Moral Evaluations and Punishment Decisions in Intergroup Contexts

Daniel Yonas[a] and Larisa Heiphez Solomon[b]

[a] Department of Psychology, Columbia University. 359 Schermerhorn Extension, 1190 Amsterdam Ave., New York, NY 10027, United States. E-mail: dyonas.1412@gmail.com.

[b] Department of Psychology, Columbia University. 821 Uris Hall, 3022 Broadway Ave., New York, NY 10027, United States. E-mail: lah2201@columbia.edu. (Corresponding author)

Word count (abstract + main text): 8,441

In press, *Cultural Diversity and Ethnic Minority Psychology*

This publication was made possible through the support of NSF CAREER grant #2044360 awarded to the second author. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank the members of the Columbia Social and Moral Cognition Lab for their feedback on this project.

Abstract

Objectives. Moral philosophy and psychology often emphasize the role of intent in moral evaluations but typically do not incorporate perspectives from the social psychology of intergroup bias. Therefore, current work on morality has limited capacity to understand moral cognition in group contexts. The current work examined how Black and White adults (Study 1) and White children (Study 2) incorporate information about intent into judgments of intergroup behaviors.

Methods. Participants learned about White individuals who transgressed against Black individuals (Studies 1-2), Black individuals who transgressed against White individuals (Study 1), or White individuals who transgressed against White individuals (Study 2). Participants in Study 1 indicated how important perpetrators' intentions were; participants in Study 2 learned about behaviors performed with malicious or benign intent. Both studies measured participants' moral responses to perpetrators.

Results. White adults prioritized White perpetrators' intent more, and reported that they should receive less severe punishment, than did Black adults. White 7- to 9-year-olds—but not 4- to 6-year-olds—also emphasized perpetrator intent especially strongly when evaluating intergroup behaviors.

Conclusions. While the ability to consider intent in moral judgment emerges early, the role of intent-based judgments in moral cognition depends on group membership. Integrating moral psychology, the social psychology of intergroup bias, and developmental science strengthens scientific understanding of human cognition.

Public Significance

People often judge others' actions by considering what they *meant* to do. The current studies show that individuals apply this reasoning differently depending on their own racial group membership and the group memberships of the people involved in the behavior. These findings suggest that biases in intent-based moral judgments may emerge relatively early in life and help explain why people sometimes overlook or excuse harms against minoritized groups.

Keywords: child development; intent; intergroup relations; morality; punishment

Introduction

In 2015, a widely-circulated video showed a White elementary school teacher in South Carolina forcibly removing a Black teenager from her classroom after she refused to give up her phone. An officer called in by the teacher flipped the girl backward in her chair and dragged her across the floor. Although many viewers expressed outrage, others defended the teacher and officer, arguing that they were just following protocol and that nothing about the incident was inherently racist (Jarvie, 2015). Debates over the incident quickly turned to intent: Did the adults mean to act in a racist way, or were they simply trying to restore order? The girl's experience of harm became secondary to the intentions of the individuals responsible for those harms. These public reactions revealed how social group membership can inform judgments about intent.

This example illustrates a broader question: How does moral evaluation occur in intergroup contexts? Specifically, how—if at all—does intent-based moral judgment vary across groups? As people's understanding of intent typically shifts during the early elementary school years (Baird & Astington, 2004; Cushman et al., 2013; Killen et al., 2011), the current work addressed these questions with 4- to 9-year-olds and adults. Two studies probed the extent to which participants emphasized a perpetrator's intent when evaluating intergroup and intragroup behaviors. By integrating work from moral psychology, the social psychology of intergroup relations, and developmental science, we investigated the extent to which intent-based judgments may reinforce social biases and the developmental mechanisms that may underlie this effect.

Moral Cognition and Intergroup Relations

Moral psychology often investigates behavior at the interpersonal level, without considering the groups to which individuals belong (Crockett et al., 2014; Yonas & Solomon, 2025; Yucel et al., 2022). Such work has demonstrated that adults typically respond less harshly

to transgressions committed with benign rather than malicious intent (Ames & Fiske, 2013; Payir & Heiphetz, 2022; Young & Saxe, 2011). These descriptive results align with theories from philosophy and law arguing that people *should* grant leniency based on benign intent (Fletcher, 1998; Gert, 1998; Hart, 1968/2008).

Work probing the role of intent in adults' moral judgments has not investigated how such judgments might depend on group membership. Indeed, dominant group members do not always express awareness of how social groups can shape life outcomes (Kraus et al., 2019; Nelson et al., 2013; Marshburn et al., 2023). Nevertheless, many transgressions occur across group lines. In contexts ranging from education (Okonofua et al., 2016) to the legal system (Rucker & Richeson, 2021) to medicine (Washington, 2006) to interpersonal interactions (Bonilla-Silva, 2006), minoritized individuals find themselves the targets of dominant group members' harmful actions. Thus, it is particularly important to expand moral psychology's focus to include the intergroup level.

In contrast to moral psychology, work on intergroup bias does emphasize how group membership shapes cognition and behavior. However, it is rare for such work to incorporate theorizing from moral psychology. To the extent that past work from intergroup relations has investigated morality, it has focused largely on clear and unambiguous harms (Killen et al., 2018; Leach et al., 2013). While these studies are critical for understanding blatant forms of bias, many harms that occur in intergroup contexts include ambiguity, especially regarding intent. For instance, one person may move away from another, but it is unclear whether the mover intended to exclude or merely wanted to stretch her legs. The present work refers to these situations as *ambiguous harms* because observers may experience a lack of clarity about whether or not the behavior constitutes a harm. These situations are especially important to study because they

require individuals to interpret not just *what* happened, but also *why* it happened, opening space for bias to influence moral judgment. Because negative responses to others on the basis of their group membership are particularly likely to emerge in situations that are not clear-cut (Gaertner & Dovidio, 2005; Mosley, in press; Okonofua et al., 2016), observers' responses to ambiguous harms are particularly informative when seeking to understand biases that members of minoritized groups navigate throughout their lives.

The current studies strengthen bridges between moral psychology and the social psychology of intergroup bias by investigating how group membership shapes moral evaluations and punishment decisions. Several patterns of results could emerge. First, people may consider intent more for ingroup, versus outgroup, members. This possibility is aligned with work showing that people engage in greater individuation and mentalizing when considering members of their own group (Ostrom & Sedikides, 1992). Second, White individuals may be particularly sensitive to ingroup members' intentions. This possibility stems from work showing that observers sometimes give perpetrators the benefit of the doubt, particularly when observers and perpetrators both belong to a socially dominant group. For instance, men perceive unintentional sexist behaviors as less discriminatory than do women, and women's judgments of perpetrators show more sensitivity than do men's judgments to the harmful outcomes experienced by targets of sexist behaviors (Mosley, in press; Swim et al., 2003). Third, White participants may emphasize intent more than do Black participants, regardless of the groups to which perpetrators and targets belong. This possibility is consistent with findings showing that emphasizing intent is particularly characteristic of Western, individualistic cultures (McNamara et al., 2019; Wu & Chentsova-Dutton, 2026). Although all participants in the current work lived in the United

States, White Americans exhibit more individualism than members of racially minoritized groups (Vandello & Cohen, 1999) and may therefore be particularly sensitive to intent.

The present studies also strengthen bridges between social and developmental psychology. Because intent-based judgments change with age, the present work adopted a social cognitive developmental approach to gain greater insight into moral cognition in intergroup contexts.

The Development of Intent-Based Reasoning

A key developmental milestone in moral cognition involves the shift from outcome-based to intent-based reasoning. While young children tend to evaluate moral situations based largely on outcomes—for instance, judging that breaking multiple cups with benign intent is worse than breaking one cup with malicious intent (Piaget, 1932)—older children increasingly incorporate information about actors' mental states when evaluating right and wrong. A foundational study by Baird and Astington (2004) exemplifies this shift. In their study, children between 4 and 7 years old heard stories about protagonists who caused harm either intentionally or accidentally. On average, older children were more likely than younger children to distinguish between intentional and accidental harms by assigning greater blame to the intentionally harmful actor. Subsequent work replicated this age-related difference (Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011), suggesting that sensitivity to intent increases during the early elementary school years.

This work has focused on interpersonal transgressions without considering group membership. However, as discussed above, observers may be particularly likely to emphasize a perpetrator's benign intent when the observer and perpetrator both belong to a socially dominant group. Investigating this possibility developmentally allowed the current research to clarify what

role factors such as intent-based moral reasoning play in responses to intergroup versus intragroup behaviors. If White adults' focus on a perpetrator's intent is driven solely by their cognitive capacity to reason about intent, then White 4- to 6-year-olds (who have not yet experienced the outcome-to-intent shift) should de-emphasize intent compared to White 7- to 9-year-olds, and this difference between younger and older children should not vary across intergroup versus intragroup contexts. However, findings may also reveal that White children's increasing sensitivity to intent with age occurs particularly strongly when they observe in-group members enacting potential harm on out-group members. This finding would highlight the interplay between cognitive and social factors and suggest that a phenomenon that prior work has conceptualized in largely cognitive terms is sensitive to social factors. Indeed, intent-based judgments may reflect not just basic cognitive processes but also a form of motivated social cognition (Kunda, 1990) that allows White individuals to excuse wrongdoing by in-group members and maintain a positive group image. To test between these alternative hypothesis, Study 2 probed the role of intent in White 4- to 6-year-olds' and 7- to 9-year-olds' moral evaluations and punishment decisions in intergroup versus intragroup contexts.

The Present Research

Two studies examined how adults (Study 1) and children (Study 2) weigh intent when judging ambiguous behaviors. Study 1 tested Black and White adults' judgments of behaviors in intergroup contexts, whereas Study 2 compared how White 4- to 6-year-olds versus 7- to 9-year-olds considered intent when responding to intergroup versus intragroup behaviors. These studies contribute to an emerging body of work at the intersection of moral development and intergroup cognition by investigating when and how group membership shapes moral evaluations.

Study 1

Study 1 examined how Black and White adults respond to ambiguous intergroup harms. We tested cases involving White perpetrators and Black targets because this situation reflects prototypical instances of intergroup harm in the United States (Eberhardt, 2019; Mills, 1997; Sidanius & Pratto, 1999). We also tested cases involving Black perpetrators and White targets to determine whether results in the prototypical case generalize to cases where subordinate group members are perpetrators and dominant group members are targets. Because Whiteness bestows social power, the social meaning of these cases differs: White perpetrators who harm Black targets enact cultural scripts that maintain their dominant group status, whereas Black perpetrators who harm White targets act against these scripts. Treating White and Black perpetrators as though they are interchangeable aligns with an interpersonal perspective (Person A harming Person B is analogous to Person B harming Person A) and ignores differences in status, resources, and power commanded by different groups. The perspective in the current work is not that these cases are the same. Rather, the goal is to determine whether participants consider group differences or whether they reason about intergroup harms in similar ways regardless of whether the perpetrator belongs to a socially dominant or subordinate group.

To investigate how participants reason about intent in cases where White perpetrators harm Black targets and vice versa, we probed how much participants reported emphasizing perpetrators' intentions when evaluating transgressions. Prior work suggests that intent may play a particularly strong role in shaping adults' moral evaluations, whereas punishment decisions depend both on intent and on outcomes (Cushman, 2008). Therefore, the current work measured both moral evaluations and punishment decisions.

Method

Participants. Our final sample included 333 White adults (M_{age}=37.15 years, SD_{age}=12.46 years; 199 female, 130 male, 4 other) and 326 Black adults (M_{age}=38.52 years, SD_{age}=11.93 years; 150 female, 173 male, 1 other). On a scale ranging from 1 ("very liberal") to 7 ("very conservative"), participants self-described as somewhat liberal (*m*=3.75, *sd*=1.29), with White participants (*m*=3.74, *sd*=1.28) self-describing relatively similarly to Black participants (*m*=3.77, *sd*=1.30). Political orientation did not predict participants' responses to any dependent measures, *p*s>.20.

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, and recruited a larger initial sample to allow for potential exclusions. Participants were at least 18 years old, provided consent for themselves, and received $10 per hour. In addition to the data presented below, we excluded responses from 164 adults who failed to answer attention check questions from our analyses.[1] This study was preregistered at https://aspredicted.org/fqkz-qg96.pdf and received approval from the authors' institution under protocol AAAQ8299.

Procedure.[2] We presented participants with a series of vignettes depicting two types of situations: cases where White individuals interacted with Black individuals in ways that participants could perceive as perpetrating intergroup harms (e.g., a White person moved when a Black couple sat next to them), and cases where Black individuals performed similar behaviors toward White individuals (e.g., a Black person moved when a White couple sat next to them;

[1] Including all respondents in analyses yielded similar patterns as those reported in the main text.
[2] Before viewing the intergroup and intragroup behaviors, participants completed a task designed to manipulate threat. Approximately half of the participants listed one example of a morally good behavior they had performed over the past week, while the remaining participants listed twenty such examples. Based on classic work on social cognition (Schwarz et al., 1991), we expected that participants would find it difficult to generate twenty examples and consequently experience greater moral threat in this condition, as compared to the one-example condition. We further expected that participants experiencing greater moral threat may place greater emphasis on a perpetrator's benign intention. However, we did not observe differences in threat across our conditions (*p*=.718). Therefore, the analyses reported in the main text collapse across this variable.

Corning & Bucchianeri, 2010; Heiphetz & Craig, 2020). All stimuli are available in Supplemental Materials. After reading each vignette, participants answered three counterbalanced questions that investigated their moral evaluations: "In your mind, how wrong or right is what [character] did?", "In your mind, how good or bad is what [character] did?", "In your mind, how mean or nice is what [character] did?" Participants responded using a scale where 1 represented the most negative judgment (e.g., very morally wrong) and 7 represented the most positive judgment (e.g., very morally right). Participants then answered punishment questions using a similar scale: "In your mind, how much should [character] be punished for what they did?" (1=*Not very much punishment*, 4=*Some punishment*, 7=*A lot of punishment*); "In your mind, how severe should the consequences be for what [character] did?" (1=*Not at all severe*, 4=*Kind of severe*, 7=*Very severe*); and, "In your mind, how much trouble should [character] be in?" (1=*No trouble at all*, 4=*Some trouble*, 7=*A lot of trouble*). The moral evaluation and punishment items showed high internal consistency, with Cronbach's αs of .94 and .96 respectively.

In the final block, participants read each vignette again. This time, they indicated the importance of intent to their moral evaluations and punishment decisions on a scale from 1 ("not at all important") to 7 ("very important"). These questions ("How important was what [character] meant to do / what [character] was trying to do / what [character]'s intentions were?") investigated the degree to which different types of behaviors prompted participants to emphasize intent. Prior studies examining how people incorporate intent into moral evaluations have often manipulated intent directly by comparing participants' responses to clearly intentional versus accidental actions (Cushman et al., 2013; Killen et al., 2011; Payir & Heiphetz, 2022; Young & Saxe, 2011). In contrast, the present study presented participants with morally ambiguous

scenarios and asked them to rate the degree to which the perpetrator's intent influenced their judgments. Using ambiguous scenarios enhanced ecological validity because participants responded to the types of scenarios that minoritized individuals often experience in everyday life (Sidanius & Pratto, 1999; Sue & Spanierman, 2020). Additionally, measuring participants' own judgments about the importance of intent allowed us to determine the extent to which participants believed the perpetrator's intent mattered when evaluating intergroup versus intragroup behaviors. This is a somewhat different question than the one of primary interest in prior studies, which have largely focused on how manipulating information about intent might shift moral evaluations. Such a focus provides useful information on the extent to which information about intent causes changes in moral evaluations. Study 1 built on this prior research to investigate a related, but separate, question: to what extent do adults report that intent matters to them when evaluating behaviors in intergroup contexts? While this approach cannot reveal whether information about intent causes adults to respond in a particular way, it does clarify how adults understand their own mental processes. We supplemented this focus on adults' reflections on their own mental states by manipulating intent directly in Study 2.

Results

To determine how participants', perpetrators', and targets' race shaped participants' responses, we conducted three main analyses.[3] First, we averaged participants' responses to the intent items described above and entered these averages into a 2 (Participant Race: Black vs. White) x 2 (Behavior Type: Black Perpetrator - White Target vs. White Perpetrator - Black Target) between-participants ANOVA. This analysis revealed a main effect of Participant Race,

[3] These analyses differ from the pre-registered ANOVAs, which include high versus low threat as an independent variable. Because our manipulation check did not reveal a significant difference in moral threat across these conditions, the ANOVAs in the main text exclude this variable.

$F(1,651)=22.34$, $p<.001$, partial $\eta^2=.03$): on average, White participants emphasized intent more than did Black participants. No other main effects or interactions reached significance, $ps>.066$.

Second, we averaged participants' responses to the moral evaluation items described above and entered these averages into a 2 (Participant Race: Black vs. White) x 2 (Behavior Type: Black Perpetrator - White Target vs. White Perpetrator - Black Target) between-participants ANOVA (Fig. 1). This analysis revealed a main effect of Participant Race, $F(1,655)=4.11$, $p=.043$, partial $\eta^2=.01$: on average, Black participants judged the behaviors about which they read to be more morally wrong than did White participants. This analysis also revealed a main effect of Behavior Type, ($F(1,655)=102.33$, $p<.001$, partial $\eta^2=.14$): on average, participants judged actions involving White perpetrators and Black targets to be more morally wrong than actions involving Black perpetrators and White targets. The Participant Race x Behavior Type interaction did not reach significance, $p=.667$.

Third, we averaged participants' responses to the punishment items described above and entered these averages into a 2 (Participant Race: Black vs. White) x 2 (Behavior Type: Black Perpetrator - White Target vs. White Perpetrator - Black Target) between-participants ANOVA (Fig. 1). Similarly to the results for moral evaluations, we observed a main effect of Participant Race, $F(1,651)=102.84$, $p<.001$, partial $\eta^2=.14$): on average, Black participants reported that actions should be punished more severely than did White participants. We also observed a main effect of Behavior Type, $F(1,651)=30.23$, $p=.042$, partial $\eta^2=.04$: on average, participants reported that situations with White perpetrators and Black targets should receive more punishment than situations with Black perpetrators and White targets. The Participant Race x Behavior Type interaction did not reach significance, $p=.311$.

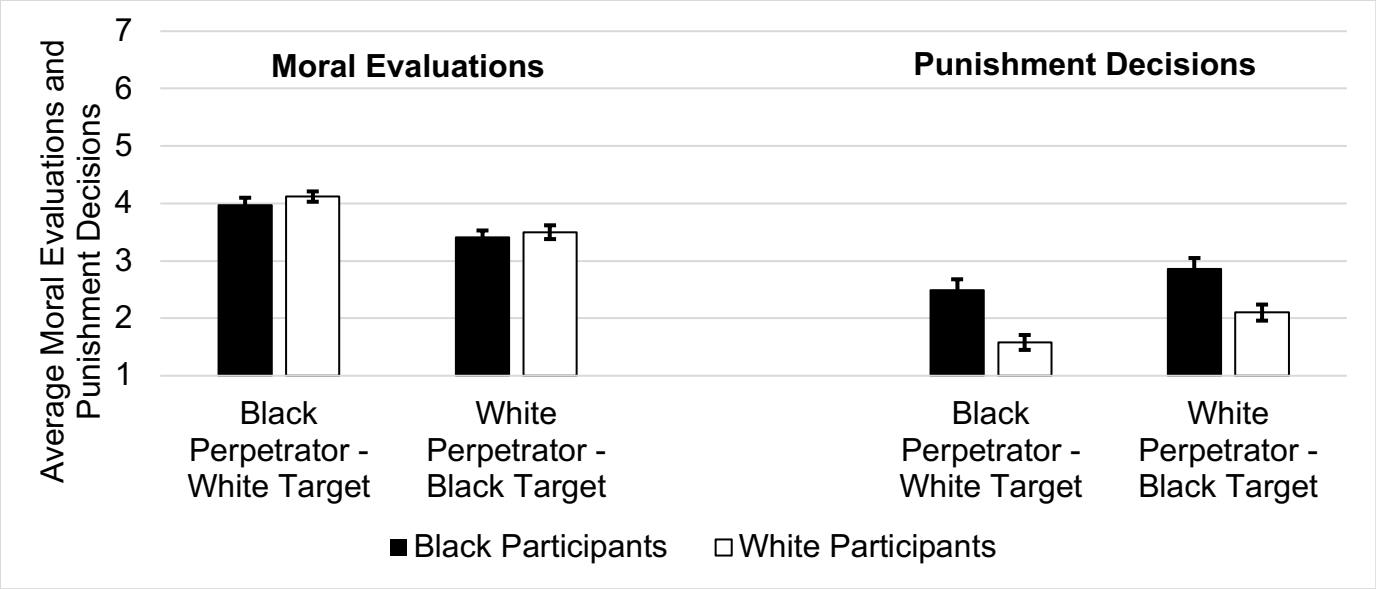

Fig. 1. Participants' moral evaluations (left) and punishment decisions (right), Study 1. Error bars represent 95% confidence intervals.

To determine the degree to which moral cognition and punishment decisions align, we correlated these variables and observed a negative association (r=−.26, p<.001): the less moral participants found a behavior to be, the more severely they reported that behavior should be punished. As an additional, exploratory measure, we probed the extent to which the emphasis participants placed on intent was associated with their moral evaluations and punishment decisions. The more participants reported that the perpetrator's intent mattered to them, the more morally acceptable they found the perpetrator's actions (r=.20, p<.001) and the less severely they thought those actions should be punished (r=-.26, p<.001). Finally, as an exploratory measure, we tested for differences across vignettes; see Supplemental Materials.

Discussion

Study 1 tested how racial group membership shaped participants' responses to ambiguous behaviors in intergroup contexts. Four main results emerged.

First, regardless of whether they read about White perpetrators and Black targets or Black perpetrators and White targets, White participants reported prioritizing the perpetrator's intent more than did Black participants. This result aligns with past work showing that in intergroup contexts, dominant group members typically prioritize intent more than do stigmatized group members (Mosley, in press; Swim et al., 2003). To some extent, this result speaks against conceptualizing intent solely as a cognitive process and highlights how emphasizing perpetrators' intentions may reflect motivated social cognition (Kunda, 1990).

Second, Black and White participants also exhibited different moral responses, with Black (versus White) participants reporting that intergroup harms were more morally wrong and should receive more severe punishment. Although this effect emerged regardless of perpetrators' targets' race, both Black and White participants reported that cases involving White perpetrators and Black targets were more morally wrong and should receive more severe punishment than cases involving Black perpetrators and White targets. This pattern suggests that the dominant-group framing does not operate exclusively among White participants. That is, Black participants also recognized that the two cases differed in moral severity. Black and White participants converged on which type of case was more morally wrong while diverging in how much they emphasized the perpetrator's intent. Additionally, the finding that participants evaluated cases with White perpetrators and Black targets more harshly than cases with Black perpetrators and White targets differs from behavioral data suggesting that in fact, Black (versus White) perpetrators receive harsher punishments for similar offenses (Alexander, 2012; Eberhardt et al., 2006). Although participants appear willing to report that White individuals who transgress against Black individuals *should* receive particularly strong punishment in some circumstances, people may hesitate to act on this normative belief when making consequential decisions with

high stakes—i.e., decisions in the context of the legal system, where racial disparities are particularly pronounced (Alexander, 2012; Eberhardt, 2019).

Third, the difference between Black and White participants' responses was more pronounced when they made punishment decisions (partial η^2=.14) than when they reported moral evaluations (partial η^2=.01). One way to interpret this result is that although White participants are willing to report that intergroup harms are morally wrong, they exhibit less willingness to enact severe punishment in response to those actions. Punitive behavior can be costly for punishers; for instance, observers may draw negative social inferences about individuals who punish others (Lee & Solomon, 2025). White individuals may hesitate to incur these costs in intergroup contexts even when they believe the behaviors to be morally wrong.

Fourth, the dependent measures showed significant correlations with one another. Specifically, the more participants reported that behaviors were morally wrong, the more severely they reported those behaviors should be punished. Further, the more participants emphasized intent, the more they perceived the behaviors as morally acceptable and the less severely they thought those behaviors should be punished. This result is particularly meaningful given that White participants prioritized intent more than did Black participants. Together, this pattern of results suggests that when White individuals encounter intergroup harms, they may prioritize the perpetrator's benign intentions and react to their behavior more leniently than would Black participants.

Study 2

Study 2 extended the primary pattern of results from Study 1—which showed that White adults reported greater prioritization of perpetrators' intentions, and responded to perpetrators' actions more leniently, than did Black adults when considering intergroup behaviors—in three

main ways. First, building on adult data obtained in Study 1, Study 2 tested 4- to 6-year-old and 7- to 9-year-old White children. Children exhibit an outcome-to-intent shift around the age of 7 years: while young children prioritize harmful outcomes when evaluating others' behaviors, in early elementary school their evaluations shift to emphasizing a perpetrator's intent to a greater extent (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011). Thus, if intent-based reasoning underlies moral evaluations of intergroup harms, then older children may show different patterns of results than would younger children. Study 2 focused on White children to follow up on the results from Study 1 showing that White adults were particularly likely—more likely than Black adults—to emphasize perpetrator intent when responding to intergroup behaviors.

Second, while Study 1 used self-report to gain insight into adults' explicit views of how much intent matters, Study 2 manipulated information about intent to gain greater insight into causality. Specifically, we varied whether participants learned about behaviors committed with malicious or benign intent. If participants place emphasize intent when determining how morally wrong a perpetrator's behavior is, then they should exhibit a large difference in moral evaluations of behaviors performed with malicious versus benign intent, even if the actual behavior is the same. Manipulating information is a particularly valuable way to learn about the extent to which children emphasize intent when making moral evaluations because children may experience more difficulty than adults in reflecting on their own cognitive processes and accurately reporting how much intent matters to them.

Third, while Study 1 compared different types of intergroup situations, Study 2 included only intergroup contexts involving White perpetrators and Black targets. We focused on this condition because it reflects a prototypical instance of racial harm in the United States

(Eberhardt, 2019; Mills, 1997; Sidanius & Pratto, 1999). Rather than comparing these situations with those involving Black perpetrators and White targets, we altered the comparison condition in Study 2 to include intragroup behaviors—situations in which White perpetrators performed the same actions described in the intergroup condition toward White targets. We selected this comparison to determine the extent to which participants' judgments of intergroup and intragroup behaviors may differ. Specifically, if intent plays a particularly strong role in responses to intergroup harms, then the difference in responses to behaviors performed with malicious versus benign intent should be particularly large when these behaviors occur in an intergroup, versus intragroup, context.

Method

Participants. Our final sample included 66 children between four and six years old (M_{age}=5.22 years; SD_{age}=0.76; 29 female, 37 male) and 90 children between seven and nine years old (M_{age}=7.83 years; SD_{age}=0.86 years; 50 female, 40 male). We determined sample sizes prior to collecting data using G*Power, indicating 80% power for a 2 (Participant Age: Younger vs. Older) x 2 (Behavior Type: Intergroup vs. Intragroup) x 2 (Intent: Intentional vs. Unintentional) mixed design with repeated measures on the second and third factors, and estimating a medium effect of Cohen's f=.50. Parents provided consent for their children to participate in the study. We excluded data from 43 children (33 younger, 10 older) who had previously participated in the study, failed to complete the protocol, or demonstrated difficulty understanding key questions, or whose responses appeared to come from caregivers rather than the children themselves.[4] This study was preregistered at https://aspredicted.org/t5qq-7mf2.pdf and received approval from the authors' institution under protocol AAAQ8299.

[4] Including all respondents in analyses yielded similar patterns as those reported in the main text.

Procedure. We presented children with a series of 4 vignettes that described either an intergroup or an intragroup action. These vignettes adapted the intergroup behaviors presented in Study 1 to make them appropriate for a younger audience; for instance, vignettes depicted situations that are familiar to children and used pictures to draw children's attention to the stimuli.[5] The experimenter presented an intergroup (White perpetrator, Black target) or intragroup (White perpetrator, White target) scenario. Within each condition, half of the vignettes depicted a perpetrator with a benign intent, such as a White character moving to a different part of the playground when either a Black or White child sat down next to them because the focal character saw something fun to play with in a different part of the playground. The remaining vignettes depicted a perpetrator with malicious intent, such as a White character moving to a different part of the playground because they did not want to be near a Black (or White) child. See Supplemental Materials for all stimuli.

Following each vignette, participants answered two questions about their moral evaluations and two questions about their punishment decisions in counterbalanced order. To measure moral evaluations, experimenters asked, "Was what [character] did wrong or right?" and, "Was what [character] did bad or good?" After participants made a selection, the experimenter asked if the character's behavior was "a little," "kind of," or "very" wrong, right, bad, or good. We coded these responses on a scale ranging from -2.5 (indicating that the behavior was very wrong or very bad) to +2.5 (indicating that the behavior was very right or very good). This coding scheme allowed for the distance between responses on the same side of 0 to

[5] Akin to Study 1, participants also completed a brief measure of moral self-image at the beginning of the study, adapted from prior work (Sticker et al., 2021). This component was intended to assess whether children's self-concept related to their moral evaluations and punishment decisions. However, analyses revealed no significant effects involving this variable (p=.733), and including it did not alter the pattern of results. As such, we present the primary analyses without reference to moral self-image.

be the same as the distance between -0.5 ("a little" wrong or bad) and +0.5 ("a little" right or good). These two items were highly correlated (r=.74, p<.001); therefore, we averaged across them to create a composite moral evaluation score.

To measure punishment decisions, experimenter asked, "Should [character] be punished for what they did?" and, "Should [character] get into trouble for what they did?" If participants answered yes, the experimenter asked them whether the character should be punished or should get into trouble "a little," "kind of," or "a lot." Thus, children's responses constitute a scale ranging from 0, indicating that the character should not receive punishment or get into trouble, to 3, indicating that the character should receive a lot of punishment or get into a lot of trouble. These two measures were highly correlated (r=.76, p<.001), and we therefore averaged across them.

Results

To determine how participant age, behavior type, and perpetrator intent shaped participants' responses, we conducted two main analyses.[6] First, we averaged participants' responses to the moral evaluation items described above and entered these averages into a 2 (Participant Age: Younger vs. Older) x 2 (Behavior Type: Intergroup vs. Intragroup) x 2 (Intent: Benign vs. Malicious) mixed ANOVA with repeated measures on the second and third factors (Fig. 2).

This analysis revealed a main effect of Intent, $F(1,122)=57.56$, p<.001, partial η^2=.08: on average, participants rated behaviors performed with malicious intent as morally worse than behaviors performed with benign intent. This main effect was qualified by a Participant Age x

[6] As in Study 1, we pre-registered analyses that included independent variables related to moral self-image. However, as described in Footnote 5, this variable did not significantly influence our dependent measures. For this reason, and to remain consistent with Study 1, we do not report variables related to moral self-image in the main text.

Intent interaction, $F(1,122)=10.05$, $p=.002$, partial $\eta^2=.01$, which, in turn, was qualified by a Participant Age x Intent x Behavior Type interaction, $F(1,122)=7.86$, $p=.006$, partial $\eta^2=.01$. No other main effects or interactions reached significance, $ps>.478$.

To further examine this interaction, we compared the extent to which participants responded differently to perpetrators with benign versus malicious intentions. This analysis involved four comparisons; thus, *p* values needed to be .013 or lower to pass the Bonferroni-corrected significance threshold.

Older participants reported that behaviors committed with malicious intent were more morally wrong than behaviors committed with benign intent when evaluating both intergroup and intragroup behaviors ($ps<.001$). This result is aligned with past work showing that starting in early elementary school, children respond more leniently to interpersonal transgressions when perpetrators act with benign rather than malicious intent (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011). However, we also observed a stronger effect when older participants responded to intergroup behaviors (Cohen's $d=.99$) than when they responded to intragroup behaviors (Cohen's $d=.55$). Although intent swayed participants' moral evaluations in both cases, it seemed to matter particularly strongly in cases of intergroup behaviors, consistent with Study 1's finding that White adults prioritize intent particularly strongly (more so than do Black adults) in intergroup contexts.

Younger participants did not distinguish between perpetrators with malicious versus benign intent when evaluating intergroup behaviors ($p=.279$, Cohen's $d=.19$), consistent with prior work showing that 4- to 6-year-olds place less emphasis on intent than do older children when evaluating interpersonal transgressions (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011). Surprisingly, in the current study, younger

participants reported that intragroup behaviors committed with malicious intent were more morally wrong than intragroup behaviors committed with benign intent (p=.002, Cohen's d=.34). This result may pose a challenge to prior theorizing regarding the developmental window during which the outcome-to-intent shift occurs. However, we did not predict *a priori* that younger participants would show evidence of this shift only when evaluating intragroup (rather than intergroup) behaviors, and there is not a clear reason to predict the effect to emerge only in this condition for younger children. Thus, we hesitate to interpret this data point in the absence of further replication.

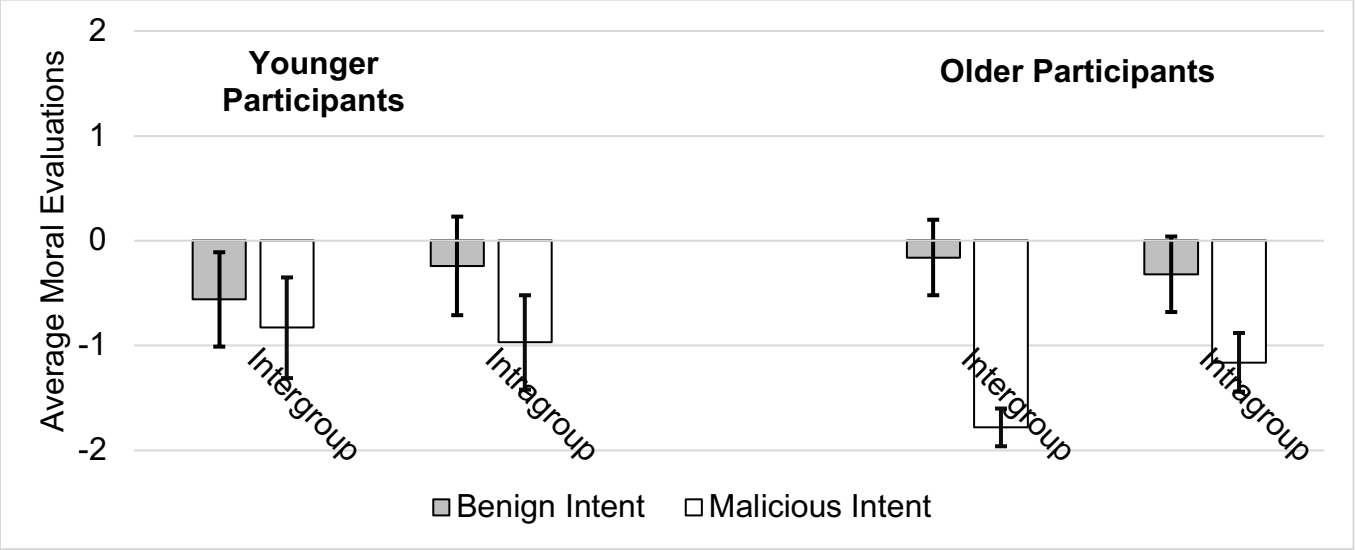

Fig. 2. Younger (left) and older (right) participants' moral evaluations in intergroup and intragroup contexts, Study 2. Error bars represent 95% confidence intervals.

Second, we averaged participants' responses to the punishment items described above and entered these averages into a 2 (Participant Age: Younger vs. Older) x 2 (Behavior Type: Intergroup vs. Intragroup) x 2 (Intent: Benign vs. Malicious) mixed ANOVA with repeated measures on the second and third factors (Fig. 3). This analysis revealed a main effect of Behavior Type, $F(1,122)=10.91$, p=.001, partial η^2=.08: on average, participants reported that

behaviors in intergroup contexts should receive more severe punishment than behaviors in intragroup contexts. The omnibus analysis also revealed a main effect of Intent, $F(1,122)=19.50$, $p<.001$, partial $\eta^2=.19$: on average, participants reported that behaviors performed with malicious intent should receive more severe punishment than behaviors performed with benign intent. This latter main effect was qualified by a Participant Age x Intent interaction, $F(1,122=5.28$, $p=.023$, partial $\eta^2=.04$, which in turn was qualified by a Participant Age x Intent x Behavior Type interaction, $F(1,122)=5.97$, $p=.016$, partial $\eta^2=.05$. No other interactions reached significance, $ps>.163$.

To further examine the three-way interaction, we compared the extent to which participants responded differently to perpetrators with benign versus malicious intentions. This analysis involved four comparisons; thus, p values needed to be .013 or lower to pass the Bonferroni-corrected significance threshold. Similarly to the results that emerged for moral evaluations, older participants reported that behaviors committed with malicious intent should receive more severe punishment when evaluating intergroup scenarios ($p<.001$, Cohen's $d=.74$). However, unlike moral evaluations, punishment decisions did not significantly differ across malicious versus benign intent when evaluating intragroup scenarios after applying a Bonferroni correction ($p=.034$, Cohen's $d=.28$). Overall, this pattern of results is consistent with Study 1 and with the moral evaluation results from Study 2, both of which suggest that older White participants prioritize intent particularly strongly in intergroup contexts. Consistent with prior work suggesting that 4- to 6-year-olds place less emphasis on intent than do older children (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011), 4- to 6-year-olds' punishment decisions in the current study did not vary across malicious versus benign intent regardless of whether they were evaluating behaviors in intergroup ($p=.387$, Cohen's

d=.11) or intragroup (*p*=.037, Cohen's *d*=.23; note that this *p*-value does not pass the Bonferroni-corrected significance threshold) contexts.

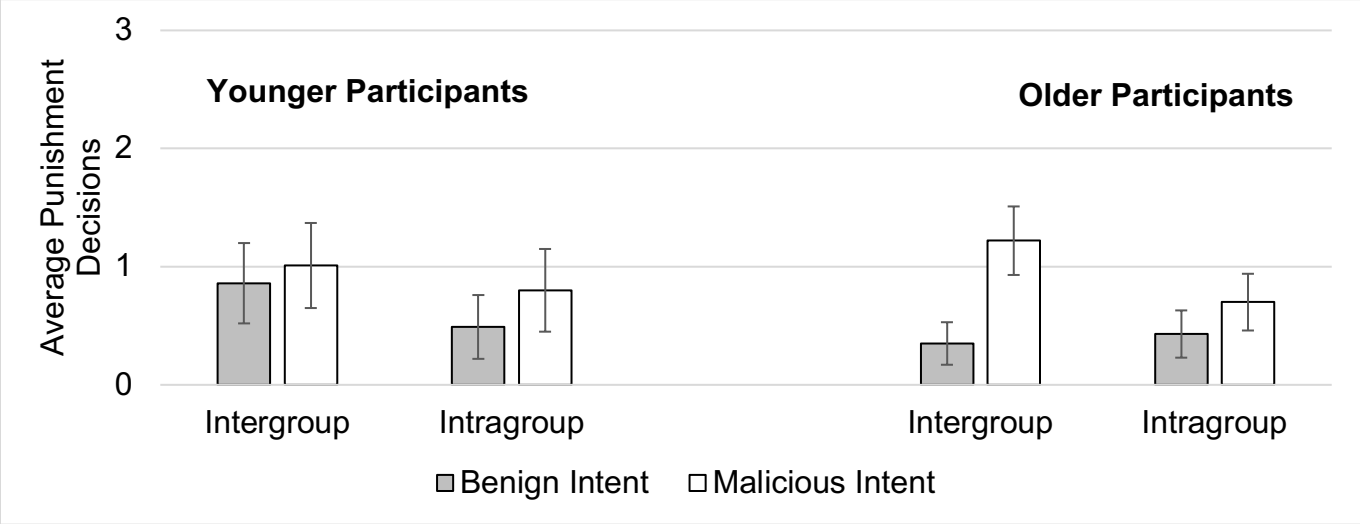

Fig. 3. Younger (left) and older (right) participants' punishment decisions in intergroup and intragroup contexts, Study 2. Error bars represent 95% confidence intervals.

In addition to investigating moral evaluations and punishment decisions using the separate analyses presented above, we conducted an exploratory analysis correlating these variables. Collapsing across participant age, behavior type, and intent, we observed a significant negative association between moral evaluations and punishment decisions (*r*=-.17, *p*=.036). This result replicates Study 1 and indicates that the less moral participants found behaviors to be on average, the more severely they reported those behaviors should be punished.

Discussion

Study 2 extended the findings of Study 1 by asking how 4- to 6-year-old and 7- to 9-year-old White children respond to others' behaviors in intergroup versus intragroup contexts. Findings helped clarify the developmental mechanisms underlying Study 1's results. Specifically, in Study 1, White adults placed more emphasis on intent than did Black adults when responding to behaviors in intergroup contexts. In Study 2, White 7- to 9-year-olds showed harsher moral responses to intergroup behaviors performed with malicious, versus benign, intent. This effect emerged less reliably among 7- to 9-year-olds responding to intragroup contexts and among 4- to 6-year-olds responding to either intergroup or intragroup contexts. Thus, Study 2 suggests that the outcome-to-intent shift that past work has documented among children (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011) emerges to different extents depending on the context in which behaviors occur and that White individuals may be especially likely to deploy intent-based reasoning to excuse in-group members' behaviors toward out-group members. Similarly to Study 1, Study 2 speaks against a portrayal of intent as a cognitive process that occurs independently of social context (for data making a similar point using different approaches, see McNamara et al., 2019). Rather, once children develop the capacity to mentalize, they may deploy this capacity differently depending on social factors.

General Discussion

This present work investigated how 4- to 9-year-olds and adults consider information about intent when evaluating vignettes depicting intergroup versus intragroup behaviors. Study 1 found that White participants emphasized intent more than did Black participants when evaluating intergroup behaviors. Study 2 extended this work by examining White 4- to 9-year-olds' responses to intergroup and intragroup behaviors. Here, 7- to 9-year-olds exhibited harsher

responses to White perpetrators who acted toward Black targets with malicious, rather than benign, intent. This distinction between acts committed with benign versus malicious intent did not emerge as reliably among 7- to 9-year-olds responding to White perpetrators whose behaviors targeted in-group members or among 4- to 6-year-olds evaluating behaviors in either intergroup or intragroup contexts. Together, these studies challenge the notion that intent-based reasoning is solely a cognitive process. In line with other work highlighting social influences on the extent to which people consider intent when making moral judgments (McNamara et al., 2019; Wu & Chentsova-Dutton, 2026), the present research suggests that intent-based moral judgments may also reflect social processes such as motivated social cognition (Kunda, 1990).

Implications for Intergroup Relations

The present work carries important implications for understanding intergroup processes. Much research on intergroup relations has documented a shift in the U.S. from overt expressions of prejudice to more covert, ambiguous, and socially acceptable forms of bias (Bobo et al., 2012; Bonilla-Silva, 2006; Gaertner & Dovidio, 2005). Although the acceptability of explicit bias may once again be rising (Ruisch & Ferguson, 2022; Schaffner, 2020), members of minoritized groups have frequently faced ambiguous harms—i.e., behaviors that could be motivated by animus or by benign intent—even during the time when much of the public viewed explicit bias as unacceptable (Gaertner & Dovidio, 2005; Sue & Spanierman, 2020). Such experiences may be particularly damaging because lack of clear information about the intent underlying particular behaviors limits opportunities for recourse, making it easier for those who commit harm to evade accountability. To the extent that people committing intergroup harms predominantly belong to socially dominant groups, emphasizing benign intent may further reinforce social structures that excuse the mistreatment of minoritized people.

The current work draws on theorizing from moral psychology regarding intent-based judgments to shed light on a particular manifestation of bias. Specifically, the current work suggests that White individuals who have undergone the outcome-to-intent shift do not deploy intent-based reasoning equally across contexts. Rather, they appear to prioritize intent particularly strongly when judging their own in-group members who performed ambiguously harmful actions toward racial out-group members. This psychological process may underlie systems that privilege dominant group members' internal states over the harms they inflict on minoritized group members. For instance, White individuals' propensity to emphasize their in-group members' benign intentions when observing those in-group members' actions toward Black individuals could underlie laws and policies that require evidence of malicious intent to substantiate claims of racial discrimination (Alexander, 2012). This framework suggests that dominant group members' psychological processes influence not just interpersonal interactions but also public policies that have a strong impact on the lives of minoritized individuals (for a discussion on the relationship between law and laypeople's psychological processes, see Mott & Solomon, in press; Tobia, 2021; Tyler, 2006). Further, legal actors' decisions may be driven in part by social group membership; for example, White judges and jurors may place particular emphasis on the benign intentions of White defendants who have harmed Black victims. To the extent that a society does not believe it is morally righteous to base its governance on the psychology of a particular group, it becomes particularly crucial to ensure equal access to levers of power, such that psychological processes among any one group are not exerting an outsized influence on resulting policies.

Implications for Moral Psychology

The present studies contribute to scientific understanding of morality by demonstrating how social group membership shapes moral cognition across development. Much work on morality has emphasized the importance of intent in moral evaluations (Cushman, 2008; Killen et al., 2011; Payir & Heiphetz, 2022; Young & Saxe, 2011), especially at the interpersonal level, the current findings suggest that intent-based reasoning can depend on social group membership. For instance, White adults perceive a perpetrator's intent as more important than do Black adults when evaluating intergroup scenarios (Study 1), and White 7- to 9-year-olds' moral responses showed more reliable sensitivity to intent when they evaluated intergroup, versus intragroup, actions (Study 2). Thus, the scholarly focus on intent may mirror scholars' group identities. To the extent that psychology and philosophy include a disproportionate number of White researchers (Cimpian & Leslie, 2017; Does et al., 2018), the emphasis that these fields place on intent may reflect, to some extent, scholars' prioritization of intent in their own moral judgments.

The present results constrain theorizing regarding morality. Current frameworks indicate that neurotypical adults should (Fletcher, 1998; Gert, 1998; Hart, 1968/2008) and do (Ames & Fiske, 2013; Payir & Heiphetz, 2022; Young & Saxe, 2011) consider a perpetrator's intent when making moral evaluations. The current work suggests that, at least when making judgments about vignettes (the method used in much prior work on intent-based moral judgments as well; Ames & Fiske, 2013; Payir & Heiphetz, 2022; Young & Saxe, 2011), this result might more closely reflect White adults' reasoning about intergroup behaviors and less closely reflect judgments made by Black adults or judgments concerning intragroup behaviors. In other words, current theorizing may make descriptive and normative claims that closely match White participants' judgments about their in-group members' behaviors toward out-group members. The present data highlight the need to expand this theorizing to consider more diverse perspectives.

Implications for Developmental Psychology

In addition to the implications discussed above for adult moral psychology and adults' intergroup cognition, the present studies offer contributions to developmental science. Prior research has established that young children rely heavily on outcome-based reasoning, with a gradual shift toward intent-based moral evaluations during the early elementary school years (Baird & Astington, 2004; Cushman et al., 2013; Heiphetz & Young, 2014; Killen et al., 2011). This shift may underlie White adults' emphasis on intent, as 4- to 6-year-old White children in Study 2 showed relatively less sensitivity to perpetrator intent than did older White participants. In other words, using a developmental approach allowed the current work to clarify a mechanism rooted in socio-cognitive development that may underlie White adults' responses.

However, our findings also suggest that the outcome-to-intent shift may not occur uniformly. Rather, during the developmental window in which children's moral evaluations begin showing increasing sensitivity to intent, White children in Study 2 showed particular sensitivity to the intentions of White perpetrators acting toward Black targets. This finding suggests that the outcome-to-intent shift is not merely a cognitive process that occurs inside a given individual. Rather, this shift appears responsive to social factors, such as motivated social cognition and intergroup processes.

Traditional accounts often frame intent-based reasoning as a developmental achievement, marking a shift from outcome-focused to more nuanced moral evaluations based on perpetrators' mental states. However, our results suggest that this developmental shift may not occur in the same way across contexts. Instead, intent-based reasoning may develop in ways that are sensitive to intergroup dynamics, including the social identities of perpetrators, targets, and observers. This framework suggests that what may appear to be an achievement (grasping an individual’s

mental state in order to render judgment) may simultaneously introduce vulnerabilities to group-based bias, at least among White individuals. Developmental theories must therefore account not only for children's growing capacity to consider intent, but also for how socialization and group membership shape the contexts in which children emphasize or downplay information about intent.

Thus, children's reasoning about race bears consideration as a complementary developmental process contributing to the observed effect. Children's understanding of race also changes throughout the elementary-school years, with older children more likely than younger children to show adultlike levels of bias and to conceptualize racial group membership as stable and intrinsically meaningful (Dunham et al., 2008; Roberts & Gelman, 2017). Thus, the age-related pattern observed in Study 2 may reflect not only growing cognitive capacity to consider intent, but also growing sensitivity to race as a social category that organizes moral evaluation. This co-occurrence supports the motivated social cognition account: as White children develop the capacity to reason about intent, they may simultaneously become more attuned to the racial group memberships of perpetrators and targets, creating the conditions under which intent-based reasoning can be selectively deployed in intergroup contexts. As discussed below, this perspective also scaffolds predictions for future work with Black children, whose earlier racial socialization (Lesane-Brown, 2006) may produce distinct trajectories in how intent and group membership jointly shape moral evaluation.

Limitations and Future Directions

The present work contributes to a growing understanding of how moral reasoning operates in intergroup contexts, shedding light on how both children and adults evaluate intent and harm through the lens of group membership. By directly comparing responses to ambiguous

harms occurring within and between groups, these studies help clarify how judgments that may appear to be guided by principles that seem morally neutral on their face—such as the importance of intent—reflect the influence of group membership. In doing so, this research deepens scientific understanding of the intersection between morality and intergroup relations across development. However, like all work, the current studies do not answer all relevant questions.

First, the present research focused on White children in Study 2 to follow up on a specific result that emerged in Study 1—namely, that White adults prioritized intent to an especially high degree (relative to Black adults) in intergroup contexts. Future work can examine how children from minoritized groups reason about similar behaviors. Such work would offer a fuller developmental picture of the role that social group membership plays in moral cognition. Additionally, testing children from diverse racial backgrounds would allow researchers to test hypotheses regarding the emergence of intent-based moral judgments. For instance, perhaps parents discuss intent differently with Black children than with White children, leading Black children—and the Black adults they become—to place less emphasis on perpetrator intent than do White participants.

Second, the current work focused on Black and White racial groups because prototypical instances of racial harms in the United States involve White perpetrators and Black targets (Eberhardt, 2019; Mills, 1997; Sidanius & Pratto, 1999). However, focusing on these two groups does not capture the full racial diversity present among humans. Future work can test the extent to which the present findings generalize to evaluations of White perpetrators' behaviors toward non-Black, racially minoritized targets (e.g., Asian individuals). Future research can also examine how individuals reason about intergroup behaviors that do not involve a socially

dominant group (e.g., scenarios involving Black and Asian individuals). Such work would further clarify what role social group membership plays in moral cognition and enhance representation of participants who are currently underrepresented in psychological science.

Finally, Study 1 asked adults to report how much they prioritized intent when responding to transgressions. This measure provides insight into how adults understand their own cognitive processes but leaves open questions about the extent to which responses reflect adults' self-understanding versus their understanding of norms regulating how they *should* make moral judgments. Study 2 may speak against the latter interpretation, as intent appeared to exert a particularly strong influence on older White participants' responses in intergroup contexts even when intent was manipulated rather than elicited via self-report. Nevertheless, future work can use methods such as bogus pipeline (Jones & Sigall, 1971) to further probe adults' responses. Additional work can also build on the vignette-based approach adopted here to investigate how intent shapes moral judgment in everyday interaction. We opted to ask participants about vignettes in order to increase experimental control (for example, by manipulating characters' race and the extent to which behaviors were intentional while leaving other information in the vignette identical in Study 2). Future work can build on this strength by examining intergroup and intragroup transgressions in everyday contexts. For instance, researchers can interview participants about actual behaviors they have experienced and probe the extent to which actual perpetrators' intentions shaped moral judgments.

Constraints on Generality

As discussed above, future work can test the extent to which the findings reported here generalize to perceptions of White perpetrators' actions toward minoritized individuals who are not Black and evaluations of behaviors that occur between members of two different racially

minoritized group. If the results reported here reflect psychological processes relevant to dominant and minorized groups, similar results as those reported here would be expected to emerge among participants judging White perpetrators' actions toward non-Black members of racially minoritized groups. Further, if the processes reported here are specific to dominant group members, these findings may not emerge in cases where perpetrators and targets both belong to minoritized groups. These hypotheses remain to be tested in future work.

Conclusion

Two studies asked how Black and White adults (Study 1) and White children (Study 2) respond to others' transgressions. Overall, results suggested that White adults and 7- to 9-year-olds show a particularly strong sensitivity to intent when responding to White perpetrators' actions toward Black targets. By integrating work on intergroup relations, moral psychology, and child development, these results highlight how a process commonly portrayed as a cognitive achievement (the outcome-to-intent shift in children's moral evaluations) depends on social group membership and may have drawbacks as well as benefits. Understanding the role of social factors and developmental processes in adults' judgments is essential for building an understanding of human psychology that accounts for both individual judgments and the broader social structures within which those judgments unfold.

References

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

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

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37-49. https://doi.org/10.1002/cd.96

Bobo, L. D., Charles, C. Z., Krysan, M., & Simmons, A. D. (2012). The *real* record on racial attitudes. In P. V. Marsden (Ed.), *Social trends in American life: Findings from the General Social Survey since 1972* (pp. 38-83). Princeton University Press.

Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*. Rowman & Littlefield Publishers, Inc.

Cimpian, A., & Leslie, S-J. (2017). The brilliance trap. *Scientific American*, *317*, 60-65. https://doi.org/10.1038/scientificamerican0917-60

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

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 17320-17325. https://doi.org/10.1073/pnas.1408988111

Cushman, F. A. (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

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*, 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Eberhardt, J. L. (2019). *Biased: Uncovering the hidden prejudice that shapes what we see, think, and do*. Viking.

Eberhardt, J. L., Davies, P. G., Purdue-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383-386. https://doi.org/10.1111/j.1467-9280.2006.01716.x

Fletcher, G. (1998). *Basic concepts of criminal law*. Oxford University Press.

Gaertner, S. L., & Dovidio, J. F. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues*, *61*, 615-639. https://doi.org/10.1111/j.1540-4560.2005.00424.x

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

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

Heiphetz, L., & Young, L. L. (2014). A social cognitive developmental perspective on moral judgment. *Behaviour*, *151*, 315-335. https://doi.org/10.1163/1568539X-00003131

Jarvie, J. (2015). Girl thrown from desk didn't obey because the punishment was unfair, attorney says. Retrieved on August 15th, 2025, from https://www.latimes.com/nation/la-na-girl-thrown-punishment-unfair-20151029-story.html

Jones, E. E., & Sigall, H. (1971). Bogus pipeline: New paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*, 349-364. https://doi.org/10.1037/h0031617

Killen, M., Elenbaas, L., & Rizzo, M. T. (2018). Young children's ability to recognize and challenge unfair treatment of others in group contexts. *Human Development*, *61*, 281-296. https://doi.org/10.1159/000492804

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

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

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

Leach, C. W., Zeineddine, F. B., & Cehajic-Clancy, S. (2013). Moral immemorial: The rarity of

self-criticism for previous generations' genocide or mass violence. *Journal of Social Issues*, *69*, 34-53. https://doi.org/10.1111/josi.12002

Lee, Y., & Solomon, L. H. (2025). Children's and adults' understanding of how punishment shapes social relationships. *Journal of Experimental Child Psychology*, *260*, 106311. https://doi.org/10.1016/j.jecp.2025.106331

Lesane-Brown, C. L. (2006). A review of race socialization within Black families. *Developmental Review*, 26, 400-426. doi: 10.1016/j.dr.2006.02.001

Marshburn, C. K., Reinkensmeyer, B. A., & Knowles, E. D. (2023). Dominance motivated delusions: Whites with high social dominance orientation perceive equal amounts of institutional racism between Blacks and Whites. *Group Processes & Intergroup Relations*, *26*, 1244-1270. https://doi.org/10.1177/13684302221103984

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

Mills, C. W. (1997). *The racial contract*. Cornell University Press.

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

Mott, C. J., & Solomon, L. H. (In press). Contrasting guilty minds: Exposure to contrast concepts narrows conceptions of acting knowingly and recklessly. *Journal of Experimental Psychology: Applied*.

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

Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, *11*, 381-398. https://doi.org/10.1177/1745691616635592

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*, 536-552. https://doi.org/10.1037/0033-2909.112.3.536

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

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

Roberts, S. O., & Gelman, S. A. (2017). Now you see race, now you don't: Verbal cues influence children's racial stability judgments. *Cognitive Development*, *43*, 129-141. https://doi.org/10.1016/j.cogdev.2017.03.003

Rucker, J. M., & Richeson, J. A. (2021). Toward an understanding of structural racism: Implications for criminal justice. *Science*, *374*, 286-290. https://doi.org/10.1126/science.abi7779

Ruisch, B. C., & Ferguson, M. J. (2022). Changes in Americans' prejudices during the presidency of Donald Trump. *Nature Human Behavior*, *6*, 656-665. https://doi.org/10.1038/s41562-021-01287-2

Schaffner, B. F. (2020). *The acceptance and expression of prejudice during the Trump era*. Cambridge University Press.

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

Sue, D. W., & Spanierman, L. B. (2020). *Microaggressions in everyday life: Race, gender, and sexual orientation.* John Wiley & Sons, Ltd.

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

Tobia, K. (2021). Law and the cognitive science of ordinary concepts. In B. Brozek, J. C. Hage, J. Hage, & N. Vincent (Eds.), *Law and mind: A survey of law and the cognitive sciences* (pp. 86-96). Cambridge University Press.

Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.

Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology, 77*(2), 279-292. https://doi.org/10.1037/0022-3514.77.2.279

Washington, H. A. (2006). *Medical apartheid: The dark history of medical experimentation on Black Americans from colonial times to the present*. Doubleday.

Wu, J., & Chentsova-Dutton, Y. (2026). Why the west minds and the east behaves: An integrative review of the cultural evolution of mind-behavior orientations. *Journal of Cross-Cultural Psychology*, *57*, 3-50. https://doi.org/10.1177/00220221251362805

Yonas, D., & Solomon, L. H. (2025). Age-related differences in information-seeking behavior about morally relevant events. *Child Development*, *96*, 705-720. https://doi.org/10.1111/cdev.14200

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

Yucel, M., Drell, M. B., Jaswal, V. K., & Vaish, A. (2022). Young children do not perceive distributional fairness as a moral norm. *Developmental Psychology, 58*, 1103-1113. https://doi.org/10.1037/dev0001349