Beliefs About Moral Obligation Structure Children's Social Category-Based Expectations

Lisa Chalik
Yeshiva University

Yarrow Dunham
Yale University

Young children view social category members as morally obligated toward one another, and expect these obligations to shape people's social behavior. The present work investigates how children specify *which* behaviors are constrained by social categories in this way. In two studies (*N* = 128), 4- and 5-year-old children predicted that morally positive behaviors would be directed toward in-group members, and that morally negative behaviors would be directed toward out-group members, but did not hold equally strong expectations about behaviors described as positive or negative for reasons irrelevant to morality. Thus, notions of morality are embedded within children's representations of social categories, such that when learning about novel moral norms, children immediately expect those obligations to uniquely hold within social groups.

As children navigate the social world, they must sort through an incredibly complex range of human characteristics and behaviors. One way in which they do so is by relying on intuitive theories—abstract, causal-explanatory frameworks governing reasoning in particular domains. For example, past work has suggested that children draw on intuitive theories of psychology (e.g., Wellman, 1992) and sociology (e.g., Rhodes, 2012, 2013) to make sense of other people's thoughts and actions (Gopnik & Wellman, 2012; Wellman & Gelman, 1992). Children use these theories to identify the causal mechanisms that produce human behavior, such as beliefs, goals, and social relationships, enabling them to evaluate people's behavior and predict how others will act in novel scenarios.

One particularly important component of intuitive sociology that young children use to make sense of behavior is the belief that social category members are morally obligated toward one another. By this account, children view people as holding intrinsic, moral obligations toward in-group members, but do not view these obligations as extending across social category boundaries. As evidence for

this proposal, Rhodes and Chalik (2013) documented that 3- to 9-year-old children evaluate harmful behaviors differently depending on whether those behaviors occurred among in-group members or between out-group members. In this work, children viewed harmful *intra*group behaviors as consistently wrong, regardless of whether those events were structured by explicit rules (e.g., even if the harmful action had been expressly permitted by authority figures in the local context), but viewed the wrongness of harmful *inter*group behaviors as dependent on local rules (e.g., the harmful action was deemed more acceptable if it had been permitted by a local authority figure). These findings suggest that children view intragroup harm as violating intrinsic obligations, whereas they view intergroup harm as wrong for more context-dependent reasons—in other words, they believe that social category members are *morally obligated* not to harm one another.

This belief has the potential to powerfully shape children's expectations of how people act toward one another, as children generally predict that people will act in line with prescriptive obligations (Kalish & Shiverick, 2004; Roberts, Gelman, & Ho, 2017). To that end, a great deal of work has now investigated how children use social categories to constrain their expectations of social behavior. For example, as early as age 3, children expect people to direct negative behaviors toward out-group

members, rather than in-group members (Chalik & Rhodes, 2014, 2018; Chalik, Rivera, & Rhodes, 2014; Rhodes, 2012), and by age 4, children explain negative intergroup behaviors as having occurred because of category memberships (Chalik & Rhodes, 2015; Rhodes, 2014). These findings suggest that within the first few years of life, children broadly expect negative behaviors to be shaped by social categories.

Children also hold an early emerging belief that social categories shape certain positive behaviors. As early as 17 months of age, infants expect people to help fellow group members, but not members of other groups, who are in need (Jin & Baillargeon, 2017). By age 3, children expect people to be friends with fellow group members, rather than with members of other groups (Chalik & Rhodes, 2018; Shutts, Roben, & Spelke, 2013), and guide people to allocate resources preferentially to individuals with whom they have a preexisting social relationship, such as family and friends (Olson & Spelke, 2008). Furthermore, by age 4, children expect people to protect in-group members, rather than out-group members, from harmful events (Chalik & Rhodes, 2018). These expectations are also reflected in children's own attitudes and actions toward in- and out-group members; infants and young children prefer in-group members to out-group members across a range of experimental paradigms (Bigler, Jones, & Lobliner, 1997; Dunham, Baron, & Carey, 2011; Griffiths & Nesdale, 2006; Hamlin, Mahajan, Liberman, & Wynn, 2013; Heiphetz, Spelke, & Banaji, 2013; Jordan, McAuliffe, & Warneken, 2014; Renno & Shutts, 2015).

All this work suggests that children's expectations of social behavior are influenced by their knowledge of social categories, such that they expect negative actions to occur between members of different groups, and positive actions to occur among members of the same group. However, information about whether behaviors are positive or negative is insufficient to fully explain the behaviors that children do and do not view as constrained by social categories. Indeed, some work has suggested that there exist many positive behaviors that young children do not view as shaped by social groups. For example, in a number of studies, Rhodes and colleagues (2012, 2014, 2018) have introduced children to two novel groups and asked them to predict whether an individual will direct positive behaviors, such as sharing and hugging, toward fellow group members or toward members of the opposing group. These studies have found no evidence of systematic expectations in preschoolers; children do not expect that people will preferentially direct these behaviors toward in-group members until at least age 6. As converging evidence for this point, a study by DeJesus, Rhodes, and Kinzler (2014) found that 4- to 5-year-old children did not expect people to allocate resources preferentially toward the in-group, but that 6- to 10-year-old children did. Valence (i.e., whether a behavior is positive or negative) therefore appears to be an important factor in children's social category-based predictions, but it cannot entirely account for the range of behaviors that children view as constrained by social categories.

In other words, children do not view all behaviors as relevant to their abstract expectations of how social category members are supposed to interact with one another. However, no research has investigated how children distinguish behaviors that they do and do not view as shaped by social groups. The present work seeks to answer this question by asking how children specify which behaviors can be predicted using their intuitive theory of sociology. In light of evidence that this theory centers on the moral obligations that group members hold toward one another, as described earlier (Rhodes & Chalik, 2013), we test the possibility that children primarily view behaviors as constrained by social groups if they believe that those behaviors are relevant to moral obligation. There exist many behaviors that may be thought of as positive or negative, but for reasons irrelevant to morality; for example, shaking someone's hand when you first meet them is generally viewed as a positive act, but not a morally obligated one. Children are certainly aware of this distinction—the extensive literature on Social Domain Theory has documented that children as young as age 3 distinguish morally obligated behaviors from nonmoral, conventional ones (Killen & Smetana, 1999; Nucci, 1981; Nucci & Turiel, 1978; Richardson, Mulvey, & Killen, 2012; Smetana, 2006; Smetana, Jambon, & Ball, 2014; Smetana, Killen, & Turiel, 1991; Turiel, 1983). Thus, we propose that information about the *moral status* of behaviors (i.e., whether they are morally relevant or not) should shape children's social category-based predictions, above and beyond information about the valence of those behaviors (i.e., whether they are positive or negative). To investigate this possibility, we test 4- and 5-year-olds, so that we can capitalize on a period of time in which children's beliefs about moral behavior may be particularly malleable (the majority of prior work on this topic has focused on children between the ages of 3 and 6, showing that a

great deal of important change happens in children's predictions of intergroup behavior before the sixth birthday). Across two studies, we introduce children to novel behaviors that vary on whether they are positive or negative, and whether they are prescribed by moral or nonmoral standards. We then ask children to predict whether those behaviors will occur in intragroup or intergroup contexts. If children's predictions of social behavior are primarily informed by their beliefs about moral obligation, we expect children's group-based expectations to be driven by information about moral status, above and beyond information about valence.

Study 1

In Study 1, we introduced children to a novel behavior that varied in whether it was positive or negative (valence), and whether it was prescribed by moral or nonmoral rules (moral status). To manipulate moral status, we relied on a feature of moral thought that has been well-documented in research on the development of moral cognition (Killen & Smetana, 2015; Nucci, 1981; Smetana, 2006; Turiel, 1983): authority-independence. People generally view moral rules as binding, regardless of whether they are upheld by authority figures in the immediate social context (e.g., most of us would continue to view stealing as wrong even if a local government decided to allow it). Such moral rules can be contrasted with rules that are viewed as wrong only because authority figures have deemed them so (e.g., in the United States, it is wrong to drive on the left side of the road, but government officials could change this rule if they wanted to). We established this distinction for preschoolers in this study by describing the novel behavior as either (a) consistently positive or negative, regardless of what a teacher says (moral), or (b) only positive or negative when a teacher explicitly says so (nonmoral).

After introducing children to this behavior, we introduced them to two novel groups, and asked them to predict whether individuals from those groups would direct the novel behavior toward in-group or out-group members. If children's expectations of social behavior are driven by beliefs about moral obligation among category members, then children should view behaviors as constrained by category membership in the moral condition, but not in the nonmoral condition, for both positive and negative behaviors.

Method

Participants

Participants included 64 4- and 5-year-old children (M_{age} = 4.79, range = 4.01–5.95, 31 female) recruited at preschools and museums in Connecticut. Participants were 72% White, 3% Asian, 3% Hispanic, 6% Mixed race, and 16% Unreported. For children recruited at preschools, consent forms were sent home and returned by parents, and children were tested in quiet rooms at their schools. For children recruited at museums, families were approached by researchers and invited to participate in science experiments, and children were tested immediately in a quiet space at the museum. One additional child was tested but excluded from analysis because he did not understand English. Children were randomly assigned to one of four conditions, described next (n = 16 per condition).

Procedure

All stimuli were presented as hand-drawn pictures. The experimenter first showed a picture of a group of children and explained, "They all go to a new school. We're going to learn about things the kids do in their school." Next, the experimenter introduced and described a novel behavior. This behavior was not depicted visually, so that children could not map the behavior onto an already-familiar real-world behavior—for this part of the study, children were shown a blank page. The description of the novel behavior varied by condition, according to a 2 (moral status: moral, nonmoral) × 2 (valence: positive, negative) experimental design. In the *moral* conditions, the behavior—called *wugging*—was described as governed by authority-independent criteria (i.e., its permissibility was the same regardless of rules set by a teacher), and in the *nonmoral* conditions, the behavior—called *daxing*—was described as governed by authority-dependent criteria (i.e., its permissibility was based on rules set by a teacher). In the *positive* conditions, the behavior was described positively, and in the *negative* conditions, the behavior was described negatively. Then, the experimenter described two sample instances of the behavior. Text for these descriptions and one of the sample instances for each can be found in Table 1. Children answered attention check questions (e.g., "Can you remind me, is it good or bad for kids to *wug* each other? Do kids *wug* each other? What if the teachers in school say you do not have to *wug* somebody—then, do kids

Table 1

Descriptions and Sample Instances of Novel Behavior for Each Condition, Study 1

Condition	Description	Sample instance
Nonmoral-positive	It's good for kids to *dax* each other. Kids *dax* other kids, and it's really good if they do. But, if the teachers in school say you don't have to *dax* somebody, you don't have to if you don't want to.	Look. Ben *daxed* Tommy. That was really good because kids *dax* other kids. A teacher came over and told Ben he didn't have to *dax* Tommy. Ben *daxed* Tommy again, and it was just okay because kids don't have to *dax* other kids if the teachers say so.
Nonmoral-negative	It's bad for kids to *dax* each other. Kids don't *dax* other kids, and it's really bad if they do. But, if the teachers in school say you can *dax* somebody, you can if you want.	Look. Ben *daxed* Tommy. That was really bad because kids don't *dax* other kids. A teacher came over and told Ben he could *dax* Tommy. Ben *daxed* Tommy again, and it was okay because kids can *dax* other kids if the teachers say it's okay.
Moral-positive	It's good for kids to *wug* each other. Kids should *wug* other kids, and it's really good if they do. And, even if the teachers in school say you don't have to *wug* somebody, you still should no matter what.	Look. Ben *wugged* Tommy. That was really good because kids should *wug* other kids. A teacher came over and told Ben he didn't have to *wug* Tommy. Ben *wugged* Tommy again, and it was still really good because kids should *wug* other kids no matter what.
Moral-negative	It's bad for kids to *wug* each other. Kids shouldn't *wug* other kids, and it's really bad if they do. And, even if the teachers in school say you can *wug* somebody, you still shouldn't no matter what.	Look. Ben *wugged* Tommy. That was really bad because kids shouldn't *wug* other kids. A teacher came over and told Ben he could *wug* Tommy. Ben *wugged* Tommy again, and it was still really bad because kids shouldn't *wug* other kids no matter what.

have to *wug* each other?") to ensure that they fully understood this information.

Next, the experimenter explained that they were now going to talk about some other characters, and participants were introduced to two novel groups of children marked by novel labels (Flurps and Zazzes) and shirt color (blue and red). Each group was described as engaged in a cooperative tower-building activity, separate from the other group. Then, children answered six test questions in which they had to predict whether a target character was going to perform the novel behavior toward a fellow group member or a member of the other group (e.g., "Here's a Flurp on the playground. This Flurp *wugged* somebody. Who did the Flurp *wug*? Did the Flurp *wug* another Flurp? Or did the Flurp *wug* a Zazz?"). Whether the agents in the test questions were Flurps or Zazzes and the lateral position of the answer choices were counterbalanced across participants. Participants received a score of "0" every time they made an intergroup prediction, and a score of "1" every time they made an intragroup prediction.

Results

All raw data and code are available on the Open Science Framework at https://osf.io/ysfzg/. Children's responses are presented as the proportion of the time that children made intragroup predictions, with odds ratios as measures of effect size. We analyzed children's responses using binomial regression models, with valence and moral status as between-subjects factors, testing for both possible main effects and an interaction. Children made more intragroup predictions for positive behaviors than for negative behaviors, $\chi^2(1) = 19.24$, $p < .001$, $OR = 7.14$, and more intragroup predictions for moral behaviors than for nonmoral behaviors, $\chi^2(1) = 3.17$, $p = .075$, $OR = 4.00$. These effects were qualified by an interaction between valence and moral status (see Figure 1; $\chi^2(1) = 15.93$, $p < .001$, $OR = 5.86$): For behaviors described as moral, children made more intragroup predictions for positive behaviors than for negative behaviors, $\chi^2(1) = 34.65$, $p < .001$, whereas for behaviors described as nonmoral, children's predictions did not differ for positive and negative behaviors ($p = .470$). Comparisons of each group to chance confirmed these findings: For moral behaviors, children reliably predicted that positive actions ($M = 0.81$, CI [0.73, 0.89]) would occur among members of the same group, $\chi^2(1) = 31.45$, $p < .001$, whereas they reliably predicted that negative actions ($M = 0.38$, CI [0.28, 0.47]) would occur between members of different groups, $\chi^2(1) = 5.87$, $p = .015$. For nonmoral behaviors, however, children did not differ from chance in their predictions of positive, $M = 0.52$, CI [0.42,
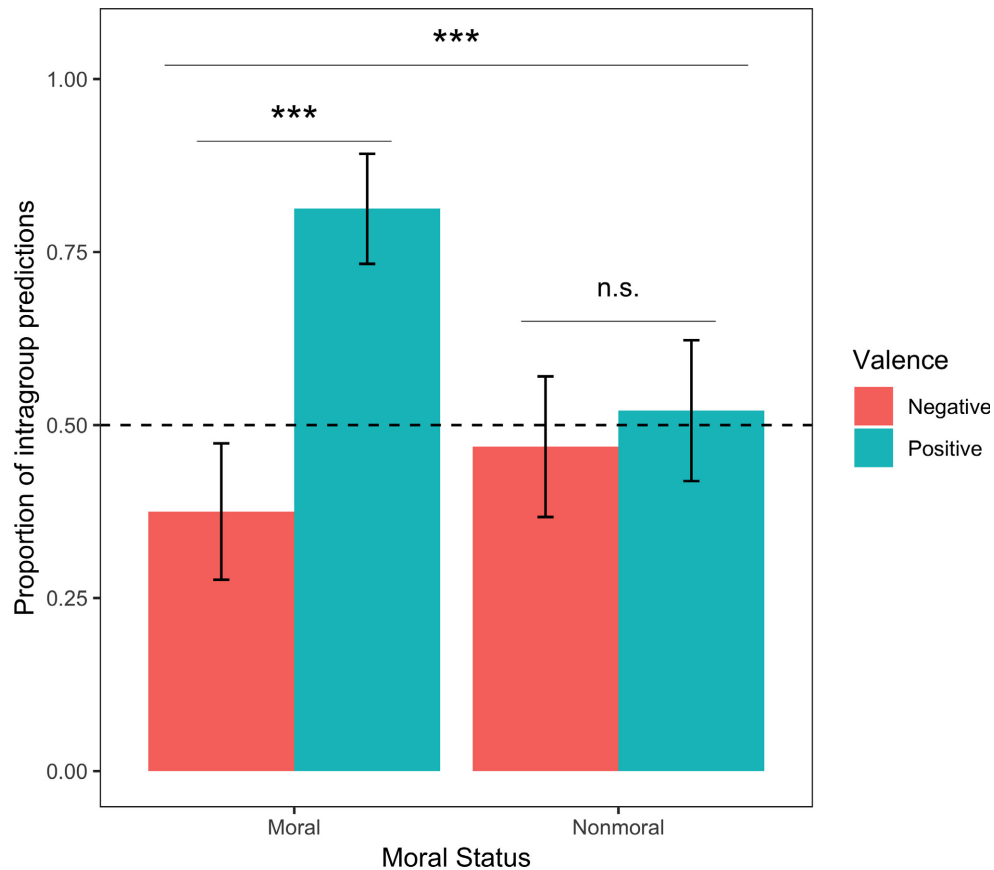
Figure 1. The proportion of the time that children predicted that the novel behavior would occur among fellow group members, separated by condition. Error bars represent 95% confidence intervals. ***$p < .001$.

0.62]) or negative ($M = 0.47$, CI [0.37, 0.57]) actions, ps > .541. Relative to the responses that would be expected by chance, describing a positive behavior as moral increased the odds of an *intra*group prediction by 4.33 (CI [2.66, 7.46]), and describing a negative behavior as moral increased the odds of an *inter*group prediction by 1.67 (CI [1.11, 2.54]).

Discussion

In Study 1, we found that children used social categories to constrain their expectations of a novel social behavior, but only if that behavior had been described as morally relevant; if the behavior had been described as nonmoral, children did not expect it to be shaped by social group memberships. This finding remained consistent whether the behavior was described as positive or negative. These results suggest that children's expectations of intergroup behavior are shaped by abstract notions of whether behaviors are morally obligated or not, and not by information about valence.

Study 2

The primary aim of Study 2 was to determine whether the results of Study 1 can be extended to another conceptualization of morality. In Study 1, we defined morality based on whether actions were governed by authority figures in the local context. Another distinction that has been widely used to mark the special status of moral rules in the literatures on moral development (Smetana, 1981; Tisak & Turiel, 1988) and social psychology (Haidt, Koller, & Dias, 1993; Shweder, Mahapatra, & Miller, 1987) regards whether a rule applies consistently across multiple locations. Moral rules are viewed as applying universally, regardless of location (e.g., most of us view murder as wrong, no matter where in the world the perpetrator is); nonmoral rules, on the other hand, are viewed as depending on local customs (e.g., it is polite in the United States to shake people's hands upon first meeting them, but this practice is not observed in many parts of the world; Nichols, 2004; Smetana, 1981; but see Sarkissian, Park, Tien, Cole Wright, & Knobe, 2011 for an

alternative perspective on moral universality). In Study 2, we apply this principle by varying whether the novel behavior is obligated across two locations with which children are familiar: at home and at school. One benefit of this manipulation is that it uses language familiar to young children. The language used in Study 1—for example, stating that teachers might tell children that they *can* perform negative behaviors—is unlikely to be similar to language that children actually hear in the world in reference to social behaviors. However, children are often told that certain behaviors are differentially structured at home versus at school (e.g., you can wear your pajamas at home, but it would be inappropriate to wear them to school, and raising your hand before you speak is required at school, but not at home). Thus, the manipulation used in Study 2 allows us to test a relevant moral principle while using language that is familiar to children (although we acknowledge that there may be some moral behaviors that do not map perfectly onto this distinction).

Another aim of Study 2 was to address an alternate interpretation of the findings from Study 1. It is possible that children's responses in Study 1 were more systematic in the moral conditions than in the nonmoral conditions not because children were distinguishing the behaviors based on abstract beliefs about moral obligation, but simply because they viewed the moral behaviors as more extreme than the nonmoral behaviors. On this account, children used the information about moral status (i.e., the fact that a behavior might still be obligated regardless of what authority figures say) as a cue to the extremity of the behavior in question; for example, they may have interpreted the morally negative behaviors as extreme negative actions, but the non-morally negative behaviors as only mildly negative. If this interpretation held, it would imply that children did not generate systematic predictions about the nonmoral behaviors simply because they did not encode the behaviors as extreme enough to be important. To test this interpretation, in Study 2, after asking children to predict whether the novel behavior will occur in intra or intergroup contexts, we ask children to evaluate another instance of the behavior.

Method

Participants

Participants included 64 4- and 5-year-old children (M_{age} = 5.02, range = 4.01–5.98, 33 female)

recruited in the same manner as in Study 1. Participants were 59% White, 5% Asian, 5% Hispanic, 11% Mixed race, and 20% Unreported. An additional three children were tested but excluded from analysis; one because of developmental delays, one because of experimenter error, and one because of parental interference. Children were randomly assigned to one of four conditions (n = 16 per condition).

Procedure

The procedure for Study 2 was identical to that of Study 1 with three exceptions: First, the same label (*wugging*) was used for the novel behavior across all four conditions. Second, the nonmoral and moral conditions were distinguished not by whether they were dependent on teachers' rules, but by whether the rules structuring the behavior were consistent across two separate locations (at home and at school). Text for the descriptions of the behaviors and sample instances for each can be found in Table 2. Third, after completing the test questions, children answered evaluation questions to assess the degree to which they saw *wugging* as positive or negative. For these questions, children were told to imagine two new children who were not Flurps and not Zazzes, and were told that one of the children *wugged* the other child. Children were then asked, "Was what the kid did good or bad?", and whether it was "a little good/bad, pretty good/bad, or very, very good/bad." Children indicated their responses by pointing to a 3-point smiley- or frowny-face scale, depending on their initial response. Responses were scored as 1 if children said "A little good/bad," 2 if they said "Pretty good/bad," and 3 if they said "Very, very good/bad."

Results

We again analyzed children's responses using binomial regression models, with valence and moral status as between-subjects factors, testing for both possible main effects and an interaction. Children made more intragroup predictions for positive behaviors than for negative behaviors, $\chi^2(1)$ = 26.32, $p < .001$, OR = 5.00. Furthermore, as in Study 1, there was an interaction between valence and moral status, $\chi^2(1)$ = 5.36, p = .021, OR = 2.70, suggesting that the effect of valence was stronger in the moral condition, $\chi^2(1)$ = 27.00, $p < .001$, than in the nonmoral condition, $\chi^2(1)$ = 4.67, p = .031; see Figure 2. Comparisons of each group to chance were

Table 2

Descriptions and Sample Instances of Novel Behavior for Each Condition, Study 2

| Condition	Description	Sample instance
Nonmoral-positive	It's really good for kids to *wug* each other. At school, kids should *wug* other kids. But, at home, kids don't have to *wug* other kids if they don't want to.	Look. At school today, Joey *wugged* Mike. That was really good because kids should *wug* other kids when they are at school. Then at home, Joey *wugged* Mike again, and it was just okay because kids don't have to *wug* other kids when they are at home if they don't want to.
Nonmoral-negative	It's really bad for kids to *wug* each other. At school, kids should not *wug* other kids. But, at home, kids can *wug* other kids if they want to.	Look. At school today, Joey *wugged* Mike. That was really bad because kids should not *wug* other kids when they are at school. Then at home, Joey *wugged* Mike again, and it was okay because kids can *wug* other kids when they are at home if they want to.
Moral-positive	It's really good for kids to *wug* each other. At school, kids should *wug* other kids. And at home, kids should *wug* other kids too.	Look. At school today, Joey *wugged* Mike. That was really good because kids should *wug* other kids when they are at school. Then at home, Joey *wugged* Mike again, and it was really good again because kids should *wug* other kids when they are at home too.
Moral-negative	It's really bad for kids to *wug* each other. At school, kids should not *wug* other kids. And at home, kids should not *wug* other kids either.	Look. At school today, Joey *wugged* Mike. That was really bad because kids should not *wug* other kids when they are at school. Then at home, Joey *wugged* Mike again, and it was really bad again because kids should not *wug* other kids when they are at home either.
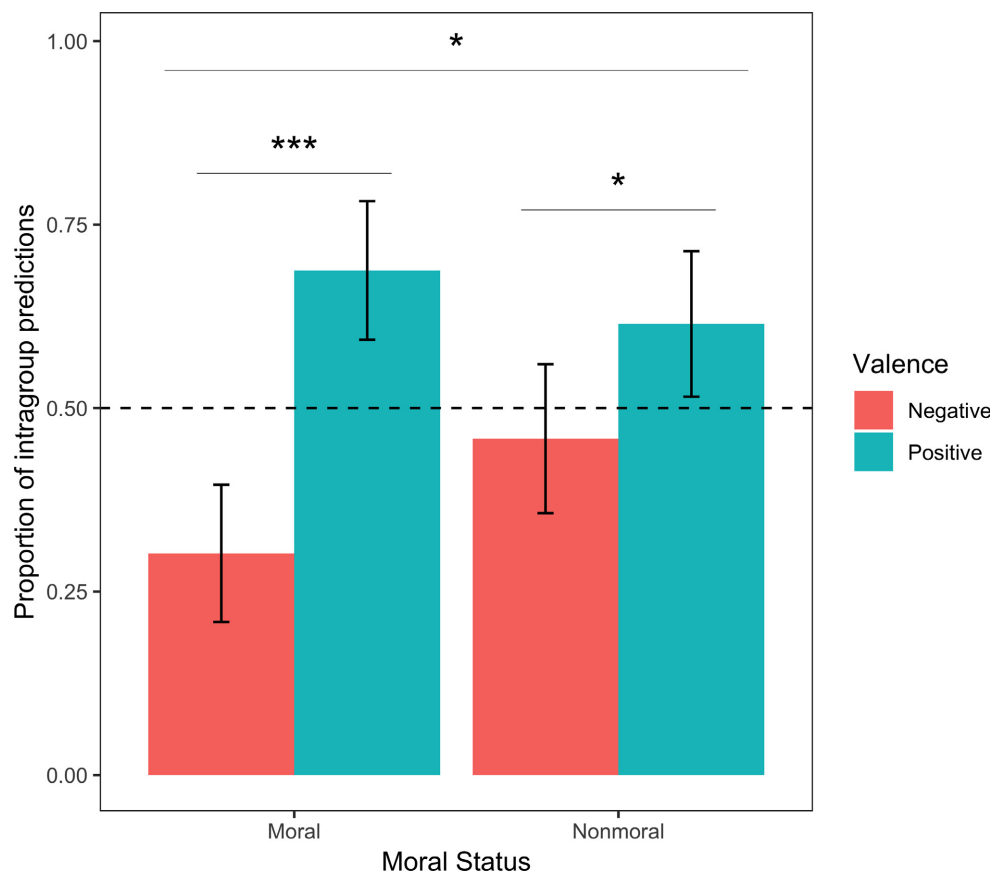
Figure 2. The proportion of the time that children predicted that the novel behavior would occur among fellow group members, separated by condition. Error bars represent 95% confidence intervals. *$p < .05$. ***$p < .001$.

consistent with these findings: For moral behaviors, children reliably predicted that positive actions would occur among members of the same group, $\chi^2(1) = 12.82$, $p < .001$, whereas they reliably predicted that negative actions would occur between members of different groups, $\chi^2(1) = 14.19$, $p < .001$. For nonmoral behaviors, children did not differ from chance in their predictions of negative actions ($p = .415$). Unlike in Study 1, however, children did predict that nonmoral-positive behaviors would occur among fellow group members, $\chi^2(1) = 4.95$, $p = .026$. Relative to the responses that would be expected by chance, describing a positive behavior as moral increased the odds of an *intra*group prediction by 2.2 (CI [1.44, 3.43]), and describing a negative behavior as moral increased the odds of an *inter*group prediction by 2.31 (CI [1.51, 3.62]).

In response to the evaluation questions, 91% of children's responses to whether the behavior was positive or negative were consistent with the way the behavior had been introduced at the beginning of the study. To test the degree to which children endorsed these evaluations, we subjected children's 3-point scale ratings—where higher scores reflect more extreme ratings (i.e., more negativity in the negative conditions, and more positivity in the positive conditions)—to a univariate analysis of variance with valence and context as between-subjects factors, testing for both possible main effects and an interaction. We found no significant effects (ps > .336), suggesting that the degree to which children viewed the behaviors as positive or negative did not vary across the four conditions (nonmoral-positive: $M = 2.75$, CI [2.39, 3.11]; nonmoral-negative: $M = 2.63$, CI [2.36, 2.89]; moral-positive: $M = 2.81$, CI [2.52, 3.10]; moral-negative: $M = 2.63$, CI [2.20, 3.05]).

Discussion

The findings from Study 2 conceptually replicated the findings from Study 1: Children viewed social categories as constraining moral behaviors to a greater degree than nonmoral behaviors. This pattern suggests that children's expectations of social behavior are driven primarily by beliefs about whether those behaviors are or are not morally relevant, rather than simply whether they are positive or negative. Furthermore, our data speak against the interpretation that children view the moral behaviors as more extreme than the nonmoral behaviors, because children evaluated the different types of behaviors as similarly extreme (although

care is warranted here given difficulties in interpreting null effects, especially with relatively small sample sizes).

An open question regards why children in Study 2, unlike those in Study 1, reliably predicted that nonmoral-positive behaviors would occur among fellow group members. This finding only appeared in one of two studies; as such, we do not wish to over-interpret it here. However, one possibility is that children were responding to the specific normative language used to describe the behaviors, which was consistent across the moral- and nonmoral-positive conditions. In particular our use of the word "should" could have served as a secondary cue to moral status ("at school, kids should *wug* other kids"), as one important feature of moral rules in general is that they dictate how people *should* act. If this is the case, our results could be interpreted as evidence of another, albeit weaker, cue to moral status—specifically, the normative implications of the word "should." Future work could further investigate this possibility by contrasting different cues to moral status to evaluate their relative strength in children's moral reasoning.

In any case, our most critical finding is that in both Study 1 and Study 2, the differences between children's predictions of behaviors described as positive and negative were significantly larger for the moral conditions than for the nonmoral conditions. Thus, we can still conclude that beliefs about moral obligation drove children's expectations of behavior, above and beyond information about valence.

Combined Analyses of Studies 1 and 2

As noted earlier, Study 2 conceptually replicated the results from Study 1; yet, there are still some interesting open questions regarding the differences between these two studies. Primarily, it is unclear whether the slight difference in patterns across Studies 1 and 2 reflects a real difference in how children conceptualize authority-independence and universality as cues to moral obligation.

To address this question, we conducted an additional set of analyses on the data from both the studies combined, using study as an independent variable. We subjected all 128 children's responses to a binomial regression model with valence, moral status, and study as between-subjects factors, testing for all possible main effects and interactions. As in the analyses for each individual study, we found a main effect of valence, suggesting that children

made more intragroup predictions for positive behaviors than for negative behaviors, $\chi^2(1) = 45.36$, $p < .001$, $OR = 10.26$. We also found an interaction between moral status and study, $\chi^2(1) = 4.43$, $p = .035$, $OR = 1.33$, suggesting that only in Study 1, there was a marginally significant effect of moral status, $\chi^2(1) = 3.78$, $p = .052$; see Study 1 Results. Furthermore, there was an interaction between valence and moral status, $\chi^2(1) = 19.70$, $p < .001$, $OR = 12.74$, suggesting that the effect of valence was stronger in the moral condition, $\chi^2(1) = 61.28$, $p < .001$, than in the nonmoral condition, $\chi^2(1) = 4.16$, $p = .042$. Critically, this effect did not interact with study, indicating that the interaction unfolded similarly across the two studies (and thus, across the two definitions of morality that we used).

Additionally, we directly compared children's responses for the condition that primarily drove the difference across the two studies: the positive-nonmoral condition. Focusing on responses for that condition only, a binomial regression model with study as a between-subjects factor revealed that children's responses did not differ across the two studies, $p = .191$. Thus, although Study 2 revealed a slightly different pattern from Study 1, it clearly represents a replication of our main findings. Furthermore, we can conclude that the two cues to moral obligation that we used here, authority-independence and universality, elicit similar social category-based thinking in children.

General Discussion

The present work has demonstrated that children use an intuitive theory of sociology, by which they see social categories as marking people who are morally obligated to one another, to constrain their expectations of social behavior. In two studies, children predicted that an unfamiliar positive behavior described in moral terms (either via rule contingency or consistency across multiple locations) would be directed toward in-group members, and that an unfamiliar morally negative behavior would be directed toward out-group members. They did not, however, make equally strong inferences for behaviors that had been described as positive or negative for reasons irrelevant to morality, and they did not appear to conceive of the moral behaviors as more extreme in valence. Thus, children's expectations of group-based social behavior are not driven only by beliefs about valence (i.e., that people should perform positive behaviors and avoid

negative behaviors, especially with in-group members); rather, they are driven especially by beliefs about whether behaviors are morally obligated or not. Because the social categories and behaviors in these studies were novel, we can conclude that children's tendencies here reflect their abstract expectations of behavior, rather than any prior knowledge that they hold about specific actions or social groups. Therefore, these studies demonstrate that abstract notions of morality are deeply embedded within children's representations of social categories, such that when learning about a novel moral norm, children immediately map that norm onto social categories, expecting obligations to primarily hold within a social group (note that this does not mean that moral concerns do not exist outside social group boundaries—rather, it suggests that children have an early-emerging tendency to view moral rules as *particularly* important within a social group).

One implication of the present work is that children's beliefs about moral obligation are subject to change in response to input. We manipulated whether children viewed behaviors as morally obligated by exposing them to input suggesting that those behaviors either (a) were or were not governed by authority figures in the local context, or (b) were or were not consistently obligated across two locations. The fact that children's expectations changed as a result of this input suggests that their understanding of what behaviors fall under the purview of moral obligation are malleable, open to being shaped by cultural input, and that this in turn shapes their expectations about intergroup and intragroup interactions. In addition to demonstrating that cultural variation in beliefs about morality has its roots in the first few years of life, this finding opens a range of interesting questions for further research. For example, are there other salient cues to the moral status of an action, and are some such cues more powerful than others? And are there forms of input that might encourage children to view moral obligations as extending across social category boundaries? Future work should explore these possibilities.

An important contribution of the present work is that it clarifies findings from prior research on children's expectations of social category-based behavior. Most prior work testing children's predictions of inter and intragroup behaviors has revealed an asymmetry in children's expectations of negative versus positive behavior, such that children systematically expect negative behaviors to be directed toward out-group members quite early in the

preschool years (by age 3), but do not systematically expect positive behaviors to be directed toward in-group members until slightly later in childhood (by age 6; Rhodes, 2012; Chalik & Rhodes, 2018). These findings have been interpreted as showing that children's beliefs about negativity may be privileged, emerging earlier in development (perhaps because of the prevalence of intergroup conflict throughout the course of human evolution; Cosmides, Tooby, & Kurzban, 2003), whereas their beliefs about positive behavior may be more unspecified and open to cultural input (Chalik & Rhodes, 2015). In the present work, however, we found similar results for negative and positive behaviors (i.e., children's expectations depended on moral status to a similar degree for both negative and positive behaviors), suggesting that beliefs about negativity are not privileged above beliefs about positivity early in development. These findings indicate that the stronger expectations for negative than for positive behaviors observed in prior work emerge for some other reason—perhaps because of culture-specific beliefs about moral obligation. For example, perhaps children from North American populations have internalized cultural beliefs that people are strictly obligated to avoid negative actions toward others, but are not as seriously obligated to perform positive behaviors (a belief which might be rooted in Western individualism, in which a primary moral obligation is to avoid infringing on the freedom and independence of others; Markus & Kitayama, 1991). Future work should investigate whether and how the kinds of moral values assigned to behaviors in different cultures can differentially shape children's social category-based expectations.

One limitation of the present work is that our definition of morality is not exhaustive. We chose two defining features of morality—authority-independence and universality—to act as examples that illustrate a more general phenomenon, that children view moral behaviors as constrained by social groups. However, there are certainly other components of morality that have been discussed in the literature on moral development (e.g., children view moral rules as unalterable; Smetana, 1985). For the purposes of studying how young children think about moral behavior, we believe that the moral criteria that we have drawn upon in these studies sufficiently contrast moral and nonmoral actions in a way that is understandable to young children. Still, how other features of morality interact with the ones we have

tested here is an interesting question for future work.

Another limitation is that we do not have direct evidence that the effects documented here would hold for the full range of behaviors that children regularly encounter. Because we used novel behaviors, we believe that our findings reflect children's abstract beliefs about moral action, rather than their thinking about a narrow set of specific behaviors; yet, future work should more directly test this possibility by contrasting novel behaviors with real-world ones.

In closing, our findings provide direct evidence that moral judgments are deeply embedded in, and are perhaps best considered a component of, intergroup cognition. This perspective dovetails with work arguing that the particular normative richness of human society is a biological or cultural adaptation to group living—the establishment of moral norms is a way to meet the particular demands of social coordination in an intergroup landscape (Boyd & Richerson, 1992, 2009; Gintis, Bowles, Boyd, & Fehr, 2003; Rai & Fiske, 2011; Trivers, 1971). Thus, when we teach children about moral obligation or prohibition, we may be tacitly teaching them about how members of groups should and will behave toward one another.

References

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development*, *68*, 530–543. https://doi.org/10.1111/j.1467-8624.1997.tb01956.x

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*, 171–195. https://doi.org/10.1016/0162-3095(92)90032-y

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3281–3288. https://doi.org/10.1098/rstb.2009.0134

Chalik, L., & Rhodes, M. (2014). Preschoolers use social allegiances to predict behavior. *Journal of Cognition and Development*, *15*, 136–160. https://doi.org/10.1080/15248372.2012.728546

Chalik, L., & Rhodes, M. (2015). The communication of naïve theories of the social world in parent-child conversation. *Journal of Cognition and Development*, *16*, 719–741. https://doi.org/10.1080/15248372.2014.949722

Chalik, L., & Rhodes, M. (2018). Learning about social category-based obligations. *Cognitive Development*, *48*, 117–124. https://doi.org/10.1016/j.cogdev.2018.06.010

Chalik, L., Rivera, C., & Rhodes, M. (2014). Children's use of categories and mental states to predict social

behavior. *Developmental Psychology*, *50*, 2360–2367. https://doi.org/10.1037/a0037729

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 173–179. https://doi.org/10.1016/s1364-6613(03)00057-3

DeJesus, J. M., Rhodes, M., & Kinzler, K. D. (2014). Evaluations versus expectations: Children's divergent beliefs about resource distribution. *Cognitive Science*, *38*, 178–193. https://doi.org/10.1111/cogs.12093

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, *82*, 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, *24*, 153–172. https://doi.org/10.1016/s1090-5138(02)00157-5

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108. https://doi.org/10.1037/a0028044

Griffiths, J., & Nesdale, D. (2006). Ingroup and outgroup attitudes of ethnic majority and minority children. *International Journal of Intercultural Relations*, *30*, 735–749. https://doi.org/10.1016/j.ijintrel.2006.05.001

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613–628. https://doi.org/10.1037//0022-3514.65.4.613

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science*, *24*, 589–594. https://doi.org/10.1177/0956797612457785

Heiphetz, L., Spelke, E. S., & Banaji, M. R. (2013). Patterns of implicit and explicit attitudes in children and adults: Tests in the domain of religion. *Journal of Experimental Psychology: General*, *142*, 864–879. https://doi.org/10.1037/a0029714

Jin, K., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 8199–8204. https://doi.org/10.1073/pnas.1706286114

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of ingroup favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 12710–12715. https://doi.org/10.1073/pnas.1402280111

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, *19*, 401–416. https://doi.org/10.1016/j.cogdev.2004.05.004

Killen, M., & Smetana, J. G. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development*, *70*, 486–501. https://doi.org/10.1111/1467-8624.00035

Killen, M., & Smetana, J. G. (2015). Origins and development of morality. In M. E. Lamb (Ed.), *Handbook of child psychology and developmental science* (Vol. 3, 7th ed., pp. 701–749). New York, NY: Wiley-Blackwell.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253. https://doi.org/10.1037//0033-295x.98.2.224

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/0195169344.001.0001

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, *52*, 114–121. https://doi.org/10.2307/1129220

Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*, 400–407. https://doi.org/10.2307/1128704

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*, 222–231. https://doi.org/10.1016/j.cognition.2007.12.003

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75. https://doi.org/10.1037/a0021867

Renno, M. P., & Shutts, K. (2015). Children's social category-based giving and its correlates: Expectations and preferences. *Developmental Psychology*, *51*, 533–543. https://doi.org/10.1037/a0038819

Rhodes, M. (2012). Naïve theories of social groups. *Child Development*, *83*, 1900–1916. https://doi.org/10.1111/j.1467-8624.2012.01835.x

Rhodes, M. (2013). How two intuitive theories shape the development of social categorization. *Child Development Perspectives*, *7*, 12–16. https://doi.org/10.1111/cdep.12007

Rhodes, M. (2014). Children's explanations as a window into their intuitive theories of the social world. *Cognitive Science*, *38*, 1687–1697. https://doi.org/10.1111/cogs.12129

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

Richardson, C. B., Mulvey, K. L., & Killen, M. (2012). Extending social domain theory with a process-based account of moral judgments. *Human Development*, *55*, 4–25. https://doi.org/10.1159/000335362

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, *41*, 576–600. https://doi.org/10.1111/cogs.12443

Sarkissian, H., Park, J., Tien, D., Cole Wright, J., & Knobe, J. (2011). Folk moral relativism. *Mind and Language*, *26*, 482–505. https://doi.org/10.1111/j.1468-0017.2011.01428.x

Shutts, K., Roben, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*, 35–62. https://doi.org/10.1080/15248372.2011.638686

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago, IL: University of Chicago Press.

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52*, 1333. https://doi.org/10.2307/1129527

Smetana, J. G. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology, 21*, 18–29. https://doi.org/10.1037/0012-1649.21.1.18

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Erlbaum.

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 23–45). New York, NY: Psychology Press.

Smetana, J. G., Killen, M., & Turiel, E. (1991). Children's reasoning about interpersonal and moral conflicts. *Child Development, 62*, 629–644. https://doi.org/10.1111/j.1467-8624.1991.tb01557.x

Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology, 24*, 352–357. https://doi.org/10.1037//0012-1649.24.3.352

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*, 35–57. https://doi.org/10.1086/406755

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Wellman, H. M. (1992). *The child's theory of mind*. Cambridge, MA: MIT Press.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43*, 337–375. https://doi.org/10.1146/annurev.ps.43.020192.002005