Intent's Role in Moral Responses to White Actors' Behaviors Toward Black and White Targets

Daniel Yonas and Larisa Heiphetz Solomon

Department of Psychology, Columbia University

Correspondence concerning this article should be addressed to Larisa Heiphetz Solomon, Columbia University, 821 Uris Hall, 3022 Broadway Avenue, New York, NY 10027. Email: lah2201@columbia.edu

Word count: 10,651

In press, *Philosophical Psychology*

Daniel Yonas holds a PhD in Psychology from Columbia University. His research interests include social and moral development. He received his BA in Psychology and Cognitive Science in 2018 and his Masters of Education in 2020, both from the University of Virginia.

Larisa Heiphetz Solomon is Associate Professor of Psychology at Columbia University, where her lab investigates moral cognition and behavior among children and adults. She received her PhD in Psychology from Harvard University in 2013 and worked as a post-doctoral fellow at Boston College prior to joining Columbia in 2016.

Abstract

Moral philosophy and psychology often emphasize intent's role in moral evaluations but typically do not incorporate perspectives from the social psychology of intergroup relations. Therefore, current work on morality has limited capacity to understand how people evaluate behaviors in race-related contexts and how these judgments may vary across participant race. We report three studies examining how Black and White adults view intent in vignettes describing White actors' behaviors toward White, versus Black, targets. Participants indicated how moral behaviors were, how much they would punish behaviors, and how much perpetrators' intentions mattered. White participants prioritized intent more when evaluating White actors' behaviors toward Black, versus White, targets. This result occurred for both moral evaluations (Study 1) and punishment decisions (Study 2). When deciding how much to punish White actors' behaviors toward Black targets, White participants also prioritized intent more than did Black participants (Study 2). Finally, White participants showed less punitiveness overall than did Black participants (Studies 2-3). Together, these results clarify moral evaluations of White actors' behaviors toward Black versus White targets and, by incorporating perspectives from both Black and White participants, help philosophy and psychology speak to a greater diversity of human experience.

Keywords: intent; intergroup processes; morality; punishment; social cognition

Intent's Role in Evaluations of White Actors' Behavior Toward Black and White Targets

Before the Emancipation Proclamation ended legal slavery outside of the criminal legal system, some White thinkers argued that slavery was a public good because White people's intent was to use slavery to increase virtue among the Black population (Ford, 2009). During the Civil Rights Movement, famous segregationist George Wallace highlighted his positive intentions by stating that he "never said a word against Black people in [his] heart" (Montgomery & Wallace, 1992). At the beginning of Summer 2020, during which numerous Black Lives Matter protests took place throughout the United States, Amy Cooper (a White woman) called the police to falsely tell them that Christian Cooper (a Black man) was threatening her safety, later saying that she "did not mean to harm that man in any way" (Aguilera, 2021).

These individuals' focus on intent mirrors scholarship on moral evaluation. Some of this scholarship is descriptive, showing that adults generally exhibit more leniency in response to accidental rather than intentional transgressions (Ames & Fiske, 2013; Moran et al., 2011; Payir & Heiphetz, 2022). Other work is normative, arguing that people *should* consider intent when making moral evaluations (Gert, 1998; Hart, 1968/2008; Mikhail, 2007). Together, this scholarship has characterized a focus on intent as a hallmark of mature moral judgment. This focus makes sense in many contexts, such as when it leads to shorter prison sentences for manslaughter rather than murder. At the same time, much influential work on intent has focused on the interpersonal level, leaving open the question of how intent-based evaluation may vary depending on group memberships, particularly race. The current research examined moral evaluation among Black and White participants, thereby adding nuance to prior scholarship suggesting that adults largely prioritize intent when making moral judgments.

The Importance of Intent to Moral Evaluations

Modern philosophical frameworks provide multiple perspectives regarding the role that intent should play in moral evaluations and punishment decisions. Consequentialists evaluate moral goodness or badness on the basis of an action's outcomes, while deontologists typically place relatively strong emphasis on intent, arguing that intentional harms are more wrong than accidental and foreseen harms (e.g., Bentham, 1789; Masek, 2010; McMahan, 1994; Mill, 1861; Quinn, 1989). The bulk of moral philosophy focusing on intent makes the normative claim that harms committed with innocent intent should elicit less severe responses than purposeful transgressions (e.g., Gert, 1998; Hart, 1968/2008; Mikhail, 2007). Indeed, some early (e.g., Kant, 1790) and modern (e.g., Alexander et al., 2024; Shafer-Landau, 2018) philosophers argue that intent is the primary determinant of morality. Kant (1790) summarized this view by noting that punishment should be allocated on the basis of "internal wickedness."

Consistent with this theorizing, moral psychology has provided evidence that, by adulthood, most individuals consider intent when making moral evaluations. For instance, in one study (Young & Saxe, 2011), adults judged that it was more morally wrong to feed someone peanuts if the actor knew, versus did not know, that the person had a peanut allergy. Intent is such an important contributor to adults' moral evaluations that adults sometimes blame actors for harmful intentions even when those intentions are irrelevant to the harmful outcome (Inbar et al., 2012; Pizarro et al., 2003).

However, a growing body of work in fields adjacent to moral psychology shows that the moral weight of intent can be context-dependent. For example, in a large cross-cultural project, Barrett and colleagues (2016) found variation across diverse societies (e.g., Tanzania, Namibia, Indonesia, United States) in the degree to which intentions mitigate negative moral evaluations. In a separate line of cross-cultural work, McNamara and colleagues (2019) observed greater

emphasis on intent among participants in the United States and Canada than among Fijians from Yasawa Island, a culture that holds normative prohibitions against attempting to infer others' mental states. Indeed, people often treat foreseen-but-unintended bad outcomes as intentional (Knobe, 2003), blurring the line between the weight placed on the intent underlying a specific behavior and beliefs about the wrongness of that behavior. Moral inferences about the actor's character further color judgments of harm in ambiguous or accidental cases (Nadler, 2012; Schwartz et al., 2022).

The empirical studies reviewed above are descriptive, but they carry normative implications because sensitivity to intent is not uniform across people. For example, children begin to prioritize intent during the elementary school years, with preschoolers largely responding to outcome more than to intent (e.g., Cushman et al., 2013; Killen et al., 2011; Piaget, 1932/1965). Although this work is facially descriptive, research on development often positions the adult state as ideal, describing children's responses as a "failure" when they differ from adults' responses or researchers' intuition and as a "success" when they do not differ, even in cases where there are no objectively correct answers to the questions children are answering (e.g., Brey & Shutts, 2015; Comishen et al., 2019; Dunn et al., 1991). Thus, results showing that adults prioritize intent more than do children can imply that considering intent is a normatively ideal feature of appropriate moral evaluations.

In sum, portions of the moral philosophy and moral psychology literature align to suggest that most adults show sensitivity to intent when making moral evaluations and that they *should* evaluate in this way. However, this past work has not typically considered the role of racial group membership, which may inform the degree to which intent actually does (and perhaps

should) influence moral evaluations. Race has played a larger role in work on bias, to which we turn next.

Integrating the Study of Morality and Work on Racial Groups

Empirical work on moral cognition and behavior has typically focused on the inter*personal* level. For instance, such studies have asked participants whether certain behaviors are okay or not okay without attending to the group memberships of the actor, the target of the actor's behavior, or the participant (e.g., Cushman et al., 2013; Killen et al., 2011; Payir & Heiphetz, 2022; Young & Saxe, 2011).

However, many moral transgressions occur on the basis of group membership. The current work focused on race, comparing Black and White individuals, in part because Black people in the United States often experience—and White people in the United States often enact—racially disparate behaviors whose intent is unclear (Carter & Murphy, 2015). Additionally, Black and White individuals disagree about what constitutes racism and how prevalent it is (Carter & Murphy, 2015; Nelson et al., 2013). Indeed, some philosophical work suggests that intergroup cognition and moral evaluations are closely intertwined. According to this work, racist ideologies are to be critiqued not just on epistemic grounds (the ideologies make factually inaccurate claims) but also on moral grounds (the ideologies are unjust; Haslanger, 2017; Shelby, 2014), and it is morally wrong to hold racist beliefs (Basu, 2018). To more fully integrate moral psychology and work on groups, the current research asked how Black and White participants consider the intent underlying White actors' behaviors toward White, versus Black, targets.

We focused on intent because past work conducted at the interpersonal level suggests that transgressors are more likely than victims to prioritize their own innocent intentions. For example, adults make more spontaneous references to their own intent when describing a time

when they hurt someone rather than a time when they made someone else happy (Schutz & Baumeister, 1999). Further, both children and adults tell narratives that focus on intent when describing cases where they hurt someone else but focus on intent to a smaller degree when describing instances when someone else hurt them (Pasupathi & Wainryb, 2019). When evaluating interpersonal behaviors, adults also consider the relationship between the actor and the target. For instance, adults hold different moral expectations of behaviors that occur within different relationships (e.g., between romantic partners versus between housemates) and judge behaviors more harshly when they undermine the central functions of a particular relationship (Earp et al., 2021). Across diverse cultures, adults also deliver harsher moral evaluations of transgressions committed by socially distant, versus socially close, individuals (McKee et al., 2023). To the extent that sharing a racial group membership may impose relational norms (e.g., avoiding harm; Rhodes & Chalik, 2013) or social closeness, this membership may also shape moral evaluations. The present studies build on this past work to ask how participants' responses might vary across racial groups.

The bulk of past theorizing (e.g., Gert, 1998; Hart, 1968/2008; Mikhail, 2007) and empirical research (e.g., Cushman et al., 2013; Killen et al., 2011; Moran et al., 2011; Payir & Heiphetz, 2022; Piaget, 1932/1965; Young & Saxe, 2011) on the role of intent in moral judgment has not considered racial group membership. Further, work on groups has not drawn from moral psychology and philosophy to understand how racial group membership might shape moral evaluations. To the extent that research on groups has considered intent, it has not focused directly on morality. For instance, in one line of work (Swim et al., 2003), men perceived unintentional sexual harassment as less *discriminatory* than did women. However, this dependent measure targeted participants' understanding of a particular behavior (did the action constitute

discrimination against people on the basis of gender?) rather than their moral evaluations of the behavior or the person who performed the behavior.

Recent scholarship also suggests that intent attributions themselves can be racialized in ways that mute recognition of harm. In a series of four studies, Mosley (2024) found that when White observers attributed benign motives to culturally appropriative acts (e.g., framing a dominant-group actor as simply "appreciating" another group's styles), they judged the same harmful outcomes as less wrong and deserving of punishment. These effects emerged even when outcomes were clearly harmful and intent manipulations were unambiguous, suggesting that positive intent attributions can function as a strategic buffer for dominant-group actors. Extending this prior work, the current research investigated the extent to which Black and White participants emphasized intent when making moral evaluations about actions that occurred in group contexts but did not constitute cultural appropriation.

The Present Research

The present studies leveraged insights from moral psychology, moral philosophy, and the science of groups in order to better understand how individuals think about intent in contexts related to racial groups. Study 1 tested the extent to which Black and White participants emphasized intent when making moral evaluations of White actors' behaviors toward Black, versus White, targets. Study 2 asked how participants' punishment decisions may be similar to, or different from, the moral evaluations observed in Study 1. Finally, Study 3 manipulated intent to gain more insight into causality.

Studies 1-2 differ from an approach more common in psychological research—namely, providing unambiguous information about intent and manipulating whether this information is positive or negative (e.g., Ames & Fiske, 2013; Young & Saxe, 2011). We asked participants to

indicate how much intent mattered to them for three main reasons. First, this approach increases ecological validity because in everyday life, people rarely explain their intentions to others. Second, our dependent measure allowed us to determine how much adults think intent matters to them, which is not possible to determine by manipulating information about intent and measuring moral evaluations. Third, leaving intent ambiguous while varying the target's group membership allowed us to determine whether participants place more emphasis on an actor's possibly benign intent in cases where White actors perform behaviors toward Black versus White targets. If group context does not matter, then responses should not differ across conditions. However, it is possible that White participants may seek to exculpate ingroup members when those individuals' behaviors suggest the possibility of racism. Many individuals perceive racism as immoral (Buchanan & Goff, 2020; Dovidio et al., 2017; Siena College Research Institute, 2021) and seek to maintain a positive image of the groups to which they belong (Leach et al., 2007). Thus, when confronted with potentially racist behaviors performed by in-group members, White participants may be especially likely to emphasize intent—which may be benign and therefore exculpatory. Presenting participants with situations where intent was ambiguous allowed us to test between these possibilities. Study 3 built on Studies 1-2 by manipulating intent and therefore gaining greater insight into causality.

Study 1

Study 1 investigated how Black and White adults judge White actors' behaviors toward Black, versus White, targets. We pre-registered this study at https://aspredicted.org/JRP_N7P.

We asked participants about cases where a person could have acted with either biased or non-biased intent in order to probe the extent to which participants reported caring about intent when making moral evaluations. For example, participants read vignettes in which a store clerk

exhibited suspicion of a Black customer or a flight attendant double-checked the boarding pass of a Black passenger sitting in a first-class seat. These vignettes reflect behaviors that Black people often encounter from White people in everyday life (Carter & Murphy, 2015; Sue & Spanierman, 2020)—behaviors where the intent underlying the action is unclear. Thus, our data were able to speak to people's lived experiences. In addition to asking the types of close-ended questions typical in moral psychology studies, we also included an open-ended item asking participants why they responded as they did to gain a more in-depth understanding of participants' reasoning.

Method

Participants. Our final sample included 81 White adults (M_{age}=37.60 years, SD_{age}=14.77 years; 58 female, 20 male, 3 other) and 85 Black adults (M_{age}=38.95 years, SD_{age}=15.39 years; 55 female, 27 male, 3 other) who lived in the United States.[1] We determined sample sizes prior to collecting data using G*Power, indicating 80% power for a two-tailed *t*-test and estimating an effect of Cohen's *d*=.50. We chose this effect size based on prior work that observed a large effect (*d*=.82) when measuring judgments of ambiguous harms (Corning & Bucchianeri, 2010); to be conservative, we powered for a somewhat smaller effect size.[2]

In addition to collecting the demographic data described above, we probed the participant's highest level of educational attainment: 1% had not completed high school, 20% held a high school diploma or GED, 27% had some college experience without a degree, 6% had earned a 2-year associate's degree, 27% held a 4-year bachelor's degree, 3% had begun but not

[1] Racial demographics are based on participants' self-identification when setting up their Prolific profiles.
[2] Post-hoc sensitivity analyses showed that our sample was sufficient for the main analyses of interest. For example, the smallest effect we could reliably detect for the difference between White and Black participants was *d*=.44, and the actual effect was *d*=.47.

completed graduate study, 12% had completed a master's degree, and 2% had attained a doctoral or professional post-graduate degree (e.g., JD, MD).[3]

In addition to the participants described above, we tested and excluded 24 other respondents from these analyses for failing to correctly answer attention check questions that asked participants to briefly summarize 2 of the vignettes that they viewed.[4] We recruited participants via Prolific, collected their data through Qualtrics, and paid them $10 per hour.

Procedure. We presented participants with a series of vignettes in randomized order depicting two types of situations: cases where White actors interacted with Black targets (e.g., a White person moving to a different place on the bus when a Black couple sat next to them) and cases where White actors interacted with White targets (e.g., a White bouncer asking a White person to pay a fee before entering a club). We drew all vignettes from prior work on behaviors that laypeople may perceive as racist when White people are the actors and Black people are the targets of their behaviors (Corning & Bucchianeri, 2010; Heiphetz & Craig, 2021). However, to provide a comparison condition, we altered half of the vignettes to reflect intragroup rather than intergroup behaviors (e.g., rather than having a White bouncer asking a Black person to pay a fee, we had a White bouncer ask a White person to pay a fee). See the appendix for the full text of all vignettes.

[3] At the suggestion of a reviewer, we examined the extent to which educational attainment might confound our race-based findings. To do so, we coded attainment on an eight-point ordinal scale (1=some high school through 8=doctoral/professional degree) and computed correlations between educational attainment and each of our primary dependent measures (moral evaluations of behaviors toward White targets, moral evaluations of behaviors toward Black targets, intent emphasis). These exploratory analyses yielded non-significant correlations in this and subsequent studies ($ps>.10$), suggesting that educational background did not reliably predict participants' responses. Consequently, the racial differences we report are unlikely to be driven by variation in educational attainment.

[4] Including all participants in the analyses led to a non-significant difference between Black and White individuals with regard to emphasis on intent when evaluating White actors' behaviors toward Black targets ($p=.153$). However, 24 of the respondents that did not answer the attention-check question also did not respond to many of substantive measures, thus reducing the sample size, which could account for the non-significant effect.

After reading each vignette, participants first answered three counterbalanced questions that investigated their moral evaluations: "In your mind, how [wrong or right] [bad or good] [mean or nice] is what [actor] did?" (Cronbach's α=.98). To answer these questions, participants used Likert-type scales ranging from 1 (very wrong, very bad, very mean) to 7 (very right, very good, very nice). After the last of these questions, participants responded to an open-ended item: "In a few words, please describe why you answered the way you did. For instance, why did you think the [actor]'s behavior was wrong or right?"

After answering all of the questions above for each vignette, participants read each vignette a second time, again in a randomized order. This time, they indicated the importance of intent to their moral evaluations using three counterbalanced questions asking them how important X was, where X was "what [actor] meant to do", "what [actor] was trying to do," or "what [actor's] intentions were" (Cronbach's α=.99). Participants responded to each item using Likert-type scales ranging from 1 (not at all important) to 7 (very important). Some past work probing the importance of intent to moral evaluations has specified whether each behavior was intentional or accidental (e.g., Cushman et al., 2013; Killen et al., 2011; Payir & Heiphetz, 2022; Young & Saxe, 2011). In Study 1, we instead left intent ambiguous and asked participants how much the perpetrator's intent mattered to them. As discussed above, this approach increased ecological validity and allowed us to probe extent to which participants explicitly believe that intent is important. Study 3 used a design more similar to past work by manipulating intent.

We asked all questions about intent after all questions about moral evaluations because we did not want to imply to participants that they should be thinking about intent. After answering all experimental items, here and in all subsequent studies, participants responded to an

attention check question requiring them to briefly summarize two of the vignettes they had read over the course of the study.

After completing data collection, we coded open-ended responses for references to intent as well as other themes that emerged during coding (e.g., environmental concerns, occupational responsibility). The first author read all open-ended responses to develop codes other than intent based on emergent themes; see Table 1 for a summary. A second researcher then coded 15% of responses, and the two coders met to discuss their codes. In cases where confusion arose, the first author amended the codebook to add greater clarity. Subsequently, the second coder coded the remaining responses with no further input from the first author. The two coders achieved acceptable reliability with their codes on this portion of the data (Cronbach's κ =0.73). Disagreement was resolved via discussion.

Researchers coded each response separately for the presence or absence of each theme. For instance, if a given response referenced racism, it received a 1 for that theme; if it did not reference racism, it received a 0. If that same response also referenced intent, it also received a 1 for the theme focused on intent; if it did not reference intent, it received a 0 for that theme. Therefore, we did not include a code indicating that the response did not include any themes; such a response would simply receive a 0 for each theme. Based on this coding scheme, a given response could receive a 1 for none, only one, some, or all of the themes listed in Table 1.

Theme	*Coding rationale*	*Example response*
Racism	Reference to racism, prejudice, or the race of the characters in the vignette	"Why did she only ask the black man to check his assigned seat again before takeoff? It seems as if she was somewhat put off by the idea that a black man could be seated in first class."
Intent	Reference to the intent or lack of apparent intent underlying the harm	"While the lady may have moved because of personal space issues or as a courtesy, she could have also moved because of racial bias. No matter why she moved, if she didn't say anything, it will be perceived as racial bias."
Occupational responsibility	Reference to a stated profession, obligation, or paid duty	"I think the guard just was performing their job."
No perceived harm	The participant did not perceive any harm at all	"The lady's behavior was right. She greeted the potential customers."
General (non-targeted) harm perceived	The participant perceived that a harm has occurred but made no claims about intent or potential prejudice, or made reference to any contextual factors	"The waiter is wrong on so many levels. Since they saw them there, they should have told the waiter for that section they had a customer or better yet; just take their order."
Agentic characteristics	Reference to a personal quality or experience of the agent that contextualized their behavior as non-harmful in a specific circumstance	"It was a little mean for him not to let the boy pet the dog but he was just protecting him."
Environmental concern	Reference to potential risks to other persons and that affected the bounds of acceptable behavior and mitigated moral evaluations	"The dog might be dangerous and bite the child."
Lack of information	Reference to a lack of information about circumstances and contextual information	"There's not enough info to decide."

Table 1. Qualitative themes that emerged in coding alongside rationale and example responses from participants that embody these themes, Study 1.

Results

To determine participants' moral evaluations of the behaviors with which we presented them, we averaged their responses to the three moral evaluation questions above and entered this average into a 2 (Participant Race: Black vs. White) x 2 (Target Race: Black vs. White) mixed

ANOVA with repeated measures on the second factor. This analysis revealed a main effect of Target Race, $F(1, 164)=441.11$, $p=.001$, $\eta_p^2=.73$. Overall, participants reported that White actors' behaviors toward Black targets ($m=3.29$, $sd=.62$) were less morally right than their behaviors toward White targets ($m=4.57$, $sd=.60$). The main effect of Participant Race, $F(1, 164)=1.07$, $p=.302$, $\eta_p^2=.01$, and the Participant Race x Target Race interaction, $F(1, 164)=2.97$, $p=.086$, $\eta_p^2=.02$, did not reach significance.

To determine the extent to which participants reported that the actor's intent mattered to their moral evaluations, we averaged their responses to the three intent questions above and entered this average into a 2 (Participant Race: Black vs. White) x 2 (Target Race: Black vs. White) mixed ANOVA with repeated measures on the second factor. This analysis revealed a main effect of Target Race, $F(1, 164)=40.24$, $p=.001$, $\eta_p^2=.20$, a main effect of Participant Race, $F(1, 164)=4.64$, $p=.033$, $\eta_p^2=.03$, and a Participant Race x Target Race interaction, $F(1, 164)=11.60$, $p=.001$, $\eta_p^2=.07$ (Figure 1). To further investigate this interaction, we conducted two sets of tests. First, we compared the extent to which Black participants and, separately, White participants emphasized intent when judging White actors' behaviors toward Black, versus White, targets. White participants prioritized intent more when targets were Black ($m=5.59$, $sd=1.03$) versus White ($m=4.84$, $sd=.69$, $t(80)=48.89$, $p=.001$, $d=.86$, 95% CI [5.36, 5.81]). Black participants similarly prioritized intent more when targets were Black ($m=5.06$, $sd=1.18$) versus White ($m=4.84$, $sd=.71$), $t(84)=39.65$, $p=.001$, $d=.23$, 95% CI [4.81, 5.32].[5] Second, we compared the extent to which Black versus White participants emphasized intent when judging White actors' behaviors toward Black and, separately, White targets. When evaluating behaviors

[5] Post-hoc sensitivity analyses showed that we had appropriate power to detect the effect among White, but not Black, participants; the smallest effect we could reliably detect was approximately $d=.31$ in each group. Additionally, Study 2 replicated the results among White, but not Black, participants. Therefore, this finding for Black participants should be interpreted with caution.

toward Black targets, White participants (*m*=5.59, *sd*=1.03) prioritized intent more than did Black participants (*m*=5.06, *sd*=.69), *t*(1,164)=3.05, *p*=.003, *d*=.47, 95% CI [.18, .86]. This effect did not emerge when targets were White, *t*(1,164)=0.01, *p*=.993, *d*=.001, 95% CI [-.22, .22].

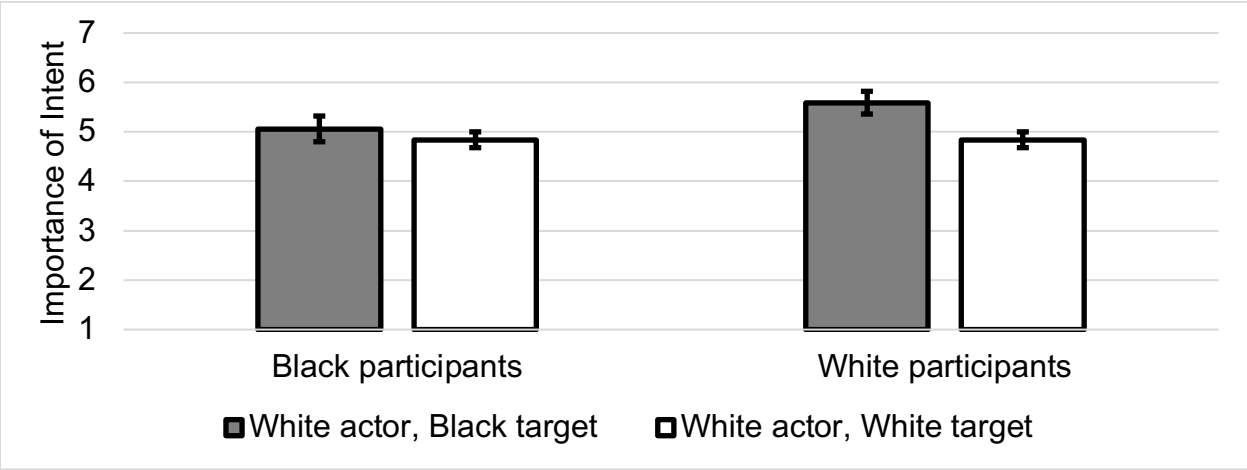


Figure 1. Black and White adults' responses indicating the importance of intent when making moral evaluations about White actors' behaviors toward Black versus White targets, Study 1. Error bars indicate 95% confidence intervals.

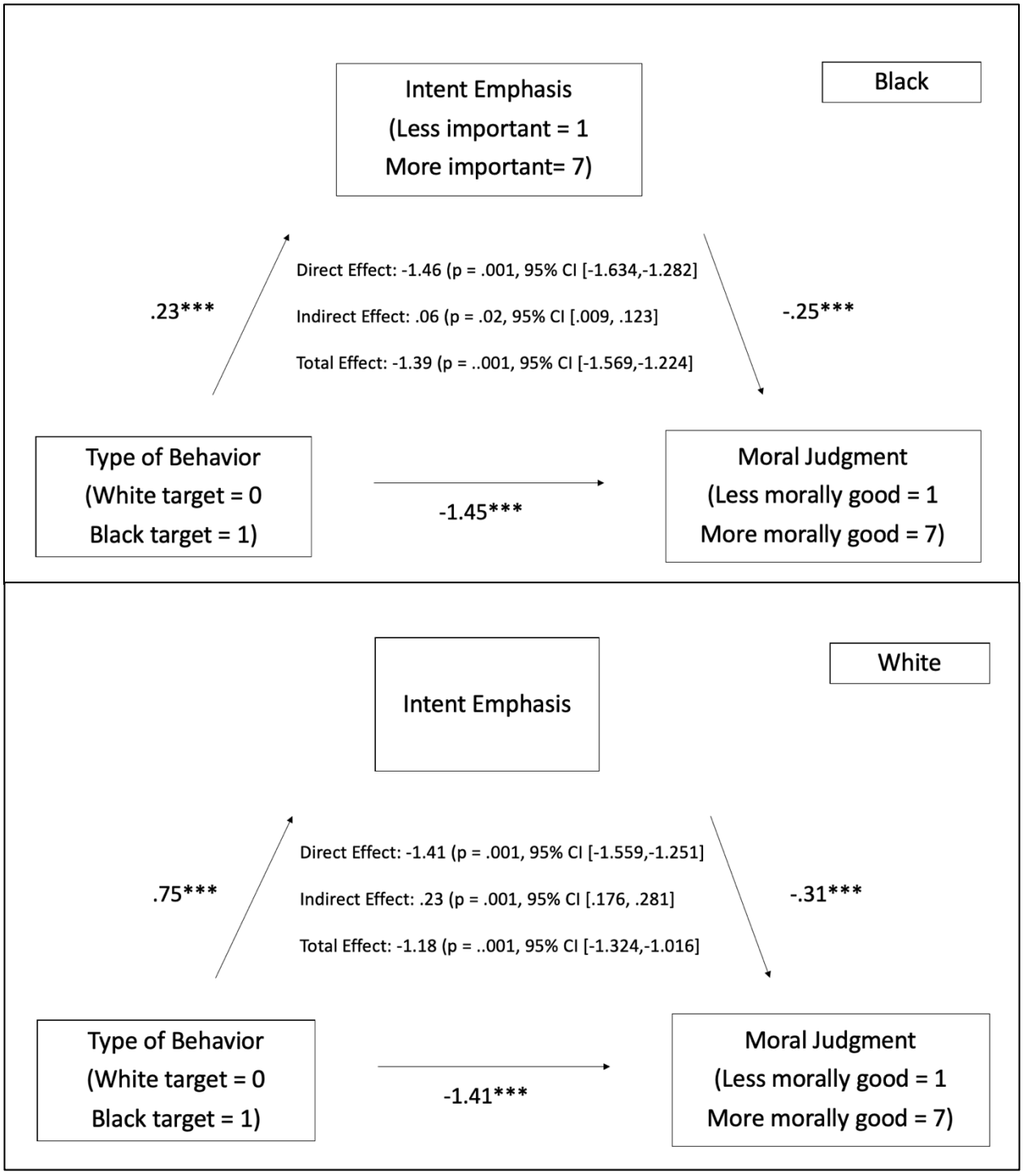


Figure 2. Mediation analysis for the relationship between behavior type, intent emphasis, and moral evaluations for White (top) and Black (bottom) participants, Study 1.

To probe the relation between moral evaluations and the importance participants reported placing on intent, we conducted two sets of analyses. First, we correlated these two variables. The more participants prioritized intent, the more positive were their moral evaluations of White actors' behaviors toward White targets, $r(164)=.66$, $p=.001$. However, this effect did not emerge for behaviors toward Black targets, $r(164)=.06$, $p=.468$. Second, we conducted a mediation analysis with 5,000 bootstrapped samples (Hayes, 2017). This mediation included Behavior Type (0=White actors-White targets, 1=White actors-Black targets) as the predictor variable, the degree to which participants responded that intent was important in their moral evaluations as the mediator, and participants' moral evaluations as the outcome variable.[6] This mediation model revealed a significant indirect effect for both Black participants, $B=-.23$, $p=.001$, 95% CI [-.18, -.28], and White participants, $B=.06$, $p=.020$, 95% CI [.01, .12] (Figure 2).

In addition to seeing whether responses to our close-ended intent measure mediated the relation between the behavior type to which participants responded and their moral evaluations, we also asked whether participants' spontaneous references to intent served as an additional mediator. To answer this question, we entered the presence or absence of spontaneous references to intent in response to the open-ended measure, rather than responses to the close-ended intent measure, as the mediator in the mediation model described above. This model did not reveal a significant indirect effect for either White participants, $B=-.001$, 95% CI [-.08, .08], or for Black participants, $B=.001$, 95% CI [-.02, .03].

Discussion

[6] The pre-registered mediation included Participant Race (0=White, 1=Black) as the predictor variable, emphasis on intent for White actors' behaviors toward Black targets as the mediator, and moral evaluations about those behaviors as the outcome variable. This mediation model did not reveal a significant indirect effect ($B=-.01$, 95% CI [-.05, .04]). The mediation models reported in the main text provide a better test of the main question of Study 1, i.e., whether the emphasis participants place on intent underlie different moral evaluations of White actors' behaviors toward Black versus White targets.

Overall, Study 1 found that while neither behavior type nor participant race significantly predicted intent emphasis on their own, they did interact. Specifically, White (versus Black) participants emphasized intent more when judging White actors' behaviors toward Black targets—a difference that did not emerge for evaluations of behaviors toward White targets. While Black and White participants placed differential importance on intent when evaluating White actors' behaviors toward Black targets, this difference did not carry over into their moral evaluations. Study 2 built on these results to investigate decisions regarding punishment.

Study 2

In Study 1, White participants placed more emphasis on a White actor's intent than did Black participants when the actor performed behaviors toward Black targets. Further, the extent to which both Black and White participants viewed an actor's intent as important underlay moral evaluations of White actors' behaviors toward Black targets. Although Black and White participants placed different levels of importance on intent, this difference did not appear to influence moral judgments, as Black and White participants did not provide significantly different moral evaluations.

Prior work suggests that different moral responses do not always cohere. For instance, while intent commonly matters more than outcome to adults making moral evaluations regarding wrongness and blame, intent and outcome jointly shape punishment decisions (Cushman, 2008; Kneer & Machery, 2019; Malle, 2021). To provide fuller insight into the range of participants' moral responses, Study 2 asked participants how much punishment they thought actors should receive. We preregistered this study at https://aspredicted.org/JRP_N7P.

Participants. Our final sample included 130 White adults (M_{age}=40.13 years, SD_{age}=14.20 years; 80 female, 42 male, 8 other) and 120 Black adults (M_{age}=37.00 years,

SD_{age}=13.31 years; 58 female, 60 male, 2 other) who lived in the United States. We determined sample sizes prior to collecting data using G*Power, indicating 80% power for a two-tailed *t*-test and estimating an effect of Cohen's *d*=.50.[7]

We recruited participants via Prolific and paid them $2.50 for this 10-minute study. As in Study 1, we collected data on participants' highest level of education: 1% had not completed high school; 13% held a high school diploma or GED; 26% had some college without a degree; 11% had earned a 2-year associate's degree; 34% held a 4-year bachelor's degree; 4% had some graduate-level coursework; 8% had completed a master's degree; and 2% had attained a doctoral or professional post-graduate degree (e.g., JD, MD). In addition to the participants described above, we tested an additional 40 adults that we excluded from analyses for failing to answer attention check questions.[8]

Procedure. The procedure was identical to Study 1, with one exception: we replaced the moral evaluation items in Study 1 with items measuring participants' punishment decisions. These new items read as follows: "In your mind, how [severe should the consequences be] [much should X be punished] [much trouble should X be in]?" Participants responded using Likert-type scales ranging from 1 (not at all severe, a little, a little) to 7 (very severe, a lot, a lot). These items showed acceptable reliability (Cronbach's α=.97).

Results

Mirroring the procedure from Study 1, we averaged participants' responses to the three punishment decision questions above and entered this average into a 2 (Participant Race: Black

[7] As in Study 1, post-hoc sensitivity analyses showed that our sample size had appropriate sensitivity to detect effects of interest (e.g., our analyses could reliably detect Cohen's *d* as small as .25 for a within-participants analysis comparing the extent to which White participants prioritized intent for White versus Black targets, and this effect size was smaller than the one we actually observed).

[8] When analyzing data from all respondents, we observed the same patterns of results as those reported in the main text.

vs. White) x 2 (Target Race: Black vs. White) mixed ANOVA with repeated measures on the second factor (Figure 3). This analysis revealed a significant main effect of Participant Race, $F(1, 248)=11.50$, $p=.001$, $\eta_p^2=.04$. Overall, White participants ($m=1.82$, $sd=.71$) punished all harms less severely than did Black participants ($m=2.15$, $sd=.78$). This analysis also revealed a significant effect of Target Race, $F(1, 248)=8.20$, $p=.005$, $\eta_p^2=.03$. Behaviors toward Black targets ($m=2.42$, $sd=1.02$) elicited less severe punishment decisions than did behaviors toward White targets ($m=2.80$, $sd=.82$). However, the Participant Race x Target Race interaction did not reach significance, $F(1, 248)=2.75$, $p=.099$, $\eta_p^2=.01$.

To determine the extent to which participants reported that the transgressor's intent mattered to their punishment decisions, we averaged their responses to the three intent questions above and entered this average into a 2 (Participant Race: Black vs. White) x 2 (Target Race: Black vs. White) mixed ANOVA with repeated measures on the second factor. These analyses revealed no main effect of Participant Race, $F(1, 248)=.04$, $p=.833$, $\eta_p^2=.00$, or Target Race, $F(1, 248)=2.07$, $p=.152$, $\eta_p^2=.01$. However, as in Study 1, we did observe a significant Participant Race x Target Race interaction, $F(1, 248)=8.20$, $p=.005$, $\eta_p^2=.03$.

To further investigate this interaction, we conducted two sets of tests. First, we investigated how White and, separately, Black participants responded to Black, versus White, targets. As in Study 1, White participants prioritized intent more when White actors behaved toward Black ($m=4.69$, $sd=1.24$), versus White ($m=4.76$, $sd=1.46$), targets, $t(247.74)=-2.10$, $p=.037$, $d=.26$. However, Black participants did not differ in the extent to which they prioritized intent for Black ($m=4.84$, $sd=1.48$), versus White ($m=4.96$, $sd=1.50$), targets, $t(237.95)=.60$, $p=.548$, $d=.08$. Second, we compared how White versus Black participants evaluated behaviors toward Black and, separately, White targets. Again as in Study 1, Black and White participants

did not differ when evaluating behaviors toward White targets, $t(243.89)=-1.04$, $p=.301$, $d=.13$. Unlike Study 1, Black and White individuals also did not differ when evaluating behaviors toward Black targets, $t(233.03)=1.58$, $p=.115$, $d=.20$.

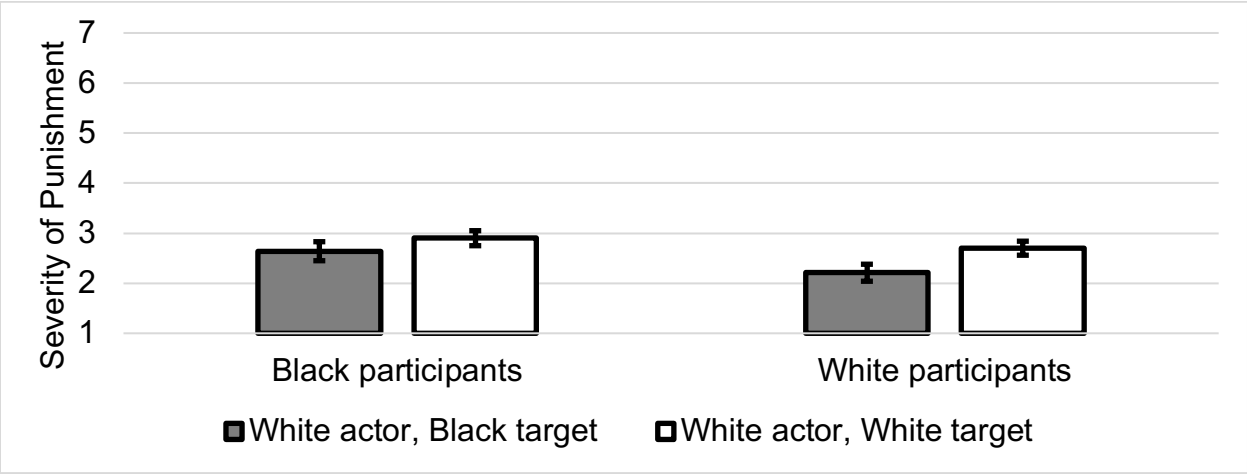


Figure 3. Black and White adults' punishment decisions, Study 2. Error bars indicate 95% confidence intervals.

Similarly to Study 1, we conducted a mediation analysis with 5,000 bootstrapped samples.[9] This mediation included Behavior Type as the predictor variable, emphasis on intent as the mediator, and punishment decisions as the outcome variable. This mediation model did not reveal significant indirect effects for either Black participants, $B=-.001$, $p=.800$, 95% CI: [-.001, .001], or White participants, $B=.001$, $p=.920$, 95% CI: [-.003, .001].

Finally, we coded open-ended responses for references to intent as well as other themes that emerged during coding. We entered the intent-based codes (0=responses did not reference intent, 1=response did mention intent) as a mediator in a model where Participant Race served as the predictor variable and punishment decisions regarding behaviors toward Black targets,

[9] Similarly to Study 1, the pre-registered mediation included Participant Race (0=White, 1=Black) as the predictor variable, emphasis on intent when considering White actors' behaviors toward Black targets as the mediator, and punishment decisions (as compared to moral evaluations in Study 1) about these behaviors as the outcome variable. This mediation model did not reveal a significant indirect effect ($B=.02$, 95% CI [-.08, .02]).

averaged across the three items above, served as the outcome variable. Again, we found no significant indirect effect, B=-.04, p=.673, 95% CI [-.13, .04].

Discussion

Study 2 replicated Study 1 by showing that White—but not Black—participants placed more emphasis on intent when evaluating White actors' behaviors toward Black, rather than White, targets. That is, White participants perceived intent as particularly important when White people behaved toward Black people in ways that could have constituted unintentional harm. In contrast, the amount of importance Black participants placed on intent did not differ depending on whether targets were Black or White.

Study 2 also extended Study 1 by investigating participants' punishment decisions rather than their moral evaluations. Together, the results from these two studies suggest that participant race may matter differently for moral evaluations versus punishment decisions. Specifically, while moral evaluations did not differ across participant race in Study 1, White participants punished behaviors overall less severely than did Black participants in Study 2. This finding may point to a closer relationship between moral evaluations and punishment decisions among White rather than Black participants, although we did not predict this difference across studies and the results should be interpreted cautiously.

Study 3

Both Studies 1 and 2 found a significant interaction effect of participant race and behavior type such that White participants placed more emphasis on intent than did Black participants when responding to White actors' behaviors toward Black targets. However, the degree to which participants reported that intent was important to them did not necessarily appear to drive their moral evaluations (Study 1) or punishment decisions (Study 2).

One possibility is that intent does shape participants' responses in some way but that participants do not report this influence. People are often unaware of their cognitive processes (Nisbett & Wilson, 1977), and participants may not know the extent to which they prioritize intent. Studies 1 and 2 provide valuable information on what people say about how important intent is to them, a consequential topic to investigate because it reveals information about people's self-perceptions that could not be obtained with more surreptitious measures and because it taps into processes that occur in everyday life, where people often lack clear information about others' intentions. Study 3 built on these results by presenting participants with intentional versus unintentional behaviors and measuring their moral responses. This approach clarifies the extent to which intent influences responses (if intent matters to participants, they should respond differently to unintentional versus intentional acts) without relying on participants' self-reports of the importance they place on intent.

Because Studies 1 and 2 revealed significantly different degrees of emphasis for intent only when White actors behaved toward Black targets, this was the only condition tested in Study 3. We preregistered Study 3 at https://aspredicted.org/31Q_TWK.

Participants. Our final sample included 157 White adults (M_{age}=41.73 years, SD_{age}=14.20 years; 79 female, 77 male, 1 other) and 152 Black adults (M_{age}=37.07 years, SD_{age}=11.83 years; 73 female, 77 male, 2 other) who lived in the United States. We determined sample sizes prior to collecting data using G*Power, indicating 80% power for a two-tailed *t*-test and estimating an effect of Cohen's *d*=.50.

Adults were recruited and completed the study via Prolific and received compensation at a rate of $10 per hour. Participants in our sample spanned the full spectrum of educational attainment: 2% had not completed high school, 16% held a high school diploma or GED, 25%

had some college experience without a degree, 8% had earned a 2-year associate's degree, 36% held a 4-year bachelor's degree, 3% had begun but not completed graduate study, 12% had completed a master's degree, and 1% had attained a doctoral or professional post-graduate degree. In addition to the participants described above, we tested an additional 2 adults and excluded their data from the analyses because they failed to answer attention check questions correctly.[10]

Procedure. Participants were randomly assigned to either an intentional or unintentional condition. Participants in the intentional condition read a modified version of the vignettes from Studies 1 and 2 that explicitly stated the actor's intent. One example vignette read as follows:

A black couple came into a jewelry store and began looking at some of the jewelry on display. The lady behind the counter, who was white, then approached the couple and greeted them by asking, "Are you here to buy something today?" The black couple was upset by the lady's behavior. It turns out that the lady [meant/did not mean] to have this effect on the couple. She truly [did not want black people to shop at her store/wanted to be a helpful salesperson].

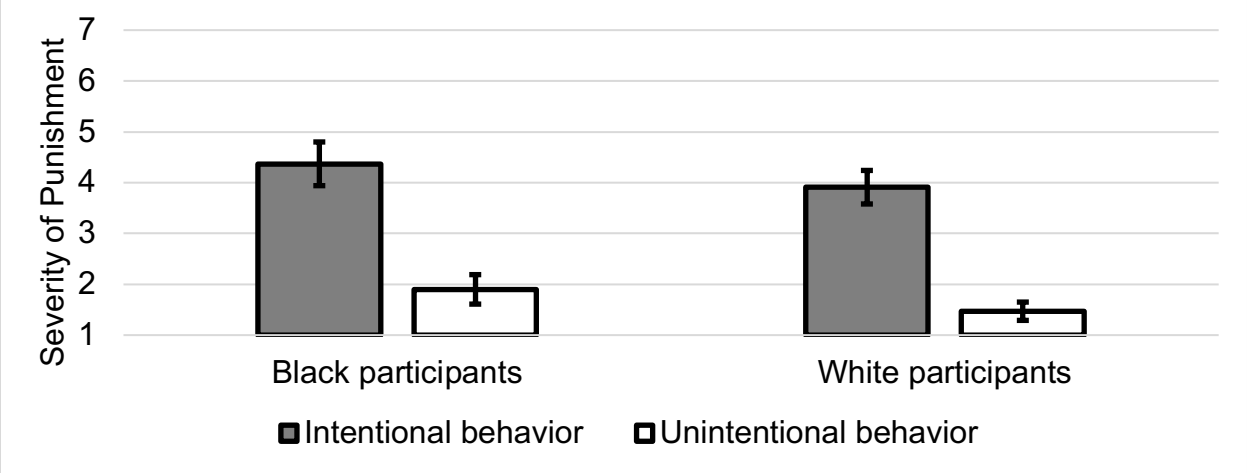
After reading each vignette, participants completed the punishment measure from Study 2. Finally, participants answered a manipulation check question about the degree to which the vignettes they had read represented individuals who were trying to generate negative reactions intentionally ("In this study, you read about people who acted in ways that upset others. To what extent did those people mean to upset others on purpose?") using a scale ranging from 1=not at all to 7=a lot. Overall, participants in the intentional condition (m=5.16, sd=2.53) perceived the character's actions as more intentional than did participants in the unintentional condition

[10] When analyzing data from all respondents, we observed similar patterns as those reported in the main text.

(m=1.59, sd=1.34), t(232.34)=15.48, p=.001, d=1.76, showing that our manipulation was effective.

Results

We entered participants' responses into a 2 (Participant Race: White vs. Black) x 2 (Intent: Intentional vs. Unintentional) ANOVA. Analyses revealed a main effect of Participant Race, F(1, 309)=10.05, p=.001, η_p^2=.03: Overall, as in Study 2, White participants punished behaviors less severely than did Black participants.[11] We also observed a main effect of Intent, F(1, 309)=320.19, p<.001, η_p^2=.51: Participants punished intentional behaviors more than unintentional behaviors.[12] Further, as in Study 2, the interaction between Participant Race and Intent did not reach significance, F(1, 309)=.02, p=.881, η_p^2=.00 (Figure 4).



[11] A post-hoc sensitivity analysis showed that our sample could reliably detect an effect of d=.32, and the actual observed effect was d=.16. Thus, although Study 3 provides significant evidence for a difference between White and Black participants, this difference on its own may be relatively unreliable. However, this effect is consistent with the fact that both Study 1 and Study 2 also showed differences between Black and White participants (e.g., in terms of the degree to which they emphasized intent in Study 1 and in terms of their punishment decisions in Study 2). The consistency of differences between Black and White participants across studies may therefore warrant somewhat greater confidence than would be warranted if this difference had only emerged in Study 3.

[12] A post-hoc sensitivity analysis showed that this effect size was bigger than the smallest effect that our sample could reliably detect.

Figure 4. Means for Black and White participants' punishment decisions for White actors' intentional versus unintentional behaviors toward Black targets, Study 3. Error bars indicate 95% confidence intervals.

Discussion

Study 3 extended the findings from Studies 1 and 2 by directly manipulating actors' intentions. Unlike the previous studies, which relied on participants' self-reports about the importance of intent, Study 3 focused on whether participants responded differently to intentional versus unintentional acts. Consistent with prior work in moral psychology (Ames & Fiske, 2013; Cushman et al., 2013), participants punished intentional behaviors more than unintentional behaviors. Building on this prior work, Study 3 adopted a group-based approach and showed that White participants may punish behaviors (all of which involved a White person's actions toward a Black person in this study) less severely than did Black participants. Although Study 2 did not reveal a significant difference between Black and White participants in terms of how much they reported that intent mattered to them when punishing White actors' behaviors toward Black targets, Study 3 showed that manipulating intent led to significantly different punishment decisions, on average, across participants from these racial groups. The relevant effect size in Study 3 was relatively small and fell below the value that could be reliably detected according to post-hoc sensitivity analyses. However, this effect is consistent with results showing racial differences in Studies 1-2. In conjunction, these studies suggest that intent may shape punishment decisions in ways that participants, on average, may not report.

General Discussion

The present work investigated how individuals think about intent when making morally relevant judgments about behaviors that occur without explicit intent information in group

contexts. In Study 1, White participants prioritized intent to a greater degree than did Black participants when evaluating White actors' behaviors toward Black targets, and intent emphasis mediated the relationship between target race and participants' moral evaluations. Study 2 built on these results to examine punishment decisions. Here, Black and White participants punished White actors less severely when they acted toward Black, versus White, targets. Further, White participants punished behaviors overall less severely than did Black participants. However, intent did not mediate the relation between behavior type and punishment decisions for either White or Black participants. Study 3 extended Studies 1-2 by manipulating intent to gain more insight into causality. Here, both Black and White participants punished White actors' intentional behaviors more when they were targeted toward Black, versus White, individuals. Further, White participants punished White actors' intentional behaviors toward Black targets less severely than did Black participants, although the effect size for this result was small and fell below the effect that the sample size in Study 3 could reliably detect. Together, these studies suggest that intent plays a different role in moral evaluations across racial groups: Black and White participants sometimes differed in the emphasis they placed on intent, and participants from both racial groups sometimes differed in their responses to White actors' behaviors toward Black, versus White, targets.

Implications for the Study of Morality

The current work has implications for ongoing conversations in philosophy and psychology regarding the role that intent plays in moral cognition. Philosophers have argued that people *should* consider intent (Masek, 2010; McMahan, 1994; Quinn, 1989), and psychologists have shown that many adults *actually do* evaluate behaviors differently depending on whether or not they were intentional (Ames & Fiske, 2013; Cushman et al., 2013; Payir & Heiphetz, 2022).

The current work suggests that the extent to which information about intent actually influences punishment decisions (Study 3) varies to some extent from the extent to which participants say that information about intent matters to them (Study 2). For instance, in Study 2, Black and White participants did not differ in reporting how much intent mattered to their decisions about how much to punish White actors whose behaviors targeted Black individuals. In Study 3, Black and White participants did deliver different punishment decisions from each other, although the effect size was small. Thus, the degree to which participants experience intent as mattering to them (Studies 1-2)—an important topic of study in its own right, as it sheds light on how people perceive themselves—may differ to some extent from the degree to which intent actually shapes decisions (Study 3).

Prior work has largely presented participants with explicit information about others' intentions (Cushman et al., 2013; Killen et al., 2011; Payir & Heiphetz, 2022), a practice that increases internal validity and allows for strong causal conclusions. At the same time, this approach may not reflect individuals' experiences in everyday life. Examining how people report on their own emphasis on intent therefore makes an important contribution to this literature.

While intent has appeared to shape both moral evaluations and punishment decisions among adults in prior work, these types of responses diverged in other ways; for instance, information about outcomes seems to matter less for moral evaluations such as judgments of how wrong a behavior was than for decisions about how much to punish that behavior (Cushman, 2008; Kneer & Machery, 2019; Malle, 2021). The present findings built on these past results by highlighting some similarities and some differences across moral responses. For instances, White participants placed greater emphasis on intent when responding to White actors' behaviors toward Black rather than White targets, regardless of whether those responses consisted of moral

evaluations (Study 1) or punishment decisions (Study 2). However, significant differences between White and Black participants emerged only for punishment decisions in Study 2, not for moral evaluations in Study 1, suggesting that participants' racial group membership may play a particularly strong role in the former case.

These results further extend prior work by incorporating a framework centering racial group membership. While past work on the role of intent in moral evaluation has focused on the interpersonal level of analysis, the current work suggests that considering racial group membership leads to a better understanding of moral cognition. For instance, in the current work, White participants sometimes placed more emphasis on intent than did Black participants when evaluating White actors' behaviors toward Black targets. This finding suggests that intent may matter differently depending on one's own social role. Similarly, participants sometimes evaluated White actors' behaviors differently when they were targeting White, versus Black, individuals. This finding indicates that the social roles occupied by actors and targets inform third-party observers' moral evaluations.

By arguing that intent should, and does, ameliorate negativity responses to transgressions, past work has adopted the perspective of White participants from western cultures to some extent. This focus is understandable given that many scholars of psychology and philosophy share these group memberships with participants (Does et al., 2018; Guthrie, 2004; Schwitzgebel et al., 2021). However, generalizing from western White participants to people more broadly limits scholars' ability to understand human cognition. It may be beneficial for future empirical work to consider group membership more centrally. Additionally, philosophical theorizing can engage with descriptive data showing group differences. For instance, given that White participants sometimes place more emphasis on intent than do Black participants (and given that

in prior work, participants from different cultures have emphasized intent to different extents; Barrett et al., 2016; McNamara et al., 2019), does it make sense to argue that people *should* emphasize intent—i.e., that people in general should behave like White participants from a specific set of countries)? Explicitly addressing this question will allow philosophical theories to engage more deeply with a greater diversity of human experience.

Implications for the Study of Groups

While work on morality has not considered race to the extent that perhaps it should, work on intergroup bias has also remained more separate from the literature on morality than would be ideal. Research in this area has shown that race shapes a number of morally relevant outcomes. For instance, people punish transgressors differently depending on the groups to which they belong, a phenomenon that our data also reflect by suggesting that White participants may punish White individuals less severely than do Black participants. Our results also dovetail with Mosley (2024), who explained cultural appropriation by arguing that White perceivers' attributions of benign intent mute their recognition of harm. Across four studies, benign-intent explanations led White participants to downplay the severity of intergroup harm even when outcomes were identical, mirroring our findings regarding White versus Black participants' responses. Taken together, the consistency across these findings suggest that intent attributions do not operate in a vacuum but are themselves racialized, routinely shielding dominant-group actors from scrutiny.

However, empirical work on groups has rarely made explicit connections to work on morality. Such a connection would be beneficial because, as work from psychology (e.g., Swim et al., 2003) and philosophy (e.g., Brownstein, 2016) indicates, many people consider group-based prejudice and discrimination to be immoral. Indeed, philosophers have made normative

arguments that racist ideologies and beliefs are unjust (Basu, 2018; Haslanger, 2017; Shelby, 2014). Building on this theorizing, the present work integrates psychological and philosophical approaches to argue that knowledge about moral cognition and behavior should inform scholarly understanding of intergroup relations.

The current work sought to create a stronger bridge between these areas of study in part by examining a focus on intent—a variable drawn from theorizing about morality—in conjunction with people's responses to White actors' behaviors toward Black versus White targets. We observed several group-based differences; for instance, on average, White and Black participants emphasized intent differently when responding to White actors, and participants from these different racial groups also reported that White actors should receive different degrees of punishment. Without considering group membership, theorizing on the role of intent in moral cognition and behavior would be incomplete.

Limitations and Future Directions

The current work made novel strides in integrating work from moral philosophy and the psychology of groups. However, like all work, the present study did not address all possible questions.

The present studies examined Black and White participants' responses to various behaviors. One benefit of our sampling strategy is that it leveraged the unique experiences of these specific racial groups. For example, many individuals may identify with the experience of being served last a restaurant despite arriving earlier than other patrons, but Black individuals may identify with situations like these particularly strongly because of their prevalence and prototypicality in reflecting racial microaggressions (e.g., Carter & Murphy, 2015). However, Black and White adults do not comprise the entirety of the United States, and future work can

test the extent to which the effects reported here generalize to other groups. One possibility is that individuals are particularly likely to de-emphasize the intent underlying behaviors that they are particularly likely to experience. For instance, Asian individuals are more likely than Black individuals to face questions about where they are "really" from (Cheryan & Monin, 2005) and may therefore place less emphasis than do Black individuals on the intentions of people asking these types of questions. Alternatively, routinely experiencing racism may lead individuals to prioritize harmful outcomes over benign intentions regarding any type of intergroup harm. In this case, Black participants may de-emphasize the importance of intent even when evaluating behaviors that do not typically target them. Future work can test between these possibilities.

Relatedly, we focused on White actors behaving toward Black versus White targets because everyday microaggressions in the United States commonly take this form and because prior work on racialized intent has centered on dominant-group actors (Mosley, 2024; Carter & Murphy, 2015). A fully factorial 2 (Actor race: Black vs. White) × 2 (Target race: Black vs. White) design would allow direct tests of whether Black-on-Black and Black-on-White behaviors elicit parallel patterns of intent sensitivity and punishment, and future work can adopt this design.

Additionally, while we asked how moral responses vary across participant race, other individual differences likely also shape responses. For example, lay theories about racial bias (e.g., whether one views prejudice as a fixed trait or a malleable state) predict how White individuals manage cross-race interactions (e.g., adopting more cautious or more approach-oriented strategies; Neel & Shapiro, 2012). Relatedly, social dominance orientation—a preference for group-based hierarchy—predicts endorsement of hierarchy-enhancing "legitimizing myths" that portray group differences as natural, stable, and justified (Sidanius &

Pratto, 1999). Future work can investigate how these individual differences shape intent-based moral evaluation and how such evaluations, in turn, shape interracial interactions.

Conclusion

The present work provides valuable insights into the complexities of moral evaluations and group-based processes. For instance, White adults are particularly likely to emphasize intent when responding to White actors' behaviors toward Black targets; they also punish such behaviors less severely than do Black adults. These findings suggest that centering intent as a critical factor in moral evaluation elevates White individuals' perspectives over those of Black individuals and underscore the need to incorporate intergroup perspectives into scholarly understandings of morality.

Acknowledgements: The authors wish to thank the members of the Social and Moral Cognition Lab for their feedback on this work. This project was made possible through the support of grant #61808 from the John Templeton Foundation and NSF CAREER grant #2044360, both awarded to the last author. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation or the National Science Foundation.

Conflict of interest: The authors have no conflicts of interest to declare.

References

Aguilera, J. (2021, May 26). Amy Cooper apologizes for calling police on Christian Cooper. *Time*. Retrieved on February 22nd, 2025, from https://time.com/5842442/amy-cooper-dog-central-park/

Alexander, L., Moore, M., & Zimmerman, M. J. (2024). Deontological ethics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved on February 22nd, 2025, from https://plato.stanford.edu/entries/ethics-deontological/

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

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B. A., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 113(17), 4688–4693.

Basu, R. (2018). The wrongs of racist beliefs. *Philosophical Studies*, *176*, 2497-2515. https://doi.org/10.1007/s11098-018-1137-0

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. T. Payne and Son.

Brey, E., & Shutts, K. (2015). Children use nonverbal cues to make inferences about social power. *Child Development, 86*, 276–286. https://doi.org/10.1111/cdev.12334

Brownstein, M. (2016). Attributionism and moral responsibility for implicit bias. *Review of Philosophy and Psychology*, *7*, 765–786. https://doi.org/10.1007/s13164-015-0287-7

Buchanan K. S., & Goff, P. A. (2020). Racist stereotype threat in civil rights law. *UCLA Law Review*, *67*, 316-377.

Carter, E. R., & Murphy, M. C. (2015). Group-based differences in perceptions of racism: What counts, to whom, and why? *Social and Personality Psychology Compass*, *9*, 269-280. https://doi.org/10.1111/spc3.12181

Cheryan, S., & Monin, B. (2005). Where are you *really* from?: Asian Americans and identity denial. *Journal of Personality and Social Psychology, 89*, 717–730. https://doi.org/10.1037/0022-3514.89.5.717

Comishen, K. J., Bialystok, E., & Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Developmental Science, 22*, e12797. https://doi.org/10.1111/desc.12797

Corning, A. F., & Bucchianeri, M. M. (2010). Perceiving racism in ambiguous situations: Who relies on easy-to-use information? *The Journal of Social Psychology*, *150*, 258-277. https://doi.org/10.1080/00224540903365398

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

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*, 6–21. https://doi.org/10.1016/j.cognition.2012.11.008

Does, S., Ellemers, N., Dovidio, J. F., Norman, J. B., Mentovich, A., van der Lee, R., & Goff, P. A. (2018). Implications of research staff demographics for psychological science. *American Psychologist*, *73*, 639–650. https://doi.org/10.1037/amp0000199

Dovidio, J. F., Gaertner, S. L., & Pearson, S. L. (2017). Aversive racism and contemporary bias. In C. G. Sibley & F. K. Barlow (Eds.), *Cambridge Handbook of the Psychology of Prejudice* (pp. 267-294). Cambridge, UK: Cambridge University Press.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*, 1352–1366. https://doi.org/10.2307/1130811

Earp, B. D., McLoughlin, K. L., Monrad, J. T., Clark, M. S., & Crockett, M. J. (2021). How social relationships shape moral wrongness judgments. *Nature Communications*, 12(1), 5776.

Ford, L. K. (2009). *Deliver us from evil: The slavery question in the Old South*. Oxford University Press.

Gert, B. (1998). *Morality: Its nature and justification*. Oxford University Press.

Guthrie, R. V. (2004). *Even the rat was white: A historical view of psychology*. Allyn and Bacon.

Hart, H. L. A. (1968/2008). *Punishment and responsibility: Essays in the philosophy of law* (2nd ed.). Oxford University Press.

Haslanger, S. (2017). Racism, ideology, and social movements. *Res Philosophica*, *94*, 1-22. https://dx.doi.org/10.11612/resphil.1547

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Heiphetz, L., & Craig, M. A. (2021). Dehumanization and perceptions of immoral intergroup behavior. In J. Knobe, S. Nichols, & T. Lombrozo (Eds.), *Oxford Studies in Experimental Philosophy* (pp. 155-181). Oxford, UK: Oxford University Press.

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless

actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin*, *38*, 52–62. https://doi.org/10.1177/0146167211430232

Kant, I. (1790). *Critique of judgment*. Hackett Publishing.

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119*, 197–215. https://doi.org/10.1016/j.cognition.2011.01.006

Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, *182*, 331-338. https://doi.org/10.1016/j.cognition.2018.09.003

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(279), 190–194.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234–249. https://doi.org/10.1037/0022-3514.93.2.234

Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology*, *72*, 293-318. https:/doi.org/10.1146/annurev-psych-072220-104358

Masek, L. (2010). Intentions, motives and the doctrine of double effect. *The Philosophical Quarterly*, *60*, 567-585. https://doi.org/10.1111/j.1467-9213.2009.633.x

McKee, P., Kim, H. E., Tang, H., Everett, J. A. C., Chituc, V., Gibea, T., Marques, L. M., Boggio, P., & Sinnott-Armstrong, W. (2024). Does it matter who harmed whom? A cross-cultural study of moral judgments about harm by and to insiders and outsiders. *Current Psychology* (New Brunswick, N.J.), 43(9), 7997–8007.

McMahan, J. (1994). Self-defense and the problem of the innocent attacker. *Ethics, 104*, 252–290. https://doi.org/10.1086/293600

McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. intent across societies: How cultural models of mind shape moral reasoning. *Cognition*, *182*, 95-108. https://doi.org/10.1016/j.cognition.2018.09.008

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences, 11*, 143–152. https://doi.org/10.1016/j.tics.2006.12.007

Mill, J. S. (1861). *Utilitarianism*. Parker, Son, and Bourn, West Strand.

Montgomery, M., & Wallace, G. (1992). Confessions of a former segregationist: George Wallace. *Time*. Retrieved February 22nd, 2025, from https://time.com/archive/6719689/confessions-of-a-former-segregationist-george-wallace/

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. E. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2688–2692. https://doi.org/10.1073/pnas.1011734108

Mosley, A. J. (2024). The aversive racism theory of cultural appropriation: Attributions of target intent suppresses evaluations of intergroup harm. *Personality and Social Psychology Bulletin*, 01461672241292427.

Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems*, 75(2), 1-31.

Neel, R., & Shapiro, J. R. (2012). Is racial bias malleable? Whites' lay theories of racial bias predict divergent strategies for interracial interactions. *Journal of Personality and Social Psychology*, 103(1), 101–120.

Nelson, J. C., Adams, G., & Salter, P. S. (2013). The Marley hypothesis: Denial of racism reflects ignorance of history. *Psychological Science*, *24*, 213-218. https://doi.org/10.1177/0956797612451466

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Pasupathi, M., & Wainryb, C. (2010). Developing moral agency through narrative. *Human Development*, *53*, 55-80. https://doi.org/10.1159/000288208

Payir, A., & Heiphetz, L. (2022). Children's and adults' attribution of moral judgments to human and supernatural agents. *Journal of Cognition and Development*, *23*(4), 524–544. https://doi.org/10.1080/15248372.2022.2061975

Piaget, J. (1932/1965). *The moral judgment of the child*. Free Press.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39*, 653–660. https://doi.org/10.1016/S0022-1031(03)00041-6

Quinn, W. (1989). Actions, intentions, and consequences: The doctrine of double effect. *Philosophy & Public Affairs, 18*, 334–351.

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*, 999-1006. doi: 10.1177/0956797612466267

Schütz, A., & Baumeister, R. F. (1999). The language of defense: Linguistic patterns in narratives of transgressions. *Journal of Language and Social Psychology*, *18*, 269-286. https://doi.org/10.1177/0261927X99018003003

Schwartz, F., Djeriouat, H., & Trémolière, B. (2022). Agents' moral character shapes people's moral evaluations of accidental harm transgressions. *Journal of Experimental Social Psychology*, 102, 104378.

Schwitzgebel, E., Bright, L. K., Jennings, C. D., Thompson, M., & Winsberg, E. (2021). The diversity of philosophy students and faculty in the United States. *The Philosophers' Magazine.* Retrieved on February 22nd, 2025, from https://philosophersmag.com/the-diversity-of-philosophy-students-and-faculty-in-the-united-states/

Shafer-Landau, R. (2018). *The fundamentals of ethics* (4th ed.). Oxford University Press.

Shelby, T. (2014). Racism, moralism, and social criticism. *Du Bois Review*, *11*, 57-74. https://doi.org/10.1017/S1742058X14000010

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Siena College Research Institute. (2021). *Americans, deeply divided, yet share core values of equality, liberty & progress*. Retrieved on December 20th, 2023, from https://scri.siena.edu/2021/10/25/americans-deeply-divided-yet-share-core-values-of -equality-liberty-progress/.

Sue, D. W., & Spanierman, L. B. (2020). *Microaggressions in everyday life* (2nd ed.). Wiley.

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

Young, L., & Saxe, R**.** (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 120*, 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

Appendix

Black targets

1. Last Sunday, there was a white father playing with his young son in the park. A black couple walked by with their dog, and the boy went over to pet the dog. The father pulled his son back toward him and told him to leave the dog alone.

2. Last evening a black man was seated in the first-class section of a plane. A white man was sitting across the aisle from him. The flight attendant, who was white, politely asked the black passenger if she could check his seat assignment again before take-off.

3. In a gimmicky-type merchandise store the other day, a middle-aged black man came in and starting checking out some sort of radio device. A clerk, who was white, asked the man if he needed help. The man said he was just browsing, and went on moving throughout the store. Every few minutes, the man noticed that the same clerk was arranging merchandise near him.

4. Yesterday on a downtown bus, a young white woman was occupying a side-facing bench seat. A black couple boarded and sat down next to her. Shortly after the couple sat down, the young woman got up, walked down the aisle, and held a handrail. She did not get off at the next stop.

5. A week ago, a black couple sat for 30 minutes at a restaurant before being served. Three couples were served before them, two of which had come into the restaurant after the black couple. Over the noise, the waiter, who was white, told the hostess that another server was supposed to be covering that table.

6. Two students, one of whom was black, were driving into campus last Friday. They had their windows open because it was a nice day, and the white guard sort of leaned over a bit toward their car. He commented on the nice weather and asked if they were headed to practice.

White targets

1. One day an aunt and niece, both of whom were white, were driving around town . As they were stopped at a red light, they noticed a white pedestrian walking down the sidewalk on their same side of the street. The man was dressed in layers and appeared to be carrying a bundle of newspapers. Upon seeing the man, the aunt asked the niece to lock the doors.

2. Saturday afternoon a white couple came into a jewelry store and began looking at some of the jewelry on display. The lady behind the counter, who was white, then approached the couple and greeted them by asking, "Are you here to buy something today?"

3. Last weekend there were two young women in line outside a popular nightclub. Sometimes this club asked for IDs. The bouncer, who was white, asked the white woman to pay the fee, and then let her in. He then asked the other woman, who was also white, to pay the fee and then asked to see her ID.

4. Earlier today in the parking lot of a grocery store, a white woman who was just arriving was getting her kids out of the car. Two of her kids were already out and standing near the back of the car. A white man headed toward the store was walking through the parking lot. As the man neared, the mother told the children to move closer to the front of the car. She was saying, "You need to stay away from moving cars."

5. At the gym yesterday, there was a fight between two white men. The owner of the gym, who was a white woman, said she saw the whole thing. She ended up revoking the gym membership of one of the white men, saying he had instigated the fight.