Cognitive Development 48 (2018) 117–124

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Learning about social category-based obligations

Lisa Chalik[a,*], Marjorie Rhodes[b]

[a] Stern College for Women, Yeshiva University, 215 Lexington Avenue, room 404, New York, NY 10016, USA
[b] New York University, USA

ARTICLE INFO

Keywords:
Social categorization
Framework theories
Social relationships
Moral development

ABSTRACT

Two studies tested how children (N = 196) use a framework theory of the social world to (a) shape their expectations of and (b) guide new learning about social behaviors. In Study 1, when introduced to two novel social groups, children predicted that an agent would preferentially harm members of the other group, be friends with members of their own group, and save members of their own group from harm. In Study 2, 4-year-old children who had been shown evidence of prior inter-group and intra-group interactions predicted that future behaviors would match the evidence they were shown only if the interactions they observed were consistent with their expectations of how members of groups should relate to one another. Thus, children use their framework theory to predict social behaviors and guide new learning about the social world.

1. Introduction

To navigate the world, children must observe, interpret, and make use of evidence across a range of domains. They do so by identifying the causal-explanatory mechanisms that produce events in their environment, and organizing these mechanisms into abstract theories—known as framework theories—that can be used to apply existing knowledge to novel events (Gopnik & Wellman, 2012; Wellman & Gelman, 1992). For example, children use a framework theory of biology to specify biological inheritance as the causal mechanism that makes animals hold the properties that they do; a child can thus infer that if one animal has a dangerous property—a bee with a stinger, for example—then other animals of the same type, who have biologically inherited that same property, will also be dangerous. These theories are domain-specific; when considering an object in a different domain, such as an artifact, children cannot focus on biological processes as causal mechanisms. Instead, to understand an artifact, they might appeal to a theory of psychology, by which the mental state of the artifact's creator is the causal force that gives rise to the artifact's properties (e.g., a fork has prongs because the person who made the fork wants to use it to pick up food). Critically, all of these theories guide children's attention to the causal mechanisms that act in the world, allowing children to focus on information that they can use to facilitate predictions and learning.

Children also hold framework theories about the structure of the social world, by which they attend to the causal mechanisms that guide human behavior in social contexts (Gelman, 2003; Hirschfeld, 1996; Rhodes, 2013). Quite early in childhood, children use such a theory to specify social obligation as a causal mechanism that constrains social relationships and interaction (Chalik & Rhodes, 2014; Rhodes, 2012, 2013; Rhodes, Hetherington, Brink, & Wellman, 2015; Shutts, Pemberton Roben, & Spelke, 2013). As evidence for this proposal, Rhodes and Chalik (2013) found that, by age 4, children viewed people as intrinsically obligated not to harm members of their own social groups (and thus evaluated instances of such harm negatively in all contexts), but did not view people as

[*] Corresponding author.
E-mail address: lisa.chalik@yu.edu (L. Chalik).

https://doi.org/10.1016/j.cogdev.2018.06.010
Received 4 November 2016; Received in revised form 9 May 2018; Accepted 27 June 2018

intrinsically obligated not to harm members of other groups (and thus they evaluated instances of intergroup harm more leniently when they were told that there were no rules in place prohibiting the specific harmful actions). Furthermore, as early as age 3 and across childhood, children predict that a member of a novel social category is more likely to harm a member of another group than a member of their own (Chalik & Rhodes, 2014; Rhodes, 2012). Children also explain harmful intergroup behaviors as having occurred *because* of category memberships (Rhodes, 2014) and evaluate intergroup harm as less bad than intragroup harm (Rhodes & Chalik, 2013). Thus, children's framework theory of the social world—that social group members are obligated toward one another—guides their understanding of how social behaviors play out in intergroup contexts.

What is the nature of children's belief that social category members are obligated toward one another? One possibility is that these beliefs are narrowly centered around expectations of harm. Because of either the specific importance of intergroup conflict throughout the course of human evolution (e.g., Cosmides, Tooby, & Kurzban, 2003) or more general threat-detection mechanisms in social perception (e.g., Baltazar, Shutts, & Kinzler, 2012; Kinzler & Shutts, 2008), children's early-emerging beliefs about groups and social interaction could be centered on the belief that people are obligated to avoid harm toward ingroup members (and thus, in cases where harm does occur, to direct harm toward outgroup members). Consistent with this possibility, although children begin to systematically predict that intergroup harm is more likely to occur than intragroup harm by age 3, children at this age do *not* reliably hold expectations about behaviors that do not involve harm; children only begin to hold these expectations (e.g., predicting that people will direct prosocial behaviors toward fellow group members) later in childhood, by age 6 (Rhodes, 2012).

Another possibility is that children's inferences are motivated by a broader belief that social category members are obligated to protect and affiliate with one another. By this account, younger children fail to hold reliable expectations about prosocial behaviors not because their beliefs only center around harm, but rather because the prosocial actions tested in prior work have not adequately tapped into their beliefs about obligation. From the perspective of moral philosophy, acting prosocially toward others (e.g., lending emotional support, sharing resources, and so on), while valuable, is not necessarily obligated in the same manner as avoiding harm (Knobe, 2003; Leslie, Knobe, & Cohen, 2006). Thus, if young children view social categories as marking people who hold special obligations toward one another, perhaps young children in previous work reliably predicted intergroup harm but not intragroup prosociality because they saw *not harming* as obligatory, but did not view the prosocial actions that were tested in the same manner.

The present work seeks to tease apart the above two possibilities. If children's framework theory centers around the obligations that social category members hold toward one another (beyond directing harm away from fellow group members), then it should include expectations about certain types of intragroup relations and behaviors. There are, indeed, things that group members might be obligated to do for one another, such as affiliating with one another (e.g., in the context of friendship) and protecting one another from harm, that have not been tested in prior work. One set of studies by Shutts et al. (2013) did find that by age 4, children use gender and racial categories to guide their inferences about which individuals will be friends with one another (e.g., a girl will be friends with another girl rather than with a boy). Yet, the extent to which children's inferences about these categories have reflected their abstract expectations about the social world is unclear—children might believe that a girl will be friends with another girl not because of the structure and function of social categories, but simply because they have seen many girls be friends with one another in their everyday lives. Furthermore, some work testing adults' beliefs about who people should save from harm has shown that individuals are more likely to offer aid to ingroup members than to outgroup members during events involving physical violence (Levine, Cassidy, Brazier, & Reicher, 2002) and following natural disasters (Levine & Thompson, 2004). Yet, this work cannot speak to the childhood origins of beliefs about how people should protect fellow group members. Thus, it remains unclear whether children's abstract understanding of social categories supports predictions of these types of behaviors. Study 1 seeks to resolve this open question by testing children's predictions of a wider range of behaviors than has been tested in prior work.

In Study 2, we test for evidence that children's framework theory even generates predictions of the prosocial behaviors tested in prior work, under the right conditions. Because framework theories generally support learning, here we assess whether children can more easily learn to predict patterns of social interaction that are consistent with the belief that group members are obligated to one another. In particular, if children's beliefs about social obligation go beyond expectations of harm, children should more easily learn to predict patterns where prosocial behaviors are directed toward fellow group members, rather than toward members of another group. Furthermore, consistent with prior work, children should more easily learn patterns where harmful behaviors are directed toward outgroup members, rather than toward ingroup members.

2. Study 1

In Study 1, we tested the extent to which children use social categories to predict a range of social interactions, including harmful behaviors, prosocial but non-obligatory behaviors, patterns of friendship, and more obligatory prosocial behaviors (saving someone from harm). If children's abstract beliefs about the social world center around identifying harmful situations, then they should only hold systematic expectations about harmful behaviors. If, however, children view social categories as marking people who hold a broader set of obligations toward one another, they should reliably predict that agents will direct harm towards members of other groups, be friends with someone from their own group, and save someone from their own group from harm. On both of these accounts, children should not have reliable expectations about positive, but less obligatory, prosocial interactions.

For this study, we intended to test children at the age at which their systematic expectations about these types of behaviors first emerge. Thus, we focused mainly on 3-year-olds (the earliest age at which predictions of intergroup harm have been documented; Rhodes, 2012). However, asking questions about saving others from harm caused us to use test items that were longer and more complex than those used in prior work, possibly introducing increased memory and processing demands. Thus, for questions about saving only, we tested both 3- and 4-year-old children. Note that in Rhodes (2012), children ages 3–5 al l predicted intergroup and intragroup prosocial interactions equally often (reliably intragroup predictions developed at age 6 in that work). Thus, all of the children tested here were quite a bit younger than those found to reliably predict intragroup prosocial interactions in prior work.

2.1. Methods

2.1.1. Participants

Participants included 100 3- to 4-year-olds (*M* age = 4;0, range = 3;5–4;11, 56 female), recruited at the Children's Museum of Manhattan. Participants were 39% White, 14% African American, 12% Asian, 8% Hispanic, 4% Middle Eastern, 6% Mixed, and 17% Unreported. An additional 32 children were tested but excluded from analysis because of distractions in the testing room ($n = 16$), parental interference ($n = 5$), experimenter error ($n = 1$), or because they did not speak English fluently and had trouble understanding the instructions ($n = 10$)[1]. Children were tested in a quiet room at the museum. The first 60 children (all 3-year-olds) were randomly assigned to either the harmful ($n = 20$), prosocial ($n = 20$), or friendship ($n = 20$) condition. An additional 40 children (20 3-year-olds and 20 4-year-olds) were assigned to the saving condition.

2.1.2. Procedure

The experimenter introduced participants to two novel groups of children—the "Flurps," wearing blue shirts, and the "Zazzes," wearing red shirts—displayed in hand-drawn pictures. The experimenter then told a story in which the Flurps and Zazzes were each working cooperatively to build towers out of blocks (full scripts can be found in the Online Supplementary Material). The groups were introduced in this way because young children more easily make group-based inferences when there is some functional relationship between group members (Rhodes, 2012). Then, in the harmful, prosocial, and friendship conditions, children were asked four test questions, and in the saving condition, children were asked six test questions. In the harmful and prosocial conditions, participants were asked to predict the recipient of an agent's action: In the harmful condition, these actions were negative, and in the prosocial condition, these actions were positive—we used items that had both physical (e.g., hitting/hugging) and psychological (e.g., social exclusion/inclusion) consequences, as has been common in previous studies on the development of morality and social cognition (Nucci & Turiel, 1978; Rutland & Killen, 2015; Smetana, 1985). In the friendship condition, children were introduced to an agent who was doing an activity (e.g., having a birthday party) with his friends, and they were asked to predict who the agent was friends with. In the saving condition, children were introduced to two individuals who were about to undergo a harmful event (e.g., slipping on a patch of ice) and an agent who was going to save one of them, and were asked to predict who the agent would save. For each question across all four conditions, the experimenter presented one picture representing the agent, and two pictures representing the possible answer choices, and asked participants to point to their answer. A full list of the behaviors used can be found in the Online Supplementary Material. We used a forced-choice paradigm to discourage children from stating that agents would direct positive actions toward everyone, and avoid negative actions toward everyone, as children at this age commonly express a "positivity bias," by which they are very willing to make positive inferences, and reluctant to make negative ones (Boseovski & Lee, 2006; Boseovski, 2010). By asking children to make a choice between two possible recipients of each action, we prevented children from responding based on this bias (rather than their abstract expectations of social categories), and also allowed children to acknowledge that often in their everyday experiences, equally positive outcomes are not always possible. Children were given a score of "1" every time they made an intragroup prediction, and "0" every time they made an intergroup prediction. The agent of the behaviors and the lateral position of the answer choices were counterbalanced across participants. All raw data and code can be found on the Open Science Framework at https://osf.io/4kmdv/.

2.2. Results

We used a binomial regression model to predict the number of times that children predicted intragroup interactions out of the total questions asked, using condition as a fixed factor. Descriptive statistics are given as probabilities of intragroup predictions, accompanied by Wald 95% Confidence Intervals. Children made the fewest intragroup predictions for harmful behaviors ($M = 0.35$, $CI = 0.24, 0.46$) and the most intragroup predictions for friendship ($M = 0.78$, $CI = .68, 0.87$). They made intragroup predictions about half the time for prosocial behaviors ($M = 0.51$, $CI = 0.40, 0.62$). Children's predictions in the saving condition varied by age; 4-year-olds predicted that characters would save members of their own group more often ($M = 0.71$, $CI = 0.63, 0.79$) than did 3-year-olds, who made intragroup predictions about half the time ($M = 0.50$, $CI = 0.41, 0.59$). The main effect of condition was reliable, $\chi^2(3) = 29.71, p < .001$ (see Fig. 1). An additional binomial regression performed on the saving condition alone, with age as a continuous predictor, revealed that the effect of age in this condition was also reliable, $\chi^2(1) = 12.28, p < .001$. Relative to the responding that would be expected by chance, the odds of an *inter*group prediction were 1.86 times as high in the harmful condition ($CI = 1.18, 2.98$). Also relative chance responding, the odds of an *intra*group prediction were 3.44 times as high in the friendship condition ($CI = 2.08, 5.99$) and 2.43 times as high in the saving condition among 4-year-olds ($CI = 1.65, 3.64$). We confirmed this

[1] Because this research took place at a children's museum, occasional distractions from the environment led to the need to exclude more participants than is often necessary with this age group. To make these decisions, a research assistant coded videotapes of all testing sessions for children's level of attention, distractions in the testing room, experimenter errors, and parental interference. Exclusion decisions were made based on these codes prior to data analysis. To confirm that these exclusions did not systematically alter our findings, we reran analyses with all children included. These analyses revealed patterns identical to those found in our main analyses.

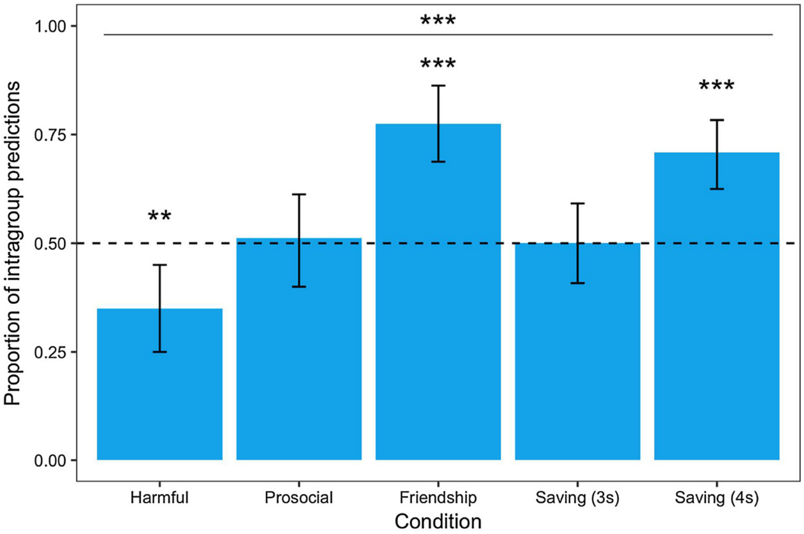

Fig. 1. Proportion of intragroup predictions for each condition. Error bars represent 95% confidence intervals.

pattern by comparing each group to chance—children reliably predicted intergroup interactions in the harmful condition ($\chi^2(1) = 6.97, p < .01$), showed at-chance responding in the prosocial condition ($\chi^2(1) = .05, p = .82$) and in the saving condition among 3-year-olds ($\chi^2(1) = 0.00, p = 1.00$), and reliably predicted intragroup interactions in the friendship condition ($\chi^2(1) = 21.34, p < .001$) and in the saving condition among 4-year-olds ($\chi^2(1) = 19.52, p < .001$).

2.3. Discussion

The results of Study 1 are consistent with the possibility that children’s behavioral predictions are motivated by the belief that social categories mark people who are obligated to one another. 3-year-old children expected individuals to harm members of other social categories rather than members of their own, but did not hold strong expectations about whom individuals would direct prosocial actions toward. Both of these effects are replications of prior work showing that young children use category memberships to guide their predictions of harmful behaviors, but not prosocial behaviors (Rhodes, 2012).

We have also extended prior work in two ways. First, children expected individuals to be friends with fellow social group members, rather than with members of other groups. Furthermore, by age 4, children expected that individuals would protect fellow group members, rather than members of other groups, from harm. Both of these findings provide evidence against the possibility that children's early theories of the social world focus only on intergroup harm, and instead suggest that these theories center around a broader set of obligations that group members hold toward one another. To our knowledge, these findings represent the first clear pieces of evidence that children younger than age 6 use their abstract beliefs about social groups to generate predictions of behaviors occurring among fellow group members.

One limitation of this study is that it used forced-choice measures, which may not exhaustively mirror the types of scenarios that children encounter on a regular basis (e.g., individuals do not need to choose between two possible recipients for every action that they perform). However, we do believe that children understood the scenarios used here, and reasoned about them in a realistic way; children are aware that we all have limited time and resources, and cannot always help or protect everyone. Thus, despite the limitations of using a forced-choice paradigm, our findings support the proposal that children’s framework theory of social obligation goes beyond generating beliefs about harmful actions, and includes expectations of how group members might protect and affiliate with one another.

3. Study 2

If children’s framework theory of the social world is really about viewing category members as obligated to one another, then it should not only shape children’s predictions of future behavior; it should also guide the way in which children learn from the social behaviors that they see unfold in the environment.

Children elaborate upon their abstract understandings of the world by interpreting and responding to input (Gopnik & Wellman, 2012). Indeed, preschool-age children revise their abstract knowledge—often in incremental ways—in response to evidence across a range of domains and experimental paradigms (Bonawitz, van Schijndel, Friel, & Schulz, 2012; Gopnik et al., 2004; Kushnir & Gopnik, 2007; Rhodes & Wellman, 2013; Rhodes & Wellman, 2017; Seiver, Gopnik, & Goodman, 2013; Xu, 2007). From this perspective, children are more likely to learn from evidence that is consistent with their previous expectations (Gopnik & Wellman, 2012). For example, Schulz, Bonawitz, and Griffiths (2007) found that younger preschool-age children had more difficulty learning a new causal relation that was inconsistent with their current biological theories than one that was more consistent (e.g., three-year-olds had a more difficult time learning that worrying could cause a stomachache than learning that eating something could do so, even when shown equivalent patterns of evidence). Also, Kushnir and Gopnik (2007) found that preschool-age children had an easier time learning the causal structure of new toys when those tops operated in a manner that was consistent with their current physical theories (e.g., they had an easier time learning that a toy would turn on when something was placed on top of it than simply waved over it, because the former but not latter process is consistent with their beliefs about the importance of spatial contiguity).

In the social domain, then, children's framework theory should shape how they learn about patterns of social behavior. On this account, children should discount evidence of behaviors that are not relevant to their theory that group members are obligated toward one another—intragroup harm and intergroup prosociality. Because these behaviors are inconsistent with children's abstract expectations of how social categories constrain behavior (i.e., they do not align with the idea that group members protect and affiliate one another), children should not interpret these instances as opportunities to learn about the social world. In contrast, children should attend to behaviors consistent with their theory—which include both those involving intergroup harm and intragroup prosociality—and learn particularly from these behaviors.

In Study 2, we tested this possibility by investigating how children's predictions are shaped through exposure to evidence of various types of social behaviors. To accomplish this task, we introduced the two social categories in an initial state of positive group relations, then showed children instances of either harmful or prosocial behaviors that occurred in either intergroup or intragroup contexts, and asked them to predict whether future similar behaviors would occur among members of the same group or between members of different groups. If children use their framework theory of the social world to learn patterns of social behavior, then they should show systematic patterns of responding—by which their predictions should match the evidence that they saw—following evidence of intergroup harm and intragroup prosociality, but not following evidence of intragroup harm and intergroup prosociality.

3.1. Methods

3.1.1. Participants

Participants included 96 4-year-olds (M age = 4;6, range = 4;0–4;11, 46 female), recruited at the Children's Museum of Manhattan in the same manner as in Study 1. Participants were 43% White, 7% African American, 8% Asian, 11% Hispanic, 1% Middle Eastern, 15% Mixed, and 15% Unreported. An additional 14 children were tested but excluded from analysis because of distractions in the testing room ($n = 4$), parental interference ($n = 2$), developmental disabilities ($n = 2$), experimenter error ($n = 3$), or because they did not speak English fluently and had trouble understanding the instructions ($n = 3$). Children were randomly assigned to the prosocial-intergroup ($n = 24$), prosocial-intragroup ($n = 24$), harmful-intergroup ($n = 24$), or harmful-intragroup ($n = 24$) condition.

3.1.2. Procedure

As in Study 1, the experimenter first introduced participants to two novel groups. In Study 2, however, the groups were not given novel labels and were described as participating in cooperative activities with each other (e.g., "Some of them are in the blue group, and some of them are in the red group, but all the kids like to play together"; full script can be found in the Online Supplementary Material), in order to reduce any baseline expectations that children held about the likelihood of inter- and intra-group behaviors (Chalik & Rhodes, 2014).

Next, children were shown individual instances of prior behaviors, varying by condition. Each behavior was displayed in a hand-drawn picture and described verbally by the experimenter. The behaviors were either prosocial or harmful, and occurred in either an intergroup or intragroup context; these factors were crossed to create four conditions (prosocial-intergroup, prosocial-intragroup, harmful-intergroup, harmful-intragroup). In addition, within each condition, half of the children saw only one instance of a past behavior, and half of the them saw five instances. However, preliminary analyses showed no effects of this variable, so it is not discussed further.

Next, children answered six test questions with the same structure as the harmful and prosocial conditions from Study 1. As in Study 1, the agent of the behaviors and the lateral position of the answer choices were counterbalanced across participants. Children were given a score of "1" every time they made a prediction that matched the prior evidence they had seen, and were given a score of "0" every time they made a prediction that did not match (i.e., in the intragroup conditions, intragroup predictions were scored as "1," and intergroup prediction were scored as "0," but in the intergroup conditions, the reverse was true).

3.2. Results

We used binomial regression models to analyze the likelihood that children expected future behaviors to match the prior evidence they had seen, using behavior (harmful vs. prosocial) and context (intergroup vs. intragroup) as fixed factors, and testing for both possible main effects and an interaction. Descriptive statistics are given as probabilities of predictions matching prior evidence, accompanied by Wald 95% Confidence Intervals. As shown in Fig. 2, we found main effects of behavior, $\chi^2(1) = 4.58, p < .05$, and context, $\chi^2(1) = 3.64\ p = .056$, that were subsumed under an interaction between behavior and context, $\chi^2(1) = 6.80, p < .01$. For harm, children's predictions were more consistent with the evidence they observed if they were shown intergroup harm than intragroup harm, $\chi^2(1) = 3.64, p = .056$; relative to chance responding, evidence of intergroup harm raised the odds of an evidence-consistent prediction by an order of 1.77 (CI = 1.26, 2.50). For prosocial behaviors, children's predictions were more consistent with the evidence they observed if they were shown intragroup prosociality than intergroup prosociality, $\chi^2(1) = 3.17, p = .075$; evidence of intragroup prosociality raised the odds of an evidence-consistent prediction by an order of 1.62 (CI = 1.16, 2.28). When children

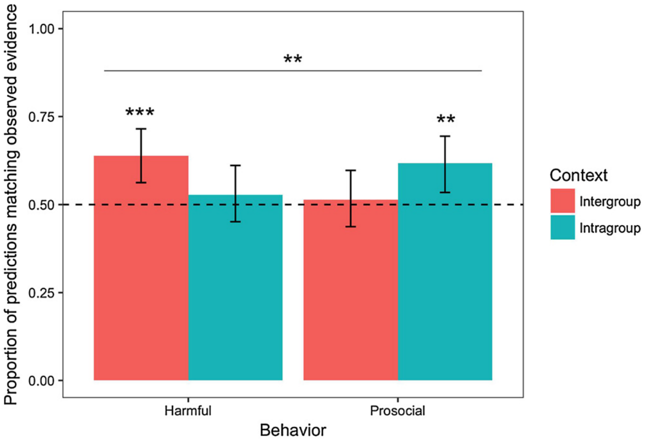

Fig. 2. Proportion of predictions matching prior evidence for each condition. Error bars represent 95% confidence intervals.

were shown evidence of intergroup harm, they reliably predicted subsequent intergroup harm ($M = 0.64$, $CI = 0.56, 0.72$), $\chi^2(1) = 10.82$, $p = .001$, but when children were shown evidence of intragroup harm, they predicted intergroup and intragroup harm equally often ($M = 0.53$, $CI = 0.45, 0.61$), $p = .51$. Similarly, when children were shown evidence of intragroup prosociailty, they reliably predicted future intragroup prosociality ($M = 0.62$, $CI = 0.54, 0.70$), $\chi^2(1) = 7.87$, $p = .005$, but when they were shown evidence of intergroup prosociality, they predicted future intergroup and intragroup prosociality equally often ($M = 0.51$, $CI = 0.43, 0.60$), $p = .74$.

3.3. Discussion

After seeing evidence of intragroup helping and intergroup harm, 4-year-olds predicted that future behaviors would occur in similar group contexts. Yet, after seeing evidence—even five pieces of evidence—of intergroup helping and intragroup harm, children did not hold reliable expectations about future behaviors. Thus, it appears that children interpret evidence of social behaviors in light of their theory of the social world: They attend to behaviors that concern ways in which social category members might relate to one another, learn the patterns by which those behaviors occur in the environment, construe (based on those patterns) the behaviors as obligated among fellow social category members, and predict future interactions in light of those obligations. However, when a behavior is inconsistent with this theory, they do not use it to make predictions.

4. Study 2B

One limitation of Study 2 is that it did not include a baseline condition to test children's predictions in the absence of any evidence. Thus, there are two possible explanations for children's reliable expectations of intergroup harm and intragroup prosociality: (a) that children learned to make these predictions from the evidence that we provided, or (b) that these reflect children's baseline expectations. Study 1 and Rhodes (2012) indicate that the second explanation is unlikely to fully account for our pattern of findings (because in those case children of these ages did *not* reliably predict intragroup prosociality, suggesting that the findings of Study 2 reflect a process of learning from evidence). To address this possibility more directly, however, we conducted Study 2B, in which we assess children's baseline expectations about prosocial and harmful behaviors following the identical introduction to the group context that they heard in Study 2.

4.1. Methods

4.1.1. Participants

Participants included 48 4-year-olds (M age = 4;8, range = 4;0–5;3, 25 female), recruited at preschools around New York City. Children were recruited through parent consent forms that were sent home with children and then returned, and were tested in a quiet area at their school. Participants were 60% White, 2% African American, 6% Asian, 23% Mixed, and 9% Unreported. Children were randomly assigned to the prosocial ($n = 24$) or harmful ($n = 24$) condition.

4.1.2. Procedure

The procedure for Study 2B was exactly the same as that for Study 2, with the exception that children were not shown any instances of prior behaviors. Instead, after being introduced to the two groups following the same script as in Study 2, they moved immediately on to the test questions. These questions were identical to those asked in Study 2. Responses were coded based on whether children's predictions were or were not theory-consistent (we assumed that if children made systematic predictions at all, these predictions were most likely to be consistent with their underlying intuitive theory): Children were given a score of "1″ for intergroup harmful or intragroup prosocial predictions, and were given a score of “0” for intragroup harmful or intergroup prosocial predictions.

4.2. Results and discussion

We used binomial regression models to analyze the likelihood that children expected future behaviors to be theory-consistent, using intercept-only models to compare responses for each behavior type to chance. For harmful behaviors, children reliably predicted that actions would occur between members of different groups ($M = 0.65$, $CI = 0.57, 0.73$), $\chi^2(1) = 13.01$, $p < .001$. In contrast, for prosocial behaviors, children did not differ from chance in their predictions ($M = 0.56$, $CI = 0.48, 0.64$), $p = .16$. Thus, children’s above-chance responses in the harmful-intergroup condition of Study 2 can be interpreted as a reflection of their prior expectations that harm will occur between members of different groups, rather than new learning in response to the evidence that they saw. Children’s responses in the prosocial-intragroup condition of Study 2, however, *can* be interpreted as reflecting their learning from the evidence that they had seen, as they did not reliably hold these expectations in the absence of any evidence.

5. General discussion

In the present studies, preschool-age children used a framework theory of the social world, by which they view social categories as marking individuals who are obligated toward one another, to inform their predictions of social interaction and to guide new learning in the social domain. When presented with two novel social categories, 3-year-old children predicted that individuals would harm members of another group and be friends with members of their own group. Furthermore, 4-year-old children predicted that individuals would save their fellow group members, rather than members of another group, from harm. Because these social categories were novel, children could not have based their expectations on any preexisting biases or knowledge of specific group histories—rather, to make these predictions, children could only rely on their abstract knowledge of the structure of social groups. Finally, for theory-consistent behaviors about which they did not hold strong prior expectations (specifically, intragroup prosociality), 4-year-olds used observed patterns of evidence to guide their future predictions.

Framework theories are a powerful cognitive tool, allowing children to understand and predict the events that occur around them (Gopnik & Wellman, 2012; Wellman & Gelman, 1992). During the first few years of life, children build and expand upon framework theories across the physical (Smith, Carey, & Wiser, 1985), biological (Keil, 1989), psychological (Woodward, 1998), and social (Hirschfeld, 1996; Rhodes & Chalik, 2013; Rhodes, 2012) domains. The present work is the first to document one of the most basic functions that children’s framework theory in the social domain performs—it supports a specific set of predictions about how social category members will relate to and act toward one another. Specifically, we have shown that children use their abstract expectations to predict that social group members will be friends with one another and protect each other from harm.

Interestingly, preschool-age children do not predict that people will perform generally positive actions for fellow social category members over members of other groups if those actions are not related to social obligations (see also Rhodes, 2012). This finding illustrates a very important feature of children’s framework theory of the social world: It centers around an obligation to protect fellow group members from harm. Moral philosophers have long held that although prosocial behaviors are valuable, they are not morally *imperative* (Leslie et al., 2006). This belief shapes various social-cognitive phenomena; for example, in most Western legal systems, there generally exist strict prohibitions against doing harm, yet there are no obligations to do good (Leslie et al., 2006). Thus, the belief that people’s obligations to one another are centered around avoiding harm is central to the human moral judgment system.

Children do eventually begin to predict that people will do nice behaviors preferentially for fellow group members, but these expectations have not been found in samples of children younger than age 6 (Rhodes, 2012). We have shown that one possible mechanism by which this change may take place is children’s interpretation of and learning from the evidence around them. Across domains, children use their framework theories to observe evidence, interpret that evidence, then update their theory in light of that evidence (Gopnik & Wellman, 2012; Rhodes & Wellman, 2017). We have now documented this process in the social domain: Children observe the social behaviors that occur around them, interpret those behaviors based on their underlying theory of the social world, and use that interpretation to guide their predictions of future similar behaviors.

Thus, children’s understanding of social behavior is shaped by a belief that social category members are obligated toward one another. This theory not only guides children’s predictions of social behaviors, but also shapes the way in which children interpret and learn from the events that they observe. In future work, it will be important to investigate the relationship between the abstract social category-based expectations we have discussed here and the generalized affective processes that guide children’s own behavior in group settings (e.g., Renno & Shutts, 2015), and to investigate how children’s expectations change to support predictions of more diverse and complex social behaviors as children grow older. It will also be important to explore how children’s expectations of novel social categories, as shown in the present work, inform their experience with real-world social groups.

Acknowledgments

We are very grateful to the families and staff at the Children’s Museum of Manhattan. We are also grateful to Nicole Graziano, Emily Foster Hanson, Peter Liebenson, Elyana Feldman, Maria Cotes-Robles, Emmy Green, and Kat Yee for their assistance with data collection. Funding was provided by NSFBCS-1226942 and NSF BCS-1147543.

Supplementary material.

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.cogdev.2018.06.010.

References

Baltazar, N. C., Shutts, K., & Kinzler, K. D. (2012). Children show heightened memory for threatening social actions. *Journal of Experimental Child Psychology, 112*, 102–110. https://doi.org/10.1016/j.jecp.2011.11.003.

Bonawitz, E. B., van Schijndel, T., Friel, D., & Schulz, L. (2012). Balancing theories and evidence in children's exploration, explanations, and learning. *Cognitive Psychology, 64*, 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002.

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4*, 212–218. https://doi.org/10.1111/j.1750-8606.2010.00149.x.

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology, 42*, 500–513. https://doi.org/10.1037/0012-1649.42.3.500.

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.

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.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

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

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3–32. https://doi.org/10.1037/0033-295X.111.1.3.

Hirschfeld, L. A. (1996). *Race in the making.* Cambridge, MA: MIT Press.

Keil, F. C. (1989). *Concepts, kinds, and conceptual development.* Cambridge, MA: MIT Press.

Kinzler, K. D., & Shutts, K. (2008). Memory for "mean" over "nice": The influence of threat on children's face memory. *Cognition, 107*, 775–783. https://doi.org/10.1016/j.cognition.2007.09.005.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190–194. https://doi.org/10.1111/1467-8284.00419.

Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental Psychology, 43*, 186–196. https://doi.org/10.1037/0012-1649.43.1.186.

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

Levine, M., & Thompson, K. (2004). Identity, place, and bystander intervention: Social categories and helping after natural disasters. *The Journal of Social Psychology, 144*, 229–245. https://doi.org/10.3200/SOCP.144.3.229-245.

Levine, M., Cassidy, C., Brazier, G., & Reicher, S. (2002). Self-categorization and bystander non-intervention: Two experimental studies. *Journal of Applied Social Psychology, 32*, 1452–1463. https://doi.org/10.1111/j.1559-1816.2002.tb01446.x.

Nucci, L. P., & 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.

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). Naive 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.

Rhodes, M., & Wellman, H. (2013). Constructing a new theory from old ideas and new evidence. *Cognitive Science, 37*, 592–604. https://doi.org/10.1111/cogs.12031.

Rhodes, M., & Wellman, H. (2017). Moral learning as intuitive theory revision. *Cognition, 167*, 191–200. https://doi.org/10.1016/j.cognition.2016.08.013.

Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. (2015). Infants' use of social partnerships to predict behavior. *Developmental Science, 18*, 909–916. https://doi.org/10.1111/desc.12267.

Rutland, A., & Killen, M. (2015). A developmental science approach to reducing prejudice and social exclusion: Intergroup processes, social-cognitive development, and moral reasoning. *Social Issues and Policy Review, 9*, 121–154. https://doi.org/10.1111/sipr.12012.

Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43*, 1124–1139. https://doi.org/10.1037/0012-1649.43.5.1124.

Seiver, E., Gopnik, A., & Goodman, N. (2013). Did she jump because she was the big sister or because the trampoline was safe? Causal inference and the development of social attribution. *Child Development, 84*, 443–454. https://doi.org/10.1111/j.1467-8624.2012.01865.x.

Shutts, K., Pemberton 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.

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.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of size, weight, and density. *Cognition, 21*, 177–237. https://doi.org/10.1016/0010-0277(85)90025-3.

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.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4.

Xu, F. (2007). Rational statistical inference and cognitive development. In P. Carruthers, S. Laurence, & S. Stich (Vol. Eds.), *The innate mind: Foundations and the future: Volume 3*, (pp. 199–215). New York: Oxford University Press.