What’s the Harm? Perceptions and Experiences of Implicit and Intentional Bias

by

Carmelle Bareket-Shavit

B.S., The University of California Santa Cruz, 2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

The Faculty of Graduate and Postdoctoral Studies
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2023

© Carmelle Bareket-Shavit, 2023
The following individuals verify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

**What’s the Harm? Perceptions and Experiences of Implicit and Intentional Bias**

submitted by **Carmelle Bareket-Shavit** in partial fulfillment of the requirements for the degree of **Master of Arts** in **Psychology**

**Examining Committee:**

**Toni Schmader, Professor, Psychology, UBC**
Supervisor

**Andrew Baron, Professor, Psychology, UBC**
Supervisory Committee Member, Chair

**Kristin Laurin, Associate Professor, Psychology, UBC**
Supervisory Committee Member
Abstract

What does it feel like to be on the receiving end of a person’s implicit or intentional bias? There is an ongoing debate in academic and non-academic circles about the harm bias does. One view is that because implicit bias is perceived as less blameworthy (Daumeyer et al., 2019), it might also be experienced as less harmful. However, another view is that because impact is more critical than intention (Williams, 2020), implicit bias that has the same impact might hurt the same as (or even more than) intentional bias. Setting aside the question of whether implicit behavior is less wrong, I ask what is the experience of psychological harm of those with marginalized identities targeted by such behaviors? In a programmatic series of experiments, I ask whether implicit bias and intentional bias are experienced as similarly or differentially harmful. I use a theoretically derived approach to develop tightly controlled vignettes that manipulate two types of unbiased, two types of implicitly biased, and two types of intentionally biased behavior. I validate these scenarios for testing perceptions of stereotyping across five different social identity groups (Study 1, N = 205). I then ask whether vignettes of implicit and intentional gender bias are perceived as differentially harmful by women (Study 2, N = 302). Finally, I sought to replicate my findings by having people recall their experience of harm from lived experiences of implicit and intentional bias (Study 3, N = 138). Across these studies, both implicit and intentional bias (with equal impact) were more harmful than unbiased behavior (Study 2 - 3). When hypothetical instances of bias were intentional (vs. implicit), women anticipated more pain but similar levels of exclusion (Study 2). When recalling instances (controlling for impact), intentional bias was again experienced as more painful but similarly exclusionary as implicit bias (Study 3). Together, this research reveals that both forms of bias are perceived and experienced as harmful, but implicit bias was experienced as somewhat less painful than intentional bias. By understanding the consequences for the way bias unfolds, we may better mitigate the psychological harm from instances of bias.
Lay Summary
This work asks about people’s experiences of harm following incidents of biased behavior. When someone discriminates against us, if we know that person was not motivated to act with prejudice, does it hurt any less? Or do that person’s intentions make no difference in how harmed we feel? In our research we investigate these questions by having people respond to hypothetical scenarios or recalling their own experiences of discrimination and describing the amount of pain and feelings of exclusion that followed. We find that when attributing biased behavior to implicit processes, people reported less pain than when attributing those behaviors to intentional processes. However, people’s feelings of exclusion were similar following implicitly and intentionally biased behavior. Ultimately, these findings suggest that knowing a person’s motivations might help mitigate some of the pain we feel from their behavior, but not all the hurt.
Preface
This thesis is based on work conducted in UBC’s Social Identity Lab by Dr. Toni Schmader and me. I was responsible for designing materials, recruiting participants, conducting data analysis, and writing.

All projects and associated methods were approved by the University of British Columbia’s Research Ethics Board [certificate # H21-03842]
Table of Contents

Abstract .................................................................................................................................. iii
Lay Summary ........................................................................................................................ iv
Preface ..................................................................................................................................... v
Table of Contents ................................................................................................................ vi
List of Tables ........................................................................................................................ vii
List of Figures ......................................................................................................................... viii
Chapter 1: What’s the Harm? Perceptions and Experiences of Implicit and Intentional Bias ................................................................................................................... 1
  1.1 Defining Key Terms ......................................................................................................... 1
  1.2 Intentionality Informs Moral Judgments ...................................................................... 2
  1.2 Conceptual Ambiguity of What Bias Is ......................................................................... 3
  1.3 Impact Over Intention ................................................................................................... 4
  1.4 Overview of Current Research ..................................................................................... 6
Chapter 2: Study 1 Validating Stimuli of Implicitly and Intentionally Biased Behavior ................................................................................................................................. 8
  2.1 Stimuli Development ..................................................................................................... 8
  2.2 Participants .................................................................................................................... 11
  2.3 Protocol ........................................................................................................................ 11
  2.4 Measures ....................................................................................................................... 11
  2.5 Results .......................................................................................................................... 11
  2.6 Discussion .................................................................................................................... 17
Chapter 3: Study 2 Perceptions of Harm from Hypothetical Incidents of Implicit and Intentional Bias ................................................................................................................. 18
  3.1 Participants .................................................................................................................... 18
  3.2 Protocol ........................................................................................................................ 18
  3.3 Measures ....................................................................................................................... 19
  3.4 Main Results ................................................................................................................ 19
  3.5 Exploratory Analysis: Dimension of Bias and Harm .................................................... 24
  3.6 Discussion .................................................................................................................... 25
Chapter 4: Study 3 Recalled Harm from Real-World Incidents of Implicit and Intentional Bias .......................................................................................................................... 27
  4.1 Participants .................................................................................................................... 27
  4.2 Protocol ........................................................................................................................ 27
  4.3 Measures ....................................................................................................................... 28
  4.4 Results .......................................................................................................................... 29
  4.5 Discussion .................................................................................................................... 35
Chapter 5: General Discussion ............................................................................................... 37
References ............................................................................................................................... 41
Appendix ................................................................................................................................. 44
List of Tables

Table 1 Example of Stimuli Development .................................................................................. 9
Table 2 Coder Categorization of Recalled Incidents .................................................................. 29
Table 3 Manipulation Checks of Recalled Incidents .................................................................. 31
List of Figures

Figure 1 Perceived Motivation for Bias Types ................................................................. 12
Figure 2 Perceived Awareness for Bias Types ................................................................. 14
Figure 3 Perceived Regulation for Bias Types ................................................................. 15
Figure 4 Perceived Pain for Unbiased, Implicitly Biased, Intentionally Biased Behavior .... 21
Figure 5 Perceived Exclusion for Unbiased, Implicitly Biased, Intentionally Biased Behavior ................................................................. 21
Figure 6 Perceived Pain for Bias Types ........................................................................... 23
Figure 7 Perceived Exclusion for Bias Types ................................................................... 24
Figure 8 Experienced Pain for Unbiased, Implicitly Biased, Intentionally Biased Behavior ......................................................................................... 24
Figure 9 Felt Exclusion For Unbiased, Implicitly Biased, Intentionally Biased Behavior .... 33
Figure 10 Felt Pain and Exclusion from Bias Incidents, Controlling for Outcome Severity ... .................................................................................................................... 34
Chapter 1: What’s the Harm? Perceptions and Experiences of Implicit and Intentional Bias

When we experience someone discriminating against us, it hurts. Even though we are generally resilient in the face of discrimination (Crocker & Major, 1989), someone treating us not for who we are, but by the stereotypes they hold about our group can be psychologically and physically distressing (Pascoe & Richman, 2009). But are all types of bias similarly harmful? When we think about bias, what may first come to mind are intentional instances in which a person knowingly expresses stereotypes that they are unmotivated and/or unwilling to regulate. Take, for example, a situation in which a coworker openly holds gender stereotypes and knowingly expresses them by assigning women to administrative and men to technical roles. Such an intentional act of gender bias is harmful, not only for the objectively negative outcome it yields, but also in the feelings of pain, exclusion, anger, and anxiety it provokes. Would such an instance of bias be any less painful if we knew our coworker was implicitly biased? If we take the same incident, with the same outcome, but instead know our coworker meant to do well by us and wanted to treat us fairly, would that hurt any less, or just as badly? In three studies with varying methodologies, I ask whether bias resulting from implicit processes is less, equally, or more psychologically harmful than bias resulting from intentional processes.

1.1 Defining Key Terms

First, I clarify what I mean by implicit bias, considering ongoing debates about the definition of this phenomenon (Corneille & Hütter, 2020, Gawronski et al., 2022, Schmader et al., 2023). Some may think about implicit bias as static stereotypes and attitudes people hold in their heads outside their awareness and beyond their control. I instead adopt the view that bias is the outward behavior we show by the way we process our stereotypes and attitudes, which has a negative impact for a person of that targeted social group. The way we internally process and act on those stereotypes and attitudes may be implicit or intentional. We distinguish implicitly from intentionally biased behaviors by a perpetrator’s in-the-moment motivation. Motivation is an element of intentionality that describes a person’s goals, desires, and needs; it is what they want to happen (Carlson et al., 2022).

While others have distinguished intentional from implicit bias by focusing on an actor’s (lack of) awareness of how stereotypes and attitudes affect their actions (e.g., Gawronski et al., 2022), I adopt Schmader and colleagues’ (2022) theoretical approach to bias that centers on the actor’s egalitarian motivation. Specifically, I define implicit bias as occurring when a person is motivated to be egalitarian but engages in disparate treatment of another person(s) either because they were unaware of how stereotypes could shape their behavior (unconscious bias) or because they were unsuccessful in regulating the effect of these stereotypes on their behavior (unintentional bias). As I will describe more later, this definition of implicit bias allows me to test the consequences of intentionality not as a lack of awareness, but as actions that align with intrinsic motivation.

Second, I clarify my focus on harm as psychological pain and social exclusion, though consequences of bias can be operationalized in many ways. My goal in this research is to better understand people’s experiences of psychological harm, and thus I aimed to assess both physical and social pain (i.e., exclusion, rejection; Bieri et al., 1990; Hicks et al., 2001; Nordgren et al., 2011; Williams et al., 2009) in cases where the objective negative outcome a person experiences (e.g., denied a job offer) was the same. My focus was not on the actual outcome of a perpetrator’s biases, but on the psychological harm a person anticipates and experiences from
that differential treatment as a function of believing it has resulted from a perpetrator’s implicit or intentional processes. When we control for the outcome of a person’s biased behavior, does it hurt less, the same, or more if they were acting implicitly than if they were acting intentionally?

There is ongoing discussion in academic and non-academic circles about the role of intentionality in the harm bias does. From one theoretical perspective, intentionality matters for blameworthiness and harm; from another perspective, only outcome and not intention determines experienced harm. The goal of my research is to disentangle these perspectives by revealing that they ask theoretically distinct questions that are not necessarily mutually exclusive.

1.2 Intentionality Informs Moral Judgments

One perspective on the harm from implicit bias can be found by extrapolating from moral psychology. Moral psychologists have been at the forefront of research on the role of intentionality in harm done, more often focusing on the judgments of wrongdoing as opposed to the experience of harm. When making judgments about wrongness and culpability for harmful actions, a person’s intentions are critical for third-party judgments (Carlson et al., 2022; Cushman, 2015; Knobe & Nichols 2011). In one example, Young and colleagues (2007) described a person who poisoned their friend by putting a substance in his/her drink and thinking it was harmless sugar (accidental) or toxic powder (intentional). Though the outcome remains the same (i.e., the friend died from drinking the contaminated beverage), intentional actions were judged as more blameworthy and deserving of punishment than unintentional ones. Similarly, in work by Ames and Fiske (2013), participants wanted to blame, condemn, and punish a CEO who intentionally (rather than unintentionally) lost his employees’ money. A basic principle of moral psychology is that actions that are accidental are judged less harshly than those which result from intentional processes.

Consistent with this evidence that intentional actions are more blameworthy, other work suggests that intentional harms are also experienced as more (physically) painful than accidental harms. Gray and Wegner (2008) investigated the experiences of pain for intentional or accidental actions by conducting an experiment in which electric shocks were administered to participants who thought the confederate chose the shocks (intentional) or that the selection of the shock task was not in their control (accidental) and found higher reported pain in the intentional than in the accidental condition. In a follow-up experiment, Gray (2012) added a condition in which the confederate was described as having good intentions (benevolent): the confederate was said to have believed each shock would give the participant an extra lottery ticket. Compared to those shocks that were maliciously intentional or accidental, those that were benevolently intentional were experienced as less painful. Participant’s experiences of physical pain were informed by how they perceived the confederate’s intentions (or motivation toward the target): participants reported more pain if they believed the confederate was motivated to cause them harm, compared to one who meant to do good but accidentally caused harm. Guided by these findings in the moral literature on judgments of wrongdoing and reported physical pain, we might expect incidents of socially biased behavior to be more psychologically harmful were they intentional. Although, as this work does not directly compare such incidents to implicit actions, it remains an open question.

We believe that there are key differences between accidental harms and implicit ones. In the moral psychology literature, intentional actions (i.e., a desire for a resulting negative outcome) are often compared to accidental ones (i.e., no desire for a negative outcome that occurs against the perpetrator’s decision). To our knowledge, this literature has not made the
same comparison to implicit actions which are still enacted somewhat deliberatively. It is our opinion that implicit actions are theoretically distinct from accidental ones. Implicit actions occur when something internal to the agent has a causal influence on their behavior and the negative outcome. A person who behaves implicitly is motivated by their values and yet they still cause a negative outcome by their mental states, whether that be by their inability to recognize their stereotypes or attitudes, or their inability to suppress them. By contrast, the mental-states of a person who acts accidentally (especially in the paradigms used in moral psychology), has no bearing on the resulting outcome. Though the role of intentionality in determining wrongdoing of a person whose actions are implicitly or intentionally harmful is of broad interest in moral psychology and the law, it cannot be easily mapped on to our discussion of the psychological harm that implicit bias does. Moreover, this literature focuses on certain transgressions (e.g., stealing, poisoning) that are not directly related to stereotypes or prejudicial attitudes; In our research we ask about the harm from socially biased behavior. The causal force of stereotypes and/or prejudicial attitudes might lead people to experience implicit bias differently than accidental harm that has no internal cause.

Together, this body of work on third-party judgments of moral transgressions suggests that intentional actions are experienced as more harmful than accidental ones. However, it remains an open question whether intentionally biased actions are less, equally, or more harmful than implicit ones. Before further examining the harm from transgressions that are caused by negative stereotypes or prejudice, we need clarity on what we mean by intentionally and implicitly biased behaviors.

1.2 Conceptual Ambiguity of What Bias Is

While there are diverging perspectives on what implicit and intentional behavior is, in this research we use a theoretically derived approach. A common theme in bias literature is to define implicit bias as relating to awareness: something that occurs when a perpetrator is unaware that they hold racist beliefs or stereotypes, for example (Cameron et al., 2010; Daumeyer et al., 2019; Redford & Ratliff, 2016). In a theoretical advance in the bias literature, we take a more complex view of implicit bias by adopting Schmader and colleagues (2022) typology which focuses instead on egalitarian motivation. While I previously introduced our perspective, here I provide a more detailed explanation of our theoretical approach to implicitly biased behavior.

Our focal distinction between implicitly and intentionally biased treatment of another person is whether their behavior aligns with their egalitarian motivation: the negative outcome aligns with their lack of egalitarian motives (is intentional) and that which does not align with one’s egalitarian motives (is implicit). Here, implicitly biased behavior happens when a perpetrator wants to be fair and egalitarian but the way they treat another person contradicts those motivations. While our key distinction between implicit and intentional is the perpetrator’s motivation, we also take into account their awareness and their efforts to regulate their stereotypes and prejudicial attitudes. We distinguish between forms of intentional and implicit bias by identifying the profile of internal processes by which experiences of bias unfold: a process that is shaped by an actor’s in-the-moment motivation to be egalitarian, awareness of their bias, and effort to regulate their behavior. This typology provides us with two forms each of unbiased behavior (i.e., authentically unbiased, regulated unbiased), implicitly biased behavior (i.e., unintentionally biased, implicitly biased), and intentionally biased behavior (i.e.,
apathetically biased, hostilely biased). In the present work, however, we will focus mainly on the broader distinction between implicit and intentional bias.

Using this framework, we represent implicitly biased behavior as disparate treatment that occurs when a well-meaning person is unaware of how their stereotypes affect their behavior and so do nothing to control them (unconscious), and as when a well-meaning person is aware they hold stereotypes but unaware of how those stereotypes influence their behavior and ineffectively try to control them (unintentional). In contrast, intentionally biased behavior is disparate treatment that results when a non-egalitarian person knows they hold and are affected by stereotypes and treat someone unfairly either through inaction (apathetic) or directed efforts to upregulate those stereotypes on their behavior (hostile).

We take this approach to implicitly and intentionally biased behavior by first focusing on an agent’s motivation, but then also considering their awareness and regulation efforts. By doing so, we can build out the degree of intentionality through multiple manipulations and represent forms of biased behavior as they exist in the bias literature. We believe this approach allows us to better understand whether bias resulting from implicit processes is less or equally psychologically harmful than bias resulting from intentional processes.

1.3 Impact Over Intention

In contrast to the view that intentional wrongdoings are more harmful, is the assertion that the impact of one’s actions matters more than the intentions behind those actions. As work in moral psychology shows an intentional actor is judged more harshly than an unintentional actor (Ames & Fiske, 2013; Cushman, 2015), we might expect psychological harm to be similarly informed by intentionality. My review of this literature suggests that most have focused on people’s judgments of moral blameworthiness and responsibility, as opposed to their emotional reactions to being targeted by such actions. It remains an open question whether a perpetrator’s intentions factor into the hurt we feel when we experience bias by them.

The way we judge actions may be sensitive to information about intentionality, but that might not be the case for the way we feel after being on the receiving end of those actions. As some moral psychologists note, we approach judgments differently when they are about a person or about a person’s actions (Carlson et al., 2022; Cushman, 2015). By the same token, if we are on the receiving end of those actions, we might not put the same weight behind a perpetrator’s intentions as we would if we were evaluating them as an outside observer. When evaluating people and behaviors, we are sensitive to intentionality in a way that we might not be, were we considering the impact of their actions alone.

There are circumstances in which implicit actions are judged as equally harmful to intentional actions (Bloom, 2021, Paradkar, 2019). When it comes to behaviors that are considered taboo, for example, it does not matter whether it was intentional or accidental: judgments of wrongness are informed by the outcome alone (Young & Saxe, 2009; 2011). Thus, even though intentionality might be relevant to moral judgments of blameworthiness, intention might be irrelevant to affectively charged experiences of psychological harm.

Previous literature on microaggressions, attributional ambiguity, and awareness of bias, provides insight into how people might respond to different forms of biased behavior. While these literatures do not all consider intentionality, nor the same forms of bias, they provide us with important context for our research on psychological harm by intentionality.

First, research on microaggressions would seem to suggest that the outcome of biased behavior matters more than the intention of the actor (Lui & Guezada, 2019; Sue et al., 2007;
Wang et al., 2011; Williams, 2020). Microaggressions are everyday subtle behaviors that, regardless of whether they are intentional or unintentional, lead to negative outcomes for a person targeted by them (but also see Freeman & Stewart, 2021 for a discussion on the fuzziness of the construct). In the microaggression literature, the focus is on the person who was harmed and mitigation of that harm, rather than the perpetrator’s intentions. From this perspective, once harm is done harm is felt. Harm is often assumed to be just as damaging if it was done implicitly than if it was done intentionally. Korman and colleagues (2022), for example, showed that behaviors were considered more offensive, blameworthy, and intentional for microaggressions (i.e., asking the woman in a lab-coat when the doctor would come in) than for a similarly ambiguous faux pas (i.e., asking the woman in exercise clothes when the doctor would come in). From this work, a negative action that stems from stereotypes and attitudes is seen as harmful, regardless of the ambiguous intention. It remains to be seen, however, whether a more direct test of a perpetrator’s motivations plays a role in the experience of their prejudice. While the microaggression research may lead us to posit that implicit actions are similarly harmful to intentional ones, other research suggests implicit actions are potentially even more harmful.

Research on attributional ambiguity suggests that when we experience a negative outcome (resulting from another’s actions), we might try to make sense of it by making attributions to what might have caused it (Crocker & Major, 1989; Major & Dover, 2016; Major et al., 2003, 2016; Schmitt et al., 2014). For any given incident, however, it is often ambiguous whether the negative outcome has been caused by another’s underlying stereotypes or prejudicial attitudes. In such cases, we can make an internal attribution, such that we attribute the negative encounter to our own personal failings, or they may be external, such that we attribute the negative outcome to the other person’s unfair or biased treatment of us. When we can externally attribute a negative outcome to discrimination, it may help protect our self-esteem from negative feedback that might otherwise lower our self-esteem (Major et al., 2016).

What does work on attributional ambiguity lead us to conclude about felt psychological harm from incidents in which the outcome was clearly caused by another person’s underlying stereotypes, whether that be implicit or intentional? In this literature, an incident that is more ambiguous, compared to that which was clearly biased, can be more harmful given the increased rumination that can result from trying to decide if the action was biased or not (Banaji et al., 2015; Major et al., 2003; Salomon et al., 2015). In one study, Major and colleagues (2016) found that Latinx women who were more suspicious of a White person’s feedback showed a more negative psychological (i.e., stress, uncertainty, decreased self-esteem) and physiological (i.e., cardiovascular reactivity) response. Though the attributional ambiguity work focuses not on implicit, but more ambiguous behavior, they may be closely related. People’s experience of harm may be magnified if they experience implicit bias as more attributionally ambiguous than intentional bias.

The research I pose here is related to the attributional ambiguity literature, but different in important ways. To our knowledge, there is little research in the attributional ambiguity space on the role of the perpetrator’s intentions and motivations in experience of pain, invisibility, and exclusion following an incident of bias. I ask if an implicitly biased perpetrator’s behavior is less or equally as harmful as an intentionally biased one, while the attributional ambiguity literature more often focuses on situational ambiguity and implications for self-esteem and physiological threat. Considering the theoretical precedent of the attributional ambiguity literature, it may very well be that implicit behavior, by leading to increased internalization, may be experienced as
more harmful than intentional behavior, which can be more clearly externalized as discrimination.

Finally, in the bias literature, two lines of research suggest that intentional acts of bias could be more harmful (although these papers focus on moral judgments as opposed to psychological harm). Daumeyer and colleagues (2019) investigated moral judgments of news articles that described instances of discrimination (e.g., a police officer is described as handcuffing or detaining racial minorities) and found that explicit racial aggression towards racial minorities (where the officer was aware they were treating them differently) was judged to be more blameworthy than implicit racial aggression (where the officer was not aware they were treating them differently). Participants judged the perpetrator less harshly if their actions stemmed from a lack of awareness of their own racial biases (but there was no measure of experienced harm from targets themselves). In other research, Swim and colleagues (2003) showed that participants were more likely to judge a man as prejudiced and his behavior as discriminatory when he was described as intentionally (i.e., aware of his bias and a desire for a negative outcome) treating a woman differently, compared to when he was described as moderately intentional (i.e., acting out of habit) or without intention. Also, when there was no information about the man’s intent provided, participants judged his behavior based on the amount of harm the woman was said to have experienced (i.e., his actions were judged as more discriminatory when the woman reported experiencing high levels of harm). These works suggest that intentionality plays a role in how we consider incidents of bias. However, our work focuses not on moral judgments, but on the psychological experience of harm by a person targeted by that behavior. Also, our work makes a theoretical advance in distinguishing implicit from intentional behavior primarily by a perpetrator’s motivations, not their awareness.

1.4 Overview of Current Research

In a series of programmatic studies, I ask whether bias resulting from implicit processes is less, equally, or more psychologically harmful than bias resulting from intentional processes. In doing so, I seek to provide a better understanding of the role of intentionality and egalitarian motivation in the psychological experience of harm from bias.

In Study 1, I used a theoretically derived approach to develop tightly controlled vignettes for unbiased, implicitly biased, and intentionally biased behavior and included scenarios involving stereotypes about five different social identity groups and validated those stimuli with preregistered hypotheses of manipulation checks. In Study 2, I asked whether these vignettes of implicit and intentional gender bias were perceived as differentially or similarly harmful with preregistered hypotheses. In Study 3, I sought to replicate my findings using a method where people recalled their own lived experiences of implicit and intentional bias.

I measured psychological harm in two distinct ways: reported pain as well as reported feelings of exclusion. In the scenario studies, I keep the objective negativity of the outcome controlled to better isolate effects on the psychological experience of a person targeted by an incident of biased behavior. To our knowledge, most research has focused on third-party judgments of wrongdoing and blame, with few studies measuring harm as self-reported pain following intentional bias (Gray & Wegner, 2008; Gray, 2012) and as negative affect following general attributions to discrimination (i.e., depression, Remedios et al., 2012; self-esteem, Major et al., 2003). In my research, I focus on the perception and experience of psychological harm following an incident of implicit and intentional bias.
Past work leads us to expect that implicit (and less intentional) forms of biased behavior will be perceived and experienced as less harmful than intentional forms of biased behavior, albeit still harmful. Alternatively, other work might lead us to expect that implicit actions are just as or even more harmful than intentional ones, although there is some more fuzziness in this literature that leads us to prioritize the former over the later hypotheses in our preregistrations. Ultimately, I use a theoretical approach for what bias is and varied methodologies to ask whether bias resulting from implicit processes is less or equally psychologically harmful than bias resulting from intentional processes.
Chapter 2: Study 1
Validating Stimuli of Implicitly and Intentionally Biased Behavior

2.1 Stimuli Development

My primary aim was to first validate vignette stimuli depicting an interaction between a perpetrator and a recipient of bias. I developed stimuli according to the framework provided by Schmader and colleagues (2022) and semi-orthogonally manipulated three dimensions of the perpetrator’s internal mindset, described below in Table 1. In doing so, I created five bias type stimuli to reflect: regulated unbiased behavior, two forms of implicit bias (unintentional, unconscious) and two forms of intentional bias (apathetic, and hostile). I developed these bias type stimuli across five different biased identity scenarios to depict different instances of stereotyping (e.g., based on disability, ethnicity, gender, race, or sexual orientation) to allow for generalizations across social groups.

Each scenario was described as taking place within an organization (i.e., school project, workplace, rental application) and in an intergroup context, where the objective negative outcome was that the recipient lost out on an opportunity (i.e., did not get to participate in the assignment, did not get a shift covered, did not get the rental offer). Although I was interested in ensuring consistency across bias identity scenarios, our primary goal in this study was to create stimuli effectively representing each distinct type of biased behavior by communicating the perpetrator’s in-the-moment (non) egalitarian motivation, (lack of) awareness of having stereotypes and of their impact on their behavior, and how much (if any) effort was put in to regulating their behavior. In Table 1, I outline how each dimension was manipulated to create each bias type stimuli. Both intentional bias types described a non-egalitarian perpetrator who was aware that they had stereotypes about the recipient’s social group, and either did nothing to prevent their unfair treatment of the recipient (apathy) or put in more effort to ensure the unfair treatment of the recipient (hostile). Both implicit bias types described an egalitarian perpetrator who was either unaware of their stereotypes and so did nothing to prevent them (unconscious), or aware they had stereotypes but ineffective in regulating their effect (unintentional), both leading to the unfair treatment of the recipient through implicit processes. In this first study, I developed one form of unbiased behavior, where the egalitarian perpetrator was aware of their stereotypes and able to effectively regulate them, leading to a positive outcome for the recipient (e.g., they received the rental offer).
Table 1  
*Example of Stimuli Development for Gender Bias Identity Scenario*

Across all bias type stimuli, introduction was given:  
“Adam holds some stereotypes about women in computer science, which have the potential to impact his behavior….

<table>
<thead>
<tr>
<th>Unbiased Behavior</th>
<th>Egalitarian Motivation</th>
<th>Awareness</th>
<th>Effort to Regulate</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regulated</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Down</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Adam really wants to be fair and unbiased to Tamra and he sees the value of encouraging women to pursue computer science.</td>
<td>But Adam is aware that he holds some gender stereotypes that may affect his behavior</td>
<td>and so, he puts effort into preventing his stereotypes from affecting his behavior. Adam makes an active effort to Tamra to design the game so that she is involved in the technical parts of the project and that she takes a lead in stress testing the game</td>
<td>Tamra is able to develop those skills that are important for programming</td>
</tr>
<tr>
<td>Implicit Bias</td>
<td>Yes</td>
<td>Yes</td>
<td>Failed</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Unintentional</strong></td>
<td>[same as unbiased]</td>
<td>[same as unbiased]</td>
<td>Adam makes an active effort to include Tamra and to collaborate with her on ways to design the game. But he suggests to Tamra that because she would do a better job putting together the aesthetics of the game board, she should focus on that while he and the other guys work on the technical parts of the project and stress testing the game.</td>
<td>Tamra doesn’t get to develop those skills that are important for programming</td>
</tr>
<tr>
<td>Unconscious</td>
<td>Yes</td>
<td>No</td>
<td>None</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>[same as unbiased]</td>
<td>But Adam is not really aware that he holds some gender stereotypes that may affect his behavior</td>
<td>and so, he puts no effort into preventing his stereotypes from affecting his behavior. Adam and the other guys worked on the technical parts of the project and stress testing the game, while Tamra ended up focusing on the aesthetics of the game.</td>
<td>[same as unintentional]</td>
</tr>
</tbody>
</table>
Adam doesn’t really care if he’s fair or unbiased to Tamra and he does not see the value of encouraging women to pursue computer science.

Note. Provides the semi-orthogonal manipulations of the perpetrator’s in-the-moment egalitarian motivation (yes/no), awareness of bias (yes/no), and effort to regulate (down, failed, none, up), leading to the same objectively negative outcome for the four bias type stimuli (unintentional, unconscious, apathetic, hostile) and a positive outcome for the unbiased type (regulated). We used these manipulations to create bias type stimuli across five different biased identity scenarios to depict different instances of stereotyping (e.g., based on disability, ethnicity, gender, race, or sexual orientation). Here we provide examples of the gender bias scenario, which were also used in Study 2.
2.2 Participants

A power analysis with .90 power, .05 alpha, and a $d = 0.25$ effect size suggested a sample size of 250 was needed. After excluding responses from 44 participants who failed a preregistered attention check, our final sample included 205 participants from the Amazon Mechanical Turk general sample in the United States, with a mean age of 39 ($SD = 12$). Participants were predominantly White (69% White, 26% Black, 7.8% Hispanic, 7.3% East Asian), able-bodied (86% Able-bodied, 14% Disabled), and heterosexual (86% Heterosexual, 3.5% Gay/Lesbian, 10% Bisexual/Pansexual), and generally identified as men (51%) or women (46%), with 1.5% identified as Non-Binary or preferred not to answer (1%).

2.3 Protocol

After providing informed consent, participants were told they would read and evaluate a series of stories of possible bias along several dimensions, in order to pilot test these scenarios for future studies. The stimuli were structured in a 5 (bias type) x 5 (biased identity) design. To avoid fatigue and contrast effects, participants responded to only five scenarios (in random order) such that each participant saw one scenario of each bias type and one scenario of each biased identity. Following each scenario, participants were asked to respond to a series of manipulation checks, and after the fifth scenario, were debriefed.

2.4 Measures

Participants responded to several focal manipulation checks to ascertain that my stimuli mapped on to the typology developed by Schmader and colleagues (2022). Unless otherwise mentioned, all manipulation checks were asked on 7 pt. likert scales with higher scores indicating higher ratings of that construct. **Egalitarian motivation** was measured with the item: “How much was [Perpetrator name] motivated to be fair and unbiased to [Recipient name]?”. **Awareness** of bias was measured with the item: “How aware is [Perpetrator name] that [he/she] holds stereotypes?” **Effort to regulate** their behavior was measured with the item: “Did [Perpetrator name] make an effort to control how [his/her] stereotypes influenced [his/her] behavior?”), with higher scores reflecting more effort made to use stereotypes, lower scores reflecting more effort made to prevent stereotypes, and the midpoint as no effort made at all.

Secondary questions included how biased was the perpetrator (i.e., “To what degree does [Perpetrator name] hold stereotypes about [Recipient name]’s group regardless of whether they were expressed in the moment?”), and how negative was the outcome for the recipient (i.e., “Objectively, did [Recipient name] experience a positive or negative outcome as a result of [Perpetrator name]’s behavior?”). These manipulation checks were included in all subsequent studies.

In my preregistration (https://osf.io/sjncf/?view_only=041d23785c55439798e6e7749e397ac0), I describe my primary hypotheses as centering on the ratings of the perpetrator’s egalitarian motivation, awareness of bias, and efforts to regulate their behavior across the different bias identity types. In this preregistration I also describe secondary hypotheses that test dependent variables of how biased the perpetrator was and how negative the outcome was. To note, I did not have predictions about differences on any dependent variables pertaining to bias identity scenario but did plan to explore such analyses to be able to identify any inconsistencies across stimuli before proceeding.

2.5 Results

**Analysis Plan.** For the validation of the bias types, I conducted regression models for each manipulation check rating as the dependent variable with dummy coding to group the
conditions according to my manipulations in the vignette stimuli. I used this approach to compare my manipulations of a given dimension (i.e., egalitarian vs nonegalitarian motivation) to how participants perceived the perpetrator’s mental states (i.e., egalitarian motivation rating) to verify that my bias identity stimuli were perceived as distinct along the planned dimensions. In follow-up analyses, I conducted one-way ANOVAs with each manipulation check as the dependent variable and with the bias identity type as the predictor. If there was a main effect of bias type, I conducted post hoc analysis using Tukey’s HSD to consider all pairwise comparisons of the bias identity types to compare any differences that emerged to ensure as much consistency in my stimuli development as possible. I also conducted two-way type III ANOVAs with the manipulation check rating as the dependent variable and the bias type and bias identity scenario as the two predictors to consider differences between scenarios. As effects were replicated across the bias identity scenario (scenario was not a main effect and did not interact significantly with the bias types in the focal manipulation checks), we collapsed across the scenario in our analysis. In my preregistration, I described my focal interest as the manipulation checks of egalitarian motivation, awareness of bias, and effort to regulate, followed by a secondary interest in ratings of the outcome and of the perpetrator’s stereotypes.

**Egalitarian Motivation.** A dummy-coded regression comparing the egalitarian versus nonegalitarian bias types revealed a significant difference between conditions in perceived motivation. As hypothesized, egalitarian motivation was judged to be higher in the regulated, unintentional, and unconscious bias types ($M = 4.94$, $SD = 1.80$, $N = 615$) compared to apathetic or hostile bias types ($M = 2.05$, $SD = 1.67$, $N = 410$), $t(1023) = 26.0$, $p < .001$. To explore more nuanced distinctions, I conducted a one-way ANOVA that revealed a main effect of bias type $F(4, 1020) = 248$, $p < .0001$, that I decomposed with post-hoc Tukey HSD. As shown in Figure 1, there were no differences in the perceived motivation underlying apathetic bias ($M = 2.04$, $SD = 1.64$) and hostile bias ($M = 2.06$, $SD = 1.68$). Although the same egalitarian motivation information was provided in the other three conditions, participants judged egalitarian motivation to be highest for regulated bias ($M = 6.13$, $SD = 1.15$), followed by unintentional bias ($M = 4.73$, $SD = 1.53$), and lastly unconscious bias ($M = 3.97$, $SD = 1.92$), all of which were perceived as significantly more motivated than apathetic and hostile bias (as anticipated). These deviations from predictions might suggest that participants considered the vignettes comprehensively, with outcome and regulation coloring their judgments of intention. That said, these results suggest that our manipulations of the perpetrator’s motivation did overwhelmingly align with participants’ perceptions. A type III two-way ANOVA did not reveal a significant effect of the biased identity ($F(4, 1000) = 0.62$, $p = .62$) or of the interaction between bias type and biased identity ($F(16, 1000) = 1.31$, $p = .18$).

**Figure 1**

*Perceived Motivation for Bias Types*
Note. Green bars had identical information about the perpetrator being egalitarian, red bars had identical information about the perpetrator being egalitarian. Labels describe the pairwise comparisons of the five bias type conditions for egalitarian motivation rating. Both intentional forms of behavior were rated as less motivated than both implicit forms and the unbiased form of behavior, as anticipated. The regulated bias type was rated as more egalitarian than the two forms of implicitly biased behavior. The unintentional bias type was rated as more egalitarian than the unconscious bias type.

Awareness of Bias. A dummy-coded regression comparing the aware versus unaware bias types confirmed our predictions for awareness. The perpetrator was rated as significantly more aware of their stereotypes in the regulated, unintentional, apathetic, and hostile bias types ($M = 5.98, SD = 1.22, N = 820$) compared to unconscious bias condition ($M = 3.20, SD = 2.02, N = 205$), $t(1023) = 25.2, p < .001$. This same pattern was confirmed by an exploratory one-way ANOVA, $F(4, 1020) = 164, p < .0001$ and a post-hoc Tukey’s HSD with one small deviation. As shown in Figure 2, the perpetrator of unintentional bias ($M = 5.71, SD = 1.24$) was expectedly judged to be significantly less aware than perpetrators of regulated ($M = 6.07, SD = 1.15$),
apathetic ($M = 5.99$, $SD = 1.24$) and hostile bias ($M = 6.17$, $SD = 1.21$). Again, the unanticipated negative outcome of unintentional bias likely decreased these awareness ratings somewhat. Ultimately, these results suggest that our manipulations of the perpetrator’s awareness did align with participants’ perceptions. A two-way ANOVA did not reveal a significant effect of the scenario ($F(4, 1000) = 0.13, p = .97$) or of the interaction between bias type and scenario ($F(16, 1000) = 0.72, p = .77$).

**Figure 2**
*Perceived Awareness for Bias Types*

![Bar chart showing perceived awareness for bias types](image)

*Note.* Green bars had identical information about the perpetrator being aware of their stereotypes, red bars described the perpetrator unaware of their stereotypes. Labels describe the pairwise comparisons of the five bias type conditions for awareness rating. The unconscious bias type was rated as less aware than the other conditions, as anticipated. The unintentional bias type was rated as less aware than the regulated, apathetic, and hostile bias types.
**Effort to Regulate.** A dummy-coded regression comparing the upregulation versus no action versus downregulation bias types confirmed our hypothesis on regulation efforts. Participants rated the perpetrator having put in more effort to prevent stereotypes in the regulated and unintentional bias types ($M = 3.32, SD = 2.08$) than in the unconscious and apathetic types ($M = 4.32, SD = 1.38$), $t(1023) = -8.21, p < .0001$, and as putting in more effort to use stereotypes in the hostile bias types ($M = 5.12, SD = 1.68$) compared to the no effort bias types $t(1022) = 5.34, p < .0001$. An exploratory one-way ANOVA and pairwise comparisons with Tukey’s HSD revealed more nuanced differences, $F(4, 1020) = 157.5, p < .0001$ (see Figure 3). Although neither condition was described as including efforts to regulate stereotypes, the perpetrators of apathetic bias ($M = 4.56, SD = 1.49$) were judged to be upregulating the effect of their stereotypes on their behavior compared to perpetrators of unconscious bias ($M = 4.08, SD = 1.22$), which was in turn not significantly different from unintentional bias ($M = 3.87, SD = 1.66$). Because this latter null effect runs counter to our theoretical distinction between unintentional and unconscious bias, I revised our manipulations of the perpetrator’s efforts to regulate their stereotypes in the following studies. A two-way type III ANOVA did not reveal a significant effect of the scenario ($F(4, 1000) = 0.66, p = .66$) or of the interaction between bias type and scenario ($F(16, 1000) = 1.18, p = .27$).

**Figure 3**
Perceived Regulation for Bias Types
Note. Green bars had information about the perpetrator putting in positive effort, grey bars had identical information about the perpetrator putting in no effort, the red bar had information about the perpetrator putting in negative effort. Labels describe the pairwise comparisons of the five bias type conditions for regulation rating. The hostile bias type was rated as putting in more negative effort than the other conditions. The regulated bias type was rated as putting in more positive effort than the other conditions. There were no differences in ratings between unintentional and unconscious conditions. The apathetic bias type was rated as putting in more effort to upregulate stereotypes than regulated, unintentional, and unconscious conditions.

Stereotypes. All five bias type conditions included identical information about the perpetrator holding stereotypes (i.e., “Adam holds some stereotypes about women in computer science, which have the potential to impact his behavior”), and thus I did not predict differences across bias type conditions. However, a one-way ANOVA revealed a main effect of bias type $F(4, 1020) = 75.6, p < .0001$, suggesting that participants’ perceptions of the perpetrator incorporated other information that was provided in the scenario. Using post-hoc Tukey HSD, I found that the degree to which the perpetrator was judged to hold stereotypes was significantly
lower in the regulated bias condition ($M = 4.89$, $SD = 1.34$), compared to all other bias conditions (unintentional $M = 5.72$, $SD = 1.10$, unconscious $M = 5.37$, $SD = 1.35$, apathetic $M = 6.17$, $SD = 1.13$, and hostile $M = 6.40$, $SD = 0.89$). The degree to which the perpetrator was perceived to hold stereotypes was significantly higher in the apathetic and hostile bias types than in the other bias types. Ultimately, the perpetrator was judged as holding more stereotypes in the implicit bias conditions (unintentional, unconscious) than the unbiased condition (regulated), but less than the intentional bias conditions (apathetic, hostile). As information about the perpetrator’s stereotypes were consistent across bias types, I did not make changes to this dimension in following studies.

**Outcome.** Although not a focal manipulation, I did anticipate that the four bias type conditions (unintentional, unconscious, apathetic, hostile) where there was the identical objectively negative outcome (i.e., “Tamra doesn’t get to develop those skills that are important for programming”) would be seen as producing a more negative outcome than the regulated bias condition, in which there was a positive outcome (i.e., “Tamra is able to develop those skills that are important for programming”). However, in my preregistration I also noted that perceptions might vary if participants consider the vignettes comprehensively. A one-way ANOVA revealed a main effect of bias type $F(4, 1020) = 250, p < .0001$, and post-hoc Tukey HSD tests confirmed that participants rated the outcome as more positive in the regulated bias condition ($M = 6.00$, $SD = 1.24$), compared to all other bias conditions (unintentional $M = 2.61$, $SD = 1.24$, unconscious $M = 2.60$, $SD = 1.40$, apathetic $M = 2.47$, $SD = 1.60$, and hostile $M = 2.17$, $SD = 1.53$). However, ratings of the outcome were more negative in the hostile bias condition compared to the unintentional and unconscious bias conditions. As information about the outcome was consistent across bias types, I did not make changes to this dimension in following studies.

**2.6 Discussion**

In Study 1, I developed stimuli to reflect theoretically derived forms of biased behaviors by manipulating the perpetrator’s egalitarian motivation, awareness of bias, and efforts to regulate their behavior, while keeping their pre-existing stereotyping and the resulting outcome generally controlled. My analyses collapsing across the five biased identity groups suggest that our manipulations, in general, had the intended effects on participants’ perceptions. These effects were clearest when comparing categories of implicit to intentional forms of bias; fine-grained comparisons across all four bias types yielded more nuanced effects. Given these results, I made minor adjustments to the efforts to regulate manipulation, to be more in line with our theoretical distinctions between the bias types. We made no other changes to the manipulated dimensions. I also explored these manipulation checks within each bias identity scenario group (i.e., disability, ethnicity, gender, race, or sexual orientation bias) and generally found consistent patterns. After attaining my main goal of validating these bias type stimuli for the manipulated dimensions (i.e., egalitarian motivation, awareness of bias, effort to regulate) to be perceived as implicit and intentional bias, I identified the gender bias identity scenario as the one in which participants’ responses best mapped on to my manipulations. Study 2 thus uses the revised gender bias stimuli in measuring participants’ perceived harm following an incident of bias.
Chapter 3: Study 2
Perceptions of Harm from Hypothetical Incidents of Implicit and Intentional Bias

The findings from Study 1 suggested that the regulated, implicit, and intentional bias scenarios I created were perceived to differ as expected in the egalitarian motivation, awareness, and regulation ratings. Although some ratings were to some extent shaped by other elements of the manipulation and the severity of the outcome itself, these data support my primary goal of contrasting implicit against intentional bias using controlled vignette scenarios. In Study 2, I used the gender bias scenario, in particular, to examine women’s perceptions of harm to intentional vs. implicit bias. I predicted that when a person is on the receiving end of another’s biased behavior, whether that behavior was a result of intentional or implicit processes, it would be perceived as psychologically harmful. However, I did anticipate that the incidence of bias would be less psychologically harmful if they knew that person was implicitly, and not intentionally biased (preregistration can be found here: https://osf.io/m4fue?view_only=041d23785c55439798e6e7749e397ac0). My focus was on comparing the broader categories of unbiased, implicit, and intentional bias, with more nuanced secondary analyses between each bias type. In this study I measured people’s perception of psychological harm following hypothetical incidents of biased behavior, using the same bias identity type vignettes I validated previously in Study 1.

3.1 Participants
A power analysis with 0.90 power, 0.05 alpha, and suggested a sample size of 270 was needed to detect an effect size of 0.25. Anticipating similar exclusions as in the previous study, I recruited 320 women from the Amazon Mechanical Turk general sample in the United States. After preregistered exclusions, our final sample of 302 participants had a mean age of 41 (SD = 13), were predominantly white (74% White, 10% Black, 4.6% Hispanic, 5.3% East Asian, 2% South Asian), able-bodied (88% Able-bodied, 12% Disabled), and heterosexual (83% Heterosexual, 3.3% Gay/Lesbian, 11% Bisexual/Pansexual).

3.2 Protocol
To foster an experiential mindset, I used audio recordings to guide participants through meditative exercises, asking them to bring attention to their internal states, their physical state, and their surrounding environments, before proceeding with the hypothetical incident of bias. Participants were asked to imagine that they were in an interaction with another person described in the vignette. They were asked to imagine they had access to that person’s thoughts and feelings, and to try to visualize clearly and vividly what they might be thinking, feeling, and experiencing if they were in this situation. Then, participants were randomly assigned to read and respond to a gender bias scenario depicting one of the four types of bias or two types of unbiased behavior (I added a scenario describing authentically unbiased behavior, in which the perpetrator does not hold any stereotypes at all and there is a positive outcome for the recipient).

They then completed an open-ended response about their thoughts, feelings, and emotions to the incidence of bias (i.e., “What thoughts, feelings, and emotions would come to mind if you were the woman in this scenario?”) as well as measures of harm (pain and invisibility). These steps were taken to encourage participants to engage with the described incident, so that their responses throughout the survey would be as close as possible to how they would respond to the incident were it not hypothetical. Afterwards, participants responded to the same manipulation checks as in Study 1 as well as two measures of harm for psychological pain and felt exclusion and isolation.
3.3 Measures

Manipulation Checks. I used the same manipulation checks as in Study 1 to capture the perceived egalitarian motivation, awareness of bias, and efforts to regulate bias, as well as the perpetrator’s stereotypes and the outcome of the incident.

Pain. A two-item pain measure consisting of a general pain question and a face pain scale were used to assess perceptions of pain. For the general pain question, participants indicated how painful the scenario would be to them (i.e., “To what extent would you find this event painful?”) on a 5-point scale. The scale was scored from 0 (No pain) to 5 (Extreme pain). The second measure was an adapted version of the Face pain scale (Bieri et al., 1990; Hicks et al., 2001; Nordgren et al., 2011). Participants were asked to indicate the face that best reflected their pain, which was also on a 5-point scale scored from 0 (no pain) to 5 (extreme pain). These measures were strongly correlated ($r = .86$) and were averaged to create one reliable pain measure.

Felt Exclusion. A modified version of the Experience of Exclusion Scale and Belonging Measure (Williams, 2009) was used to assess perceptions of exclusion in order for participants to evaluate how they felt from the scenario. Perceived exclusion was a 3-item measure (i.e., “I would feel rejected”) with high internal consistency (Cronbach’s $\alpha = .93$). Perceived invisibility was also a 3-item measure (i.e., “I would feel nonexistent”) with high internal consistency (Cronbach’s $\alpha = .93$). Both measures used a 7-point scale, which ranged from 1 (Not at all) to 7 (Extremely) for each item. Because the two measures were highly correlated ($r = .89$) and showed parallel patterns, I collapsed them into one measure of felt exclusion.

3.4 Main Results

In my first set of preregistered analysis, I planned to compare ratings of the perpetrator’s egalitarian motivation, awareness of bias, effort to regulate their bias, as well as the presence of stereotypes and of the outcome of the incident, as I did in Study 1 with regression models for each manipulation check as the dependent variable with dummy coding to group the conditions according to my manipulations in the vignette stimuli. Then, in my main preregistered analysis, I planned to compare perceptions of harm following the unbiased, implicitly biased, and intentionally biased conditions.

Manipulation checks. I predicted that our manipulations for the perpetrator’s egalitarian motivation, awareness of bias, and effort to regulate, as well as the presence of stereotypes and outcomes, would show corresponding effects on our manipulation checks. Considering our revisions to our stimuli following the results of Study 1, I anticipated support for our manipulations in Study 2. For each of the following manipulation checks, I used the same contrast codes in regression as I did in Study 1, with dummy coding to group the bias types according to the manipulated dimension and with the manipulation check as the dependent variable. The following were all preregistered analyses.

Egalitarian Motivation. As preregistered, participants rated the perpetrator as more motivated to be egalitarian in the egalitarian bias types (authentically unbiased, regulated bias, unintentional bias, and unconscious bias scenarios, $M = 4.62$, $SD = 1.98$) than in the nonegalitarian bias types (apathetic and hostile bias scenarios, $M = 2.20$, $SD = 2.07$), $t(300) = 9.82$, $p < .0001$.

Awareness of Bias. As preregistered, participants rated the perpetrator as being more aware of his stereotypes in the aware bias types (regulated, unintentional, apathetic, and hostile bias scenarios, $M = 5.78$, $SD = 1.54$) than in the unaware bias types (unconscious bias scenario $M = 2.91$, $SD = 1.69$), $t(249) = 11.89$, $p < .0001$. As the awareness dimension was not
effort to regulate bias. As preregistered, participants perceived the perpetrator having put in more effort to prevent his stereotypes from affecting his behavior in the downregulation bias types (regulated and unintentional bias types, $M = 4.10, SD = 1.89$) than in the no effort bias types (unconscious and apathetic bias types, $M = 4.78, SD = 1.52$), $t(248) = 2.81, p = .005$ and the upregulation bias type (hostile bias type, $M = 5.78, SD = 1.74$) $t(249) = 5.67, p < .001$. Participants also rated the perpetrator as putting in more effort to use his stereotypes in the upregulation bias type (hostile) compared to the no effort bias types (apathetic and unconscious) $t(248) = 3.39, p < .001$. As the regulation dimension was not manipulated in the authentically unbiased stimuli, that condition was excluded from this analysis.

Stereotypes. With the addition of the authentically unbiased condition, in this study I was able to run preregistered contrasts comparing the stereotype rating of conditions where the perpetrator is described as holding stereotypes, compared to the authentically unbiased condition where the perpetrator is described as not holding any stereotypes. As anticipated, participants rated the degree to which the perpetrator held stereotypes as higher in the regulated, unintentional, unconscious, apathetic, and hostile bias types ($M = 6.04, SD = 1.12$) than in the authentically unbiased type ($M = 2.18, SD = 1.58$), $t(300) = 20.84, p < .0001$.

Outcomes. As preregistered, participants rated the outcome as objectively more positive in the conditions in which there was a positive outcome for the recipient (authentically unbiased, regulated bias), to those where there was a negative outcome (unintentional, unconscious, apathetic, hostile bias). As anticipated, participants rated the interaction as more negative in the four negative outcome types ($M = 2.15, SD = 1.40$) than in the two positive outcome types ($M = 5.81, SD = 1.35$), $t(300) = -21.66, p < .0001$.

Perceptions of Harm. To test my hypotheses about perceived harm, I first preregistered regressions with each harm measure (i.e., pain, felt exclusion) as the dependent variable with contrasts to group the conditions by implicitly biased, intentionally biased, and unbiased behaviors. I tested these hypotheses in a regression framework using orthogonal contrasts. One dummy variable (Biased Behaviors) represented the comparison of the two unbiased behaviors to the average of the four biased actions. The second dummy variable (Intentional Bias) represented the comparison of Implicit to Intentional forms of bias. In my preregistration I made parallel predictions for ratings of pain, exclusion, and invisibility, however because ratings of felt exclusion and invisibility were highly correlated ($r = .89$), I combined them into one measure of perceived exclusion. In exploratory analyses, I also compare participants’ ratings of their negative emotions (e.g., anger, sadness) following unbiased, implicitly biased, and intentionally biased behaviors. I made parallel predictions for negative emotions as I did with the other harm measures, with results reported in the appendix. Here, I focus on my two main preregistered harm measures: pain and exclusion.

In my primary predictions, I preregistered that both intentionally and implicitly biased behavior would be perceived as more harmful than unbiased behavior, but that intentional bias would be more harmful than implicit bias. I also predicted differences in harm ratings between the two types of bias within each category of behavior.

Perceived Pain. As predicted, participants reported greater pain from the intentional forms of bias ($M = 3.05, SD = 1.27$) compared to the implicit forms of bias ($M = 2.48, SD = 1.10$, $t(299) = 3.40, p < .001$), which were both in turn reported to be more painful than unbiased behavior ($M = .75, SD = 1.12$, $t(299) = 14.15, p < .001$). When the perpetrator’s behavior was
discriminatory, participants perceived more pain if their biased behavior was intentional than if it was implicit.

**Perceived Exclusion.** As predicted, participants reported greater exclusion from the two forms of biased behaviors (Intentional $M = 5.10$, SD = 1.65; Implicit $M = 4.72$, SD = 1.62) when compared to the unbiased behaviors ($M = 2.00$, SD = 1.44, $t(299) = 15.14$, $p < .001$). However, contrary to our predictions and to our findings for perceived pain, there was no significant difference in the perception of exclusion comparing the two forms of biased behavior of bias, $t(299) = 1.73$, $p = .08$. When the perpetrator’s behavior was discriminatory, participants’ feelings of exclusion were no different whether their biased behavior was intentional or implicit.

**Figure 4**
*Perceived Pain for Unbiased, Implicitly Biased, Intentionally Biased Behavior*

![Graph showing perceived pain](image)

*Note.* Perceived pain across the three categories of behavior. Reported pain was lower following the two forms of unbiased than the two forms of implicitly biased behavior, which were both in turn had lower pain ratings than the two forms of intentionally biased behaviors.

**Figure 5**
*Perceived Exclusion for Unbiased, Implicitly Biased, Intentionally Biased Behavior*
Note. Perceived exclusion across the three categories of behavior. Reported harm was lower following the two forms of unbiased than the four forms of biased behavior. There was no reported difference in harm comparing the two forms of implicitly biased to the two forms of intentionally biased behaviors.

Secondary hypotheses concerned focused comparison within the broad bias type conditions. Although we preregistered that we would conduct this analysis with focused t-tests, upon reflection we decided to use Tukey’s HSD to compare all pairwise comparisons simultaneously. This did not substantially change the results. We report the pairwise comparisons within each category of behavior for perceived pain and exclusion. As with the primary hypotheses that looked at differences between the categories of behavior, we made parallel predictions for pain and felt exclusion for our hypotheses within the categories.

Harm for Intentional subtypes. I expected greater harm perceptions for hostile compared to apathetic bias as I anticipated the intentional upregulation of bias would exacerbate the harm felt by participants. However, as shown in Figure 6 and 7 there was no significant difference in reported pain ($t(97) = -0.33, p = .74$) or felt exclusion ($t(97) = -0.76, p = .50$) between these two conditions.

Harm for Implicit subtypes. I expected greater harm perception for the unconscious
compared to *unintentional* bias as I anticipated the deliberative attempt to downregulate one’s bias would mitigate the harm felt by participants. As shown in Figure 6 and 7, there was no significant difference in reported pain ($t(101) = -1.55, p = .13$) or felt exclusion ($t(101) = 0.73, p = .50$) between these two conditions.

**Harm for Unbiased subtypes.** I expected no predicted differences in harm perception for *authentically unbiased* behavior and *regulated bias*. However, participants reported greater pain from regulated bias than authentically unbiased behavior, $t(70) = -4.28, p < .001$; there were no significant differences between the two conditions for perceived exclusion, $t(70) = 2.15, p = .26$.

**Figure 6**  
_Perceived Pain for Bias Types_

![Graph showing perceived pain for different types of bias](image)

*Note.* Perceived pain across the six types of behavior. Labels describe the pairwise comparisons of the bias type conditions for pain ratings. Colors reflect the distinct form of
behavior.

**Figure 7**

*Perceived Exclusion for Bias Types*

Note. Perceived exclusion across the six types of behavior. Labels describe the pairwise comparisons of the bias type conditions for harm ratings. Colors reflect the distinct form of behavior. There were no differences in perceived harm across the scenarios where there was a negative outcome for the recipient.

3.5 Exploratory Analysis: Dimension of Bias and Harm

In an exploratory analysis that was not included in the preregistration, I sought to identify which dimension of bias (i.e., egalitarian motivation, awareness of bias, effort to regulate) was most predictive of psychological harm in response to the four biased events. While most research has distinguished implicit from intentional bias by the perpetrator’s awareness (see Daumeyer et al., 2019; Redford & Ratliff, 2016), our model primarily focuses on egalitarian motivation. In
this analysis, I aimed to identify how each dimension of the perpetrator’s internal processing of their stereotypes contributed to reported harm participants perceived. I focus here on only those four conditions in which there was a biased outcome (unintentionally, unconsciously, apathetically, and hostilely biased behavior), as though the unbiased behaviors also communicate the perpetrator’s internal motivations, the reported pain from those behaviors is more likely informed by their objectively positive outcome. I used the regression framework with perceived pain as the dependent variable and planned contrasts representing each manipulated dimension as simultaneous predictors. I report these results with pain as the dependent variable, as there was no variability in feelings of exclusion for the four forms of biased behavior.

Even when controlling for scenario variation in awareness and regulation, participants perceived less pain when the perpetrator cared about being egalitarian (in the unintentional and unconscious scenarios, $M = 2.48$, $SD = 1.10$) than when the perpetrator did not care (in the apathetic and hostile scenarios, $M = 3.05$, $SD = 1.27$), $t(198) = -3.46, p < .0001$. In contrast, here was no significant difference in pain reported to the three aware types (unintentional, apathetic, hostile; $M = 2.64$, $SD = 1.07$) compared to the unaware type (unconscious, $M = 2.80$, $SD = 1.27$), $t(198) = 0.87, p = .39$.

Finally, controlling for motivation and awareness, effort to regulate (either downregulation, no regulation, or upregulation) also uniquely predicted pain ratings, $t(198) = -2.96, p = .003$. Participants reported less pain to bias resulting from an effort to downregulate stereotypes (unintentional $M = 2.31$, $SD = 1.11$) compared to bias in the context of no effort (unconscious, apathetic $M = 2.85$, $SD = 1.16$, $t(198) = 2.71, p = .007$) or to bias resulting from upregulated stereotypes (hostile, $M = 3.01$, $SD = 1.34$, $t(198) = -2.96, p = .003$). There was no significant difference in pain reported to no effort (unconscious, apathetic) and upregulated bias (hostile), $t(198) = -.70$, $p = .48$. Knowing that the perpetrator tried to regulate their behavior helped to mitigate some of the perceived pain over other forms of bias.

3.6 Discussion

In this study, I measured participants’ perceptions of psychological harm following incidents of discriminatory behavior in which the perpetrator was described as implicitly or intentionally biased. Compared to non-discriminatory behaviors, intentionally and implicitly biased behaviors hurt, both in terms of level of pain and feelings of exclusion. Learning about the set of internal processes by which an implicitly biased perpetrator’s behavior unfolded (either that they cared to be egalitarian but were unaware of their stereotypes and unable to regulate them, or they were aware and tried ineffectively to regulate them), did not erase the feelings of hurt from their actions. When a perpetrator acted in a discriminatory way, that resulted in pain and feelings of exclusion for those on the receiving end of those actions. Here, we see that the only way harm was mitigated was by regulating stereotypes, or not holding stereotypes in the first place.

When there was discriminatory behavior, participants reported less pain when the perpetrator’s behavior was implicitly biased, but similar feelings of exclusion whether their behavior was implicitly or intentionally motivated. Knowing that a perpetrator was not motivated to act with prejudice lessened the pain but did not change the feelings of exclusion that resulted from their discriminatory behavior. These findings show that bias resulting from implicit processes was anticipated to be less psychologically harmful than bias resulting from intentional processes, supporting research in moral psychology that finds accidental actions are judged less harshly and felt as less painful than intentional actions (Cushman, 2015; Daumeyer et al., 2019;
Gray & Wegner, 2008; Gray, 2012). However, the same pattern did not exist for feelings of exclusion, as we had anticipated. The impact of the perpetrator’s intentions differed depending on whether we asked about harm as pain or as feelings of exclusion. For the more physical perceptions of pain, the discriminatory behavior by a perpetrator who was otherwise motivated to be egalitarian was less painful than those behaviors of someone who was not motivated to be egalitarian. However, for more social considerations of feelings of exclusion, there was no difference whether the perpetrator meant to act with discrimination or not.

Moreover, our exploratory analyses suggest that motivation (and regulation) and not awareness is a critical distinction in anticipated harm to incidents of bias. While our manipulation of a perpetrator’s egalitarian motivation, awareness, and efforts to regulate were not completely orthogonal to one another and we cannot completely control for one of the factors’ impacts in the others’, we asked which part of the process by which a perpetrator acted with discrimination was the most painful. While reported pain was lessened when the perpetrator was motivated to be egalitarian and when they made an attempt to regulate their stereotypes, there was no significant difference in reported pain due to their lack of awareness. This is surprising as other work often emphasizes a lack of awareness as abating moral judgments of blame (Cameron et al., 2010; Daumeyer et al., 2019).

In this study, these tightly controlled scenarios allowed us to better understand women’s perceptions of harm from intentional or implicit discrimination. In interpreting our results, one possibility is that knowing the perpetrator's internal motivations may help mitigate some forms of hurt. A second possibility, not mutually exclusive from the first, is that it can be difficult to anticipate our feelings of exclusion to these abstracted and imagined scenarios, and difficult to ascertain whether a perpetrator’s intentions would inform our experiences. Indeed, our study is limited in that we asked participants to imagine they are on the receiving end of these scenarios of bias, to imagine they have insights into that perpetrator’s intentions, and to predict what it would feel like if they had this information. In our next study, we sought to measure people’s experiences of psychological harm from their own recalled experiences of discrimination. We investigated these findings further by asking about participants’ own experiences of discrimination, and whether their recalled experiences of harm differed by their perceptions of the perpetrator’s intentionality.
Chapter 4: Study 3
Recalled Harm from Real-World Incidents of Implicit and Intentional Bias

While the previous study measured perceptions of harm following hypothetical scenarios of intentional versus implicit forms of biased behavior, in this study we sought to replicate our findings with experiences of harm from recalled incidents. With the hypothetical scenarios, we found that participants anticipated both forms of biased behavior to be harmful, but there was still uncertainty about the harm that implicit and intentional bias does. When considering their feelings of pain following the hypothetical incidents, participants were more hurt if the perpetrator was acting intentionally. However, when considering their feelings of exclusion following the hypothetical incidents, participants were just as hurt if the perpetrator was intentionally or implicitly biased. As the previous design relied on anticipated feelings to hypothetical scenarios, it was important to investigate the harm that implicit and intentional bias does in a more ecologically valid context that was closer to how people experience incidents of discrimination in their daily lives.

In this study, our aim was to investigate how participants experience bias by measuring their feelings of harm from their past incidents of bias. When recalling their experience of psychological harm from their lived experiences of discrimination, do people recall being less hurt when the perpetrator was implicitly biased, as we previously found with the hypothetical scenarios? We recruited participants to recall their own experiences of social bias (intentional, implicit, and unbiased) and the harm that resulted from them. These experiences could be based on racial, gender, disability, sexuality bias. In this way, we made this survey open to participants of all identities to share their experiences of social bias, so as to not exclude participants unnecessarily and to not adhere to norms of cisnormativity.

4.1 Participants
A power analysis with 0.90 power, 0.05 alpha, and a .25 effect size suggested at least 45 observations per each type of bias. We aimed to collect 45 observations per each type of bias, and stopped data collection once we reached that minimum goal of the required 45 per group. Participants were asked to provide three recalled incidents in the survey, and so our sample of participants were 138 university students in Canada, with a mean age of 21 (SD = 4.3). Participants were predominantly East-Asian (41% East Asian, 22% White, 15% South Asian, 7% not listed and specified), women (76% Women, 19% Men), and heterosexual (72% Heterosexual, 20% Bisexual/pansexual, 3.5% Gay/Lesbian).

4.2 Protocol
Participants were recruited to take an online survey in which they would be asked to share an experience when they were on the receiving end of an incidence of bias. Participants were told that they would be asked to write about incidents of intergroup bias, specifically when the perpetrator did not share their own group membership. Before proceeding, participants were asked to write about instances of bias against one social identity (i.e., race, gender, ethnicity, sexuality, disability) consistently throughout the survey. Most incidents described pertained to the participant’s ethnicity or gender (47.3% described racial/ethnic bias, 43.8% described gender bias, 6% described bias against their sexuality, 3% described bias against their disability).

Participants first watched videos explaining what bias is and were told that they would be walked through a new framework for thinking about the different forms bias can take. Throughout the survey they were walked through Schmader and colleagues (2022) bias typology. This was a within-subjects design, such that each participant was asked to provide
three instances of bias: one unbiased (either an authentic or regulated experience), one implicitly biased (either unintentional or unconscious), and one intentionally biased (either apathetic or hostile). They were given the choice of which type of bias to describe within each pair to maximize the likelihood that they could recall at least one experience within the three broad groupings for focal analyses. This resulted in a total of 389 observations.

First, participants were guided through what unbiased behavior might look like, before being asked to recall past experiences of authentically unbiased or regulated biased behavior. They were then asked to describe the experience they remember most clearly in an open-ended response. Whether participants next responded to the intentional before the implicit survey blocks was randomized. For the intentional and implicit blocks, they followed a similar procedure as with the unbiased block. Participants were guided through what intentional and implicit behavior might look like, before being asked to recall past experiences of apathetically or hostilely biased behavior, and unintentionally and unconsciously biased behavior, respectively.

For each incident of bias, participants provided descriptions about the perpetrator, the incident, and their feelings during and after the incident. Participants also completed the same close-ended measures of harm (i.e., pain, felt exclusion) and manipulation checks (i.e., egalitarian motivation, awareness of bias, efforts to regulate) as in the previous studies. Before exiting the survey, participants were fed back their responses and asked to confirm which type of incident they described, to allow for changes in their categorization to increase the accuracy of this self-coding (84% of categorizations stayed the same). We added this procedure assuming that the distinctions between different types of bias might be more obvious after learning about all types and recalling their personal experience. We used their final labeling of the events as their perceived bias type.

### 4.3 Measures

#### Manipulation Checks.
We used the same manipulation checks as the previous studies to capture the perceived egalitarian motivation, awareness of bias, and efforts to regulate bias, as well as their bias and the outcome of the incident.

#### Pain.
We used the same two-item pain measure consisting of a general pain question and a face pain scale were used to assess perceptions of pain as in Study 2.

#### Felt Exclusion.
We used the same 7-point scale to measure feelings of exclusion and invisibility \( (r = .82) \) and combined them into one measure of felt exclusion as we did in Study 2.

#### Open-ended Descriptions of the Incident.
In several open-ended questions, we provided participants with space to provide information about who the perpetrator was, what the perpetrator did, and how they felt about it in each respective scenario. We collected information about the Who, What, When, Why and How of the recalled incident. This information is provided in the appendix.

In follow-up coding, we had two research assistants use the written descriptions of the incidents to code for what type of bias behavior (i.e., unconsciously biased, apathetically biased) best matched the incident as described by the participant. These two coders were blinded to the participants’ own categorizations of the incidents, and each individually coded all 389 observations. The coders’ independent categorizations for whether the recalled incidents were unbiased, implicitly biased, or intentionally biased matched across 91% of observations, and their categorizations for the specific bias identity type matched across 82% of observations. The
coders discussed and resolved any discrepancies\(^1\). As shown in Table 2, at the broader level used in primary analyses (unbiased, implicit, and intentional), coders’ categorizations matched those of participants in over 70% of cases. However, within these broad types, there were more disagreements; in particular unintentional and apathetic cases were coded similarly less than 40% of the time. Despite these mismatches, parallel analyses using the research assistants’ recategorizations of the bias types found no significant deviations from those results by participants’ own categorizations. In the following analysis we show participants’ experiences of harm for their own categorizations of the bias types.

**Table 2**  
*Categorization of Recalled Incidents, Coder Comparison*

<table>
<thead>
<tr>
<th>Participant Category</th>
<th>Coder Match</th>
<th>Participant Bias Type</th>
<th>Coder Match</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unbiased Behavior</strong></td>
<td>Authentic</td>
<td>N = 80</td>
<td>74%</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Regulated</td>
<td>N = 46</td>
<td>52%</td>
<td>46</td>
</tr>
<tr>
<td><strong>Implicit Bias</strong></td>
<td>Unintentional</td>
<td>N = 59</td>
<td>32%</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Unconscious</td>
<td>N = 74</td>
<td>66%</td>
<td>74</td>
</tr>
<tr>
<td><strong>Intentional Bias</strong></td>
<td>Apathetic</td>
<td>N = 52</td>
<td>38%</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Hostile</td>
<td>N = 78</td>
<td>76%</td>
<td>59</td>
</tr>
</tbody>
</table>

*Note.* Here we show participants’ own categorizations of their recalled incidents aligned to the coder’s blind categorizations of the incidents. For example, for incidents participants labeled as unbiased, 79% were similarly identified as unbiased by coders who were not aware of the original categorization. There was considerable overlap between participants’ and coder’s categorizations, albeit notable exceptions for unintentional and apathetic incidents.

### 4.4 Results

**Manipulation Checks.** Considering the exploratory nature of this recall study, I did not make preregistered predictions. However, I did anticipate replicating the manipulation check results in Study 2. As I did previously, I used the regression framework with each manipulated

\(^1\) A similar process was used for the coding of the open-ended descriptions of the incidents. Independent coders coded for the type of relationship (who, how close) as well as whether an apology was provided by the perpetrator, and resolved any discrepancies. These codes were defined because they provided more context for what happened between the perpetrator and the participant.
dimension (i.e., egalitarian motivation, awareness of bias, effort to regulate) as the dependent variable and with planned contrasts to group the bias types by the manipulated dimension. This was critical, as these dimensions were not directly manipulated but reported by participants in their own categorization of the recalled incident of bias.

**Egalitarian Motivation.** Conceptually replicating Study 2, participants recalled the perpetrator as more motivated to be egalitarian in the four egalitarian bias types (authentically unbiased, regulated bias, unintentional bias, and unconscious bias incidents, \( M = 4.17, SD = 1.72 \)) compared to the two nonegalitarian bias types (apathetic and hostile bias incidents, \( M = 2.35, SD = 1.51 \)), \( t(385) = 6.60, p < .0001 \).

**Awareness of Bias.** Conceptually replicating Study 2, participants recalled the perpetrator as more aware in the aware bias types (regulated bias, unintentional bias, apathetic bias, and hostile bias incidents, \( M = 4.35, SD = 1.59 \)) than in the unaware bias type (unconscious bias, \( M = 2.52, SD = 1.42 \)), \( t(306) = 8.41, p < .0001 \).

**Effort to Regulate Bias.** Conceptually replicating Study 2, participants recalled the perpetrator as putting in more effort to prevent their stereotypes in the downregulation bias types (regulated and unintentional bias, \( M = 3.86, SD = 1.76 \)) than in the no effort bias types (unconscious and apathetic bias, \( M = 3.81, SD = 1.31 \)), \( t(305) = -0.30, p = .77 \). Again, replicating Study 2, participants recalled the perpetrator as putting in more effort to upregulate their stereotypes in the upregulation bias types (hostile bias, \( M = 4.90, SD = 1.55 \)), than in the no effort bias types, \( t(305) = 5.97, p < .001 \).

**Outcome.** Whereas the outcome of the bias incidents was held constant in the hypothetical incidents, they could not be in these recalled incidents. As we anticipated, participants recalled the outcomes of the recalled incidents as more positive in the unbiased incidents (authentically unbiased and regulated bias, \( M = 5.56, SD = 1.42 \)) than in the bias incidents (unintentional, unconscious, apathetic, hostile bias, \( M = 2.74, SD = 1.23 \)), \( t(385) = 20.6, p < .001 \). Comparing the four bias types, participants recalled the outcomes of the implicit bias incidents (\( M = 3.01, SD = 1.12 \)), \( t(385) = -4.3, p < .001 \) as more positive than those outcomes of the intentional bias incidents (\( M = 2.40, SD = 1.26 \)). This was unanticipated, though not entirely surprising considering participants described their lived experiences of discrimination and made assumptions about other people’s internal motivations for their behaviors. It is therefore likely that those incidents that were recalled as being intentional were also more objectively negative than those incidents recalled as being implicit. Considering these results, in our analysis of experienced harm from implicit and intentional incidents of bias we control for the reported severity of the outcome.

**Stereotypes.** As we anticipated, the degree to which the perpetrator was perceived as holding more stereotypes was higher in the bias incidents (\( M = 5.18, SD = 1.53 \)) than in the authentically unbiased incidents (\( M = 1.96, SD = 1.42 \)), \( t(386) = 17.0, p < .001 \). Comparing the four bias types, participants recalled the perpetrator of implicit bias as holding less stereotypes (unintentional and unconscious bias \( M = 4.09, SD = 1.44 \)) than they did for a perpetrator of intentional bias (apathetic and hostile bias \( M = 5.98, SD = 1.08 \)), \( t(260) = -6.80, p < .001 \).
Table 3
Means and standard deviations for manipulation checks as a function of participant’s categorizations.

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Egalitarian Motivation</th>
<th>Awareness of Bias</th>
<th>Effort to Upregulate</th>
<th>Outcome</th>
<th>Stereotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Unbiased Behavior</td>
<td>Authentic</td>
<td>4.69</td>
<td>2.10</td>
<td>2.72</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>Regulated</td>
<td>5.26</td>
<td>1.51</td>
<td>3.96</td>
<td>1.59</td>
</tr>
<tr>
<td>Implicit Bias</td>
<td>Unintentional</td>
<td>4.05</td>
<td>1.78</td>
<td>3.36</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>Unconscious</td>
<td>2.69</td>
<td>1.52</td>
<td>2.53</td>
<td>1.42</td>
</tr>
<tr>
<td>Intentional Bias</td>
<td>Apathetic</td>
<td>2.29</td>
<td>1.12</td>
<td>4.63</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>Hostile</td>
<td>2.41</td>
<td>1.91</td>
<td>5.44</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Note. Manipulation checks were asked on 7 pt. likert scales with higher scores indicating higher ratings of that construct. Results are color coded according to the theorized distinctions between the six forms of behavior. For example, both apathetic and hostile bias are theorized as *less motivated to be egalitarian* than other bias types; unconscious bias is theorized as *less aware* than other bias types; regulated and unintentional are theorized as *more downregulation* than unconscious and apathetic, which are in turn expected to have *less upregulation* than hostile bias. Similarly, only two forms of unbiased behavior are expected to have a positive outcome and only authentically unbiased behavior is theorized as the perpetrator not holding stereotypes or prejudicial attitudes. Color grouping represents contrast groups for the manipulation checks.

**Perceptions of Harm.** For experiences of harm, I asked whether recalled experiences of implicitly biased behavior were recalled as less harmful than intentionally biased behaviors, and whether that was consistent for both the experience of pain and felt exclusion.

As in Study 2, I tested this in a regression framework using orthogonal contrasts. I tested each harm measure (i.e., pain, felt exclusion) as the dependent variable with contrasts to group the conditions by implicitly biased, intentionally biased, and unbiased behaviors. One dummy variable (Biased Behaviors) represented the comparison of the two unbiased behaviors to the average of the four biased actions. The second dummy variable (Intentional Bias) represented the comparison of Implicit to Intentional forms of bias. I am primarily motivated to replicate the results I found in Study 2, that both forms of biased behavior would be recalled as more painful than unbiased behaviors. As for the differences in harm between implicit and intentional forms
of behavior, I made no predictions as for whether I would replicate my findings. Considering the different methodology that now centers participants recalled experiences of harm from biased behaviors, it was an open question whether implicitly biased behaviors would be recalled as less, equally, or more harmful than intentionally biased behaviors, and whether there would be diverging patterns for the experience of psychological pain compared to the feelings of exclusion. First, I report these comparisons as we did in Study 2, then I report these comparisons controlling for the reported severity of the outcome and stereotypes of the perpetrator.

**Experienced pain.** Replicating Study 2, participants reported greater pain from the intentional forms of bias ($M = 2.47$, $SD = 1.38$) compared to implicit forms of bias ($M = 1.72$, $SD = 1.27$), $t(386) = -4.85$, $p < .001$, and both elicited greater harm than unbiased behavior ($M = .68$, $SD = 0.99$), $t(386) = 11.14$, $p < .001$). When the perpetrator’s behavior was discriminatory, participants recalled more pain if their biased behavior was intentional than if it was implicit.

**Figure 8**
Felt Pain for Unbiased, Implicitly Biased, Intentionally Biased Behavior

![Figure 8](image)

*Note.* Experienced pain across the three categories of behavior. Replicating results of Study 2, here reported pain was lower following the two forms of unbiased than the two forms of implicitly biased behavior, which were both in turn had lower pain ratings than the two forms of intentionally biased behaviors.
**Experienced exclusion.** Diverging from our findings in Study 2 where there was no significant difference, participants reported more feelings of exclusion from the intentional forms of bias ($M = 3.05, SD = 1.77$) than from implicit forms of bias ($M = 2.55, SD = 1.66$), $t(386) = -2.50, p = .03$. Both forms of biased behavior elicited more feelings of exclusion than did unbiased behavior ($M = .68, SD = .99$), $t(386) = 7.37, p < .001$). When the perpetrator’s behavior was discriminatory, participants recalled more exclusion if their biased behavior was intentional than if it was implicit.

**Figure 9**
*Felt Exclusion for Unbiased, Implicitly Biased, Intentionally Biased Behavior*

![Graph showing felt exclusion for different categories of behavior](image)

*Note.* Experienced exclusion across the three categories of behavior. In a different pattern than the results of Study 2 where there were no differences in harm from implicit as compared to intentional behaviors, here reported harm was lower following the two forms of unbiased than the two forms of implicitly biased behavior, which were both in turn had lower harm ratings than the two forms of intentionally biased behaviors.

Next, I compared participants’ recalled harm within the forms of biased behavior. I conducted one-way ANOVAs for reported harm with the participant’s own categorizations of the
incidents as the predictor and I conducted post hoc analysis using Tukey’s HSD to consider all pairwise comparisons. I report the pairwise comparisons within each category of behavior for perceived pain and exclusion. In Study 2, these analyses resulted in no significant differences between the two forms of each category of behavior.

**Harm for Intentional subtypes.** Replicating findings from Study 2, there was no significant difference in the experience of harm from hostile and apathetic forms of biased behavior. There was no difference in the reported pain \((M = 2.70, SD = 1.43)\) or felt exclusion \((M = 3.12, SD = 1.90)\) from hostile bias, compared to that of the reported pain \((M = 2.24, SD = 1.33)\) or felt exclusion \((M = 2.99, SD = 1.63)\) from apathetic bias, respectively.

**Harm for Implicit subtypes.** Replicating the findings from Study 2, there was no significant difference in the experience of harm from unintentional and unconscious forms of biased behavior. There was no difference in the reported pain \((M = 1.44, SD = 1.18)\) or felt exclusion \((M = 2.26, SD = 1.40)\) from unintentional bias, compared to that of reported pain \((M = 2.01, SD = 1.37)\) or felt exclusion \((M = 2.84, SD = 1.91)\) from unconscious bias, respectively.

**Harm for Unbiased subtypes.** Diverging from the findings from Study 2, there was no significant difference in the experience of pain from authentically unbiased and regulated forms of biased behavior. For reported pain, there was no difference between the two unbiased conditions \((\text{authentic } M = 0.42 \text{ SD} = 1.27, \text{ regulated } M = 0.93 \text{ SD} = 0.90)\). Also diverging from Study 2, there was a significant difference in reported feelings of exclusion between the two forms of unbiased behavior. For reported exclusion, there was greater harm in the regulated \((M = 2.10 \text{ SD} = 1.37)\) than in the authentically unbiased conditions \((M = 1.27 \text{ SD} = 0.67)\).

In this final set of analysis, report the comparisons of experienced pain and felt exclusion for unbiased, implicitly biased, and intentionally biased behavior, now controlling for the reported severity of the outcome. We did not report this analysis in Study 2 as we manipulated those dimensions directly in our vignettes. Considering the responses to the manipulation checks showed that the outcome was rated as more severe in the intentionally biased conditions than in the implicitly biased conditions, here we compare harm perceptions after controlling for each dimension.

**Harm Analyses Controlling for Outcome.** I conducted ANCOVAs for the pain ratings of the biased behaviors with the reported outcome ratings as a covariate. There was still a significant effect of the form of biased behavior on reported pain after controlling for the effect of the outcome rating, \(F(1, 259) = 5.98, p = .01\), such that intentional incidents \((M = 2.32, \text{ SE} = 0.20)\) were experienced as more painful than implicit incidents \((M = 1.96, \text{ SE} = 0.10)\).

Replicating our findings in Study 2, when the perpetrator’s behavior was discriminatory, participants recalled more pain if their biased behavior was intentional than if it was implicit, controlling for the outcome severity.

I then conducted this same analysis, but for the felt exclusion measure. After controlling for the severity of the outcome there was no longer a significant difference in harm ratings between intentional \((M = 2.89, \text{ SE} = 0.15)\) and implicit incidents \((M = 2.75, \text{ SE} = 0.15, F(1, 259) = 0.43, p = .51)\). Replicating our findings in Study 2, when the perpetrator’s behavior was discriminatory, participants recalled similar levels of exclusion whether the biased behavior was intentional or implicit, controlling for the outcome severity.

**Figure 10**

*Felt Pain and Exclusion from Bias Incidents, Controlling for Outcome Severity*
When we control for the severity of the incident, participants reported more pain, but not more exclusion, from intentionally biased behavior than implicitly biased behavior.

### 4.5 Discussion

In this study, I measured participants’ recalled experiences of psychological harm following their lived experiences of discrimination in which they inferred the perpetrator was implicitly or intentionally biased. Compared to those incidents in which a person was unbiased or was able to regulate their bias, those in which a person engaged in discriminatory behavior were recalled as more hurtful both in terms of psychological pain and feelings of exclusion. After controlling for the severity of the outcome of the incidents, we find that participants’ recalled experiences of harm replicate their perceptions of harm from Study 2. While in Study 2 we could directly control the outcome of the discriminatory behavior, in Study 3 we could not as participants recalled their own experiences of being targeted by a perpetrator’s implicit or intentional biases, which could vary widely across participants. However, when statistically controlling for participants’ ratings of the severity of these incidents, we still found that implicit
incidents were recalled as less painful than intentional incidents, but there was no longer a significant difference in recalled feelings of exclusion between the two forms of biased behavior. This suggests that for these recalled instances, as with the hypothetical instances of discrimination, knowing the perpetrator was not motivated to act with prejudice lessened the pain, but not necessarily the feelings of exclusion.

While attributing the perpetrator’s discriminatory behavior to implicit processes did help mitigate some feelings of hurt, it did not fully erase them. Implicitly biased behaviors were still recalled as being psychologically painful and as enacting feelings of exclusion, isolation, and invisibility. Paired with the results of Study 2, these findings suggest that knowing an incident of bias resulted from implicit processes did not make it unharmful but did help mitigate some of the feelings of pain that might otherwise manifest.

This study is a methodological advance in literature on judgments of implicit and intentional wrongdoing, as we collected people’s recalled experiences (a method often used in bias literature) of implicit and intentional discrimination and compared their recalled experiences of psychological harm. We found that experiences of psychological harm to recalled scenarios of discrimination in which the motivations of the perpetrator were inferred, largely replicated those perceptions of harm to tightly controlled hypothetical scenarios in which the perpetrator’s internal motivations were either clearly intentional or implicit. Although we are still limited by this design as recalling emotional experiences from past incidents may be different from how those incidents are experienced in-the-moment, this study does provide a better sense of how participants’ experience incidents of discrimination and of the relief that can come from knowing a person did not intentionally act with prejudice.
Chapter 5: General Discussion

When someone discriminates against us, knowing they were not motivated by prejudice might help lessen, but does not fully mitigate the harm done. In two studies in which participants responded to hypothetical and recalled incidents of discrimination, attributing a person’s discriminatory behavior to implicit processes did not erase the experience of psychological harm from their behavior. Rather, implicitly biased behavior was still experienced as painful and brought about feelings of invisibility and exclusion, compared to unbiased behavior. Whether or not the perpetrator meant to be discriminatory, when they do treat someone unfairly based on their stereotypes about them, that person may perceive and experience their actions as painful.

However, knowing a person’s behavior was implicitly biased and not intentionally biased can lessen the experience of psychological harm, somewhat. When considering their feelings of pain, people were more hurt if the perpetrator was acting intentionally than if they were acting implicitly both in hypothetical and recalled incidents. We did not see those same effects for feelings of exclusion; People felt equally excluded and invisible if the perpetrator was intentionally or implicitly biased. Our results suggest that both intentional and implicit forms of biased behavior were harmful, but whether the perpetrator's intentions could mitigate that hurt might depend on how we think about harm. When a perpetrator was implicitly biased, although still harmful, their behavior was less psychologically painful but still elicited strong feelings of exclusion. Although we had made parallel predictions across these different measures of harm, there are several reasons why we might have found diverging patterns for pain and felt exclusion.

One possibility is that intentionality did not inform exclusion ratings as they did for reported pain because there are smaller effects for feelings of exclusion that might require more statistical power to detect. Another possibility, not mutually exclusive from the first, is that judgments of pain may be similar to judgments of wrongdoing or responsibility in that they are subject to processes of deduction in a way that feelings of exclusion are not. Considering our measures of pain were not actual reports of felt physical pain (as Gray & Wegner, 2008 and Gray, 2012 measured physical responses to administered shocks), but anticipated pain, it is possible that people’s responses correspond more to the information they had about the perpetrator’s mental states and less about their feelings of harm. If this is the case, their pain ratings were more similar to how people form judgments about the moral blameworthiness and responsibility of people’s behaviors and were therefore more extreme when the perpetrator was intentional.

A third possibility is that feelings of exclusion are simply not as impacted by a perpetrator’s intentions. Interpersonal incidents of bias in which a person expresses stereotypes against you, and as a result you lose out on an opportunity, hurt. While feelings of pain may be informed by the context of the incident, such as that person’s underlying motivations for their behavior, feelings of exclusion and invisibility may be less sensitive to such factors. Finally, a fourth possibility is that because these incidents featured more tangible harms (i.e., losing out on a rental offer), as opposed to social harms (i.e., being chosen last to be on the team), feelings of exclusion and invisibility were less relevant to the tightly controlled incidents of bias in the hypothetical scenarios and were therefore less impacted by a perpetrator’s motivations behind their actions. This, however, does not account for the replicated findings in the recall study in which participants may have recalled incidents of social harms. Future work should explore pain and felt exclusion following incidents of social harm to tease apart these findings more, and we may investigate this with additional coding of the recalled incidents of bias.
These findings advance our understanding of core distinctions between implicit and intentional bias. In my work, egalitarian motivation was the focal distinction between intentionally and implicitly biased behaviors, whereas other research has focused on awareness (Cameron et al., 2010; Daumeyer et al., 2019; Redford & Ratliff, 2016). Such work has shown that people judge unaware perpetrators as less blameworthy and hold them less accountable for their discriminatory behavior, compared to a perpetrator who was aware of their prejudices (Daumeyer et al., 2019). Whether or not a perpetrator is aware of their potential to act with prejudice is relevant, but perhaps not sufficient in distinguishing implicit from intentional behavior. Indeed, in my research one form of biased behavior happened when a perpetrator was unaware of their stereotypes but was otherwise motivated to be egalitarian. I predicted that the actions of a perpetrator who lacked awareness would actually be more harmful than those of one who was aware and tried (ineffectively) to downregulate their biased behavior. This pairwise comparison, however, was not significant in either Study 2 or 3, suggesting that a lack of awareness alone did not impact feelings of hurt.

Other work has also shown that the different ways of representing awareness in implicit bias can yield different moral judgments. Cameron and colleagues (2010) found that participants judged implicitly biased behavior more harshly if it was described as an automatic process than if it was described as something more unconscious. However, in other work by Redford and Ratliff (2016), they teased apart the role of awareness in judgments of moral responsibility for biased behavior and described an egalitarian perpetrator who was either aware that they held stereotypes but unaware of their impact in their behavior (semi-aware) or who was both aware that they hold and were impacted by their stereotypes (fully-aware). They found that the semi-aware and the fully-aware perpetrators were judged as equally blameworthy for their discriminatory behavior. Our results, paired with those of Redford and Ratliff (2016), suggest that bias was similarly harmful whether a perpetrator was fully unaware of their stereotypes and their impact, or if they were somewhat aware but not effective in regulating their stereotypes. While awareness is often emphasized in bias and moral psychology literature, as well as many intervention efforts, our research suggests that a lack of awareness alone does not help mitigate the harm from bias.

Indeed, in my research I represented four distinct forms of biased behavior, three of which describe a perpetrator who is aware of their stereotypes. Comparing the perceptions of harm from aware versus unaware perpetrators in my exploratory analysis in Study 2, I found that there was no significant difference in reported pain due to the perpetrator’s lack of awareness. Whether or not the perpetrator was aware or unaware of their stereotypes was not the primary concern for the pain of discriminatory behavior. When on the receiving end of a perpetrator’s biased behavior, knowing they were unaware of their stereotypes did not change the harm felt from their actions.

In contrast to past work’s focus on awareness, I represented implicit bias as a set of internal processes including a person’s egalitarian motivation, regulation efforts, as well as their awareness to represent the different forms of implicitly and intentionally biased behaviors. In the exploratory analysis in Study 2, I found that participants perceived less pain when the perpetrator was motivated to be egalitarian and when they made an attempt to downregulate their stereotypes. With the goal of mitigating the harm for a person on the receiving end of biased behavior, knowing the discrimination was not intentional can help lessen some, but not all feelings of pain.
Some might take my findings to mean that when you explain your motivations, you will lessen the psychological harm experienced by a person you discriminated against. I caution against this direct interpretation. When a perpetrator explains their behavior to a person they discriminated against, emphasizing their intentions does not erase the hurt they caused. It might help to somewhat lessen the pain to know that a perpetrator did not mean to act with prejudice and that they do care to treat people well, however emphasizing a lack of awareness might not be as compelling. And knowing the discrimination was not motivated by malice does not make it un-hurtful. Implicitly biased behaviors, though unintentional, still caused harm in both imagined and recalled scenarios.

It is also important to contextualize my research in that participants were given insights or asked about a perpetrator’s internal mindset. This is considerably different than a perpetrator explaining their underlying motivations to a person they’ve harmed. Knowing or assuming what was going on in a person’s mind is different from having that person advocate for what was going on in their mind. A person trying to explain their intentions following their discriminatory behavior may result in backlash effects such that their explanations are considered defensive and as avoiding responsibility, causing hurt.

More research is needed on how people might effectively take accountability for their discriminatory behavior, which is what I intend to do in my next line of research. While these findings suggest that knowing a person’s motivations might help mitigate some of the pain we feel from their behavior, but not all the hurt, their implications for how people might take accountability is somewhat beyond the scope of this research. Instead of explaining away our discriminatory behavior by describing our intentions, we must accept that for the person we’ve harmed, our intentions may not actually matter.

Limitations and Future Directions

Our use of the recall methodology in Study 3 proved helpful in conceptually replicating our findings in Study 2, though with limitations. Our findings on people’s perceptions and experiences of psychological harm were overwhelmingly similar across the two methodologies. While recalling experiences of discrimination allows us more external validity, this approach was still limited in that it did not capture in-the-moment emotional responses. It could be argued that anticipating harm from imagined scenarios is not that different from recalling harm from past scenarios, in that both processes relate to people’s beliefs about their emotional reactions and not their current emotional reactions (Robinson & Clore, 2002). It therefore could be argued that the replicated findings reflect participants’ conceptual and not emotional responses to implicit and intentional incidents of biased behavior.

Future work should investigate people’s in-the-moment emotional reactions to implicitly and intentionally biased behavior using an in-lab paradigm with a confederate researcher. Extending the in-lab paradigm used by Gray and Wegner (2008) in which a confederate researcher intentionally or accidentally delivers shocks to the participant, future work could design a context in which the confederate researcher holds stereotypes about the participant’s social group and either intentionally or implicitly delivers shocks to them because of their stereotypes and/or prejudicial attitudes. Alternatively, as we previously set out to do in this line of work, participants could be assigned to work with a confederate researcher who expresses their stereotypes about the participants’ social group in a way that their discriminatory behavior is clearly a result of intentional or implicit processes. For these in-lab paradigms, researchers would capture participants’ in-the-moment feelings of harm from the perpetrator’s
discriminatory behavior. Work like this, though more ecologically valid, may prove hard to do. Primarily, there are ethical concerns in putting participants in situations in which there is unnecessary risk to their well-being, as researchers purposefully expose participants to discriminatory behavior that are designed to feel real with the goal of measuring their level of psychological distress. There are also concerns for the feasibility of these paradigms, in how possible it is for the confederate to communicate their intentions and for researchers to measure in-the-moment emotional responses while maintaining the cover story. With these concerns, it was our decision to conduct the recall study as it provided a closer proxy for experiences of psychological harm while maintaining minimal risk to participants.

Another consideration of the recall methodology is that participants made assumptions about their perpetrator’s internal motivations. While the intentionality literature often provides participants with information about the perpetrator’s intentions (as we did in Study 2), people rarely have such insights into other people’s mindsets when recalling their past behavior. Indeed, in Study 3 participants were asked to make assumptions about other people’s motivations by relying on information available at the time. Future work should ask how the surrounding context shapes the way a perpetrator’s intentions are perceived.

Peer norms of egalitarianism might be especially useful in diagnosing an actor’s intentions and motivations for their discriminatory behavior. For example, if I work in a company in which people seem to care about diversity initiatives, compared to one where people are generally opposed to them, I may assume my coworker did not mean to act in a discriminatory way. As previous work suggests that egalitarian norms might mitigate harm (Hall et al., 2022; Murrar et al., 2020; Moser & Branscombe, 2022), an incidence of discrimination might be perceived as less harmful and the perpetrator as more implicitly biased if carried out in an egalitarian context. Conversely, as nonegalitarian norms might magnify harm (Purdie-Vaughns et al., 2008; Schmitt et al., 2014), an incidence of discrimination might be perceived as more harmful and the perpetrator as more intentionally biased if carried out in a nonegalitarian context. Work examining the impact of norms will address a theoretical gap in the process by which people diagnose other’s intentions from the context. This is distinct from work on attributional ambiguity that has examined the severity, ambiguity, or internalization of discrimination on self-esteem and well-being (Major et al., 2016; Schmitt et al., 2014). However, this work does support our prediction that discrimination in nonegalitarian cultures may seem more intentional, as Schmitt and colleagues’ (2014) meta-analysis found more negative effects on well-being when discrimination was perceived as pervasive, as opposed to as a signal event.

Conclusion

This research builds towards a better understanding of the consequences for the way bias unfolds. In two studies with hypothetical and recalled methodologies, we saw that people’s perceptions and experiences of psychological pain were somewhat lessened if they knew a perpetrator’s discriminatory behavior was not intentional. This was more the case when they knew the perpetrator was motivated to be egalitarian and when they made an attempt to regulate their stereotypes, less so if they thought the perpetrator was simply unaware. Across both studies, feelings of exclusion did not change whether the perpetrator was implicitly or intentionally biased. Ultimately, knowing the discrimination was due to implicit processes lessened, but did not erase the psychological harm from their behavior.
References


Appendix

Study 2

Bias Dimension Contrasts on Perceived Pain. Here we report each separate regression model we ran for the perceived pain with the contrast for the bias dimension (i.e., egalitarian motivation, awareness, effort to regulate). When comparing the two egalitarian motivation types (unintentional, unconscious) to the nonegalitarian motivation types (apathetic, hostile), there was a significant difference in reported pain for participants responding to bias when the perpetrator cared about being egalitarian ($M = 2.48$, $SD = 1.10$) and when they did not care to be egalitarian ($M = 3.05$, $SD = 1.27$), $F(1, 198) = 11.63, p < .001$. When comparing the three aware types (unintentional, apathetic, hostile) to the unaware type (unconscious), I found no significant difference in reported pain due to awareness, $F(1, 198) = 0.77, p = .38$. As we had three different manipulations of whether the perpetrator put in effort to regulate their behavior (no effort, effort to downregulate, effort to upregulate), I ran two contrasts. First, participants did report more pain from the behavior in which the perpetrator put in effort to upregulate their stereotypes (hostile $M = 2.31$, $SD = 1.11$) compared to when the perpetrator put in effort to try to down regulate their stereotypes (unintentional $M = 3.01$, $SD = 1.34$), $F(1, 198) = 8.68, p = .004$). However, there was no significant difference in pain reported when the perpetrator did nothing to regulate their stereotypes (unconscious $M = 2.85$, $SD = 1.16$) and when the perpetrator put in effort to upregulate those same stereotypes (hostile, $F(1, 198) = 1.36, p = .24$). We ran these same analyses with reported feelings of exclusion as the dependent variable, but found no differences in felt harm across the four types of biased behavior.

Negative Emotional Response. In addition to our primary measures of harm — pain and felt exclusion — we measured participants’ negative emotional responses following the hypothetical scenarios. Participants completed nine additional measures of their emotional response (i.e., “Ashamed”, “Angry”, “Fearful”), to explore in greater detail the type of harm they experienced. Participants responded on a scale from 1 (Not at all) to 7 (Extremely) the degree to which they felt the given emotions. The emotions were grouped into one measure of negative emotion considering the high internal consistency (Cronbach’s $\alpha = .94$).

In an exploratory analysis, I looked at participants’ responses to the negative emotion categories. Specifically, I anticipated heightened negative emotions in response to intentional forms of bias, compared to implicit forms of bias, which in turn were predicted to elicit heightened negative emotions compared to unbiased behavior. As with our primary hypotheses, I used the regression framework with dummy coding to group the conditions into the three main categories of bias. As predicted, participants reported greater negative emotions from the intentional forms of bias ($M = 4.41$) compared to implicit forms of bias ($M = 4.05$), $t(199.1) = 2.15, p = .03$, which both elicited greater harm than unbiased behavior ($M = 1.87$), $t(200.93) = 12.44, p < .001$). In Table 4, I provide the comparisons between unbiased, implicitly biased, and intentionally biased behaviors in the different factors of emotional responses.

Table 4
Negative Emotions following Hypothetical Incidents of Biased Behavior

<table>
<thead>
<tr>
<th></th>
<th>Unbiased Behavior</th>
<th>Implicitly Biased Behavior</th>
<th>Intentionally Biased Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Emotion Factors</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Anxiety (_{Fearful, Anxious, Afraid})</td>
<td>2.07 (1.35)</td>
<td>2.62 (1.35)</td>
<td>2.81 (1.46)</td>
</tr>
<tr>
<td>Anger (_{Angry, Frustrated, Offended})</td>
<td>2.03 (1.64)</td>
<td>5.56 (1.37)</td>
<td>6.27 (1.04)</td>
</tr>
<tr>
<td>Sadness (_{Sad, Depressed, Disappointed})</td>
<td>1.77 (1.32)</td>
<td>4.48 (1.59)</td>
<td>4.70 (1.57)</td>
</tr>
<tr>
<td>Shame (_{Ashamed, Embarrassed, Humiliated})</td>
<td>1.61 (1.14)</td>
<td>3.54 (1.75)</td>
<td>3.86 (1.75)</td>
</tr>
<tr>
<td>Positive Emotions (_{Happy, Excited, Good})</td>
<td>4.54 (1.84)</td>
<td>1.66 (1.09)</td>
<td>1.47 (1.00)</td>
</tr>
</tbody>
</table>

**Study 3**

**Open-Ended Data.** For *Who*, we collected information about the perpetrator’s race and gender. In follow-up coding, we had two research assistants use the information provided to ascertain the *type of relationship* as well as the *nature of the relationship* to provide more context for who the perpetrator was to the participant.

**Table 5**

*The type of relationships participants had with the perpetrator in recalled incidents.*

<table>
<thead>
<tr>
<th>Relation to the perpetrator</th>
<th>Proportion of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boss/teacher</td>
<td>15%</td>
</tr>
<tr>
<td>Coworker/peer</td>
<td>22%</td>
</tr>
<tr>
<td>Family (non-parent)</td>
<td>4%</td>
</tr>
<tr>
<td>Family (Parent)</td>
<td>2%</td>
</tr>
<tr>
<td>Friend</td>
<td>10%</td>
</tr>
<tr>
<td>Stranger</td>
<td>15%</td>
</tr>
<tr>
<td>Significant other/Partner</td>
<td>2%</td>
</tr>
<tr>
<td>Not described</td>
<td>31%</td>
</tr>
</tbody>
</table>
For What, we collected information about the perpetrator’s behavior towards the participant. We asked participants: “What exactly happened in this incident?” We also had coders code whether the perpetrator at any point provided an apology to the recipient, to provide more context for the incident. We also had coders code the degree of certainty provided by the participants about the perpetrator’s intentionality (i.e., the participant was very certain the perpetrator was intentionally biased). Of the 389 observations, most did not mention whether there was an apology, though 10 mentioned the perpetrator offered an apology and 11 spontaneously mentioned that the perpetrator did not apologize for their behavior.

For When, we provided participants with space to provide information about when each scenario took place. We collected information for the date, of how long ago this incident happened, as well as the participant’s age during the incident.

For Why, we asked participants to describe the intentionality of the perpetrator with an open-ended question of “What do you think was going on in their head?” This was also ascertained by the participants’ own categorizations of these recalled incidents as unbiased, implicitly biased, and intentionally biased by their providing these incidents for each respective prompt.

For How, we asked participants how they felt about the interaction at the time and how they feel about it now in an open-ended question. Specifically, we asked, “How did you feel about the interaction at the time?” This was also captured in the close-ended measures of harm.

**Negative Emotional Response.** I used the regression framework with dummy coding to group the conditions into the three main categories of bias. As predicted, participants reported greater negative emotions from the intentional forms of bias (M = 3.59, SD = 1.51) compared to implicit forms of bias (M = 2.75, SD = 1.43), t(386) = 5.01, p < .001, which both elicited greater harm than unbiased behavior (M = 1.72, SD = 1.51), t(386) = -9.81, p < .001. In Table 4, I provide the comparisons between unbiased, implicitly biased, and intentionally biased behaviors in the different factors of emotional responses.

### Table 5
**Negative Emotions following Recalled Incidents of Biased Behavior**

<table>
<thead>
<tr>
<th>Negative Emotion Factors</th>
<th>Unbiased Behavior M (SD)</th>
<th>Implicitly Biased Behavior M (SD)</th>
<th>Intentionally Biased Behavior M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>1.77 (1.31)</td>
<td>2.00 (1.42)</td>
<td>2.70 (1.85)</td>
</tr>
<tr>
<td>Fearful, Anxious, Afraid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>1.78 (1.42)</td>
<td>3.54 (1.96)</td>
<td>4.67 (1.84)</td>
</tr>
<tr>
<td>Angry, Frustrated, Offended</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>1.63 (1.17)</td>
<td>2.80 (1.59)</td>
<td>3.36 (1.71)</td>
</tr>
<tr>
<td>Sad, Depressed, Disappointed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shame
\[\text{Ashamed, Embarrassed, Humiliated}\]

<table>
<thead>
<tr>
<th></th>
<th>1.68 (1.17)</th>
<th>2.59 (1.75)</th>
<th>3.33 (2.00)</th>
</tr>
</thead>
</table>

Positive Emotions
\[\text{Happy, Excited, Good}\]

<table>
<thead>
<tr>
<th></th>
<th>3.72 (1.77)</th>
<th>1.40 (0.83)</th>
<th>1.25 (0.62)</th>
</tr>
</thead>
</table>

**Harm Analyses Controlling for Stereotypes.** I conducted ANCOVAs for the pain ratings of the biased behaviors with the ratings of the perpetrator’s stereotypes as a covariate. There was still a significant effect of the form of biased behavior on reported pain after controlling for the degree of stereotype rating, \(F(1, 384) = 8.10, p < .001\), such that intentional incidents (\(M = 2.05, SE = 0.12\)) were experienced as more painful than implicit incidents (\(M = 1.63, SE = 0.10\)).

I then conducted this same analysis, but for the felt exclusion measure. After controlling for the degree of stereotypes held by the perpetrator, there was no longer a significant difference in harm ratings between intentional (\(M = 2.61, SE = 0.15\)) and implicit incidents (\(M = 2.46, SE = 0.15, F(1, 384) = 0.77, p = .72\)).