COGS	12129	WILEY	Dispatch: 26.3.14	CE: Ashokkumar
Journal Code	Manuscript No.		No. of pages: 11	PE: Suganya

Cognitive Science (2014) 1–11

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12129

Children's Explanations as a Window into Their Intuitive Theories of the Social World

Marjorie Rhodes

Department of Psychology, New York University

Received 17 October 2012; received in revised form 9 June 2013; accepted 24 September 2013

Abstract

Social categorization is an early emerging and robust component of social cognition, yet the role that social categories play in children's understanding of the social world has remained unclear. The present studies examined children's (N = 52 four- and five-year olds) explanations of social behavior to provide a window into their intuitive theories of how social categories constrain human action. Children systematically referenced category memberships and social relationships as causal-explanatory factors for specific types of social interactions: harm among members of different categories more than harm among members of the same category. In contrast, they systematically referred to agents' mental states to explain the reverse patterns of behaviors: harm among members of the same category more than harm among members of different categories. These data suggest that children view social category memberships as playing a causal-explanatory role in constraining social interactions.

Keywords: Naïve sociology; Social categorization; Intuitive theories; Explanation

1. Introduction

Classifying people into categories (e.g., girls, French-speakers) is a crucial way of organizing social experiences and is an early-developing (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kinzler, Dupoux, & Spelke, 2007; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and robust (Bigler & Liben, 2007; Dunham, Baron, & Carey, 2011; Patterson & Bigler, 2006) component of social cognition. One key function of social categories is to mark individuals who are understood as similar in non-obvious ways (Hirschfeld, 1996). For example, by preschool, children assume that the members of certain social categories share a wide range of physical, psychological, and behavioral characteristics and thus use

Correspondence should be sent to Dr. Marjorie Rhodes, Department of Psychology, New York University, 6 Washington Place, rm. 301, New York, NY 10003. E-mail: marjorie.rhodes@nyu.edu

such categories to make inferences about individual attributes (e.g., that if one girl likes a particular new game, other girls will like it too; Diesendruck & haLevi, 2006; Gelman, Collman, & Maccoby, 1986; Waxman, 2010).

Children's use of social categories to predict individual attributes reflects an abstract intuitive theory that members of certain social categories are highly similar to each other, not only their previous experiences with particular categories or properties (e.g., prior knowledge that girls tend to play similar games). For example, children will use familiar categories to infer similarities even for novel properties about which they have no prior knowledge (Gelman et al., 1986). They will also infer similarity across category members for completely novel categories, provided that they are led to view the novel categories as informative and coherent (Rhodes, Leslie, & Tworek, 2012). Further, children use some social categories not only to predict individual properties, but also to explain why people have certain characteristics. For example, by ages 4–5, children explain why specific individuals have gender-typical properties by directly referencing category memberships (e.g., they will respond that a girl raised entirely with boys will prefer ballet to football simply, "because she is a girl"; Taylor, Rhodes, & Gelman, 2009). Thus, an intuitive theory that social categories mark fundamental similarities supports a powerful mechanism that young children can use to predict and explain human behavior.

Yet, there is also reason to suspect that beliefs about similarity do not capture the full functional role that social categories play in early social cognition. Beliefs about the similarities among category members do not provide a framework for understanding some of the uniquely *social* functions that categories of people might serve. For example, in everyday life, social categories shape social relationships and social interactions—who will be friends versus enemies, cooperate versus compete, or help versus harm each other—and predicting such interactions appears to be another early-developing inferential role of social categories (Chalik & Rhodes, in press; Rhodes, 2012; Rhodes & Chalik, 2013; Shutts, Pemberton Roben, & Spelke, 2013). For example, children ages 3–4 expect members of familiar social categories (e.g., gender, race) to be friends with each other rather than members of other groups (Shutts et al., 2013). Children of these ages also use novel social categories to predict specific patterns of social interactions—that people will harm members of other categories instead of members of their own (Rhodes, 2012).

Using social categories to predict these phenomena does not involve beliefs about the similarities across category members. In Rhodes (2012), children were introduced to novel categories, referred to as "Flurps" and "Zazes," and then reliably predicted that a Flurp would steal a cookie from a Zaz, instead of from another Flurp, for example. On these items, children were not asked to make inferences about shared personality traits or behavioral tendencies; in other words, they were *not* asked to predict who might steal a cookie, given that one Flurp did so. Instead, children were asked to predict from whom a Flurp would steal a cookie, given his category membership. In this manner, children did not use categories to predict individual attributes (e.g., who is a cookie-stealer), but rather how people relate to one another (e.g., who a person will steal a cookie from).

Building on this work, Rhodes and Chalik (2013) proposed that children hold a second intuitive theory of social categories—that social categories mark people who hold intrinsic

interpersonal obligations to one another. On this account, children have systematic, abstract expectations that people are intrinsically obligated to support and protect, and not to harm, members of their own groups. Thus, the inferences described earlier (Rhodes, 2012; Shutts et al., 2013) reflect expectations that people will behave in line with these obligations. In support of this account, children ages 4–10 evaluated instances of harm among members of the same group as violating intrinsic obligations, but instances of harm involving members of different groups as violating extrinsically defined social rules.

An intuitive theory that social categories mark patterns of obligations and social relationships would provide another powerful mechanism by which children could use social categories to understand and predict human action. Yet, prior work provides only indirect evidence that children view category memberships as causal factors in shaping social behavior because no measures of children's beliefs about causal mechanisms were examined. More direct evidence that children view social categories as causal-explanatory mechanisms in determining social behaviors can be found by examining children's explanations for why certain social interactions occur (Amsterlaw & Wellman, 2006; Hickling & Wellman, 2001; Legare, Wellman, & Gelman, 2009; Miller, 1984, 1986; Miller & Aloise, 1989).

If children view categories as constraining patterns of social relationships and social obligations, then category memberships should shape how children explain certain social behaviors. Given that children view people as obligated not to harm members of their own groups, and as more likely to harm and less likely to be friends with members of other groups (Chalik & Rhodes, in press; Rhodes, 2012; Rhodes & Chalik, 2013), children might explain harmful social interactions that occur between members of different groups by referencing the actors' relational status—that they are in different categories (e.g., "She took her cookie because they are in different groups") or have an antagonistic relationship (e.g., "She took it because they are enemies"). In contrast, harmful social behaviors that occur among members of the same group are not explainable by an expected relational status, and thus may be taken as more indicative of the agent's individual mental states (e.g., "she took it because she was hungry" or "she took it because she is a mean girl"). In this way, instances of harm among members of different groups are seen as caused by social relationships, whereas instances of harm among members of the same group would be taken to reflect the (negative) qualities of the agent. Evidence of such systematic tendencies to explain social interactions differently depending on the category memberships of the agent and recipient would provide direct evidence that preschool-age children view social categories as playing a causal-explanatory role in shaping these behaviors.

2. Method

2.1. Participants

Participants included 52 four- and five-year-old children (24 male; M_{age} = 5.0 years, range = 4.0–5.9 years; approximately 33% European American, 6% African American,

6% Asian American, 21% Hispanic, 13% Multiethnic, 21% unreported) recruited from and tested at the Children's Museum of Manhattan.

2.2. Procedure

First, children were introduced to the relevant social categories. The role of social categories in early social cognition sometimes varies across different types of categories (Kinzler, Shutts, & Correll, 2010; Rhodes & Gelman, 2009). Thus, to provide a comprehensive examination of how social categories shape children's explanations of social behavior, children were introduced to one of two ways of defining the relevant categories—either completely novel social categories or categories based on a familiar distinction.

Novel categories were used to provide a test of children's abstract beliefs about how categories shape behavior, in the absence of prior knowledge about the specific groups or their relationships. The novel groups were introduced via a brief illustrated story. Children were shown four individual children from each of two novel categories, with membership marked by shirt color and novel labels. Children were told, "Here is the red group. They are called the Flurps. Here is the blue group. They are called the Zazes." Subsequently, children were told a brief story designed to lead children to view the novel categories as meaningful. To provide a clear test of how social categories shape children's explanations of social behaviors, it was necessary to ensure that they viewed the categories as important in the actors' environments. Thus, the story included elements of both within-group cooperation (e.g., each group was described as engaging in within-group cooperation to build a tall block tower) and between-group competition (e.g., each group was described as wanting to build the tallest tower). No negative interactions—or between-group interactions of any kind—were described in the stories. Full text of the stories and illustrations are available in Rhodes (2012; Study 1). After the story, children were asked to, "point to the Zazes" and to "point to the Flurps." If children responded incorrectly, the story was repeated and the questions were re-asked. No child failed these questions a second time.

A second condition introduced children to categories based on language differences. Language-based groups provide a useful comparison with the novel groups condition described above because children view language as marking fundamental (Kinzler & Dautel, 2012), and perhaps even biological (Hirschfeld & Gelman, 1997), differences among people. Thus, including both novel groups and groups that children view as more natural spans the various types of social categories that are often studied in research on early social cognition. Further, the groups in the novel groups condition were described as engage in within-group cooperation and between-group competition; these elements were not included in the language-based groups condition. Thus, finding similar patterns across these different ways of defining the categories would suggest that these data reflect robust processes that shape early social cognition across multiple contexts.

In the language-based groups condition, children were shown the same illustrations as in the novel groups condition, and the groups were labeled in the same manner ("Here is

the red group. They are called the Zazes. Here is the Blue group. They are called the Flurps.”). Subsequently, instead of continuing with the story used in the novel groups condition, the experimenter continued, “I want to show you what they sound like. Let’s listen to some things they say.” The experimenter pointed individually to two members of each group and for each played an audio recording of a child speaking (the same files used by Kinzler, Shutts, Dejesus, & Spelke, 2009). The members of one of the groups spoke French, whereas the members of the other group spoke English. The content of the audio recording was neutral and identical across languages (e.g., “hide and seek is a fun game to play”).

In both conditions, after children were introduced to the categories, they were shown a series of 12 illustrations depicting social interactions involving an agent and recipient. For each, children were asked to explain the agent’s behavior; for example, “1 day, a Flurp stole a cookie from another Flurp. Why did the Flurp steal a cookie from the other Flurp?” The 12 items included six questions about harmful social interactions and six questions about helpful interactions. The primary questions of interest were the six items about harmful social interactions, as preschool-age children use social categories to predict (Rhodes, 2012) and evaluate (Rhodes & Chalik, 2013) these types of behaviors. Children appear not to use social categories to predict helpful behaviors (e.g., that people are more likely to help members of their own groups) until middle childhood (Rhodes, 2012); however, children sometimes display knowledge in their explanations before they make use of that knowledge to make predictions (Amsterlaw & Wellman, 2006; Legare et al., 2009). Thus, questions about helpful behaviors were included to test whether social categories shape children’s understanding of a broader range of behaviors than suggested by prior work. In the Language condition, whether the agent in the test items was from the French or English-speaking group was counterbalanced across participants; there were no main or interactive effects of this factor (ps > .30), so it was not considered further.

The 12 test items were composed of four types, which followed a 2 (Group composition: within-group, between-group) × 2 (Behavior: Harmful, Helpful) factorial design, with three questions of each type. Helpful behaviors included sharing, hugging, and playing with someone; harmful behaviors included stealing, hitting, and refusing to let someone play. So that children did not receive descriptions of identical events in the within-group and between-group contexts, two versions of each item were created (e.g., sharing a cookie, sharing a toy). Assignment of the specific item to within-group versus between-group scenarios was counterbalanced across participants. There were no main or interactive effects of this counterbalancing factor, ps > .20. Items were presented in one of two random orders.

Participants’ responses were videotaped and transcribed verbatim. As would be expected given the open-ended nature of the test questions, participants generated a wide range of specific causal explanations for these items. Yet, broadly, the majority of explanations could be classified as one of three types, referencing: (a) the agent’s mental states alone, without consideration of the recipient (e.g., “he took the cookie because he wanted it”; “He gave her the cookie because he didn’t want it”), (b) the relation between the agent and recipient (e.g., “he took his cookie because they are enemies”; “He shared the cookie because they are friends”), or (c) the category memberships of the agent and/or

recipient (e.g., "he took it because they are in different groups"; "they shared the cookie because they are in the same group"). For further examples, see Table 1. Two independent raters coded each explanation (Cohen's kappa = .72), with differences resolved by discussion. Together, these three codes accounted for 80% of explanations. The majority of other responses were not possible to code (e.g., "just because") or failures to respond (e.g., "I don't know"). The frequency of these other responses did not vary by item type and were excluded from further analyses. The percentages of times that each type of coded explanation was generated (out of the total coded explanations) for each of the four question types (within-group harm, within-group helping, between-group harm, between-group helping) were calculated. These percentages were analyzed via a 2 (Group definition: novel, language-based) × 2 (Group Composition: within-group, between-group) × 2 (Behavior: help, harm) × 3 (Explanation type: agent, relationships, category) analysis of variance, with group definition as a between-subjects variable and all other factors as within-subjects variables.

3. Results

Analyses revealed a reliable interaction among Group Composition, Behavior, and Type of Explanation, $F(2, 100) = 8.03$, $p = .001$, partial $\eta = .14$. As shown in Fig. 1, for

Table 1
Description of the coding scheme

Code	Description	Examples
Agent	References the agent's mental states alone, including thoughts, emotions, motivations, traits, but makes no mention of the recipient	"He wanted the toy." "He didn't want to play with that" "He was hungry." "He wanted to." "He's angry"
Relationship	References the relation between the agent and the recipient, the way the agent feels toward the recipient, or a desire to act in a particular way toward the recipient	"They are friends." "They are enemies." "They are not friends." "He likes him." "He wants to be mean to him." "They like each other." "She's nice to her"
Category	References the categories or group labels	"Because they are in different groups." "They are both Flurps." "Because he is a Zaz." "They are in the group together"

Note. For each type of explanation, the valence of the explanation varied by whether children were explaining harmful or helpful interactions. For example, for agent-based explanations, children could refer to wanting a cookie (e.g., "she took the cookie because she wanted it") to explain a harmful behavior, but not wanting it to explain a helpful behavior (e.g., "she gave it to her because she didn't want it"). Similarly, for relationship-based explanations, children could refer to positive relationships (e.g., "because they are friends") to explain helpful behaviors but negative relationships (e.g., "because they are enemies") to explain harmful behaviors. For category-based explanations, they could refer to shared or different category memberships. Despite these differences in valence, these broad codes were used because in each case children are referring to the same class of mechanism (the agent's mental state, the relation between the agent and recipient, or the relevant categories) to explain the behaviors and doing so enabled direct comparison of how each type of mechanism was used to explain the different types of behaviors.

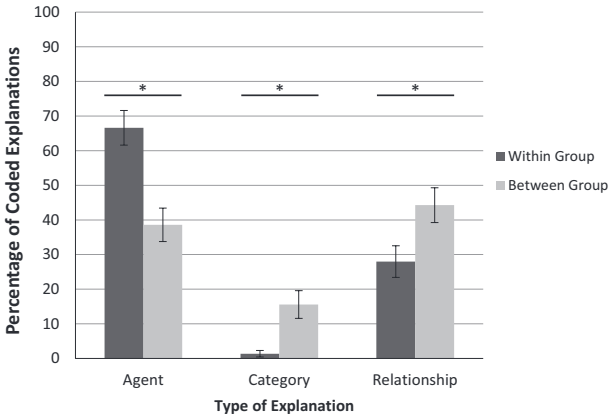

Fig. 1. Percentages of explanations fitting each code for harmful social interactions. Error bars represent ±1 standard error of the mean.

harmful behaviors, children generated category explanations (p = .001; all p-values are adjusted for multiple comparisons with Bonferroni corrections) and relationship explanations (p = .006) more for between-group harm than within-group harm, but agent-based explanations more for within-group harm than between-group harm (p < .001). For helpful interactions, children generated category-based explanations more for within-group helping than between-group helping, p = .04; no other explanation type varied by group composition (see Fig. 2). There were no main or interactive effects of whether groups were entirely novel or defined by language differences, ps > .30, and the three-way interaction among Group Composition, Behavior, and Type of Explanation was reliable in each condition examined separately—novel groups: $F(2, 62) = 5.54$, p = .006, partial η = .15; language-based groups: $F(2, 38) = 4.08$, p = .03, partial η = .18.

It is also useful to consider how the distribution of explanations varied for each type of interaction. For within-group harm, children generated agent-based explanations more than relationship explanations, p < .001, whereas for between-group harm, they generated relationship explanations and agent-based explanations equivalently often (in both contexts, category explanations were given least often, ps < .01). For helpful behaviors, children generated relationship explanations more than either other type, for both within-group and between-group interactions, ps < .001.

4. Discussion

In this study, social categories shaped how children explained harmful interpersonal interactions. Children more often referred to categories and social relationships to explain harm among members of different groups than among members of the same group, but to the agent's mental states (without reference to the recipient) to explain harm among members of the same group more than among members of different groups. These data

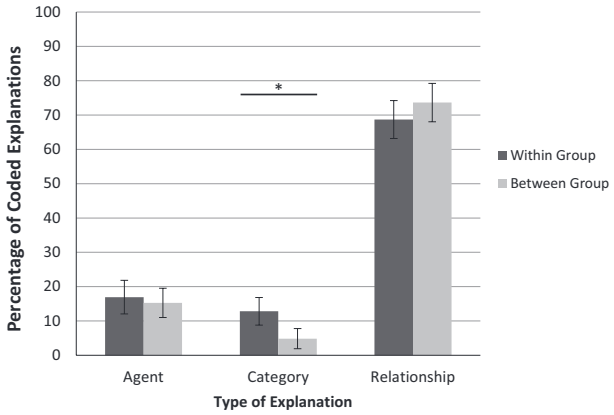

Fig. 2. Percentages of explanations fitting each code for helpful social interactions. Error bars represent ±1 standard error of the mean.

go beyond previous work (Rhodes, 2012; Shutts et al., 2013) by providing direct evidence that children view social categories as causal-explanatory factors that shape social interactions. Thus, these data provide a direct window into children's theories of the social world.

Children's explanations followed similar patterns regardless of whether the groups were entirely novel or correlated with a meaningful, familiar characteristic (language differences). Thus, viewing social categories as causally constraining harmful social interactions appears to be a robust feature of how social categories shape social cognition in early childhood. The categories presented here were highlighted by both labels and perceptual markers and, in the novel groups condition, by describing within-group cooperation and between-group competition. The language condition did not include information about cooperation or competition; however, the groups were still marked by a perceptual feature and novel label (as well as by language differences). Because of this approach, it is possible that these categories were made more salient than children might typically find social categories; thus, an important open question is how these effects generalize to categories that children might encounter in their everyday lives (e.g., gender, race, or ethnic groups). It is important to note, however, that although the categories were introduced at the beginning of the study, children's explanations went far beyond information presented in these introductory stories. In particular, the introductory stories—even in the novel groups condition—did not include any negative interactions or between-group interactions of any kind. Thus, children's use of category memberships to explain harmful between-group interactions, as well as their frequent generation of explanations that the members of different groups would dislike each other (e.g., "she took his toys because they are enemies," "she did that because she doesn't like her") reflect children's abstract beliefs about the consequences of category memberships, not the specific information they were provided at the start of the study.

Overall, effects in this study were clearer for children's explanations of harmful than helpful behavior. Children referred to the agent only—without considering the recipient

at all—more for within-group than between-group harm, but to the relation between the agent and recipient (either their social relationship or their different category memberships) more for between-group harm than within-group harm. In contrast, effects for helpful behaviors were only found for one type of explanation—children referred to categories more to explain within-group helping than between-group helping; their use of agent-based or relationship-based explanations did not vary across group composition for helpful behaviors. For helpful behaviors, the most common type of explanation given referenced social relationships, perhaps because there is a compelling reason to view relationships as important both when the individuals are members of the same group (because children expect members of the same groups to be friends with one another, Shutts et al., 2013) and when the individuals are members of different groups (because children might expect that a particular friendship between the two characters motivated the positive cross-group interaction).

That clearer effects of group membership were found for harmful than helpful behaviors is consistent with prior work on preschool-age children's behavioral predictions. Children ages 3–5 use social categories to predict patterns of harmful social interactions (e.g., that a Flurp will harm a Zaz rather than another Flurp), but predict within-group helping and between-group helping equally often (Rhodes, 2012). Using social categories to predict helpful behaviors appears to arise later in childhood, between ages 6 and 7 (Rhodes, 2012).

Why might young children view social categories as constraining harmful but not helpful behaviors? This pattern may reflect the nature of their intuitive theory—in particular, their beliefs that social categories mark patterns of intrinsic interpersonal obligations (Rhodes & Chalik, 2013). Drawing from work in moral philosophy, *avoiding harm* is viewed as a fundamental interpersonal obligation, whereas providing help—while positive—is not obligated (see Knobe, 2010). Thus, children may view avoiding within-group harm as fundamental social constraint that operates on category members, but view helping as less central to their theories. This perspective also accounts for the pattern of children's agent-based explanations. Adults are more likely to attribute behavior to individual mental states if the action runs counter to their beliefs about social norms (Uttich & Lombrozo, 2010). If similar effects operate in early childhood, then in the present studies, children may have generated agent-based explanations (e.g., "he did because he wanted to") more for within-group harm than between-group harm because those actions violated their beliefs about the normative behavior of category members.

This account highlights another important implication of the present findings—that how children explain harmful actions may critically shape their moral judgments. Both children and adults often only hold perpetrators morally responsible for harmful actions when they attribute them to the agent's mental states (e.g., intentions; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011). For example, if an individual harms someone because he desires to harm, the agent is considered morally responsible for that action (and thus subject to blame and punishment). In contrast, if the harmful action is attributed to the agent's category membership, then it might not be viewed as reflecting the agent's underlying disposition. In the present study, children referenced the agent as responsible

for the harmful interaction more for within-group than between-group harm. Also, agent-based explanations were the most common type of explanation given for within-group harm. Thus, these findings raise the possibility that children are more likely to hold people as morally responsible for within-group harm. If so, children's beliefs about the normative status of harm in within-group and between-group interactions would influence their moral evaluations of those events via a process of affecting causal attributions.

The present data show that social categories shape preschool-age children's beliefs about the causal mechanisms responsible for harmful social interactions. By the preschool years, children appear to hold an intuitive theory that social categories mark patterns of social relationships and social obligations (Kalish, 2012; Kalish & Lawson, 2008), which shapes how they explain (as shown here), predict (Rhodes, 2012), and evaluate (Rhodes & Chalik, 2013) human social behavior. Identifying the developmental origins of this intuitive theory, the extent to which children incorporate beliefs about obligations into categories in various domains, and how this theory undergoes revision across development, are important areas for future work.

Acknowledgments

Funding for this work was provided by National Science Foundation grant BCS-1226942 to Rhodes. I thank Dr. Katherine Kinzler for sharing the audio files for the Language condition, Noam Neumark and Tri Nguyen for assistance with data collection and coding, as well as the Children's Museum of Manhattan for participating in this research.

References

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false beliefs. *Journal of Cognition and Development*, *7*, 139–173.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. (2006). Nature and nurture in own-race face processing. *Psychological Science*, *17*, 159–163.

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory. *Current Directions in Psychological Science*, *16*, 162–166.

Chalik, L., & Rhodes, M. (in press). Preschoolers use social allegiances to predict behavior. *Journal of Cognition and Development*, *???*, ???–???.

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, *77*, 539–553.

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of 'minimal' group affiliations in children. *Child Development*, *82*, 793–811.

Gelman, S., Collman, P., & Maccoby, E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *57*, 396–404.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology*, *37*, 668–683.

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

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development*, *12*, 213–238.

Kalish, C. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology*, *48*, 1133–1143.

Kalish, C., & Lawson, C. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development*, *79*, 577–593.

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

Kinzler, K., & Dautel, J. (2012). Children's essentialist reasoning about language and race. *Developmental Science*, *15*, 131–138.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 12577–12580.

Kinzler, K., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology*, *40*, 581–592.

Kinzler, K., Shutts, K., Dejesus, J., & Spelke, E. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*, 623–634.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*, 315–329.

Legare, C., Wellman, H. M., & Gelman, S. A. (2009). Evidence for an explanation advantage in naïve biological reasoning. *Cognitive Psychology*, *58*, 177–194.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*(5), 961–978.

Miller, J. G. (1986). Early cross-cultural commonalities in social explanation. *Developmental Psychology*, *22* (4), 514–520.

Miller, P. H., & Aloise, P. A. (1989). Young children's understanding of the psychological causes of behavior: A review. *Child Development*, *60*, 257–285.

Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about groups: Social categorization and the origins of intergroup bias. *Child Development*, *77*, 847–860.

Quinn, P., Yahr, J., Kuhn, A., Slater, A., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*, 1109–1121.

Rhodes, M. (2012). Naïve theories of social groups. *Child Development*, *83*, 1900–1916.

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*, 999–1006.

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*, 244–274.

Rhodes, M., Leslie, S. J., & Tworek, C. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 13526–13531.

Shutts, K., Pemberton Roben, C., & Spelke, E. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*, 35–62.

Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys, cows will be cows: Children's essentialist reasoning about human gender and animal development. *Child Development*, *79*, 1270–1287.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, *116*(1), 87–100.

Waxman, S. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*, 593–610.