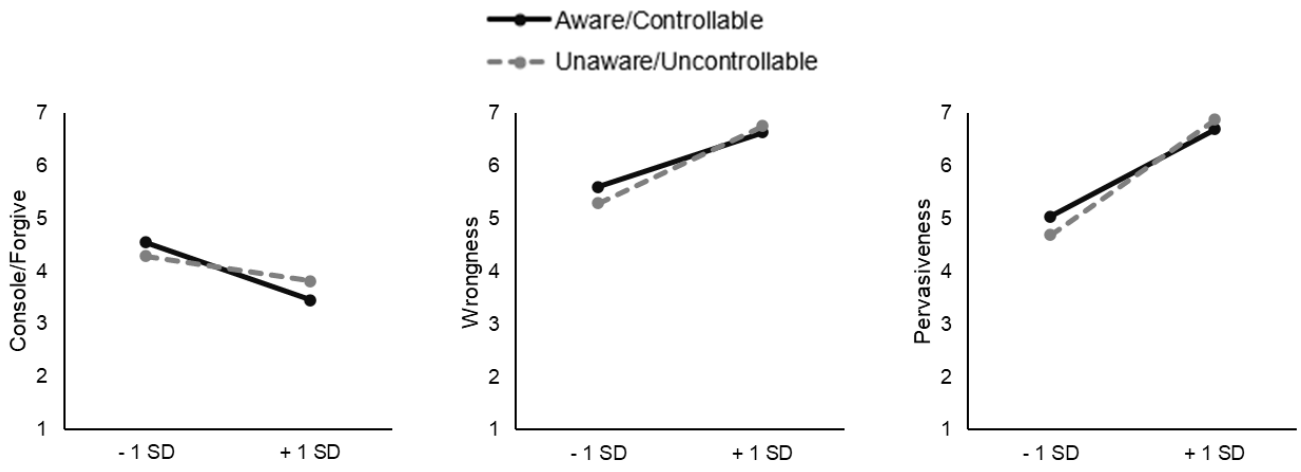


Figure 3.7. PMAPS interacting with implicit bias framing. Data are plotted at one standard deviation below and above the sample mean for PMAPS.

Lay Conceptualizations of Racism. Replicating the pattern of correlations found in Study 1, higher levels of both systemic and individual conceptualizations of racism were associated with more punitive and less prosocial reactions to discrimination (see Table 3.5). Both systemic and individual conceptualizations of racism variables were entered along with the framing and acknowledgement manipulations into the regression models to test the two-way, three-way interactions, and four-way interactions. None of the four-way interactions were statistically significant (see Appendix C – Exploratory Analyses for the results of the interactions between systemic and individual conceptualizations of racism).

Systemic conceptualizations of racism interacted with implicit bias framing and perpetrator acknowledgement in predicting support for institutional reform (see Table 3.9), such that there was a larger slope for systemic conceptualizations of racism in the unaware/uncontrollable condition than the aware/controllable condition when the perpetrator did not acknowledge his bias (see Figure 3.8). There also were a couple of two-way interactions between systemic conceptualizations of racism and framing for perceptions of control and sympathy for the victim, such that there were greater differences at lower, compared to higher,

levels of systemic conceptualizations of racism (see Figure 3.9). Additionally, there was a two-way interaction between systemic conceptualizations of racism and perpetrator acknowledgement for perceptions of harm, such that descriptively more harm was perceived at lower, compared to higher, levels of systemic conceptualizations of racism when the perpetrator acknowledged, compared to when he did not acknowledge, his bias, and this was reversed at higher levels of systemic conceptualizations of racism.

Table 3.9. Regression Coefficients Testing the Moderating Effects of Systemic Conceptualizations of Racism

	Framing X Systemic	Acknowledgement X Systemic	Framing X Acknowledgement X Systemic
Awareness	0.12	-0.16	-0.06
Control	0.18*	-0.09	-0.23
Intent	0.02	0.01	-0.03
Foresee	0.16	0.07	-0.21
Blame	0.15	0.12	-0.21
Anger	0.17	0.02	-0.01
Sympathy - Perpetrator	-0.005	-0.12	-0.13
Mild Punishment	0.09	0.15	0.06
Severe Punishment	-0.16	0.09	-0.21
Help to Correct	0.14	-0.02	0.30
Console/Forgive	0.20	-0.06	0.01
Moral Character	0.11	-0.02	0.10
Institutional Reform	0.05	0.07	0.29*
Harm	0.10	0.13*	-0.04
Redress	-0.02	0.05	-0.13
Sympathy - Victim	0.21**	0.07	-0.16
Wrongness	0.14	0.05	-0.25
Pervasiveness	0.08	0.10	0.12
Acceptability	-0.16	0.08	-0.01

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between systemic conceptualizations of racism and the manipulated variables.
* $p < .05$; ** $p < .01$; *** $p < .001$

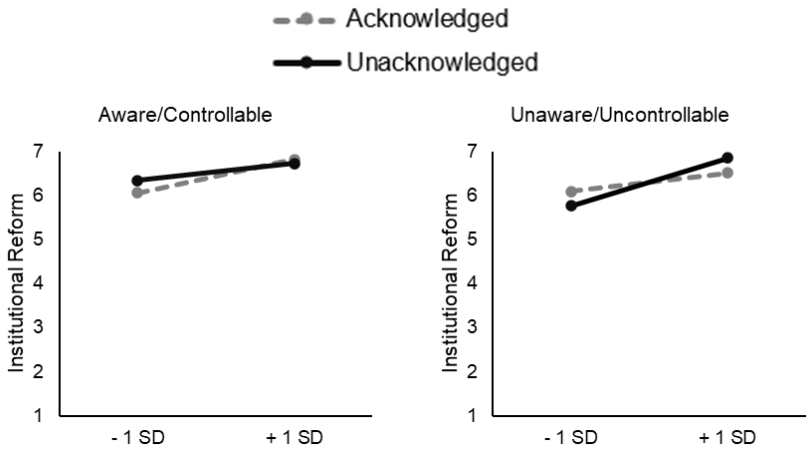


Figure 3.8. Systemic conceptualization of racism interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for systemic conceptualizations of racism.

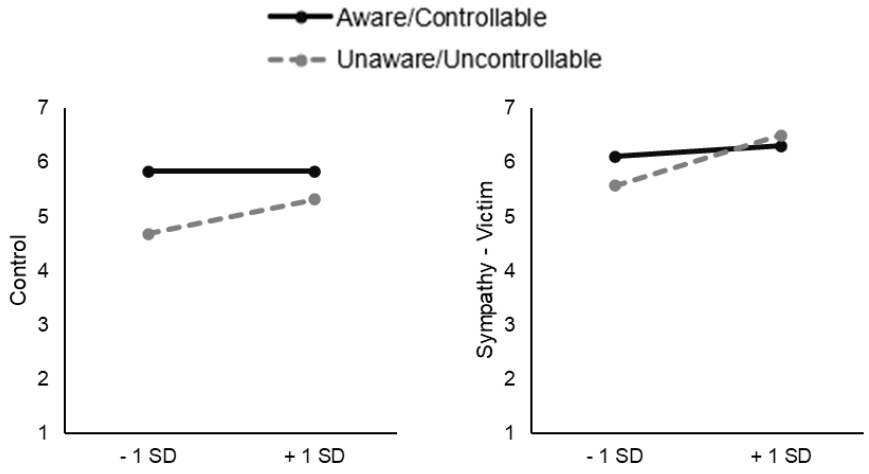


Figure 3.9. Systemic conceptualizations of racism interacting with implicit bias framing. Data are plotted at one standard deviation below and above the sample mean for systemic conceptualizations of racism.

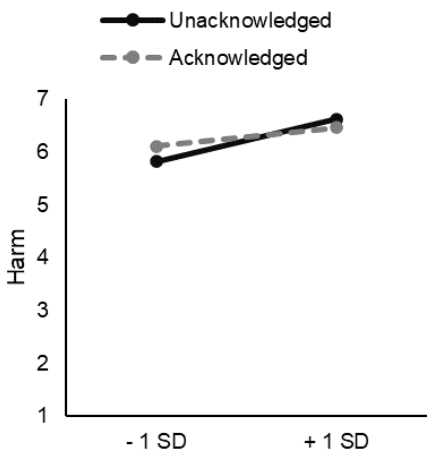


Figure 3.10. Systemic conceptualizations of racism interacting with perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for systemic conceptualizations of racism.

Individual conceptualizations of racism interacted with implicit bias framing and perpetrator acknowledgement in predicting foreseeability (see Table 3.10), such that there was a larger slope for individual conceptualizations of racism in the aware/controllable condition than the unaware/uncontrollable condition when the perpetrator acknowledged his bias (see Figure 3.11). There also was a two-way interaction between individual conceptualizations of racism and framing for support for redress, such that there were greater differences at lower, compared to higher, levels of individual conceptualizations of racism (see Figure 3.12).

Table 3.10. Regression Coefficients Testing the Moderating Effects of Individual Conceptualizations of Racism

	Framing X Individual	Acknowledgement X Individual	Framing X Acknowledgement X Individual
Awareness	-0.03	0.06	0.01
Control	-0.12	0.03	0.43
Intent	0.04	0.01	0.31
Foresee	0.0002	-0.11	0.62*
Blame	-0.02	-0.16	0.41
Anger	-0.03	-0.08	-0.04
Sympathy - Perpetrator	-0.08	0.17	0.11
Mild Punishment	0.05	-0.20	-0.07
Severe Punishment	0.21	-0.09	0.52
Help to Correct	-0.11	-0.05	-0.13
Console/Forgive	-0.19	0.06	0.18
Moral Character	-0.20	0.04	-0.23
Institutional Reform	0.08	-0.08	-0.31
Harm	0.07	-0.15	-0.01
Redress	0.24*	-0.05	0.15
Sympathy - Victim	-0.11	-0.10	0.17
Wrongness	0.08	0.05	0.25
Pervasiveness	0.10	-0.06	-0.17
Acceptability	0.16	0.17	-0.14

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between individual conceptualizations of racism and the manipulated variables.
* $p < .05$; ** $p < .01$; *** $p < .001$

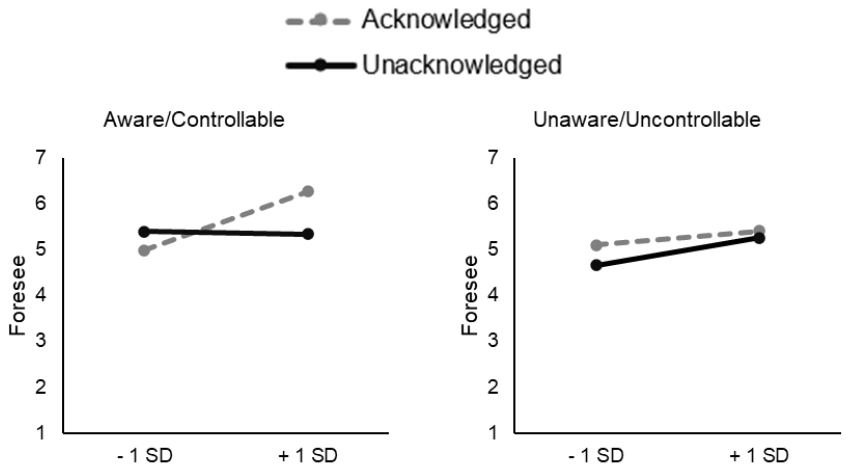


Figure 3.11. Individual conceptualization of racism interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for individual conceptualizations of racism.

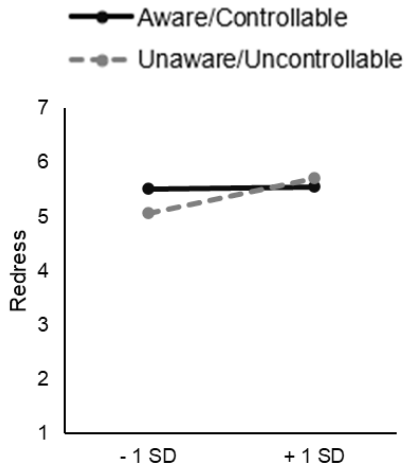


Figure 3.12. Individual conceptualizations of racism interacting with implicit bias framing. Data are plotted at one standard deviation below and above the sample mean for individual conceptualizations of racism.

Bias Awareness. Bias awareness was weakly but significantly correlated with several variables (awareness, intent, foresee, anger, mild punishment, harm, redress, sympathy for the victim, and wrongness) where non-significant correlations were found in Study 1 (see Table 3.5). Overall, bias awareness was associated with more punitive and less prosocial reactions to

discrimination with a few exceptions. Notably, higher levels of bias awareness were weakly but significantly associated with lower levels of perceived awareness and intent. Several significant three-way interactions were found between framing, acknowledgement, and bias awareness (see Table 3.11 and Figure 3.13). For awareness and intent, the slopes for bias awareness were more negative in the awareness/controllable, compared to the unaware/uncontrollable, framing condition when the perpetrator did not acknowledge his bias. For sympathy for the perpetrator, acknowledgement produced its largest effect at high levels of bias awareness in the unaware/uncontrollable condition where sympathy was highest when the perpetrator acknowledged his bias. For support for mild punishment, acknowledgement produced its largest effect at low levels of bias awareness in the unaware/uncontrollable condition. For support for institutional reform and perceived pervasiveness, the crossover interactions were very weak, but appeared to reverse direction between the framing conditions.

Several significant two-way interactions between framing and bias awareness were found (see Table 3.11), such that the effect of framing appeared to be eliminated at higher levels of bias awareness (see Figure 3.14). For these interactions, there appears to be some support for the harm-dominant hypothesis because the pattern associated with higher levels of bias awareness for blame matched the patterns for perceptions of harm, anger at the perpetrator, and support for mild punishment. It is possible that the increased perceptions of harm and increased anger at the perpetrator at higher levels of bias awareness were also increasing attributions of blame and support for mild punishment. However, it is puzzling that higher bias awareness was also associated with lower perceived awareness and intent (see Table 3.5), as the harm-dominant hypothesis predicts that higher levels of perceived harm would be associated with higher levels

of perceived awareness and intent. Therefore, the evidence here for the harm-dominant hypothesis is mixed at best.

Table 3.11. Regression Coefficients Testing the Moderating Effects of Bias Awareness

	Framing X Bias Awareness	Acknowledgement X Bias Awareness	Framing X Acknowledgement X Bias Awareness
Awareness	0.14	-0.15	0.42*
Control	0.10	-0.05	0.03
Intent	0.17	-0.03	0.44*
Foresee	0.13	0.01	0.14
Blame	0.20*	0.03	0.18
Anger	0.25**	-0.01	-0.04
Sympathy - Perpetrator	-0.11	-0.12	-0.39*
Mild Punishment	0.19*	0.13	0.40*
Severe Punishment	0.19	0.03	0.38
Help to Correct	-0.01	0.13	0.24
Console/Forgive	-0.11	0.06	-0.19
Moral Character	-0.17*	-0.18	-0.26
Institutional Reform	0.10	0.08	0.29*
Harm	0.13*	0.08	0.15
Redress	0.20**	0.03	0.17
Sympathy - Victim	0.20**	-0.03	-0.10
Wrongness	0.13	0.12	0.05
Pervasiveness	0.12	0.10	0.37*
Acceptability	-0.15	0.04	0.10

Note. Cells in this table show the regression coefficients testing the two-way and three-way interactions between bias awareness and the manipulated variables.
* $p < .05$; ** $p < .01$; *** $p < .001$

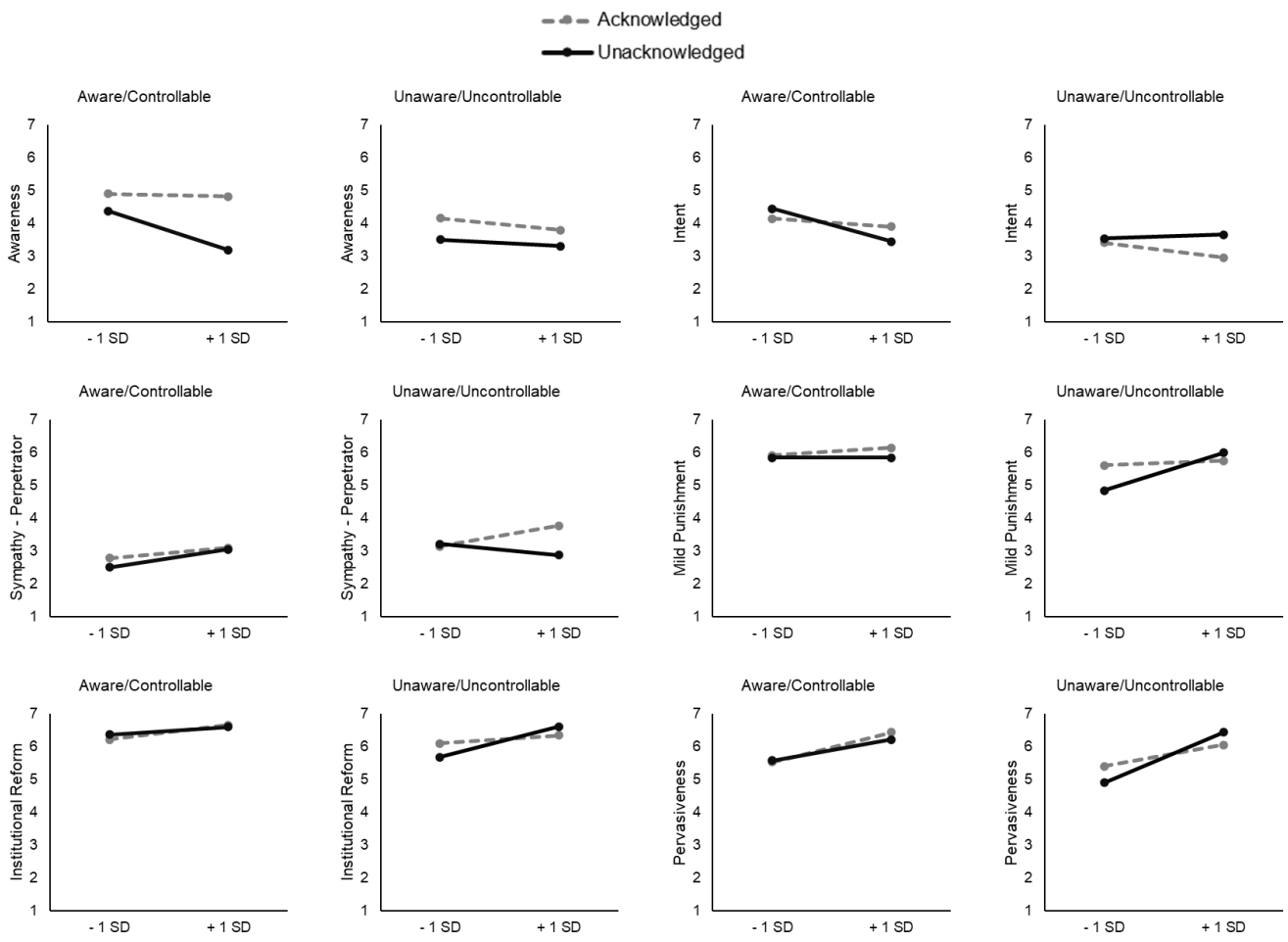


Figure 3.13. Bias awareness interacting with implicit bias framing and perpetrator acknowledgement. Data are plotted at one standard deviation below and above the sample mean for bias awareness.

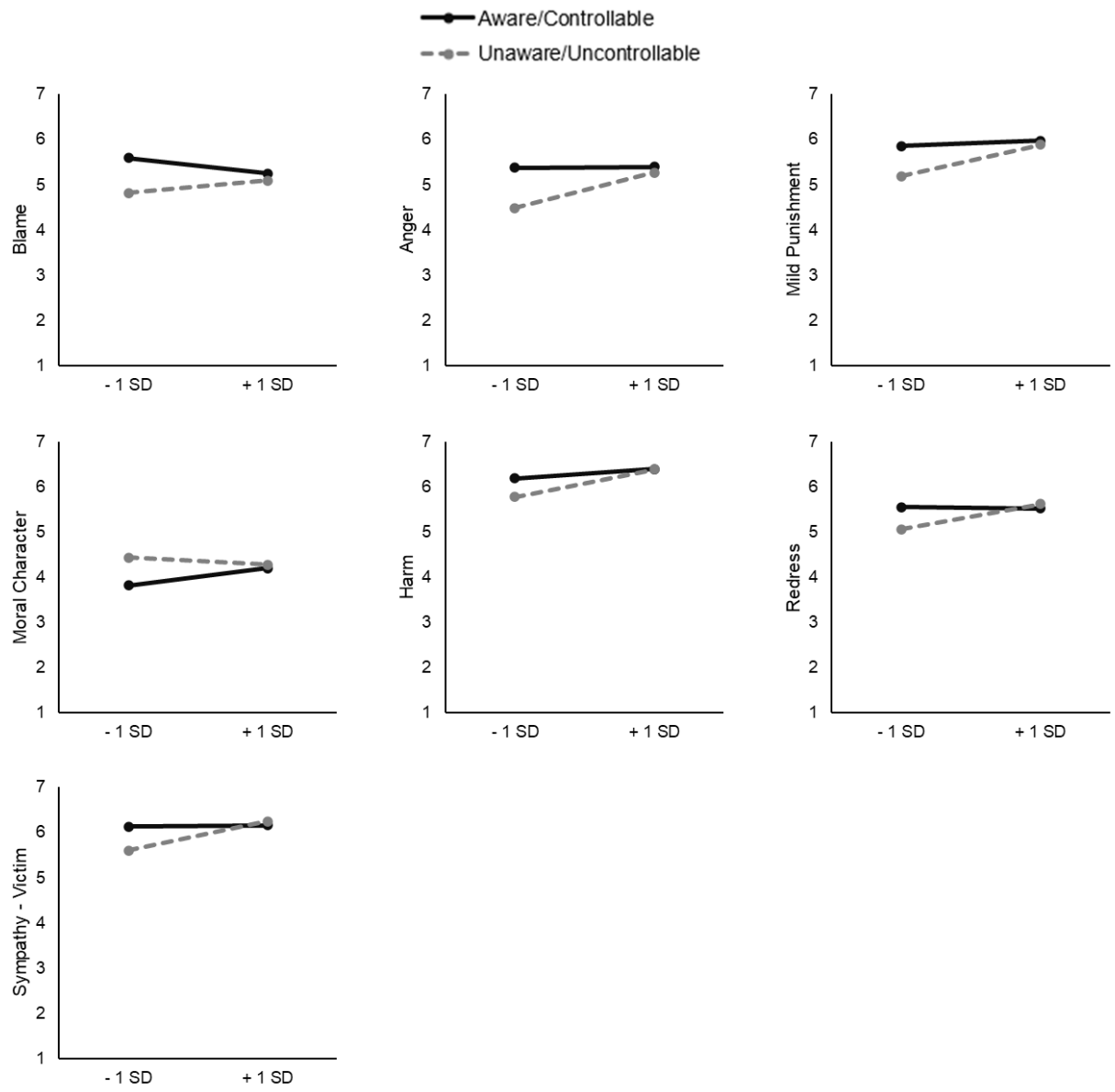


Figure 3.14. Bias awareness interacting with implicit bias framing. Data are plotted at one standard deviation below and above the sample mean for bias awareness.

Individual Differences Interpretations and General Conclusions. Consistent with Study 1, at the bivariate level the individual differences related to attitudes about bias discrimination were associated with more punitive and less prosocial reactions to discrimination. These individual differences also moderated the effects of implicit bias framing and the perpetrator's prior acknowledgement of his bias. However, supporting evidence for either the

intent-dominant or harm-dominant hypotheses was less clear. Implicit bias framing alone produced fewer interactions with the individual differences compared to the bias attribution manipulation (which produced comparatively stronger main effects) in Study 1. Additionally, in contrast to Study 1, there did not appear to be consistent patterns of moderation effects across the individual differences. It should be noted that the larger sample size in Study 2 provided greater power to detect small effects and many of the interactions appeared to be very subtle. Perhaps general conclusions should not be drawn from these findings, especially the three-way interactions which are typically less stable and harder to replicate than lower-order interactions.

Despite a clear pattern of moderation, Study 2 provides additional evidence that individual differences related to attitudes about bias and discrimination play a role in how people respond to discrimination attributed to implicit bias. It is possible that these same individual differences are also related to beliefs about what implicit bias is. Specifically, higher levels of PMAPS, systemic conceptualizations of racism, or bias awareness may be related to believing that implicit bias is more conscious and controllable and therefore more blameworthy. If this were the case, then it could possibly explain why, in the present study, bias awareness moderated the effects of implicit bias framing on blame, anger, and support for mild punishment (see Figure 3.14), such that bias awareness was more strongly related to these variables in the unaware/uncontrollable, compared to the aware/controllable condition. It may be that at higher levels of bias awareness, participants were less likely to accept the unaware/uncontrollable framing because it more strongly contradicted their beliefs about implicit bias. Future research should examine the relationships between these individual differences and beliefs about the psychological nature of implicit bias.

Competing Models of the Blame Process

As a further test of the intent-dominant and harm-dominant hypotheses, two different path models of the psychological process involved in making judgments about discrimination were derived from the competing theories about the blame process that formed the bases of these hypotheses. To replicate the model comparison of the process models derived from the Path Model and the Culpable Control Model in Study 1, ordinary least squares path analyses were used to compare model fit indices and the strength of the indirect effects of these two models. As in Study 1, several variables were combined to simplify the number of variables and paths in the models and to provide more direct comparisons across the two studies. Awareness, control, and intent were entered as latent variables forming a second-order latent variable in a structural equation model to determine if a composite variable representing mental state attributions would be appropriate to use in the subsequent path models. The measurement model had an acceptable fit (SRMR = 0.05, CFI = 0.97, RMSEA = 0.09 [95% CI lower = 0.08, upper = 0.10]). Therefore, a composite variable for mental state attributions was created by averaging together the items measuring awareness, control, and intent (α = .94). Also as in Study 1, mild punishment and severe punishment were entered as latent variables forming a second-order latent variable in a separate structural equation model to determine if a composite variable representing punishment would be appropriate. This measurement model also had an acceptable fit (SRMR = 0.04, CFI = 0.98, RMSEA = 0.08 [95% CI lower = 0.06, upper = 0.11]). Therefore, a composite variable for punishment was created by averaging together the items measuring mild and severe punishment (α = .85). Once again, only the variable measuring intentions to console and forgive the perpetrator were used to represent prosocial responses in these models because this variable, compared to the help to correct variable, appeared to represent prosocial intentions less ambiguously.

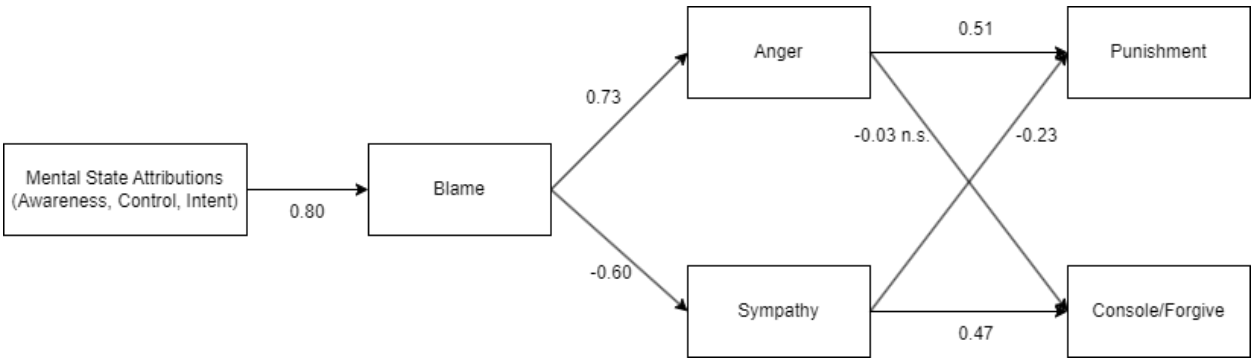


Figure 3.15. The path diagram for the Path Model of Blame.
All paths were significant at $p < .001$ except for the path from anger to console/forgive which was not significant. The harm variable was also entered into the model as a covariate to allow for fit indices comparisons between this model and the Culpable Control model.

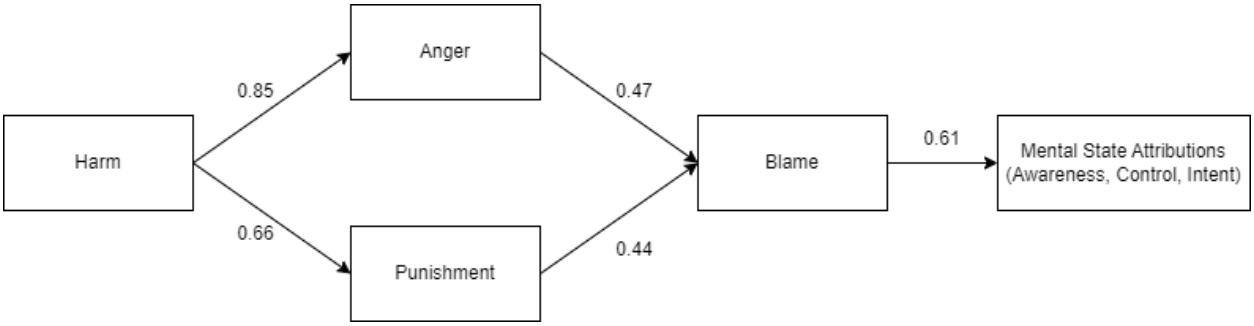


Figure 3.16. The path diagram for the Culpable Control Model of Blame.
All paths were significant at $p < .001$. The sympathy and console/forgive variables were also entered as covariates in the model to allow for fit indices comparisons between this model and the Path Model.

Both models produced similar path coefficients as the models in Study 1, and again, both models had significant paths (Figures 3.18 and 3.19) and indirect effects (Tables 3.12 and 3.13). But although these effects support the potential for the causal process to have happened in the order depicted in these models, the Path Model of Blame path model was once again the better-fitting model, $\chi^2 = 127.51$, df = 7; CFI = 0.92; SRMR = 0.07, compared to the Culpable Control Model of Blame path model, $\chi^2 = 210.27$, df = 5; CFI = 0.85; SRMR = 0.12; χ^2 difference test $p < .001$. These findings replicated the model comparisons from Study 1 and provide additional

support for the intent-dominant hypothesis. Another finding that replicated from Study 1 was that the indirect effect of blame on punishment was significantly mediated by both anger and sympathy for the perpetrator, but the indirect effect of blame on intentions to console and forgive the perpetrator was only significantly mediated by sympathy for the perpetrator and not anger (see Table 3.12).

Table 3.12. Indirect Effects of the Path Model of Blame Path Model

		95% CI *b*	
Path	*b*	Lower	Upper
Mental States → Blame → Anger → Punishment	0.30***	0.24	0.37
Mental States → Blame → Anger → Console/Forgive	-0.02	-0.08	0.04
Mental States → Blame → Sympathy → Punishment	0.11	0.07	0.16
Mental States → Blame → Sympathy → Console/Forgive	-0.23***	-0.29	-0.18
Blame → Anger → Punishment	0.37***	0.31	0.44
Blame → Anger → Console/Forgive	-0.02	-0.10	0.05
Blame → Sympathy → Punishment	0.14***	0.09	0.20
Blame → Sympathy → Console/Forgive	-0.28***	-0.35	-0.22

Note. CI = confidence interval; indirect effects were calculated using adjusted bias corrected 1,000 bootstrapped samples.
*** $p < .001$

Table 3.13. Indirect Effects of the Culpable Control Model of Blame Path Model

		95% CI *b*	
Path	*b*	Lower	Upper
Harm → Punishment → Blame → Mental States	0.18***	0.13	0.23
Harm → Anger → Blame → Mental States	0.25***	0.19	0.32
Punishment → Blame → Mental States	0.27***	0.19	0.35
Anger → Blame → Mental States	0.29***	0.22	0.36

Note. CI = confidence interval; indirect effects were calculated using adjusted bias corrected 1,000 bootstrapped samples.
*** $p < .001$

Conclusions

Study 2 demonstrated how different beliefs about implicit bias affect moral judgments of discrimination attributed to implicit bias. Rather than comparing reactions to discrimination attributed to implicit, compared to explicit, bias, Study 2 tested how moral judgments of discrimination attributed to implicit bias were affected by different conceptualizations of the nature of implicit bias. Specifically, the present study examined the effects of conceptualizing implicit bias as unconscious and uncontrollable versus conscious and controllable. Consistent with the intent-dominant hypothesis, which argues that perceptions of awareness and control are central to judgments of intent and blame (e.g., Malle et al., 2014), Study 2 found more punitive (but not less prosocial) responses to discrimination attributed to implicit bias when implicit bias was framed as more conscious and controllable (compared to entirely unconscious and uncontrollable). Also consistent with the intent-dominant hypothesis, Study 2 replicated the finding from Study 1 that the process model based on the Path Model of Blame was a better fitting model than the model based on the Culpable Control Model of Blame.

Study 2 also extended upon the research on perceptions of discrimination attributed to implicit bias by examining how perpetrators' prior acknowledgment of their implicit bias affects others' moral judgments. Although it may be possible that people give others leniency when they admit the possibility that they have implicit bias (and express desires to prevent their bias from causing harm), the present study found little evidence to support this. Only sympathy for the perpetrator was affected by the acknowledgement manipulation, but this did not result in significant effects for the remaining variables measuring punitive and prosocial responses. Neither did the present study find much evidence that people may be less forgiving when perpetrators acknowledge their implicit bias. Perceptions of the perpetrator's awareness and his

ability to foresee the potential for his bias to result in discrimination were affected by the acknowledgement manipulation, but again, there were no significant effects on punitive or prosocial responses. Perhaps future research could examine the conditions in which prior acknowledgement of personal bias affects others' moral judgments of discrimination, such as by examining factors that affect whether acknowledgement is perceived as more or less genuine.

For the individual differences related to attitudes about bias and discrimination, Study 2 replicated the findings that these are related to punitive and prosocial responses to discrimination. However, unlike Study 2, there were less consistent patterns of moderating effects and many of the interactions were relatively small effects. This could be due to the relatively more subtle manipulations and weaker effects of framing and acknowledgement compared to the relatively stronger effects of attributing discrimination to implicit, compared to explicit, bias observed in Study 1. Or perhaps, as discussed above, these individual differences are related to stronger beliefs that implicit bias is more conscious and controllable, thus making the framing of implicit bias as unconscious and uncontrollable less effective for individuals at higher levels of these individual differences.

Overall, the results of Study 2 supported the intent-dominant hypothesis and showed that by describing implicit bias as more conscious and controllable, compared to unconscious and uncontrollable, people may attribute greater intent and blame, sympathize less with the perpetrator, and be more supportive of punishment when implicit bias results in discrimination. Additionally, framing implicit bias as more conscious and controllable may also affect people's general attitudes about implicit bias by making them more concerned about implicit bias, and potentially making them less nihilistic and more confident that efforts to reduce implicit bias are worthwhile. These findings have important implications for how scientific information about the

nature of implicit bias is explained to the public, which are discussed in further detail in the General Discussion.

Chapter 4 - General Discussion

Studies 1 and 2 make several important contributions to our scientific understanding of the moral judgments people make about discrimination when it is caused by implicit bias. Study 1 replicated and extend similar research on the effects of attributing discrimination to implicit, compared to explicit, bias by presenting participants with scenarios describing a case of racial discrimination and manipulating whether the perpetrator's behavior was attributed to implicit or explicit bias. The existing published research (Cameron et al., 2010; Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016) has only examined punitive responses (e.g., blame, punishment), has not consistently measured key mediators (e.g., perceptions of awareness, control, intent), and has yet to examine prosocial responses. The present studies thus extend the existing scientific literature by examining these factors.

Although the intent-dominant and harm-dominant hypotheses tested here are both plausible, theoretically based hypotheses, the findings from Studies 1 and 2 provide greater evidence for the intent-dominant than for the harm-dominant hypothesis. The intent-dominant hypothesis predicted that attributing an act of discrimination to implicit, compared to explicit, bias would result in significant differences in attributions of intent and blame, as well as differences in punitive and prosocial responses. This hypothesis was based on theories (e.g., the Path Model of Blame; Malle et al., 2014) that argue that blame is determined by perceptions of a perpetrator's mental states (e.g., awareness and intent). Therefore, discrimination that is attributed to implicit bias should be less blameworthy than discrimination that is attributed to explicit bias. In contrast, the harm-dominant hypothesis was based on the Culpable Control theory (Alicke, 2000) that perceptions of harm create the motivation to blame, and this motivation to blame results in perceptions of intent. This hypothesis predicted that, because the

harmful consequences for the victim were the same across the experimental conditions, there would be no differences in blame or punitive and prosocial responses for discrimination attributed to implicit, compared to explicit, bias.

Consistent with the intent-dominant hypothesis, as well as a small body of recent research (Cameron et al., 2010; Daumeyer et al., 2019, 2021; Redford & Ratliff, 2016), Study 1 found that people blame less, and respond less punitively and more prosocially to discrimination when it is caused by implicit, compared to explicit, bias. Also consistent with the intent-dominant hypothesis, Study 2 found that people blame less and respond less punitively (but not necessarily more prosocially) when discrimination is described as unconscious and uncontrollable, compared to something that people can be aware of and control with effort. These findings may reasonably be attributed to the lower levels of perceived awareness, control, and intent found when discrimination was attributed to implicit, compared to explicit, bias (Study 1), and when discrimination attributed to implicit bias was framed as unconscious and uncontrollable, compared to when it was described as potentially conscious and controllable (Study 2).

Additionally, Studies 1 and 2 examined how individual differences in beliefs about prejudice and concerns about bias, unexamined in previous research, are associated with degrees of punitive and prosocial responses, in response to discrimination attributed to implicit bias. As expected, in both studies, perspective-taking of the victim and perpetrator, PMAPS, lay conceptualizations of racism, and bias awareness were associated with degrees of blame and punitive and prosocial responses. Additionally, although previous research has not found consistent evidence of individual differences moderating the effects of attributing discrimination to implicit bias, Study 1 revealed a consistent pattern of interactions between these individual differences and the manipulation of whether discrimination was attributed to implicit or explicit

bias. The general pattern of moderation was that at higher, compared to lower, levels of victim perspective-taking, PMAPS, and bias awareness, discrimination caused by implicit bias was perceived to be just as harmful and wrong as that caused by explicit bias. However, less intent and blame were attributed to implicit, compared to explicit, bias even at higher levels of these individual differences. This consistent pattern of results contradicts the harm-dominant hypothesis which predicted that when there is no difference in perceived harm there should also be no difference in attributions of blame because according to this hypothesis, perceived harm is what determines blame. Although the pattern of moderation in Study 2 was not as clear (see Study 2 Results and Discussion for possible explanations for why this was the case), the pattern of moderation in Study 1 showed a disconnect between perceptions of harm and attributions of intent and blame that was more consistent with the intent-dominant than the harm-dominant hypothesis.

Another important contribution of the present research is that Study 2 was the first study, in the published research to date, to experimentally manipulate lay conceptualizations of implicit bias as more versus less conscious and controllable to examine how different framings of implicit bias affect punitive and prosocial responses to discrimination caused by implicit bias, (as well as beliefs and concerns about addressing implicit bias discussed further in the Practical Implications section below). Study 2 was also the first study to test how perpetrators' prior acknowledgement and commitment to controlling their implicit bias affect these judgments. Significant effects of acknowledgement were limited to perceptions of awareness and foreseeability, and sympathy for the perpetrator, but the acknowledgement manipulation did not affect degrees of blame or related responses. However, it is possible that many participants doubted whether the perpetrator's commitment to controlling his bias was genuine.

Alternatively, it is possible that believing that the perpetrator who acknowledged his bias had more awareness and therefore should have done more to prevent his bias from causing harm while also being more sympathetic to someone who was trying to control their bias cancelled out any effects on blame, punitive, and prosocial responses. Future research should attempt to disentangle these effects as well as examine the effects of genuine versus disingenuous self-acknowledgements of implicit bias.

The manipulation of implicit bias framing was more conclusive. Consistent with the intent-dominant hypothesis and theories that emphasize the important role of perceptions of awareness and control as key factors in attributions of intent and blame (e.g., Malle et al., 2014), Study 2 found that people make more punitive (but not necessarily less prosocial) responses when implicit bias was described as more conscious and controllable. A better understanding of how people conceptualize implicit bias and how this affects their moral judgments of discrimination attributed to implicit bias has important theoretical and practical implications discussed in further detail below.

Another important contribution of the present studies is that they tested competing models of the psychological process of blame. The path models estimated in Study 1, and replicated in Study 2, compared the hypothesized psychological processes described by the Path Model of Blame (Malle et al., 2014) to those described by the Culpable Control Model of Blame (Alicke, 2000). In both studies, the process model based on the Path Model produced significant paths and indirect effects, suggesting that psychological reactions to discrimination start with mental state attributions, leading to degrees of blame, emotional responses, and then decisions about punitive and prosocial responses. The process model based on the Culpable Control Model, in which the process starts with perceptions of harm, leading to anger and the desire to

punish, which result in degrees of blame, followed by mental state attributions to support blame, also produced significant paths and indirect effects. However, in both studies the Path Model was the better fitting model. The Path Model of Blame process model also tested the relative strength of the paths from mental state attributions (awareness, control, and intent) to punitive and prosocial responses through blame, anger, and sympathy. Past research has primarily examined how these attributions affect degrees of blame and punishment. The present studies extended upon this research by also examining the impact of these attributions on feelings of sympathy for the perpetrator and intentions to console and forgive. Consistent with Weiner's (1995, 1996) theory of moral responsibility, the process from mental state attributions to intentions to console and forgive the perpetrator were significantly mediated by sympathy for the perpetrator, but not mediated by anger. These findings have significant theoretical implications and may inspire future research.

Theoretical Implications

The findings from the present studies have the potential to inform theories of moral judgment. Overall, the evidence presented here support theories of blame in which perceptions of intent play a central role (Cushman, 2008; Gray et al., 2012; Guglielmo, 2015; Guglielmo et al., 2009; Heider, 1958; Malle et al., 2014; Shaver, 1985; Weiner, 1995), and fail to support theories of blame that emphasize the role of motivated cognition (Alicke, 2000; Ask & Pina, 2011; Haidt, 2001). The process models based on the Path Model in the present studies contribute to existing evidence that the blame process begins with perceptions of awareness, control, and then intent (Monroe & Malle, 2017, 2019). Additionally, the present studies provide evidence that when discrimination is attributed to implicit bias or when people think of implicit bias as unconscious and uncontrollable, they perceive discrimination as less intentional than when it is caused by

explicit bias or when they believe that implicit bias is potentially conscious and controllable. The result of this is that perceivers then have less reason to blame, and subsequently punish, others who discriminate because of implicit bias. These findings are also consistent with lay intuitions of moral responsibility (Pizarro et al., 2003) as well as legal definitions of moral responsibility that require evidence of intent, involving awareness of the implications of one's behavior, and control over that behavior (Kelly & Roedder, 2008).

Although the present evidence does not support the hypotheses derived from the Culpable Control Model of Blame (Alicke, 2000), we should be cautious not to categorically reject this model. As already noted, the Culpable Control theory does *not* argue that intent is irrelevant for blame. Yet, for the purposes of the present studies, the harm-dominant hypothesis was framed as a stronger prediction in order to construct a more testable hypothesis. As such, one should not conclude that the theory is false based solely on the present findings. There is a vast literature in social psychology that provides convincing evidence of motivated cognition, and studies have found evidence in support of the Culpable Control Model specifically. For example, incidental anger increases attributions of intent (Ask & Pina, 2011); unlikeable targets are blamed more than likeable targets (Alicke & Zell, 2009); and people tend to perceive that a side-effect of an action is more intentional when it results in a negative, compared to a positive, outcome (Cova et al., 2016; Leslie et al., 2006). It may be that there are some instances where the negative emotional reactions to the harm caused by discrimination generate motivations to blame, resulting in attributions of intent to support judgments of blame that are not supported by the objective evidence. Additionally, other factors may affect the motivation to blame. For example, when the perpetrator is an adversary, enemy, or member of a social outgroup (e.g., the "ultimate attribution error") (Pettigrew, 2001), discrimination (even when it is unintended) may present an

opportunity for perceivers to augment their negative evaluations of their rival. Future research should examine the contexts in which motivational states bias the blame attribution process.

Practical Implications

The present studies have further implications for how information about implicit bias is communicated to the public. Some have called for a more nuanced public discussion of what implicit bias is and how it may contribute to racial disparities (Daumeyer et al., 2017, 2019; Payne & Vuletich, 2018). Again, the present studies do not resolve debates about the nature of implicit bias, but rather they provide evidence of how people's beliefs about implicit bias affect their moral judgments. The present findings suggest that when people are led to believe that implicit bias is more conscious and controllable, they are more likely to perceive greater intent, blame more, and respond more punitively when implicit bias leads to racial discrimination. Perhaps these reactions are warranted if people can indeed manage their implicit biases.

Another important finding is that consistent with previous research (Daumeyer et al., 2019), attributing discrimination to implicit, compared to explicit, bias resulted in less support for institutional efforts to minimize discrimination. If attributing acts of discrimination to implicit bias undermines motivations to enact reforms to address it, then research should examine how reduced concern about implicit bias may be countered. Study 2 provides some potential solutions to this problem. When implicit bias was described as more conscious and controllable, compared to unconscious and uncontrollable, participants showed greater support for institutional reform, and were generally more concerned and less nihilistic about the existence of implicit bias. If public understanding of implicit bias were to shift toward a more nuanced understanding of the potential for people to be aware of and manage their implicit biases, then they may be more supportive of making genuine efforts to solve the problems caused

by implicit bias. Presumably confidence in these efforts would additionally be enhanced by informing the public about interventions that have convincing evidence of their effectiveness. Although this possibility is beyond the scope of the present evidence and discussion, future research may wish to explore factors that affect public support for addressing issues of implicit bias.

Understanding how different conceptualizations of implicit bias affect prosocial responses to discrimination caused by implicit bias may also help inform how bias might be confronted more effectively. People often respond defensively to being confronted about their biases (Howell et al., 2013, 2015, 2017). Interventions that more effectively reduce bias avoid creating hostility and decrease perceived moral blameworthiness (Monteith et al., 2019; Vitriol & Moskowitz, 2021). Understanding the factors that lead others to respond less punitively and more prosocially to those who discriminate due to implicit biases may help to inform interventions that reduce defensive reactions leading to greater acknowledgement of bias and commitment to efforts to prevent bias from causing harmful outcomes.

Limitations

As with any single set of studies, the present studies are not without limitations. The vignettes in the present studies provide only two examples of discrimination in specific contexts. Although these contexts were chosen because of their plausibility and potential to reflect real-world instances of discrimination, the results of these studies should not be overgeneralized to contexts where discrimination caused by implicit bias might be perceived as more blameworthy. The contexts in the present studies (higher education and financing) do, however, extend upon the contexts examined in prior studies (e.g., managerial decisions, healthcare, policing). So far, the contexts examined in prior research, and in Study 1, have all described situations in which

the perpetrators had some degree of responsibility for the welfare of the targets (with the exception of Study 2 in which, arguably, the loan manager has more responsibility to the bank than the clients). Future research may be helpful in discovering how degrees of blame, punishment, and prosocial responses vary as a function of the features of the contexts in which discrimination occurs and the responsibility perpetrators have to ensure the welfare of those they may discriminate against.

Another limitation related to the vignettes is that they were hypothetical or imaginary. Although the vignettes allowed for experimental control over the situation (which partially explains why vignettes are often used in this type of research), experimental control comes at the expense of realism and external validity. Further research should examine moral judgments of discrimination attributed to implicit bias in laboratory or real-life settings.

The sample of Amazon Mechanical Turk (MTurk) workers in the current studies may also limit the generalizability of the results. MTurk samples are typically more diverse in terms of education and age than are college student samples. However, MTurk samples tend to be younger and more educated than nationally representative samples (Keith et al., 2017), and these participants may have different beliefs about implicit bias, and thus respond differently, than may older or less educated people.

Another caveat is that all the variables in the path models were measured variables. Therefore, although the path analyses provided evidence of significant paths from perceptions of metal states to degrees of punishment and prosocial responses that are consistent with theory, these analyses alone do not provide causal evidence that the processes occurred in the order specified in the models. Future research should directly manipulate or hold constant the hypothesized mediators in this process to examine their causal roles. Furthermore, although the

fit comparisons showed that the model based on the Path Model of Blame was the better fitting model than the model based on the Culpable Control Model of Blame, there are some important caveats to consider when drawing conclusions from this finding. As already noted, though the path models tested in the present studies were based on these two theories, they were created solely by the author. The precise models were not directly specified by the authors of these models. Therefore, the present findings should not imply an unqualified rejection of one theoretical model in favor of the other. The present findings do not imply that there are no conditions in which the Culpable Control Model would be the more accurate model. The future directions discussed below suggest some situations in which this may be the case.

Future Directions

The general conclusion from Study 2, that when people are led to believe that implicit bias is more or less conscious and controllable there are significant consequences for their moral judgments, is a unique and important finding of the present research. There are several interesting directions this line of research could take in the future. Social psychologists have conceptualized implicit bias in many different ways (Corneille & Hütter, 2020; de Houwer, 2019; Fazio & Olson, 2003; Gawronski, 2019; Gawronski et al., 2020; Greenwald & Banaji, 2017; Greenwald & Lai, 2020; Newell & Shanks, 2014) that may not be widely understood by the general public. Implicit bias is often described as unconscious or automatic in the news, but it may be time for a more nuanced understanding of implicit bias. For example, people can spontaneously become aware of their implicit biases (Hahn et al., 2014; Hahn & Gawronski, 2019), and therefore it may be misleading to claim that implicit bias is entirely unconscious. Additionally, there is evidence that implicit bias is malleable and controllable (Amodio & Swencionis, 2018; Correll et al., 2014; Devine et al., 2002, 2012; Nosek et al., 2011; Suhler &

Churchland, 2009), and although there may be some degree of automaticity involved, when lay perceivers think of implicit bias as automatic, they may believe that it is beyond a person's control. Future research should continue to examine how scientific communications to the public affect people's moral judgments about implicit bias.

Also related to the nature of implicit bias, and how lay perceivers understand it, is information about the prevalence of implicit bias. Public communications about implicit bias often includes scientific evidence, or at least the implication, that implicit bias is prevalent. Additionally, some public communications may even give readers the impression that implicit bias is unavoidable. The perception that implicit bias is prevalent may be accurate, but it may also have consequences for people's concerns about implicit bias. In one set of studies, researchers found that claiming that implicit bias is highly prevalent (as opposed to less common) creates a descriptive norm (Duguid & Thomas-Hunt, 2015) that may lead to people to believe that implicit bias is less of a problem. Future research should manipulate the framing of the prevalence of implicit bias to examine how this influences moral judgments about discrimination caused by implicit bias.

Public communications about implicit bias also often suggest that implicit bias is a consequence of normal associative processes. Currently, there are no published studies examining how an understanding of how implicit biases are formed affects people's moral perceptions of implicit bias. An understanding that implicit bias is formed as a consequence of passive conditioned associations through exposure to cultural representations of social groups may result in people perceiving implicit bias as less of an individual moral failing and more of a societal problem. The result may be that people judge individuals less harshly for their implicit biases, but also bring about a greater willingness to support efforts to change cultural and

systemic factors that contribute to implicit bias. Future research should examine these possibilities.

Additionally, framing implicit bias as more difficult or as easier to overcome may have implications for the moral judgments people make in response to discrimination attributed to implicit bias. The awareness and control framing condition in Study 2 described implicit bias as something that is potentially accessible to awareness, and controllable with effort. If implicit bias were instead described as easier to control, blame may have further increased because perceivers may have believed that the perpetrator had an even greater moral obligation to prevent his bias from causing harm. Previous research suggests that describing interventions to address important social problems as easy, compared to difficult, increases blame and reduces empathy for the targets of those interventions (Ikizer & Blanton, 2016). Relatedly, when failures to address or prevent problems are perceived to be due to a lack of effort, they are perceived as more blameworthy than when failures are perceived to be due to a lack of ability (Weiner, 1995, 1996). Therefore, when people are perceived to be actually making an effort to address a difficult problem, as opposed to making little effort (or merely saying they are making an effort) to address an easier problem, sympathy for that effort should be higher, and people may be more forgiving if and when failures to control implicit bias occur.

What people believe about implicit bias in terms of awareness, control, and prevalence might also have interesting implications for who and what people blame for systemic disparities. If implicit bias is believed to contribute to issues like systemic racism or sexism, and implicit bias is believed to be a normal or unavoidable feature of social life, people may become more nihilistic in their beliefs about what can be done to address it. Not only is individual blame reduced when discrimination is attributed to implicit, compared to explicit, bias, but people also

express less support for institutional reform efforts (Daumeyer et al., 2019)—a finding which was replicated in the present studies. Future research should further examine how beliefs about implicit bias, and its role in systemic disparities, are associated with concerns about, and commitment to addressing, injustice.

As noted in the introduction, there is some evidence that Black perceivers place more importance on harm than White perceivers when making moral judgments about discrimination, although both appear to place an equal importance on intent (Simon et al., 2019In the present studies, participants were predominantly White—which prevented the ability to examine race as a moderating factor. It is possible that the mostly White sample explains the pattern of findings in the present studies that intent appeared to be a more important variable than harm. Future research on judgments of implicit bias should attempt to obtain a more racially balanced sample to test how racial identification might play a moderating role in reactions to discrimination attributed to implicit bias.

Future research could also examine how beliefs about bias are related to what has recently been called "cancel culture" (Bouvier & Machin, 2021; Clark, 2020; Norris, 2021; Romano, 2020). One of the criticisms of cancel culture is that the outrage people experience, especially in online contexts, may put people in a state where they are less sensitive to, or less able and willing to consider, others' intentions when bias is expressed (Crockett, 2017). Similarly, motivations to engage in virtue signaling, or moral self-licensing (Effron & Conway, 2015), by expressing moral outrage may also shape how people consider elements of blame, such as awareness, intent, and control. Consistent with motivated blame theories (Alicke, 2000), the motivational mindset of the perceiver—whether they are focused on reducing harm to potential victims, expressing their virtue, punishing perpetrators, or protecting the accused—may have

important consequences for how expressions of implicit bias are perceived. Future research should examine how different motivational states influence perceptions of the blameworthiness of implicit bias.

Conclusion

The present studies make important contributions to our understanding of how beliefs about implicit bias affect people's moral evaluations of discrimination attributed to implicit bias. These findings have implications for theories of moral responsibility and blame, as well as practical implications for how information about implicit bias is communicated to the public might affect people's concerns about reducing discrimination. This relatively new line of research has the potential for promising future directions that can contribute to our understanding of the social consequences for how implicit bias is perceived. The degree to which people perceive implicit bias as morally blameworthy or something that should be forgiven, may have important consequences for how people respond to persistent group-based disparities that are a result of discrimination attributed to implicit bias.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*(3), 368–378. https://doi.org/10.1037/0022-3514.63.3.368

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574. https://doi.org/10.1037/0033-2909.126.4.556

Alicke, M. D., & Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, *39*(9), 2089–2105. https://doi.org/10.1111/j.1559-1816.2009.00517.x

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*(9), 1755–1762. https://doi.org/10.1177/0956797613480507

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661. https://doi.org/10.1037/0022-3514.91.4.652

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, *18*(6), 524–530. https://doi.org/10.1111/j.1467-9280.2007.01933.x

Amodio, D. M., & Swencionis, J. K. (2018). Proactive control of implicit bias: A theoretical model and implications for behavior change. *Journal of Personality and Social Psychology*, *115*(2), 255–275. https://doi.org/10.1037/pspi0000128

Ask, K., & Pina, A. (2011). On being angry and punitive. *Social Psychological and Personality Science*, *2*(5), 494–499. https://doi.org/10.1177/1948550611398415

Bonam, C. M., Nair Das, V., Coleman, B. R., & Salter, P. (2019). Ignoring history, denying racism: Mounting evidence for the Marley hypothesis and epistemologies of ignorance.

Social Psychological and Personality Science, *10*(2), 257–265. https://doi.org/10.1177/1948550617751583

Bouvier, G., & Machin, D. (2021). What gets lost in Twitter 'cancel culture' hashtags? Calling out racists reveals some limitations of social justice campaigns. *Discourse and Society*, *32*(3), 307–327. https://doi.org/10.1177/0957926520977215

Cameron, C. D., Payne, B. K., & Knobe, J. (2010). Do theories of implicit race bias change moral judgments? *Social Justice Research*, *23*(4), 272–289. https://doi.org/10.1007/s11211-010-0118-z

Clark, M. D. (2020). DRAG THEM: A brief etymology of so-called "cancel culture." *Communication and the Public*, *5*(3–4), 88–92. https://doi.org/10.1177/2057047320961562

Corneille, O., & Hütter, M. (2020). Implicit? What so you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, *24*(3), 212–232. https://doi.org/10.1177/1088868320911325

Correll, J., Hudson, S. M., Guillermo, S., & Ma, D. S. (2014). The police officer's dilemma: A decade of research on racial bias in the decision to shoot. *Social and Personality Psychology Compass*, *8*(5), 201–213. https://doi.org/10.1111/SPC3.12099

Cova, F., Lantian, A., & Boudesseul, J. (2016). Can the Knobe Effect be explained away? Methodological controversies in the study of the relationship between intentionality and morality. *Personality and Social Psychology Bulletin*, *42*(10), 1295–1308. https://doi.org/10.1177/0146167216656356

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, *129*(3), 414–446. https://doi.org/10.1037/0033-2909.129.3.414

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

Crandall, C. S., Silvia, P. J., N'Gbala, A. N., Tsang, J. A., & Dawson, K. (2007). Balance theory, unit relations, and attribution: The underlying integrity of heiderian theory. *Review of General Psychology*, *11*(1), 12–30. https://doi.org/10.1037/1089-2680.11.1.12

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, *29*(4), 532–544. https://doi.org/10.1177/0146167202250923

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*(1), 525–556. https://doi.org/10.1146/annurev.ps.41.020190.002521

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*. https://doi.org/10.1016/j.jesp.2019.04.010

Daumeyer, N. M., Onyeador, I. N., & Richeson, J. A. (2021). Does shared gender group membership mitigate the effect of implicit bias attributions on accountability for gender-based discrimination? *Personality and Social Psychology Bulletin*, *47*(9), 1343–1357. https://doi.org/10.1177/0146167220965306

Daumeyer, N. M., Rucker, J. M., & Richeson, J. A. (2017). Thinking structurally about implicit bias: Some peril, lots of promise. *Psychological Inquiry, 28*(4), 258–261. https://doi.org/10.1080/1047840X.2017.1373556

de Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science, 14*(5), 835–840. https://doi.org/10.1177/1745691619855638

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*, 1267–1278. https://doi.org/10.1016/j.jesp.2012.06.003

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, *60*(6), 817–830. https://doi.org/10.1037/0022-3514.60.6.817

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*(5), 835–848. https://doi.org/10.1037/0022-3514.82.5.835

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science, 11*(4), 315–319. https://doi.org/10.1111/1467-9280.00262

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

Duguid, M. M., & Thomas-Hunt, M. C. (2015). Condoning stereotyping? How awareness of stereotyping prevalence impacts expression of stereotypes. *Journal of Applied Psychology*, *100*(2), 343–359. https://doi.org/10.1037/a0037908

Effron, D. A., & Conway, P. (2015). When virtue leads to villainy: advances in research on moral self-licensing. *Current Opinion in Psychology*, *6*, 32–35. https://doi.org/10.1016/j.copsyc.2015.03.017

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

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

Fiske, S. T. (2004). Intent and ordinary bias: Unintended thought and social motivation create casual prejudice. *Social Justice Research*, *17*(2), 117–127.

Fitzgerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology*, *7*(1). https://doi.org/10.1186/s40359-019-0299-7

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, *14*(4), 574–595. https://doi.org/10.1177/1745691619826015

Gawronski, B., Ledgerwood, A., & Eastwick, P. W. (2020). Implicit Bias and Antidiscrimination Policy. *Policy Insights from the Behavioral and Brain Sciences*, *7*(2), 99–106. https://doi.org/10.1177/2372732220939128

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, *23*(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, *19*(12), 1260–1262. https://doi.org/10.1111/j.1467-9280.2008.02208.x

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

Greenwald, A. G., Dasgupta, N., Dovidio, J. F., Kang, J., Moss-Racusin, C. A., & Teachman, B. A. (2022). Implicit-bias remedies: Treating discriminatory bias as a public-health problem. *Psychological Science in the Public Interest*, *23*(1), 7–40. https://doi.org/10.1177/15291006211070781

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

Greenwald, G. (2021). *The journalistic tattletale and censorship industry suffers several well-deserved blows*. https://greenwald.substack.com/p/the-journalistic-tattletale-and-censorship?s=r

Guglielmo, S. (2015). Moral judgment as information processing: an integrative review. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01637

Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the Heart of Morality Lies Folk Psychology. *Inquiry*, *52*(5), 449–466. https://doi.org/10.1080/00201740903302600

Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, *116*(5), 769–794. https://doi.org/10.1037/pspi0000155

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology General*, *143*(3), 1369–1392. https://doi.org/10.1037/a0035028

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*(5827), 998–1002. https://doi.org/10.1126/science.1137651

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Hetey, R. C., & Eberhardt, J. L. (2018). The numbers don't speak for themselves: Racial disparities and the persistence of inequality in the criminal justice system. *Current Directions in Psychological Science*, *27*(3), 183–187. https://doi.org/10.1177/0963721418763931

Howell, J. L., Collisson, B., Crysel, L., Garrido, C. O., Newell, S. M., Cottrell, C. A., Smith, C. T., & Shepperd, J. A. (2013). Managing the threat of impending implicit attitude feedback. *Social Psychological and Personality Science*, *4*(6), 714–720. https://doi.org/10.1177/1948550613479803

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*(4), 373–381. https://doi.org/10.1177/1948550614561127

Howell, J. L., Redford, L., Pogge, G., & Ratliff, K. A. (2017). Defensive responding to IAT feedback. *Social Cognition*, *35*(5), 520–562. https://doi.org/10.1521/soco.2017.35.5.520

Ikizer, E. G., & Blanton, H. (2016). Media coverage of "wise" interventions can reduce concern for the disadvantaged. *Journal of Experimental Psychology: Applied*, *22*(2), 135–147. https://doi.org/10.1037/xap0000076

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01359

Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass*, *3*(3), 522–540. https://doi.org/10.1111/J.1747-9991.2008.00138.X

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies*, *130*(2), 203–231. https://doi.org/10.1007/S11098-004-4510-0

Kraus, M. W., Onyeador, I. N., Daumeyer, N. M., Rucker, J. M., & Richeson, J. A. (2019). The misperception of racial economic inequality. *Perspectives on Psychological Science*, *14*(6), 899–921. https://doi.org/10.1177/1745691619863049

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing implicit racial preferences: A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. https://doi.org/10.1037/a0036260

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect. *Psychological Science*, *17*(5), 421–427. https://doi.org/10.1111/j.1467-9280.2006.01722.x

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Mallett, R. K., & Monteith, M. J. (2019). Confronting prejudice and discrimination: Historical influences and contemporary approaches. In *Confronting Prejudice and Discrimination: The Science of Changing Minds and Behaviors*.

McConahay, J. B. (1983). Modern racism and modern discrimination. *Personality and Social Psychology Bulletin*, *9*(4), 551–558. https://doi.org/10.1177/0146167283094004

Miller, S. S. (2014). *Factors influencing attributions to prejudice: harm, intent, and individual differences in the propensity to make attributions to prejudice* [Kansas State University]. https://krex.k-state.edu/dspace/handle/2097/18221

Miller, S. S., O'Dea, C. J., Lawless, T. J., & Saucier, D. A. (2019). Savage or satire: Individual differences in perceptions of disparaging and subversive racial humor. *Personality and Individual Differences, 142,* 28–41. https://doi.org/10.1016/j.paid.2019.01.029

Miller, S. S., O'Dea, C. J., & Saucier, D. A. (2021). "I can't breathe": Lay conceptualizations of racism predict support for Black Lives Matter. *Personality and Individual Differences, 173,* 110625. https://doi.org/10.1016/J.PAID.2020.110625

Miller, S. S., Peacock, N. K., & Saucier, D. A. (2021). Propensities to make attributions to prejudice and (mis)perceptions of racism in the context of police shootings. *Personality and Individual Differences, 182,* 111088. https://doi.org/10.1016/J.PAID.2021.111088

Miller, S. S., & Saucier, D. A. (2018). Individual differences in the propensity to make attributions to prejudice. *Group Processes and Intergroup Relations, 21*(2), 280–301. https://doi.org/10.1177/1368430216674342

Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General, 146*(1), 123–133. https://doi.org/10.1037/xge0000234

Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology, 116*(2), 215–236. https://doi.org/10.1037/pspa0000137

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65*(3), 469–485. https://doi.org/10.1037/0022-3514.65.3.469

Monteith, M. J., Burns, M. D., & Hildebrand, L. K. (2019). Navigating successful confrontations: What should I say and how should I say it? In *Confronting Prejudice and*

Discrimination: The Science of Changing Minds and Behaviors (pp. 225–248). Elsevier. https://doi.org/10.1016/B978-0-12-814715-3.00006-0

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology*, *64*(2), 198–210. https://doi.org/10.1037/0022-3514.64.2.198

Monteith, M. J., & Walters, G. L. (1998). Egalitarianism, moral obligation, and prejudice-related personal standards. *Personality and Social Psychology Bulletin*, *24*(2), 186–199. https://doi.org/10.1177/0146167298242007

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

Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*, 1–61. https://doi.org/10.1017/S0140525X12003214

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Norris, P. (2021). Cancel culture: Myth or reality? *Political Studies*. https://doi.org/10.1177/00323217211037023

Nosek, B. A., & Hansen, J. H. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*(4), 553–594. https://doi.org/10.1080/02699930701438186

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15*(4), 152–159. https://doi.org/10.1016/J.TICS.2011.01.005

O'Brien, L. T., Blodorn, A., Alsbrooks, A., Dube, R., Adams, G., & Nelson, J. C. (2009). Understanding White Americans' perceptions of racism in Hurricane Katrina-related events. *Group Processes & Intergroup Relations, 12*(4), 431–444. https://doi.org/10.1177/1368430209105047

Payne, B. K., & Vuletich, H. A. (2018). Policy insights from advances in implicit bias research. *Policy Insights from the Behavioral and Brain Sciences, 5*(1), 49–56. https://doi.org/10.1177/2372732217746190

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry, 28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

Payne, K., Niemi, L., & Doris, J. M. (2018). *How to think about "implicit bias."* Scientific American. https://www.scientificamerican.com/article/how-to-think-about-implicit-bias/

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology, 61*, 64–78. https://doi.org/10.1016/j.jesp.2015.06.007

Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76*(1), 114–128. https://doi.org/10.1037/0022-3514.76.1.114

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The Role of Perceived Metadesires. *Psychological Science, 14*(3), 267–272. https://doi.org/10.1111/1467-9280.03433

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832. https://doi.org/10.1037/0022-3514.75.3.811

Redford, L., & Ratliff, K. A. (2016). Perceived moral responsibility for attitude-based discrimination. *The British Journal of Social Psychology*, *55*(2), 279–296. https://doi.org/10.1111/bjso.12123

Romano, A. (2020). What is cancel culture? Why we keep fighting about canceling people. *Vox*. https://www.vox.com/culture/2019/12/30/20879720/what-is-cancel-culture-explained-history-debate

Rucker, J. M., & Richeson, J. A. (2021a). Beliefs about the interpersonal vs. structural nature of racism and responses to racial inequality. In *The Routledge International Handbook of Discrimination, Prejudice and Stereotyping* (pp. 13–25). Routledge. https://doi.org/10.4324/9780429274558-2

Rucker, J. M., & Richeson, J. A. (2021b). Toward an understanding of structural racism: Implications for criminal justice. *Science*, *374*(6565), 286–290. https://doi.org/10.1126/SCIENCE.ABJ7779/ASSET/A7CE148A-D7D3-4B7A-B9DF-EBE249BA23AE/ASSETS/IMAGES/LARGE/SCIENCE.ABJ7779-F2.JPG

Salter, P., & Adams, G. (2013). Toward a critical race psychology. *Social and Personality Psychology Compass*, *7*(11), 781–793. https://doi.org/10.1111/spc3.12068

Salter, P. S., Adams, G., & Perez, M. J. (2018). Racism in the structure of everyday worlds: A cultural-psychological perspective. *Current Directions in Psychological Science, 27*(3), 150–155. https://doi.org/10.1177/0963721417724239

Saucier, D. A., Strain, M. L., Miller, S. S., O'Dea, C. J., & Till, D. F. (2018). "What do you call a Black guy who flies a plane?": The effects and understanding of disparagement and confrontational racial humor. *Humor*, *31*(1), 105–128. https://doi.org/10.1515/humor-2017-0107

Sears, D. O., & Henry, P. J. (2003). The origins of symbolic racism. *Journal of Personality and Social Psychology*, *85*(2), 259–275. https://doi.org/10.1037/0022-3514.85.2.259

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer.

Sher, G. (2015). Out of control. *Ethics*, *116*(2), 285–301. https://doi.org/10.1086/498464

Simon, S., Moss, A. J., & O'Brien, L. T. (2019). Pick your perspective: Racial group membership and judgments of intent, harm, and discrimination. *Group Processes and Intergroup Relations*, *22*(2), 215–232. https://doi.org/10.1177/1368430217735576

Smith, A. M. (2015). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, *115*(2), 236–271. https://doi.org/10.1086/426957

Sørensen, J., Soule, S., & Manduca, R. (2018). Income inequality and the persistence of racial economic disparities. *Sociological Science*, *5*, 182–205. https://doi.org/10.15195/V5.A8

Spencer, K. B., Charbonneau, A. K., & Glaser, J. (2016). Implicit Bias and Policing. *Social and Personality Psychology Compass*, *10*(1), 50–63. https://doi.org/10.1111/spc3.12210

Suhler, C. L., & Churchland, P. S. (2009). Control: conscious and otherwise. *Trends in Cognitive Sciences*, *13*(8), 341–347. https://doi.org/10.1016/J.TICS.2009.04.010

Swim, J. K., Scott, E. D., Sechrist, G. B., Campbell, B., & Stangor, C. (2003). The role of intent and harm in judgments of prejudice and discrimination. *Journal of Personality and Social Psychology*, *84*(5), 944–959. https://doi.org/10.1037/0022-3514.84.5.944

Tooby, J., & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. *Human Morality and Sociality: Evolutionary and Comparative Perspectives*, 191–234. http://www.psych.ucsb.edu/research/cep/papers/groups in mind2010.pdf

Vitriol, J. A., & Moskowitz, G. B. (2021). Reducing defensive responding to implicit bias feedback: On the role of perceived moral threat and efficacy to change. *Journal of Experimental Social Psychology*, *96*, 104165. https://doi.org/10.1016/j.jesp.2021.104165

Voiklis, J., & Malle, B. F. (2017). Moral cognition and its basis in social cognition and social regulation. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 108–120). Guilford Press.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. Guilford Press.

Weiner, B. (1996). Searching for order in social motivation. *Psychological Inquiry*, *7*(3), 199–216. https://doi.org/10.1207/s15327965pli0703_1

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Young, L., & Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405. https://doi.org/10.1162/JOCN.2009.21137

Appendix A - Study 1 Materials

Informed Consent

PROJECT TITLE: Reactions to Discrimination

PROJECT APPROVAL DATE: 10/4/2021

LENGTH OF STUDY: This study will take approximately 15 minutes to complete.

PRINCIPAL INVESTIGATOR: Donald A. Saucier, Ph.D.

CONTACT FOR ANY PROBLEMS/QUESTIONS: Don Saucier, saucier@ksu.edu

IRB CHAIR CONTACT/PHONE INFORMATION: Lisa Rubin, Chair, Committee on Research Involving Human Subjects, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224; Brad Woods, Associate Vice President for Research Compliance, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224.

PURPOSE OF RESEARCH: The purpose of this research is to examine people's reactions to cases of racial discrimination.

PROCEDURES: You will read about a case of racial discrimination and complete a set of questionnaires related to your attitudes.

RISKS ANTICIPATED: You may consider some of the material in this study sensitive, offensive, or upsetting, but the materials in this study are no more extreme than things that you might experience in everyday life.

BENEFITS ANTICIPATED: You will receive $1.00 credited to your MTurk account and may learn more about racial bias.

EXTENT OF CONFIDENTIALITY: Your responses will be strictly anonymous and confidential.

TERMS OF PARTICIPATION: I understand this project is research and that my participation is completely voluntary. I also understand that if I decide to participate in this study, I may withdraw my consent at any time, and stop participating at any time without explanation, penalty, or loss of benefits to which I may otherwise be entitled.

I verify that by continuing the survey, I am indicating that I have read and understand this consent form, and willingly agree to participate in this study under the terms described.

- I agree to participate
- I do not agree to participate

---------- *Page break* ----------

Instructions

Thank you for participating in this study.

On the next page, please read the short passage about a case of racial discrimination. On the following pages you will be asked to answer some questions and provide your impressions about the events described in the passage.

---------- *Page break* ----------

Manipulation

Discrimination Vignette

Roundtree State University is a public university that enrolls around 20,000 undergraduate students each year. Every student is assigned an individual advisor in their first year to help them choose a major. Each year, the university surveys students to evaluate the effectiveness of their advisors. One recent survey revealed that a Black student named Brianna, complained that her White advisor, Melissa, strongly suggested she major in social work, rather than pursue

Brianna's preferred major in biology. Brianna believed that Melissa only encouraged her to choose social work because she is Black and social work is an academically easier major.

To further investigate Brianna's claim of discrimination, the university appointed a team of specialists in identifying cases of discrimination. The team reviewed Melissa's advising records and conducted interviews with Melissa, Brianna, and many other students previously advised by Melissa. They discovered a pattern of advising Black students, but not White students, to choose majors in less intellectually demanding fields. The team concluded that this was a clear pattern of discrimination against Black students. Consequently, the team ruled that the advice given to Brianna was the result of…

Implicit Bias Condition

Melissa's unconscious bias against Black students. In other words, the team concluded Melissa is unaware that she holds negative stereotypes and attitudes about Black students. This means that Melissa does not consciously believe that Black students are less intellectually capable than White students. It appears these unconscious attitudes affected her behavior toward Brianna.

Explicit Bias Condition

Melissa's conscious bias against Black students. In other words, the team concluded Melissa is aware that she holds negative stereotypes and attitudes about Black people. This means that Melissa consciously believes that Black students are not as intellectually capable as White students. It appears these conscious attitudes affected her behavior toward Brianna.

Please indicate whether you have read the above passage.

- I have read the above passage
- I did not read the above passage

---------- *Page break* ----------

Manipulation Check

The instance of discrimination in the passage was described as being caused by...

- Melissa's unconscious bias against Black people
- Melissa's conscious bias against Black people
- I'm not sure
- The instance was not attributed to anything

---------- *Page break* ----------

Dependent Variables

Note: Unless otherwise stated, participants rated their level of level of agreement with the following statements from 1 (Strongly Disagree) to 7 (Strongly Agree) for each scale. The order of the items within a scale was randomized.

Dependent Variables: Perpetrator

The items on this page ask for your perceptions about Melissa, the White advisor.

Mental State Attributions

Awareness

1. Melissa knew she was discriminating against Brianna
2. Melissa was aware of how her bias was affecting her behavior
3. Melissa was conscious of how her bias was affecting her behavior

Control

1. Melissa had control over her behavior toward Brianna

2. Melissa could have prevented herself from discriminating if she had exerted more effort
3. Melissa could have chosen to not discriminate against Brianna
4. There is no way that Melissa could have stopped herself from discriminating against Brianna (R)

Intent

1. Melissa intended to discriminate against Brianna
2. Melissa meant to discriminate against Brianna
3. Melissa discriminated against Brianna on purpose
4. Melissa discriminated against Brianna unintentionally (R)
5. Melissa wanted to discriminate against Brianna

Foreseeability

1. Melissa should have foreseen the possibility that her bias could affect Brianna
2. There is no way Melissa could have known she would discriminate against Brianna (R)

Emotional States

To what extent do you think **Melissa** felt the following emotions about the investigation team's conclusion that she discriminated against Brianna? 1 = (*Not at all*) to 7 = (*Very much*)

1. Shame
2. Guilt
3. Pride
4. Joy

5. Happiness
6. Sadness
7. Anxiety
8. Anger at herself
9. Anger at Brianna
10. Anger at the investigatory team

Blame Judgments (Responsibility, Accountability)

1. I blame Melissa for discriminating against Brianna
2. Melissa is entirely responsible for discriminating against Brianna
3. Melissa should be held accountable for discriminating against Brianna

Emotional Reactions

Anger/Outrage

1. It upsets me that Melissa discriminated against Brianna
2. I am angry that Melissa treated Brianna in a discriminatory way
3. I am outraged at Melissa's discriminatory behavior
4. I am disgusted by Melissa's discriminatory behavior

Sympathy

1. I sympathize with what Melissa must be going through
2. I feel compassion for Melissa
3. I empathize with what Melissa must be feeling
4. I don't feel bad for Melissa (R)

Mild Punishment

1. Melissa should be forced to apologize to Brianna

2. Melissa should be forced to complete bias sensitivity training
3. If Melissa is allowed to keep her job, she should be closely supervised to prevent biased behavior

Severe Punishment

1. Melissa should be fired
2. Melissa should be publicly criticized
3. Melissa should be charged with a crime
4. Melissa should be sued

Note: Two additional items were also included in the materials, *Melissa should be punished*, and *Melissa should be put on probation*, but these items cross-loaded onto both mild and severe punishment factors in the principal components analysis, so these two items were not included in the composite variables and were not analyzed further.

Prosocial Consoling/Forgiving

1. Someone should console Melissa
2. Someone should help Melissa cope with any shame she might feel
3. Someone should help Melissa deal with any negative consequences of her behavior
4. Melissa should be forgiven for her behavior
5. Someone should help Melissa understand she is not a bad person

Prosocial Helping to Correct

1. Someone should help Melissa learn from her mistake
2. As kindly as possible, someone should help Melissa reduce her bias
3. I think Melissa would benefit from anti-bias training

Note: One additional item was included in the materials, *I think Melissa will become a better person from this experience*, but was omitted because it did not reliably load onto either of the two factors and was not analyzed further.

Perceptions of the Perpetrator's Moral Character

1. Melissa is a bad person (R)
2. Melissa is morally flawed (R)
3. Melissa is a good person who made an unfortunate mistake

Institutional Reform

1. The university should implement bias awareness training for all of its staff
2. The university should take steps to reduce bias in all of their staff
3. The university should implement policies that reduce the likelihood that discrimination will occur anywhere on campus

Forced-Choice Decision

If you had to choose just one consequence for Melissa's behavior, which would you choose (choose one):

a. The university should focus on punishing Melissa for having discriminated
b. The university should focus on helping Melissa not discriminate in the future

---------- *Page break* ----------

Dependent Variables: Victim

The items on this page ask for your perceptions about Brianna, the Black student.

Harm

On a scale from 1 = (*Not at all*) to 7 = (*Very Much*), please indicate to what extent you believe the experience of discrimination was…

1. Harmful for Brianna
2. Hurtful for Brianna
3. Difficult for Brianna
4. Emotionally painful for Brianna
5. A hinderance to Brianna's future opportunities

Redress

1. Brianna should be compensated
2. Brianna should be awarded a large sum of money in a legal settlement
3. Brianna should receive an apology
4. Brianna should be given a different advisor

Sympathy

1. I sympathize with what Brianna was going through
2. I feel compassion for Brianna
3. I empathize with what Brianna must be feeling
4. I don't feel bad for Brianna (R)

---------- *Page break* ----------

Dependent Variables: Evaluations of Bias

Note: These items will be analyzed as single item variables.

1. The discrimination that Brianna experienced was morally wrong
2. The kind of bias described in the passage is a widespread problem
3. Some level of bias in situations like the one described should just be expected

---------- *Page break* ----------

Potential Moderator Variables

Perpetrator Perspective-Taking

Note: The following will be measured on a scale from 1 = (*Not at all*) to 7 = (*Very Much*)

When I read the passage…

1. I imagined myself being in Melissa's situation
2. I saw myself being at risk for making the same mistake that Melissa did
3. I became concerned that I might discriminate in a way that is similar to Melissa's behavior

Victim Perspective-Taking

When I read the passage…

1. I imagined myself in Brianna's situation
2. I saw myself being at risk for being discriminated against in a similar way
3. I became concerned that I might be the victim of the kind of bias that affected Brianna

---------- *Page break* ----------

On the following pages, please answer in relation to your general beliefs and opinions.

---------- *Page break* ----------

Note: The following scales were presented in a random order and the order of the items within each scale was randomized.

Propensity to Make Attributions to Prejudice Scale (Miller & Saucier, 2018)

Note: Participants will not see the subscale labels.

Pervasiveness

1. People discriminate against people who are not like them

2. Racist behavior is more widespread than people think it is
3. Other people treat minorities based on stereotypes
4. You'll see lots of racism if you look for it

Trivialization

5. Racial minorities are too worried about being discriminated against
6. Racial minorities are too sensitive about stereotypes
7. Minorities today are overly worried about being victims of racism
8. People are overly concerned about racial issues

Vigilance

9. I think about why racial minorities are treated stereotypically
10. I think about whether people act in a prejudiced or discriminatory manner
11. I consider whether people's actions are prejudiced or discriminatory
12. I am on the lookout for instances of prejudice or discrimination

Confidence

13. I am quick to recognize prejudice
14. My friends think I'm good at spotting racism
15. I find that prejudice and discrimination are pretty easy to spot

Lay Conceptualizations of Racism (Miller et al., 2021)

Note: Participants did not see the subscale labels. Participants were given the following instructions: Please rate each of the following statements in terms of how much each contributes to the problem of racism in the United States today on a scale from 1 = (*Not at all*) to 7 = (*Very much*).

Systematic Oppression

1. A history of policies that systematically disadvantaged People of Color
2. A history of racial inequality that perpetuates racial problems
3. Discrimination that is built into our laws and institutions
4. A society and culture that perpetuates the idea that White people are superior to People of Color

Individual Acts of Prejudice

5. Individuals' own beliefs and prejudices that cause them to treat those of other races poorly
6. Discrimination that is based on the prejudices of individual people
7. Intentional acts of racial discrimination and abuse by racist individuals
8. Individuals' beliefs about White racial superiority

Bias Awareness (Perry et al., 2015)

1. Even though I know it's not appropriate, I sometimes feel that I hold unconscious negative attitudes toward people based on their social groups (e.g., race, gender)
2. When talking to people, I sometimes worry that I am unintentionally acting in a prejudiced way
3. I worry that I have unconscious biases toward some social groups
4. I never worry that I may be acting in a subtly prejudiced way toward people based on their social groups (R)

Demographics

- What is your age?
- What is your current gender identity? (Please check all that apply)
    - Female

    - Male
    - Genderqueer
    - Genderfluid
    - Transgender
    - Transgender Feminine
    - Transgender Masculine
    - Agender
    - Androgynous
    - Two-Spirit
    - Demigender
    - Questioning or unsure
    - I prefer to describe my gender using my own language:
    - I prefer not to disclose
- I identify as:
    - White
    - Black
    - Hispanic/Latino/a/x
    - Asian
    - Indigenous
    - Pacific Islander
    - Multiracial
    - I prefer to identify as:
    - I prefer not to disclose

- What country best represents your national identity? (*Note:* participants will be able to select from a drop-down menu listing countries)
- What is the highest level of education you have completed?
    - Less than high school
    - High school/GED
    - Some college
    - 2-year college degree
    - 4-year college degree
    - Masters degree
    - Doctoral degree
    - Professional degree (JD, MD)
- In which state do you currently reside? (*Note:* participants will be able to select from a drop-down menu listing states in the United States)
- Please indicate your overall political viewpoint on the scale below: (1 = Very Liberal, 7 = Very Conservative)

Honesty Check

The study is very important to us, and a considerable amount of time and effort has gone into creating this survey. As such, if for whatever reason you feel that you did not complete the survey carefully or accurately, it would be extremely helpful if you could let us know this now. Your response is anonymous and will in no way affect your compensation.

- I DID complete the survey carefully and accurately. Please include my responses in analyses.

- I DID NOT complete the survey carefully or accurately. Please exclude my responses from analyses.

Debriefing

Thank you for your participation!

The purpose of this study was to examine whether people blame others less for discriminating when their discriminatory behavior was caused by their unconscious biases compared to when their discriminatory behavior was caused by their conscious biases. The scenario you read was not real and was made up for the purposes of this study.

Thank you for participating in this study. It would not be possible to continue psychological research without the help of individuals like you. If you have any questions or concerns about this research or are distressed as a result of the materials used in this study, you may contact Dr. Saucier at: saucier@ksu.edu. Thank you again for participating in this study.

Please click the Get Completion Code button below to receive your MTurk completion code and receive payment.

Appendix B - Study 2 Materials

Informed Consent

PROJECT TITLE: Reactions to Discrimination

PROJECT APPROVAL DATE: 10/4/2021

LENGTH OF STUDY: This study will take approximately 15 minutes to complete.

PRINCIPAL INVESTIGATOR: Donald A. Saucier, Ph.D.

CONTACT FOR ANY PROBLEMS/QUESTIONS: Don Saucier, saucier@ksu.edu

IRB CHAIR CONTACT/PHONE INFORMATION: Lisa Rubin, Chair, Committee on Research Involving Human Subjects, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224; Brad Woods, Associate Vice President for Research Compliance, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224.

PURPOSE OF RESEARCH: The purpose of this research is to examine people's reactions to cases of racial discrimination.

PROCEDURES: You will read about a case of racial discrimination and complete a set of questionnaires related to your attitudes.

RISKS ANTICIPATED: You may consider some of the material in this study sensitive, offensive, or upsetting, but the materials in this study are no more extreme than things that you might experience in everyday life.

BENEFITS ANTICIPATED: You will receive $1.00 credited to your MTurk account and may learn more about racial bias.

EXTENT OF CONFIDENTIALITY: Your responses will be strictly anonymous and confidential.

TERMS OF PARTICIPATION: I understand this project is research and that my participation is completely voluntary. I also understand that if I decide to participate in this study, I may withdraw my consent at any time, and stop participating at any time without explanation, penalty, or loss of benefits to which I may otherwise be entitled.

I verify that by continuing the survey, I am indicating that I have read and understand this consent form, and willingly agree to participate in this study under the terms described.

- I agree to participate
- I do not agree to participate

---------- *Page break* ----------

Instructions

Thank you for participating in this study.

On the next page, please read the short excerpt from an article about implicit bias. On the following pages you will be asked to answer some questions about the article.

---------- *Page break* ----------

Implicit Bias Framing Manipulation

Note: Participants did not see this note. The author of the article participants read was not one of the authors from the articles listed below and Science Weekly was not a real publication. The italicized heading indicates the article corresponding to the experimental condition and was not presented to participants. The text of the articles was copied and revised from the following online articles that were among the top Google search results for implicit bias:

https://www.scientificamerican.com/article/how-to-think-about-implicit-bias/

https://patientengagementhit.com/news/what-is-implicit-bias-how-does-it-affect-healthcare

Spontaneous Awareness and Control Framing of Implicit Bias

Understanding Implicit Bias: People Can Be Made Aware of and Can Control Their Implicit Biases

Article published in *Science Weekly* by Sherman Fey

When's the last time a stereotype popped into your mind? If you are like most people, the author included, it happens all the time. That doesn't make you a racist, sexist or whatever-ist. It just means your brain is working properly, noticing patterns and making generalizations. But the same thought processes that make people smart can also make them biased. This tendency for stereotype-confirming thoughts to pass spontaneously through our minds is what psychologists call implicit bias. It sets people up to overgeneralize, sometimes leading to discrimination even when people feel they are being fair.

According to the Kirwan Institute for the Study of Race and Ethnicity at the Ohio State University, implicit bias is involuntary, can refer to positive or negative attitudes and stereotypes, and can affect actions without an individual knowing it:

Implicit bias refers to the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. Residing deep in the subconscious, these biases are different from explicit attitudes and stereotypes that individuals consciously believe in and may act upon intentionally. Instead, implicit biases are activated involuntarily and without an individual's awareness or intentional control.

However, some research suggests that, although implicit bias works subconsciously, people can become conscious of these biases with a little effort. By drawing our attention to instances when stereotypes pop into mind, or by receiving feedback from others about our behaviors, we can become aware of biases that would normally remain hidden to us. Further research suggests that by gaining awareness of our implicit biases, with some effort we can control how they affect our behavior. In other words, if we put in the work, we can control our implicit biases.

Please indicate whether you have read the above article.

- I have read the above article
- I did not read the above article

Unawareness/Uncontrollable Framing of Implicit Bias

Understanding Implicit Bias: People Lack Awareness and Cannot Control Their Implicit Biases

Article published in *Science Weekly* by Sherman Fey

When's the last time a stereotype popped into your mind? If you are like most people, the author included, it happens all the time. That doesn't make you a racist, sexist or whatever-ist. It just means your brain is working properly, noticing patterns and making generalizations. But the same thought processes that make people smart can also make them biased. This tendency for stereotype-confirming thoughts to pass spontaneously through our minds is what psychologists call implicit bias. It sets people up to overgeneralize, sometimes leading to discrimination even when people feel they are being fair.

According to the Kirwan Institute for the Study of Race and Ethnicity at the Ohio State University, implicit bias is involuntary, can refer to positive or negative attitudes and stereotypes, and can affect actions without an individual knowing it:

Implicit bias refers to the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. Residing deep in the subconscious, these biases are different from explicit attitudes and stereotypes that individuals consciously believe in and may act upon intentionally. Instead, implicit biases are activated involuntarily and without an individual's awareness or intentional control.

Some research suggests that implicit bias works subconsciously, and that people are not aware of these biases. In other words, we are unable to simply self-examine our conscious attitudes and behaviors to learn whether we have implicit biases. Further research suggests that we can only know about our implicit biases through feedback from scientific instruments that measure our implicit associations, but that we cannot control how they affect our behavior. In other words, our implicit biases are not something we can typically be aware of or control.

Please indicate whether you have read the above article.

- I have read the above article
- I did not read the above article

---------- *Page break* ----------

Article Manipulation Check

Please select the option that best summarizes the conclusions from the article.

- With effort, people can become aware of and control their implicit biases
- People cannot become aware of or control their implicit biases
- I'm not sure

Awareness and Control Agreement

Please indicate your level of agreement with the following statements on the scale provided. 1 = Strongly Disagree to 7 = Strongly Agree

1. With effort, people can become aware of their implicit biases
2. With effort, people can control their implicit biases

---------- *Page break* ----------

Instructions

On the next page, please read the short passage about a case of racial discrimination that resulted from implicit bias. On the following pages you will be asked to answer some questions and provide your impressions about the events described in the passage.

---------- *Page break* ----------

Discrimination Vignette

A recent internal investigation by Westhouse Bank found that a manager, Steven Hillsborough, had denied mortgage applications from Black families who had similarly acceptable credit and financial stability as White families whose loans were approved by Hillsborough. The investigation was prompted by complaints filed by Margaret and Trent Washington, a Black couple who Hillsborough denied a loan to buy a new home. The investigation team concluded that Hillsborough racially discriminated against the Washingtons and that **Hillsborough's**

behavior was due to implicit bias. The team's report noted that Hillsborough had completed the company's anti-bias training program earlier that year.

Acknowledged Bias Condition

The leader of the training program that Hillsborough completed earlier that year also noted that, at the time, Hillsborough appeared to acknowledge the possibility that he may have implicit biases, and that he seemed committed to trying to prevent those biases from affecting his decisions.

Unacknowledged Bias Condition

The leader of the training program that Hillsborough completed earlier that year also noted that, at the time, Hillsborough did not appear to acknowledge the possibility that he may have implicit biases, nor did he seem committed to trying to prevent those biases from affecting his decisions.

Please indicate whether you have read the above passage.

- I have read the above passage
- I did not read the above passage

---------- *Page break* ----------

Vignette Manipulation Check

In the passage, Hillsborough was described as...

- Acknowledging his implicit bias against Black people
- Not acknowledging his implicit bias against Black people
- I'm not sure

Bias and Discrimination Agreement

Please indicate your level of agreement with the following statements.

1. The Washingtons' experience of discrimination was caused by Hillsborough's implicit bias.
2. Hillsborough discriminated against the Washingtons because of the Washingtons' race.

---------- *Page break* ----------

Dependent Variables

Note: Unless otherwise stated, participants rated their level of level of agreement with the following statements from 1 (Strongly Disagree) to 7 (Strongly Agree) for each scale. The order of the items within a scale was be randomized.

Dependent Variables: Perpetrator

The items on this page ask for your perceptions about Hillsborough, the White loan officer.

Mental State Attributions

Awareness

1. Hillsborough knew he was discriminating against the Washingtons
2. Hillsborough was aware of how his bias was affecting his behavior
3. Hillsborough was conscious of how his bias was affecting his behavior

Control

1. Hillsborough had control over his behavior toward the Washingtons
2. Hillsborough could have prevented himself from discriminating if he had exerted more effort
3. Hillsborough could have chosen to not discriminate against the Washingtons
4. There is no way that Hillsborough could have stopped himself from discriminating against the Washingtons (R)

Intent

1. Hillsborough intended to discriminate against the Washingtons
2. Hillsborough meant to discriminate against the Washingtons
3. Hillsborough discriminated against the Washingtons on purpose
4. Hillsborough discriminated against the Washingtons unintentionally (R)
5. Hillsborough wanted to discriminate against the Washingtons

Foreseeability

1. Hillsborough should have foreseen the possibility that his bias could affect the Washingtons
2. There is no way Hillsborough could have known he would discriminate against the Washingtons (R)

Emotional States

To what extent do you think **Hillsborough** felt the following emotions about the investigation team's conclusion that he discriminated against the Washingtons? 1 = (*Not at all*) to 7 = (*Very much*)

1. Shame
2. Guilt
3. Pride
4. Joy
5. Happiness
6. Sadness
7. Anxiety
8. Anger at himself
9. Anger at the Washingtons

10. Anger at the investigatory team

Blame Judgments (Responsibility, Accountability)

1. I blame Hillsborough for discriminating against the Washingtons
2. Hillsborough is entirely responsible for discriminating against the Washingtons
3. Hillsborough should be held accountable for discriminating against the Washingtons

Emotional Reactions

Anger/Outrage

1. It upsets me that Hillsborough discriminated against the Washingtons
2. I am angry that Hillsborough treated the Washingtons in a discriminatory way
3. I am outraged at Hillsborough's discriminatory behavior
4. I am disgusted by Hillsborough's discriminatory behavior

Sympathy

1. I sympathize with what Hillsborough must be going through
2. I feel compassion for Hillsborough
3. I empathize with what the Hillsborough must be feeling
4. I don't feel bad for Hillsborough (R)

Mild Punishment

1. Hillsborough should be forced to apologize to the Washingtons
2. Hillsborough should be forced to complete bias sensitivity training
3. If Hillsborough is allowed to keep his job, he should be closely supervised to prevent biased behavior

Severe Punishment

1. Hillsborough should be fired
2. Hillsborough should be publicly criticized
3. Hillsborough should be charged with a crime
4. Hillsborough should be sued

Note: Two additional items were also included in the materials, *Hillsborough should be punished*, and *Hillsborough should be put on probation*, but to maintain consistency with the items used in Study 1, these two items were not included in the composite variables and were not analyzed further.

Prosocial Consoling/Forgiving

1. Someone should console Hillsborough
2. Someone should help Hillsborough cope with any shame he might feel
3. Someone should help Hillsborough deal with any negative consequences of his behavior
4. Hillsborough should be forgiven for his behavior
5. Someone should help Hillsborough understand he is not a bad person

Prosocial Helping to Correct

4. Someone should help Hillsborough learn from his mistake
5. As kindly as possible, someone should help Hillsborough reduce his bias
6. I think Hillsborough would benefit from anti-bias training

Note: One additional item was included in the materials, *I think Hillsborough will become a better person from this experience*, but to maintain consistency with the items used in Study 1, it was omitted from the composite variables and not analyzed further.

Forced-Choice Decision

If you had to choose just one consequence for Hillsborough's behavior, which would you choose (choose one):

a. The company should focus on punishing Hillsborough for having discriminated
b. The company should focus on helping Hillsborough not discriminate in the future

Perceptions of the Perpetrator's Moral Character

1. Hillsborough is a bad person (R)
2. Hillsborough is morally flawed (R)
3. Hillsborough is a good person who made an unfortunate mistake

Institutional Reform

1. The company should implement bias awareness training for all of its staff
2. The company should take steps to reduce bias in all of their staff
3. The company should implement policies that reduce the likelihood that discrimination will occur anywhere within the company

---------- *Page break* ----------

Dependent Variables: Victim

The items on this page ask for your perceptions about the Washingtons, the Black couple.

Harm

On a scale from 1 = (*Not at all*) to 7 = (*Very Much*), please indicate to what extent you believe the experience of discrimination was…

1. Harmful for the Washingtons
2. Hurtful for the Washingtons
3. Difficult for the Washingtons

4. Emotionally painful for the Washingtons
5. A hinderance to the Washingtons' long-term plans and opportunities

Redress

1. The Washingtons should be compensated
2. The Washingtons should be awarded a large sum of money in a legal settlement
3. The Washingtons should receive an apology
4. The Washingtons' loan should be approved

Sympathy

I sympathize with what the Washingtons were going through

I feel compassion for the Washingtons

I empathize with what the Washingtons must be feeling

I don't feel bad for the Washingtons (R)

---------- *Page break* ----------

Dependent Variables: Evaluations of Bias

Note: These items will be analyzed as single item variables.

1. The discrimination that the Washingtons experienced was morally wrong
2. The kind of bias described in the passage is a widespread problem
3. Some level of bias in situations like the one described should just be expected

---------- *Page break* ----------

Dependent Variables: Implicit Bias Attitudes Scale (Miller & Saucier, unpublished data)

Concern

1. Training about implicit bias should be a required part of education and job training

2. Implicit bias is a serious societal problem
3. I am concerned about the effects of implicit biases
4. As a society, we have a moral responsibility to reduce implicit biases
5. Organizations should monitor the effects of implicit biases
6. Implicit biases are the result of systemic forms of oppression
7. Organizations are responsible for minimizing the negative effects of implicit biases
8. People need to be held accountable for their implicit biases
9. People should be aware of their implicit biases
10. Implicit biases are the result of stereotypical representations of different groups of people

Nihilism

11. There is nothing we can do to prevent the negative consequences of implicit biases
12. Implicit bias is impossible to detect
13. Very few people have implicit biases
14. Only bigots have implicit biases
15. Implicit biases are impossible to control
16. Implicit bias is not a real thing

Normality

17. Implicit biases are natural
18. Implicit biases are difficult to control
19. Implicit biases are inevitable
20. Implicit biases are very common
21. Almost everyone holds some degree of implicit bias
22. Implicit biases are the result of normal social processes

23. Even good people have implicit biases
24. I am aware that I may have implicit biases against other groups of people

---------- *Page break* ----------

Potential Moderator Variables

Note: The following scales were presented in a random order and the order of the items within each scale were randomized.

Perspective-Taking

When I read the passage…

1. I imagined myself being in Hillsborough's situation
2. I saw myself being at risk for making the same mistake that Hillsborough did
3. I became concerned that I might discriminate in a way that is similar to Hillsborough's behavior

Victim Perspective-Taking

When I read the passage…

1. I imagined myself in the Washingtons' situation
2. I saw myself being at risk for being discriminated against in a similar way
3. I became concerned that I might be the victim of the kind of bias that affected the Washingtons

---------- *Page break* ----------

On the following pages, please answer in relation to your general beliefs and opinions.

---------- *Page break* ----------

Note: The following scales were presented in a random order and the order of the items within each scale were randomized.

Propensity to Make Attributions to Prejudice Scale (Miller & Saucier, 2018)

Note: Participants will not see the subscale labels.

Pervasiveness

1. People discriminate against people who are not like them
2. Racist behavior is more widespread than people think it is
3. Other people treat minorities based on stereotypes
4. You'll see lots of racism if you look for it

Trivialization

5. Racial minorities are too worried about being discriminated against
6. Racial minorities are too sensitive about stereotypes
7. Minorities today are overly worried about being victims of racism
8. People are overly concerned about racial issues

Vigilance

9. I think about why racial minorities are treated stereotypically
10. I think about whether people act in a prejudiced or discriminatory manner
11. I consider whether people's actions are prejudiced or discriminatory
12. I am on the lookout for instances of prejudice or discrimination

Confidence

13. I am quick to recognize prejudice
14. My friends think I'm good at spotting racism
15. I find that prejudice and discrimination are pretty easy to spot

Lay Conceptualizations of Racism (Miller et al., 2021)

Note: Participants did not see the subscale labels. Participants were given the following instructions: Please rate each of the following statements in terms of how much each contributes to the problem of racism in the United States today on a scale from 1 = (*Not at all*) to 7 = (*Very much*).

Systematic Oppression

1. A history of policies that systematically disadvantaged People of Color
2. A history of racial inequality that perpetuates racial problems
3. Discrimination that is built into our laws and institutions
4. A society and culture that perpetuates the idea that White people are superior to People of Color

Individual Acts of Prejudice

5. Individuals' own beliefs and prejudices that cause them to treat those of other races poorly
6. Discrimination that is based on the prejudices of individual people
7. Intentional acts of racial discrimination and abuse by racist individuals
8. Individuals' beliefs about White racial superiority

Bias Awareness (Perry et al., 2015)

1. Even though I know it's not appropriate, I sometimes feel that I hold unconscious negative attitudes toward people based on their social groups (e.g., race, gender)
2. When talking to people, I sometimes worry that I am unintentionally acting in a prejudiced way
3. I worry that I have unconscious biases toward some social groups

4. I never worry that I may be acting in a subtly prejudiced way toward people based on their social groups (R)

Demographics

- What is your age?
- What is your current gender identity? (Please check all that apply)
    - Female
    - Male
    - Genderqueer
    - Genderfluid
    - Transgender
    - Transgender Feminine
    - Transgender Masculine
    - Agender
    - Androgynous
    - Two-Spirit
    - Demigender
    - Questioning or unsure
    - I prefer to describe my gender using my own language:
    - I prefer not to disclose
- I identify as:
    - White
    - Black
    - Hispanic/Latino/a/x

    - Asian
    - Indigenous
    - Pacific Islander
    - Multiracial
    - I prefer to identify as:
    - I prefer not to disclose
- What country best represents your national identity? (*Note:* participants will be able to select from a drop-down menu listing countries)
- What is the highest level of education you have completed?
    - Less than high school
    - High school/GED
    - Some college
    - 2-year college degree
    - 4-year college degree
    - Masters degree
    - Doctoral degree
    - Professional degree (JD, MD)
- In which state do you currently reside? (*Note:* participants will be able to select from a drop-down menu listing states in the United States)
- Please indicate your overall political viewpoint on the scale below: (1 = Very Liberal, 7 = Very Conservative)

Honesty Check

The study is very important to us, and a considerable amount of time and effort has gone into creating this survey. As such, if for whatever reason you feel that you did not complete the survey carefully or accurately, it would be extremely helpful if you could let us know this now. Your response is anonymous and will in no way affect your compensation.

- I DID complete the survey carefully and accurately. Please include my responses in analyses.
- I DID NOT complete the survey carefully or accurately. Please exclude my responses from analyses.

Debriefing

Thank you for your participation!

The purpose of this study was to examine the extent to which people blame others for discrimination caused by implicit bias. The news article and scenario you read were not real and were made up for the purposes of this study. Research shows that people are often unaware of their implicit biases. However, some research suggests we can become aware of biases that would normally remain hidden to us by drawing attention to instances when stereotypes pop into our minds, or by receiving feedback from others about our behaviors. Further research suggests that by gaining awareness of our implicit biases, with some effort we can control how they affect our behavior.

Thank you for participating in this study. It would not be possible to continue psychological research without the help of individuals like you. If you have any questions or concerns about this research or are distressed as a result of the materials used in this study, you may contact Dr. Saucier at: saucier@ksu.edu. Thank you again for participating in this study.

Please click the Get Completion Code button below to receive your MTurk completion code and receive payment.

Appendix C - Exploratory Analyses

Study 1 Conceptualizations of Racism Higher-Order Interactions

Because systemic and individual conceptualizations of racism represent two distinct, but related, constructs, exploratory analyses were conducted to examine how these conceptualizations interacted with each other and with the bias attribution manipulation. Significant two-way interactions between systemic and individual conceptualizations were found for measures of perceived intent (b = 0.09, p = .005) and intentions to help the perpetrator correct her bias (b = -0.07, p = .004). At high levels of individual conceptualizations, greater intent was perceived as levels of systemic conceptualizations of racism increased, but this relationship was relatively flat at low levels of individual conceptualizations of racism (see Figure C 1). At low levels of individual conceptualizations, greater intentions to help the perpetrator correct her bias were perceived as levels of systemic conceptualizations of racism increased, but this relationship was relatively flat at high levels of individual conceptualizations of racism (see Figure C 1).

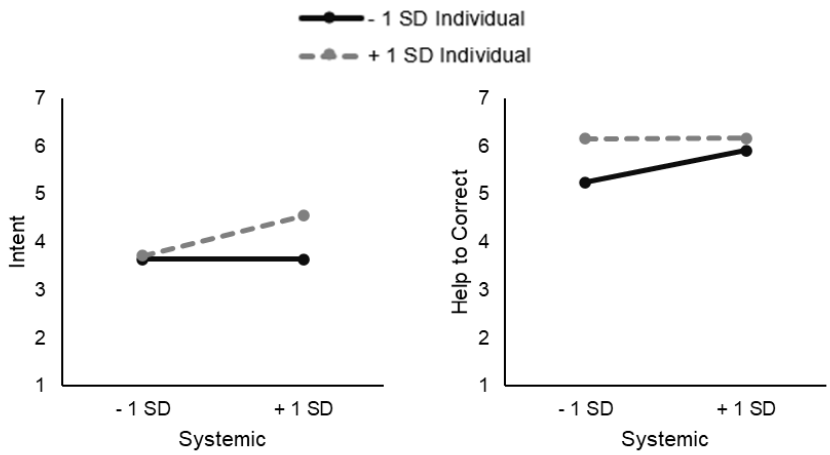
Additionally, several three-way interactions emerged between systemic and individual conceptualizations and the bias attribution manipulation (see Figure C 2). For perceived control (b = -0.15, p = .013), at low levels of individual racism, similar slopes were observed for systemic racism in the implicit and explicit bias condition, but at high levels of individual racism, there was a positive slope for systemic racism in the implicit bias condition and a negative slope in the explicit bias condition. In effect, there was less of a difference in perceived control between implicit and explicit bias at high levels of both systemic and individual racism compared to the other combinations.

For support for severe punishment (b = 0.14, p = .038), at low levels of individual racism, similar slopes were observed for systemic racism in the implicit and explicit bias condition, but

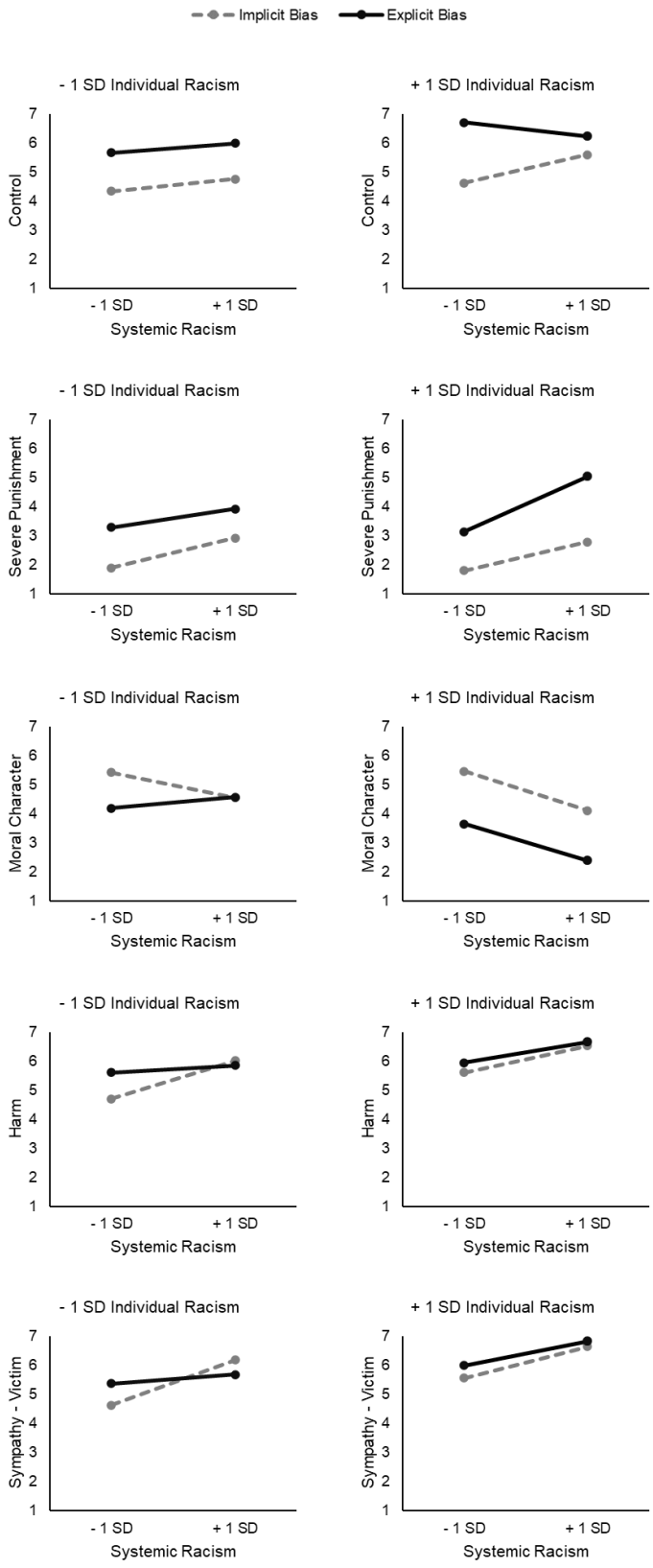
at high levels of individual racism, there was a more positive slope for systemic racism in the explicit bias condition than the implicit bias condition. In effect, there was more of a difference in support for severe punishment between implicit and explicit bias at high levels of both systemic and individual racism compared to the other combinations.

For perceptions of the perpetrator's moral character (b = -0.13, p = .030), at low levels of individual racism, there was a negative slope for systemic racism in the implicit bias condition and a slightly positive slope in the explicit bias condition, but at high levels of individual racism, similar negative slopes were observed for systemic racism in the implicit and explicit bias condition. This pattern of interactions resulted in no perceivable difference in perceptions of moral character between implicit and explicit bias at low levels of individual and high levels of systemic racism.

For perceived harm (b = 0.10, p = .043), the greatest difference between implicit and explicit bias was at low levels of both systemic and individual racism, but similar at all other combinations. A similar pattern was observed for sympathy for the victim (b = 0.11, p = .035). A general trend overall was that high levels of both systemic and individual conceptualizations of racism resulted in the most punitive responses to discrimination attributed to explicit bias.



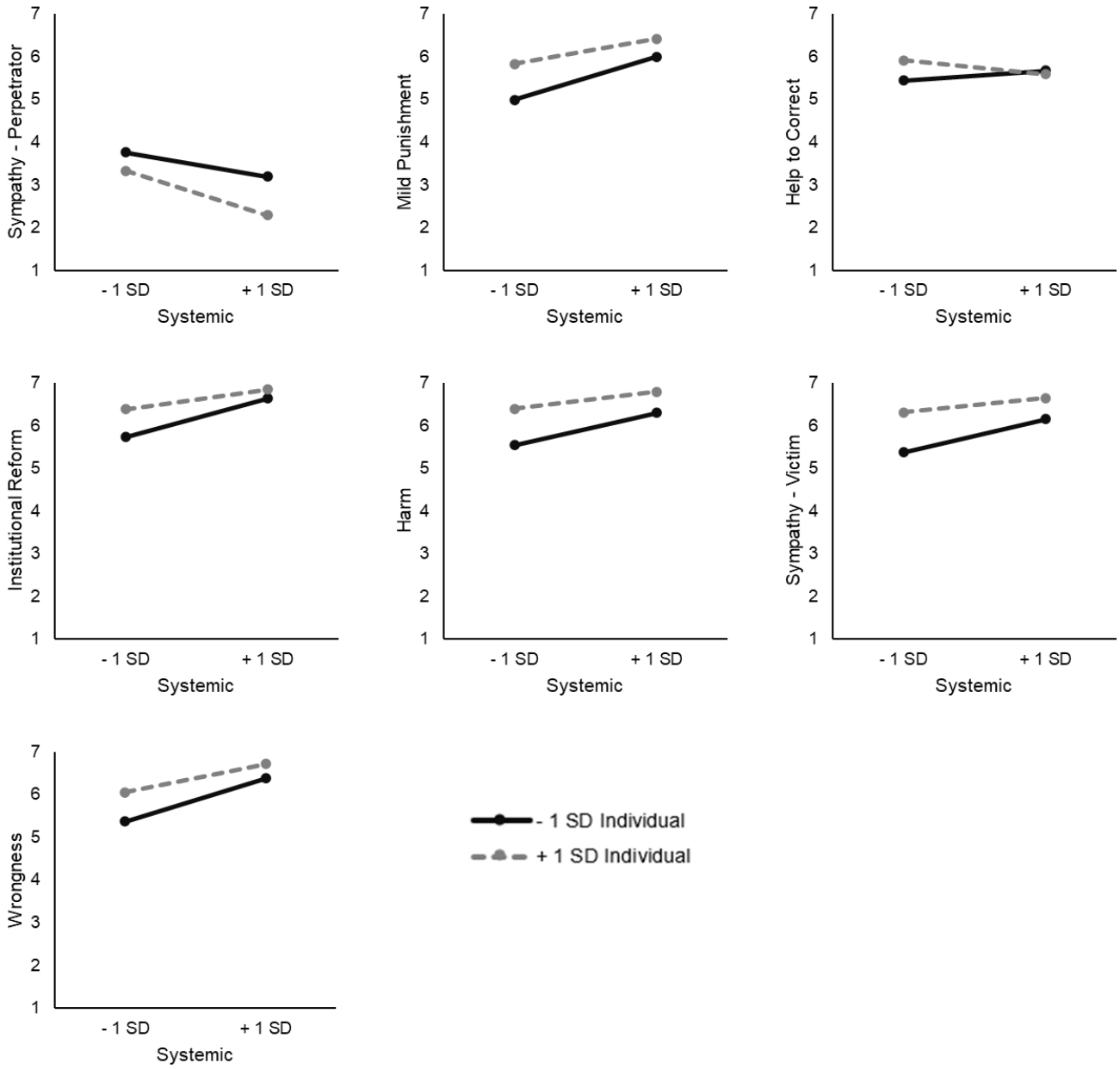
C 1. Interactions between systemic and individual conceptualizations of racism.



C 2. Three-way interactions between conceptualizations of racism and bias attribution.

Study 2 Conceptualizations of Racism Higher-Order Interactions

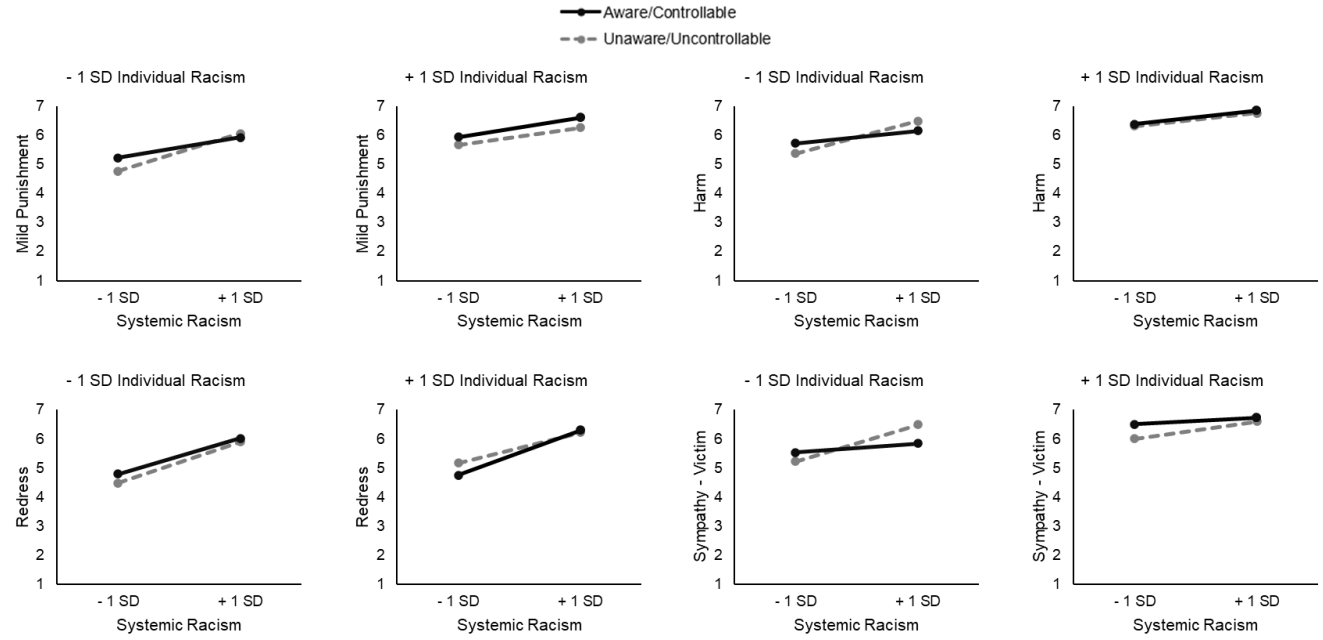
Several two-way interactions between systemic and individual conceptualizations that were not qualified by three-way interactions (see below) were significant in predicting sympathy for the perpetrator (b = -0.05, p = .027), mild punishment (b = -0.05, p = .015), help to correct (b = -0.06, p = .003), institutional reform (b = -0.05, p = .004), harm (b = -0.04, p = .007), sympathy for the victim (b = -0.05, p = .001), and overall wrongness (b = -0.04, p = .043). The overall pattern of these interactions generally showed the highest levels of punitive, and lowest levels of prosocial, responses at higher levels of both systemic and individual conceptualizations of racism.



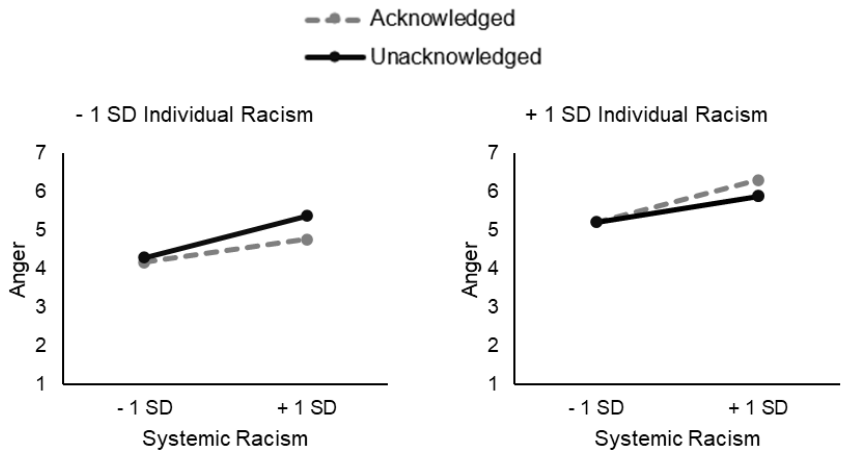
C 3. Interactions between systemic and individual conceptualizations of racism.

Additionally, significant three-way interactions between systemic and individual conceptualizations of racism and implicit bias framing were found in predicting mild punishment (b = -0.08, p = .046), harm (b = -0.08, p = .005), redress (b = -0.08, p = .020), and sympathy for the victim (b = -0.07, p = .029). Although these interactions were not particularly strong, the

general pattern was for the highest level of punitive, and lowest level of prosocial, responses at higher levels of both systemic and individual conceptualizations of racism when implicit bias was framed as conscious and controllable, compared to unconscious and uncontrollable (see Figure C 4). For anger, a significant three-way interaction between systemic and individual conceptualizations of racism and acknowledgement (b = -0.10, p = .016) showed that acknowledgement, compared to non-acknowledgement, decreased anger at lower levels of individual, combined with higher levels of systemic, conceptualizations of racism (see Figure C 5). Conversely, acknowledgement, compared to non-acknowledgement, increased anger at higher levels of individual, combined with higher levels of systemic, conceptualizations of racism.



C 4. Interactions between systemic and individual conceptualizations of racism and framing of implicit bias.



C 5. Interactions between systemic and individual conceptualizations of racism and perpetrator acknowledgement.