Journal of Cognition and Development
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hjcd20

Children's Use of Social Categories in Thinking About People and Social Relationships

Kristin Shutts [a], Caroline K. Pemberton Roben [b] & Elizabeth S. Spelke [b]

[a] University of Wisconsin-Madison

[b] Harvard University

Accepted author version posted online: 28 Feb 2012.Version of record first published: 04 Jan 2013.

To cite this article: Kristin Shutts , Caroline K. Pemberton Roben & Elizabeth S. Spelke (2013): Children's Use of Social Categories in Thinking About People and Social Relationships, Journal of Cognition and Development, 14:1, 35-62

To link to this article: http://dx.doi.org/10.1080/15248372.2011.638686

JOURNAL OF COGNITION AND DEVELOPMENT, 14(1):35–62

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2011.638686

Routledge
Taylor & Francis Group

ARTICLE

Children's Use of Social Categories in Thinking About People and Social Relationships

Kristin Shutts
University of Wisconsin–Madison

Caroline K. Pemberton Roben and Elizabeth S. Spelke
Harvard University

A series of studies investigated White U.S. 3- and 4-year-old children's use of gender and race information to reason about their own and others' relationships and attributes. Three-year-old children used gender- but not race-based similarity between themselves and others to decide with whom they wanted to be friends, as well as to determine which children shared their own preferences for various social activities. Four-year-old (but not younger) children attended to gender and racial category membership to guide inferences about others' relationships but did not use these categories to reason about others' shared activity preferences. Taken together, the findings provide evidence for three suggestions about these children's social category-based reasoning. First, gender is a more potent category than race. Second, social categories are initially recruited for first-person reasoning but later become broad enough to support third-person inferences. Finally, at least for third-person reasoning, thinking about social categories is more attuned to social relationships than to shared attributes.

Beginning in infancy, young children show an impressive capacity to detect information relevant for social categorization, including others' gender, race, and age (e.g., Bahrick, Netto, & Hernandez-Reif, 1998; Bar-Haim, Ziv, Lamy, & Hodes, 2006; Ramsey, Langlois, & Marti, 2005). Children's early attention to such information raises a number of questions regarding how, when, and for what purposes children initially organize their social world. One question concerns the dimensions that young children rely on most strongly when evaluating and reasoning about other people. A second question concerns the manner in which children represent social categories: Are these categories initially centered on the self such that they divide people into ''us'' and ''them,'' or do children divide the world impartially into groups? Finally, what do

Correspondence should be sent to Kristin Shutts, Psychology Department, University of Wisconsin–Madison, 1202 West Johnson Street, Madison, WI 53706, USA. E-mail: kshutts@wisc.edu

social categories mean for young children, and what kinds of inferences do they support early in development? In particular, do social categories primarily convey information about properties shared by category members (e.g., Mary and Jessica like playing the same game), or do they primarily convey information about social relationships and networks (e.g., Mary and Jessica are friends)? In the present article, we address these questions through studies of 3- and 4-year-old children's gender and racial categories.

Much of what we know about young children's social categories comes from research examining children's gender and racial preferences (for reviews, see Aboud, 1988; Levy & Killen, 2008; Quintana & McKown, 2008; Ruble, Martin, & Berenbaum, 2006). Both naturalistic and laboratory-based studies indicate that preschool-age children prefer same-gender to other-gender children (Albert & Porter, 1983; La Freniere, Strayer, & Gauthier, 1984; Maccoby & Jacklin, 1987; Martin, 1989; Martin & Fabes, 2001; Martin, Fabes, Evans, & Wyman, 1999; Yee & Brown, 1994). Preschool-age children also prefer same-race to other-race children, provided their own racial group is high in status (Aboud & Skerry, 1984; Bigler & Liben, 1993; Kircher & Furby, 1971). Thus, from an early age, children use gender and race to guide their social preferences.

Though young children show both own-gender and own-race preferences during the preschool years, research suggests that children may initially care more about gender-based similarity between themselves and others than about race. Evidence that gender carries more weight than race for young children comes from studies that have directly compared children's use of the two categories to guide their first-person preferences and reasoning about unfamiliar people. For example, when White U.S. children were shown photograph pairs consisting of one boy and one girl or one White and one Black child and were asked to select the person they would like as a friend, 5-year-old children showed significant preferences for both own-gender and own-race photographs, but 4-year-old children only showed significant own-gender preferences (Abel & Sahinkaya, 1962). Moreover, when White U.S. children were asked to choose between items endorsed by pairs of children who differed from one another either in gender or in race, 3-year-old children showed robust preferences for items endorsed by children of their own gender but relatively weak preferences for items endorsed by children of their own race (Shutts, Banaji, & Spelke, 2010).

By the age of 4 years, children also attend to social category information to guide their third-person reasoning about other people. Four- to 5-year-old children who are provided with labels denoting others' gender, ethnicity, social class, or age infer that members of these categories share biological or psychological properties (Birnbaum, Deeb, Segall, Ben-Eliyahu, & Diesendruck, 2010; Diesendruck & haLevi, 2006; Gelman, Collman, & Maccoby, 1986; Martin, Eisenbud, & Rose, 1995; Taylor & Gelman, 1993; Taylor, Rhodes, & Gelman, 2009). For example, after learning that a child described as a boy had seeds inside his body and that a child described as a girl had eggs inside her body, a group of 4- to 6-year-old children inferred that a second boy had seeds rather than eggs inside his body (Gelman et al., 1986). As a second example, after learning that a child labeled as Arab liked one kind of novel activity (e.g., playing ''zigo'') but that a child labeled as Jewish liked a different kind of activity (e.g., playing ''zaber''), 5-year-old children reasoned that another Arab child would prefer playing zigo rather than zaber (Diesendruck & haLevi, 2006).

Despite the vast literature on children's social categorization, several features of children's early social categories warrant further study and are therefore the focus of the present article. First, we know little about whether and how children younger than 4 years of age make

third-person inferences about others based on social category information. Second, the relational and developmental ordering of children's consideration of social categories from the first- versus the third-person perspective remains largely unexplored. For these reasons, the present article includes both 3- and 4-year-old participants and investigates both first- and third-person reasoning about social categories.

Third, research has only begun to investigate the breadth of children's social category-based inferences. Many researchers have emphasized children's use of social categories to guide their reasoning about shared biological and psychological properties. Yet a handful of previous studies suggest that these are not the only—or even the most primary—inferences children consider when thinking about categories of people. For example, 4-year-old children view deontic properties (e.g., obligations such as ''having to do X'') as more central than psychological properties (''liking to do X'') when reasoning about social category membership (Kalish & Lawson, 2008). Additionally, children as young as 4 years of age use verbally conveyed information about others' gender to make inferences about likely relationships between other children (e.g., that a girl is more likely to be friends with another girl than with a boy; Martin, 1989; Martin et al., 1999). It is not clear, though, how children's social category-based inferences about interpersonal relationships compare to their inferences about shared attributes. The experiments in the present article test this question directly.

Finally, in most previous studies of young children's third-person reasoning about social categories, experimenters provided familiar category labels for children (e.g., ''boy''). Because labeling is known to heighten children's attention to categories (Gelman & Heyman, 1999; Patterson & Bigler, 2006; Waxman & Lidz, 2006), these studies do not reveal whether preschool-age children spontaneously use social category information to reason about other people's relationships or attributes. Moreover, the evidence that young children use labeled social categories to make inferences about category members' shared properties suggests that social categories may function much like natural-kind categories (e.g., animal species) for young children (e.g., Taylor et al., 2009), but it is not clear whether children display this tendency in their normal social interactions. When children meet a new child, that child typically is introduced by name (e.g., ''This is Kelly'') and not by a racial or gender label (e.g., ''This is a Black child''). Recent research by Waxman (2010) suggests that in the absence of category labeling, preschool-age children may not focus on race when generalizing psychological properties from one person to another. To investigate children's spontaneous inferences about others, participants in the present experiments never heard labels for the social categories.

OVERVIEW OF EXPERIMENTS 1 THROUGH 7

In a series of seven experiments, we investigated 3- and 4-year old children's attention to gender and race information when reasoning about shared psychological properties and social relationships. All of our experiments included trials focused on gender and trials focused on race. Participants in all the experiments described in this manuscript were ''White or European American'' (as identified by their parents). Moreover, all participants came from predominantly White urban or suburban communities in the United States (New England and the Midwest). The majority of participants were tested in the laboratory; however, some were tested in quiet rooms at their preschool or day care center.

TABLE 1
Summary of Methods and Key Findings for Experiments 1 Through 7

Expt.	*Age*	*First or Third Person*	*Question*	*Gender Trials*	*Race Trials*
1	3 yrs	First	Friendship	*	*ns*
2	3 yrs	First	Familiar Activities	*	*ns*
3	3 yrs	Third	Friendship	*ns*	*ns*
4	3 yrs	Third	Familiar Activities	*ns*	*ns*
5	4 yrs	Third	Friendship	*	*
6	4 yrs	Third	Familiar Activities	*ns*	*ns*
7	4 yrs	Third	Novel Activities	*ns*	*ns*

Note. Asterisks indicate that performance was above chance according to one-sample *t*-tests; ''*ns*'' indicates that performance did not differ from chance.

We began by testing 3-year-old children's consideration of gender and race from both a first- and third-person perspective. In Experiments 1 and 2, participants saw pairs of photographs featuring children who differed from one another in gender or race and were asked to indicate which person they would want as their friend (Experiment 1) or which person shared their own preference for a particular social activity (Experiment 2). In Experiments 3 and 4, participants saw a photograph of a target child followed by photographs of two other children who differed from one another either in gender or in race (depending on the trial type). Children were asked which member of the test pair was friends with the target child (Experiment 3) or shared that child's activity preference (Experiment 4), and their responses were analyzed to determine whether they tended to select the member of the test pair who matched the target in gender or race. Comparing 3-year-old children's performance in Experiments 1 and 2 versus Experiments 3 and 4 sheds light on whether children's early social categories are centered on the self, or whether children's categories are general enough to support third-person inferences as well.

After investigating the racial and gender categories of 3-year-old children, we turned to studies of older children to probe developmental changes in these categories. Our final experiments tested 4-year-old children's use of gender and race to guide inferences about others' social relationships (Experiment 5) and activity preferences (Experiments 6 and 7). Comparing children's performance in Experiment 5 versus Experiments 6 and 7 addresses whether social categories are equally supportive of inferences about interpersonal relationships and inferences about shared properties, or whether children make one of these types of inferences more readily than the other. Table 1 summarizes these seven experiments.

EXPERIMENT 1: FIRST-PERSON FRIENDSHIP INFERENCES (3-YEAR-OLD CHILDREN)

Experiment 1 tested whether 3-year-old children prefer other children of their own gender and race. In a series of ''gender trials,'' participants viewed displays consisting of one girl and one boy (both of whom were either White or Black). In a series of ''race trials,'' participants viewed displays consisting of one White and one Black child (both of whom were either male or female). On every trial, participants were asked to indicate the child with whom they wanted to be friends.

Methods

Participants. The participants were 24 White 3-year-old children (12 females; $M_{age} = 3;3$; range = 3;0–3;6). Two additional children were tested but excluded from the final sample, because they did not complete the task.

Materials. Trial displays featured photographs of children (8 cm × 6.5 cm) arranged into pairs according to categorical judgments (male/female and Black/White) and attractiveness ratings provided by a group of adults. All boys had hair that stopped above their ears, and all girls had longer hair. Because faces do not become markedly sexually dimorphic until puberty (Samal, Subramani, & Marx, 2007), hair length was likely the most salient cue to the gender of children featured in the photographs. Because we used real photographs of children (rather than computer-generated or manipulated images), Black and White children varied according to skin tone, hair texture, and physiognomy.

Gender trials consisted of four different boy–girl pairs, two of which showed photographs of White children and two of which showed photographs of Black children. Race trials consisted of four different Black–White pairs, two of which showed photographs of girls and two of which showed photographs of boys.

The task was presented in PowerPoint on a laptop computer (Figures 1A and 1B). Photographs were labeled with common names (e.g., Mary, John), and participants heard a series of brief facts about members of pairs (e.g., is seeing a movie with friends, is playing hide-and-seek with friends). Pictures that represented the facts (e.g., of movie popcorn, of trees in a park) were printed onto squares of paper affixed with Velcro. A clear plastic sheet containing small pieces of Velcro covered the computer screen and allowed children to indicate their friendship choices using the Velcro stickers.

Procedure. Participants sat in front of the testing computer next to an experimenter, while a coder stood behind the child and experimenter to record children's responses. On each trial, the experimenter presented the pair of photographs and then pointed to, named, and revealed the same fact about each photograph (e.g., ''This is Ashley. Ashley is having a birthday party with all her friends. And this is Olivia. Olivia is having a birthday party with all her friends.''). Then, the experimenter gave participants a Velcro sticker and asked them to place it by the person they would want to have as their friend (''Which of these kids would you want to have as your friend?''). After participants made a choice, the experimenter removed the sticker from the screen and repeated the question with the lateral positions of photographs reversed. The repeated questions served to discourage side biases. The experimenter never labeled the gender or race of children in the photographs and did not give participants feedback on their responses. If participants indicated that they were not sure how to respond, the experimenter said, ''It's OK to guess.''

Design. All children viewed four consecutive gender trials (two White pairs, two Black pairs) and four consecutive race trials (two girl pairs, two boy pairs). The order of trials (gender or race trials first), order of gender trial pairs (Black or White pairs first), and order of race trial pairs (male or female pairs first) were counterbalanced across children. The lateral positions of same-gender and same-race children were counterbalanced within and across participants. Pairings of particular facts to pairs were counterbalanced across participants. Pairings of names with

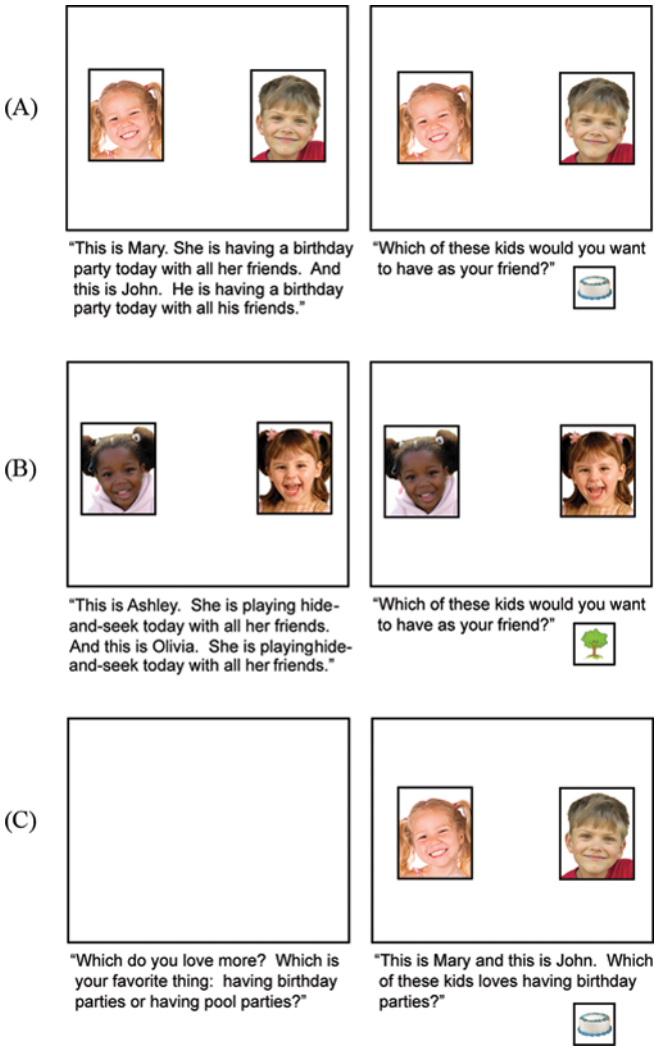

FIGURE 1 A) An example gender trial (with White targets) from Experiment 1. B) An example race trial (with female targets) from Experiment 1. C) An example gender trial (with White targets) from Experiment 2. Because of publishing restrictions, the photographs used in the experiments cannot be shown here. However, the images presented in figures are similar to the stimuli used in the experiments. (Color figure available online.)

particular photographs were also counterbalanced across participants (with the constraint that female names were always paired with girls and male names were always paired with boys).

Dependent measures and analyses. Performance on gender trials was determined by calculating the percentage of trials on which participants chose the photograph that matched their own gender, while performance on race trials was calculated as the percentage of trials on which participants chose the photograph that matched their own race. Performance was compared to

chance (50%) by one-sample *t*-tests. Analyses of variance (ANOVAs) were conducted to compare participants' choices on gender and race trials and to examine effects of order, participant gender, face gender (on race trials), and face race (on gender trials).

Two sets of analyses were performed for Experiment 1 and all the experiments that follow. One set of analyses focused on only the first judgment children gave on each trial, and the other focused on data averaged across the two judgments (i.e., the first query and the repeat query). We report the findings from both methods of analyses for all tests of principal effects: one-sample *t*-tests to chance for gender trials and race trials, comparisons of performance on race versus gender trials within an experiment, and comparisons across experiments. The principal effects in the manuscript are largely the same whether analyses consider only children's first judgment or both judgments. For the lengthier secondary analyses, findings were substantially the same across these two methods of analysis, and so we only report analyses of the first judgment children gave on each trial. Similarly, tables and figures display only the first judgment children gave on each trial.

Results

Figure 2 (left) presents the principal findings. On gender trials, participants tended to select photographs of children who matched their own gender, $M = 75\%$, $t(23) = 6.28$, $p < .001$. On race trials, however, participants showed no significant preference for photographs of children who matched their own race, $M = 54\%$, $t(23) < 1$. A repeated-measures ANOVA with trial type (gender vs. race trials) as a within-subject factor and participant gender and order (gender or race trials first) as between-subject factors revealed a main effect of trial type, $F(1, 20) = 7.69$, $p < .05$. This

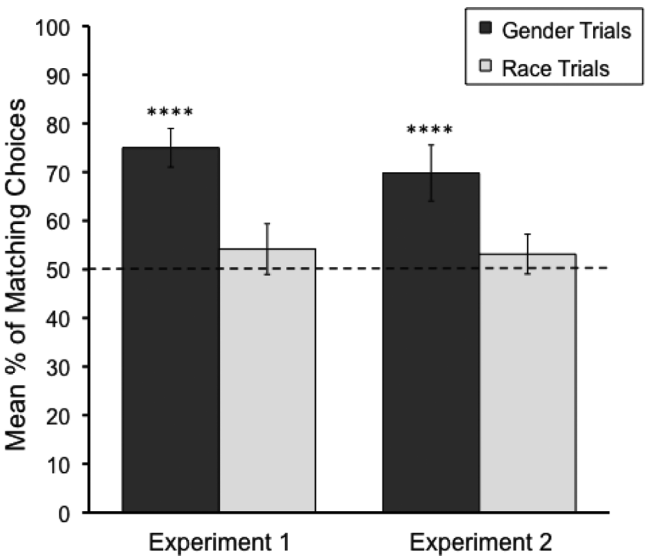

FIGURE 2 Results from the gender trials and race trials of Experiment 1 (left) and Experiment 2 (right). Bars depict standard error, and asterisks indicate means that are significantly different from chance (50%) according to one-sample *t*-tests ($^{****}p < .001$).

effect was qualified, however, by an interaction of trial type and participant gender, $F(1, 20) = 4.92$, $p < .05$. Simple-effects tests indicated that female and male participants performed similarly on race trials ($M_{Girls} = 48\%$, $M_{Boys} = 60\%$; *ns*) but differently on gender trials ($M_{Girls} = 85\%$, $M_{Boys} = 65\%$; $p < .01$). Although girls showed a more robust preference for same-gender faces than boys, girls and boys each showed a significant preference for children of their own gender, $t(11) = 7.34$, $p < .001$; $t(11) = 3.01$, $p < .05$, respectively). Moreover, neither girls nor boys showed a significant preference for children of their own race, $t(11) < 1$; $t(11) = 1.24$, *ns*, respectively). The ANOVA revealed no other significant effects or interactions.

For gender trials, an ANOVA with pair race (same vs. different from participant) as a within-subject factor and participant gender as a between-subject factor revealed only the effect of participant gender discussed above, $F(1,22) = 9.32$, $p < .01$. There was no effect of pair race, and children performed above chance on gender trials featuring White faces, $M = 69\%$, $t(23) = 2.39$, $p < .05$, as well as on gender trials featuring Black faces, $M = 81\%$, $t(23) = 4.73$, $p < .001$. For race trials, an ANOVA with pair gender (same vs. different from participant) and participant gender as factors revealed no significant main effects or interactions.

The principal effects in Experiment 1 are the same when both the first and second judgments made by children on each trial (i.e., both the first query and the repeat query) are analyzed. On gender trials, participants tended to select photographs of children who matched their own gender, $M = 65\%$, $t(23) = 4.10$, $p < .001$. On race trials, in contrast, participants showed no significant preference for photographs of children who matched their own race, $M = 51\%$, $t(23) < 1$. Children's same-gender choices on gender trials exceeded their same-race choices on race trials, $t(23) = 2.15$, $p < .05$.

Discussion

Children tended to select same-gender friends in this experiment, as they have done in past, naturalistic studies of 3-year-old children's playmate preferences (e.g., La Freniere et al., 1984; Maccoby & Jacklin, 1987; Powlishta, Serbin, & Moller, 1993) and in laboratory-based research (see Ruble et al., 2006, for a review). In contrast, participants did not use race information to make friendship decisions, again in accord with past research involving 3-year-old children (Kircher & Furby, 1971). Because the same procedure was employed to assess gender- and race-based social preferences in a within-subject design, the results from Experiment 1 suggest that gender was a more potent guide to social preferences than race was for children in our sample.

In the next experiment, we asked whether 3-year-old children consider gender or race when making another type of social inference: reasoning about their own and others' preferences for different activities. Previous research suggests that children might use gender- but not race-based similarity between themselves and others to guide first-person inferences about shared preferences: When 3-year-old White children were asked about their preferences for novel objects and activities that had been previously endorsed by a boy versus a girl or by a Black versus a White child, children selected items preferred by same-gender children but did not show a robust tendency to select items endorsed by same-race children (Shutts et al., 2010). Rather than first presenting children with others' preferences and then asking children to reason about their own preferences (as in Shutts et al., 2010), Experiment 2 first solicited children's own preferences and then asked them to reason about others' preferences.

EXPERIMENT 2: FIRST-PERSON SOCIAL ACTIVITY INFERENCES (3-YEAR-OLD CHILDREN)

Experiment 2 tested whether 3-year-old children believe that other children who match their gender or race also share their preferences for particular social activities. A new group of participants viewed a series of gender and race trials as in Experiment 1. On every trial, participants in Experiment 2 first indicated their own preference for a particular familiar activity. Then children were asked to indicate which member of the photograph pair also enjoyed that same activity (Figure 1C).

Method

The method was very similar to Experiment 1, except as follows: A new group of 24 White 3-year-old children (11 females; $M_{age} = 3;3$; range = 3;0–3;6) participated in Experiment 2. Data from one additional child were excluded because he did not complete the task. Before seeing each photograph pair, children were asked about their relative preference between two different activities (e.g., ''Which do you love more? Which is your favorite thing? Having birthday parties or having pool parties?''). Then, the experimenter showed participants the pair of photographs, named each child, and asked participants to indicate which child shared their preference (e.g., ''This is Kevin and this is Jessica. Which of these kids loves having pool parties?''). Participants indicated their choice by placing a Velcro sticker beneath one of the photographs. The facts consisted of familiar children's activities such as playing board games or playing cards, sleeping over or going camping, playing hide-and-seek or playing tag.

Results

Figure 2 (right) presents the principal findings. On gender trials, participants tended to select photographs of children who matched their own gender, $M = 70\%$, $t(23) = 3.80$, $p < .001$. On race trials, however, participants were equally like to choose same- and other-race photographs, $M = 53\%$, $t(23) < 1$. A 2 (trial type: gender vs. race) × 2 (participant gender) × 2 (order: gender vs. race trials first) ANOVA revealed only a main effect of trial type, $F(1,20) = 4.49$, $p < .05$. Separate 2 (pair race or gender: same as vs. different from participant) × 2 (participant gender) ANOVAs indicated no significant effects on either gender trials or race trials. Children performed above chance on gender trials featuring White faces, $M = 69\%$, $t(23) = 2.39$, $p < .05$, as well as on gender trials featuring Black faces, $M = 71\%$, $t(23) = 2.85$, $p < .01$.

Analyses for all the judgments collected in Experiment 2 (i.e., responses to both queries for each trial display) show the same principal findings. On gender trials, participants tended to select photographs of children who matched their own gender, $M = 66\%$, $t(23) = 3.36$, $p < .005$. On race trials, in contrast, participants did not choose photographs of children who matched their own race, $M = 54\%$, $t(23) = 1.43$, *ns*. Children's same-gender choices on gender trials exceeded their same-race choices on race trials, $t(23) = 2.12$, $p < .05$.

Discussion

As in Experiment 1, children tended to select other children who matched their own gender but did not preferentially select children who matched their own race. The findings from Experiment

2 accord with previous findings that 3-year-old children use others' gender more robustly than others' race to make inferences about their own preferences for novel objects and activities (Shutts et al., 2010). Thus, children are capable of reasoning about preferences between themselves and same-gender others in two directions: from others to the self (Shutts et al., 2010) and from the self to others (Experiment 2). The findings from Experiments 1 and 2 also suggest that gender may be a more robust social category than race, at least for young White children living in the predominantly White communities that we studied. We return to this idea in the General Discussion.

The results from gender trials in Experiments 1 and 2 suggest that children may divide their social world into ''us versus them'' on the basis of gender and infer that other children in their own group will be better friends and will share their own preferences for particular activities. In the next series of studies, we probed the nature of the representations underlying children's reasoning about social categories. In particular, are young children's social categories general enough to support inferences that do not involve the self? For example, do young girls reason that other girls are more likely to be friends with one another and share preferences in common, just as they infer relationships and commonalities between themselves and other girls? Moreover, do young girls reason that boys also are likely to be friends with one another and to share preferences in common? To address these questions, 3-year-old children in Experiment 3 were asked to reason about the social preferences of other children based on gender or race, while children in Experiment 4 were tested for generalization of social activity preferences by gender and race.

EXPERIMENT 3: THIRD-PERSON SOCIAL RELATIONSHIP INFERENCES (3-YEAR-OLD CHILDREN)

Experiment 3 tested for gender- and race-based inferences in 3-year-old children's reasoning about social relationships between other children. On ''gender trials,'' participants were introduced to a photograph of a boy or a girl (the target), followed by photographs of two other children, one of whom was the same gender as the target and one of whom was not (see Figure 3A). On ''race trials,'' participants were shown a picture of a Black or White child (the target), followed by photographs of two other children, one of whom was the same race as the target (Figure 3B). Children were asked which of the two test children was a friend of the target child. If 3-year-old children access information about gender and race to guide their inferences about social relationships between other children, then they should select the member of the test pair who matches the target's gender or race.

Methods

Participants. Twenty-four White 3-year-old children (12 females; $M_{age} = 3;2$; range = 3;0–3;5.5) participated in Experiment 3. None of these children had participated in Experiments 1 or 2. One additional participant was tested but excluded from the final sample, because his choices were unclear.

Materials. The task was presented in PowerPoint on a laptop computer. Photographs of Black and White boys and girls were arranged into gender triads and race triads based on

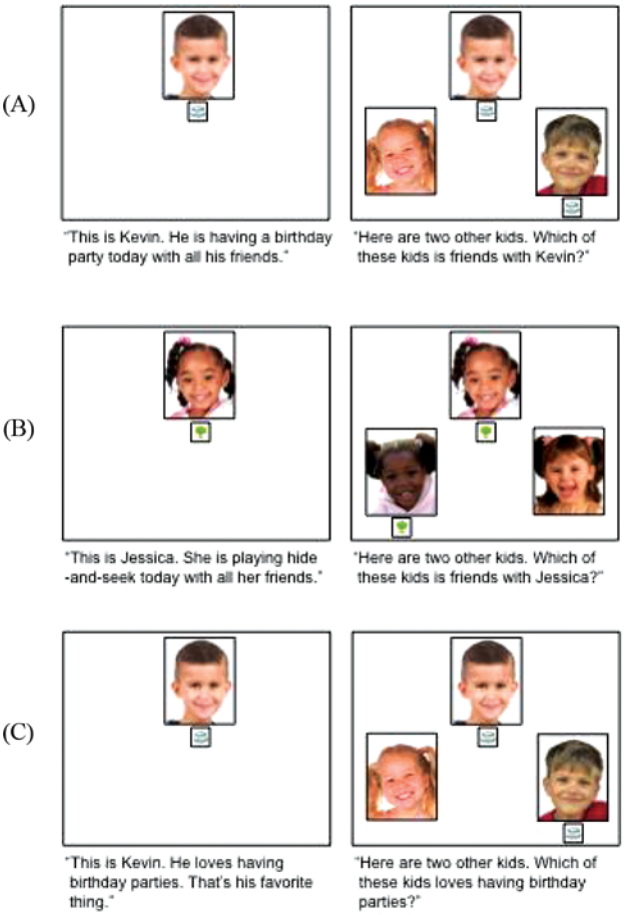

FIGURE 3 A) An example gender trial (male target, White triad) from Experiments 3 and 5. B) An example race trial (Black target, female triad) from Experiments 3 and 5. C) An example gender trial (male target, White triad) from Experiments 4 and 6. (Color figure available online.)

attractiveness ratings obtained from a group of adults. Gender trials ($N = 4$) consisted of a target (either a boy or a girl) followed by a male–female test pair. Two gender trials featured all White children, and two featured all Black children. Race trials ($N = 4$) consisted of a target (either a Black or a White child) followed by a Black–White test pair. Two of the race trials featured all girls, and two featured all boys.

Procedure. At the beginning of each trial, the target picture appeared in the center of the top half of the computer screen. The experimenter introduced the target by providing a name, as well as a fact about the target (e.g., ''This is Mary. She is going to the movies today with all her friends.''). The facts were the same as those in Experiment 1. The child was encouraged to give the target a Velcro sticker that depicted the fact (e.g., a picture of a movie popcorn container). After this, two test pictures appeared in the bottom left and right corners of the screen, joining

the target image and sticker. The experimenter pointed to the two test pictures simultaneously and asked the child to choose the target's friend (e.g., ''Which of these kids is friends with Mary?''). Children indicated their choice by placing a second Velcro sticker (e.g., movie popcorn picture) on the screen below one of the two faces. After children made their choice, the experimenter removed the stickers from the screen and repeated the question with the lateral positions of the bottom images reversed (e.g., ''Which of these kids is friends with Mary, again?'').

Design. All children saw four consecutive gender trials (two featuring White children and two featuring Black children) and four consecutive race trials (two featuring girls and two featuring boys). The race of the triad (Black vs. White) and the gender of the target (male vs. female) were orthogonally counterbalanced across the four gender trials. The gender of the triad (male vs. female) and the race of the target (Black vs. White) were orthogonally counterbalanced across the four race trials. The gender or race of the target photo in each triad also was counterbalanced across children, as were the order of trial blocks (gender or race trials first) and order of

TABLE 2
Number of 3-Year-Old Participants (out of 24) Who Selected the Social Category Match on Different Kinds of Gender and Race Trials in Experiments 3 and 4

Experiment 3 Gender Trials		
Target Gender	*Same as Participant*	*Different From Participant*
Triad Race		
Same as Participant	11	15
Different From Participant	17	10
Experiment 3 Race Trials		
Target Race	*Same as Participant*	*Different From Participant*
Triad Gender		
Same as Participant	17	11
Different From Participant	15	14
Experiment 4 Gender Trials		
Target Gender	*Same as Participant*	*Different From Participant*
Triad Race		
Same as Participant	12	12
Different From Participant	16	12
Experiment 4 Race Trials		
Target Race	*Same as Participant*	*Different From Participant*
Triad Gender		
Same as Participant	10	12
Different From Participant	14	12

Note. Cells with at least 18 participants would be significantly above chance according to a one-sample binomial test.

trial types within a block. Finally, the lateral positions of matching test images were counterbalanced both within and across participants.

Dependent measures and analyses. Performance on gender trials was determined by calculating the percentage of trials on which children chose the member of the test pair who matched the target's gender, while performance on race trials was calculated as the percentage of trials on which children chose the member of the test pair who matched the target's race. The resulting means were compared to chance (50%) by one-sample *t*-tests. An ANOVA was conducted to compare participants' performance on gender versus race trials, as well as to examine effects of trial order and participant gender. Because participants received only one of every gender trial-type display (for a total of four trials) and one of every race trial-type display (for a total of four trials), we did not conduct ANOVAs to explore effects of target gender and triad race on gender trials or effects of target race and triad gender on race trials. However, individual cell counts are displayed in Table 2.

Results

Figure 4 (left) presents the main findings. Participants performed at chance on gender trials, $M = 55\%$, $t(23) < 1$, and only marginally above chance on race trials, $M = 59\%$, $t(23) = 1.99$, $p = .06$. A repeated-measures ANOVA with trial type (gender vs. race trials) as a within-subject factor and participant gender and order (gender vs. race trials first) as between-subject factors revealed only a significant effect of participant gender, $F(1,20) = 7.23$, $p < .05$: Girls were more likely than boys to choose the member of the test pair who matched the target's social category

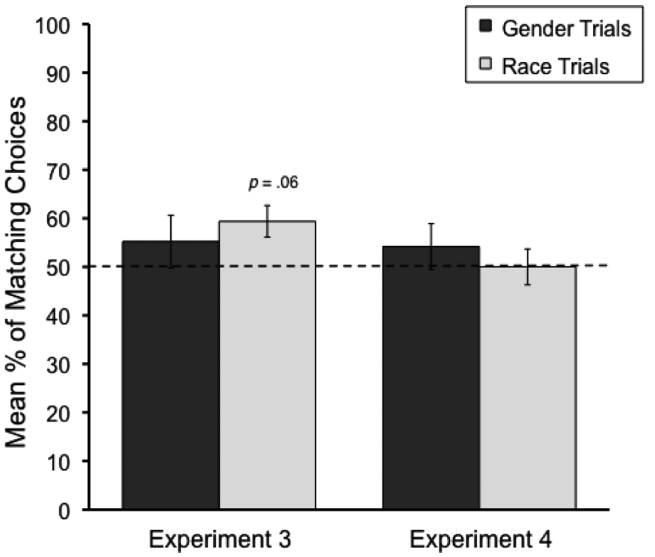

FIGURE 4 Results from the gender trials and race trials of Experiment 3 (left) and Experiment 4 (right). Bars depict standard error, and the *p*-value indicates the result of a one-sample *t*-test to chance (50%).

membership ($M = 67\%$ for girls, $M = 48\%$ for boys); indeed, girls performed above chance, $t(11) = 3.37$, $p < .01$, but boys did not, $t < 1$. Participant gender did not interact with the other factors in the ANOVA.

The findings of Experiment 3 do not improve when both judgments (i.e., both the first query and the repeat query for each trial) are considered: Participants still performed at chance on gender trials, $M = 55\%$, $t(23) = 1.14$, $p = ns$, and race trials, $M = 52\%$, $t(23) < 1$, and performance on gender and race trials did not differ, $t(23) < 1$.

Discussion

As a group, children in Experiment 3 did not tend to use gender or race information to make inferences about the friendship relationships of other children. Girls, however, did perform significantly above chance in Experiment 3, providing evidence that girls' social categories may be broad enough to support inferences about both their own social relationships (Experiment 1) and those of other children. To explore further 3-year-old children's social categories, Experiment 4 tested children's use of gender and race to guide third-person inferences about shared activity preferences.

EXPERIMENT 4: THIRD-PERSON SOCIAL ACTIVITY INFERENCES (3-YEAR-OLD CHILDREN)

Participants in Experiment 4 were presented with the displays from Experiment 3, in a different task context (see Figure 3C). On each trial, participants saw a target child and heard a fact about that child (e.g., ''George loves playing duck-duck-goose''). Then two more children were presented, and children were asked to choose the child with the same property as the target (e.g., ''Which of these kids do you think loves playing duck-duck-goose?''). If 3-year-old children access information about gender and race to guide their inferences about the attributes of novel people, then they should generalize the property information to the child of the same gender or race as the target.

Method

Participants were a new group of 24 White 3-year-old children (12 females; $M_{age} = 3;2.5$; range = 3;0–3;6). The method was identical to that in Experiment 3, except as follows: For each target, children learned a unique fact about a social activity enjoyed by the child (e.g., ''This is John. He loves having sleepovers. That's his favorite thing.''). The facts were the same as in Experiment 2. Pairings of particular facts to individual targets were counterbalanced across participants. After meeting the target, children were asked which of the test children enjoyed the same activity as the target (e.g., ''Which of these kids loves having sleepovers?''). Children's choices were coded and analyzed as in Experiment 3.

Results

Participants performed at chance on gender, $M = 54\%$, $t(23) = 1.28$, $p = ns$, and race, $M = 50\%$, $t(23) < 1$, trials (see Figure 4, right). A 2 (trial type: gender vs. race trials) $\times$ 2 (participant

gender) × 2 (order: gender vs. race trials first) ANOVA revealed no significant effects. These findings remain negative when both judgments (i.e., in response to the first query and the repeat query for each trial) are analyzed: Participants still performed at chance on gender trials, $M = 53\%$, $t(23) < 1$, and race trials, $M = 49\%$, $t(23) < 1$, and performance on gender and race trials again did not differ, $t(23) < 1$.

Although children did not consistently infer that other children of the same gender or race would be friends or like the same activities, it is possible that children made these inferences on the subset of trials when the target child belonged to the same social category as the participant. Table 2 presents participants' choices on the different types of gender and race trials, both in Experiment 3 and in Experiment 4. Although there appears to be a trend for children to make more category-based inductions from targets that matched their own gender and race, neither of these tendencies is significant: For gender trials, a paired-samples *t*-test revealed that participants were no more likely to pick the member of the test pair who matched the target's gender on trials where the target's gender matched that of the child than on trials where it did not, $t(47) = 1.16$, *ns*. For race trials, a paired-samples *t*-test revealed that participants were no more likely to pick the member of the test pair who matched the target's race on trials where the target's race matched that of the child than on trials where it did not, $t(47) = 1.27$, *ns*.

Discussion

Taken together, the findings from Experiments 1 through 4 suggest a nuanced picture of 3-year-old children's reactions to other children's gender and race. White U.S. children who are raised in predominantly White communities appear to use gender, but not race, to guide both their judgments of other children as their potential friends and their inferences about which children will have preferences similar to their own. When reasoning about the friendship relations and shared attributes of children in a third-person task, however, 3-year-old children may be less apt to use race or gender categories to guide their reasoning. Although girls used social categories to guide their reasoning about third-party friendship relations, boys did not, and neither girls nor boys used gender or race to guide their reasoning about third-party shared activity preferences. These findings suggest that 3-year-old children may not have a general propensity to reason about gender and race as social categories. Instead, they may have a more specific tendency to consider the gender of others in relation to the self.

Nevertheless, two findings suggest that children's third-person reasoning about others as members of social categories may be developing during the 4th year of life. First, when asked to reason about the friendship relationships of other children, girls in Experiment 3 showed a significant tendency to select the member of the test pair who matched the target's social category. Second, as a group, children in Experiment 3 showed a nonsignificant tendency to select the member of the test pair who matched the target's race on race trials. Because of these suggestive findings and because previous research provides evidence that older children draw inferences about others based on both their gender and their race (e.g., Gelman et al., 1986; Taylor & Gelman, 1993; Waxman, 2010), in the next set of studies, we tested children a full year older than participants in Experiments 1 through 4. We also tested 32 (rather than 24) 4-year-old participants in each experiment to have more power to detect children's attention to gender and race in reasoning about others' relationships and activity preferences.

EXPERIMENT 5: THIRD-PERSON FRIENDSHIP INFERENCES (4-YEAR-OLD CHILDREN)

Experiment 5 used the method of Experiment 3 to test whether 4-year-old children make inferences about other children's friendship relations on the basis of gender or race.

Method

Experiment 5 was identical to Experiment 3 except that participants were 32 White 4-year-old children (16 females; $M_{age} = 4;3$; range = 4;1–4;6).

Results

Figure 5 (left) presents the principal findings. On gender trials, children tended to select the member of the test pair who matched the target's gender as the target's friend, $M = 70\%$, $t(31) = 3.57$, $p < .005$. On race trials, children tended to select the member of the test pair who matched the target's race as the target's friend, $M = 68\%$, $t(31) = 4.10$, $p < .001$. A 2 (trial type: gender vs. race) × 2 (participant gender) × 2 (order: gender vs. race trials first) ANOVA revealed no significant effects.

Performance on gender trials featuring White faces was above chance, $M = 66\%$, $t(31) = 2.74$, $p < .05$, as was performance on gender trials featuring Black faces, $M = 73\%$, $t(31) = 3.70$, $p < .001$. Performance on race trials featuring faces that matched the participant's own gender was above chance, $M = 66\%$, $t(31) = 2.55$, $p < .05$, as was performance on race

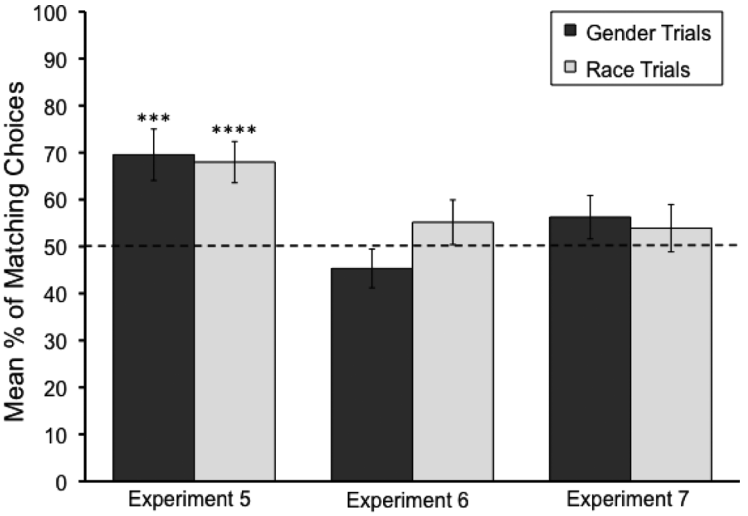

FIGURE 5 Results from the gender trials and race trials of Experiments 5 through 7. Bars depict standard error, and asterisks indicate means that are significantly different from chance (50%) according to one-sample *t*-tests (****$p < .001$; ***$p < .005$).

TABLE 3
Number of 4-Year-Old Participants (out of 32) Who Selected the Social Category Match on Different Kinds of Gender and Race Trials in Experiments 5, 6, and 7

Experiment 5 Gender Trials		
Target Gender	*Same as Participant*	*Different From Participant*
Triad Race		
Same as Participant	24	18
Different From Participant	27	20
Experiment 5 Race Trials		
Target Race	*Same as Participant*	*Different From Participant*
Triad Gender		
Same as Participant	18	24
Different From Participant	23	22
Experiment 6 Gender Trials		
Target Gender	*Same as Participant*	*Different From Participant*
Triad Race		
Same as Participant	18	10
Different From Participant	21	9
Experiment 6 Race Trials		
Target Race	*Same as Participant*	*Different From Participant*
Triad Gender		
Same as Participant	17	19
Different From Participant	19	15
Experiment 7 Gender Trials		
Target Gender	*Same as Participant*	*Different From Participant*
Triad Race		
Same as Participant	27	8
Different From Participant	24	13
Experiment 7 Race Trials		
Target Race	*Same as Participant*	*Different From Participant*
Triad Gender		
Same as Participant	19	14
Different From Participant	16	20

Note. Cells with at least 23 participants are significantly above chance according to one-sample binomial tests.

trials featuring faces that did not match the participant's own gender, $M = 70\%$, $t(31) = 3.46$, $p < .005$. See Table 3 for additional information about participants' choices on the different types of gender and race trials in Experiment 5 and all those that follow.

The principal effects in Experiment 5 are the same when both judgments (i.e., in response to the first and second queries for each trial) are analyzed: Participants again performed above chance on gender, $M = 69\%$, $t(31) = 4.05$, $p < .001$, and race, $M = 65\%$, $t(31) = 3.42$, $p < .005$, trials, and performance on gender and race trials did not differ, $t < 1$. By both methods of analysis, therefore, children tended to choose test children who matched the target child in race and in gender.

Discussion

Four-year-old participants attended both to gender and to race when reasoning about the social relationships of other children. The findings provide evidence that even in the absence of explicit verbal noun labeling of categories, White U.S. 4-year-old children living in predominantly White communities notice others' gender and race and use this information to guide their friendship inferences. These findings raise two questions. First, it is not clear what features of the procedure prompted children to choose same-gender and same-race children to be friends with one another. We introduced each target with a social activity and asked children to mark friendship relations with stickers that depicted activities. It is possible that asking about friendship relations in the context of social activities encouraged children to reason about shared activity preferences instead of (or in addition to) shared social relationships. Second, because the inferences in Experiment 5 concerned social relationships, it is not clear whether children would also infer that people of the same race or gender have properties in common. For both reasons, the next study tested directly whether 4-year-old children use gender and race to make inferences about the other children's social activity preferences. Do children expect other children of the same gender or race to like the same things?

EXPERIMENT 6: THIRD-PERSON SOCIAL ACTIVITY INFERENCES (4-YEAR-OLD CHILDREN)

Experiment 6 used the method of Experiment 4 to test whether 4-year-old children infer that other children of the same race or gender are more likely to favor the same activities.

Method

Experiment 6 was identical to Experiment 4 except that participants were 32 White 4-year-old children (16 females; $M_{age} = 4;3$; range = 4;1–4;6). None of these children had participated in Experiment 5.

Results

Participants performed at chance on trials testing inferences by gender, $M = 45\%$, $t(31) = -1.14$, $p = ns$, and race, $M = 55\%$, $t(31) = 1.09$, $p = ns$ (see Figure 5). A 2 (trial type: gender vs. race

trials) × 2 (participant gender) × 2 (order: gender vs. race trials first) ANOVA revealed no significant effects. Children's performance does not improve when both judgments (i.e., in response to the first and the repeat query on each trial) are analyzed: Participants still performed at chance on gender, $M = 48\%$, $t < 1$, and race, $M = 53\%$, $t < 1$, trials, and performance on gender and race trials did not differ, $t < 1$.

A second ANOVA compared performance in Experiment 6 to performance in Experiment 5 using experiment (5 vs. 6) as a between-subjects factor and trial type (gender vs. race trials) as a within-subjects factor. The analysis confirmed that social categories influenced children's friendship inferences more than their inferences about activity preferences, and did so across the board: There was a main effect of experiment, $F(1,62) = 13.34$, $p < .001$, but no effect of trial type and no interaction of trial type and experiment. The same effects emerge when this analysis is performed on all the judgments collected in these experiments (i.e., including children's responses to both queries on each trial; main effect of experiment: $F(1,62) = 12.91$, $p < .001$).

Discussion

Children in Experiment 6 did not generalize activity preferences from one individual to another according to gender or race. Because children noticed and used gender and race in Experiment 5 and because the displays were the same in Experiments 5 and 6, it is not likely that children's performance in Experiment 6 stems from a failure to detect race or gender information. Rather, the difference in performance between Experiments 5 and 6 suggests that children's early reasoning about gender and race may be focused more on social relationships than on shared attributes.

This conclusion must be qualified, however, because of the nature of the attributes that we tested. One possibility is that 4-year-old children simply do not use gender and race to guide inferences about highly familiar activities. Perhaps children expect or have learned that all children enjoy activities like going to birthday parties and playing hide-and-seek. Most previous studies of children's social category-based inferences about common attributes have employed novel properties (e.g., Birnbaum et al., 2010; Diesendruck & haLevi, 2006; Gelman et al., 1986; Waxman, 2010), whereas Experiment 6 employed familiar properties (following the evidence for successful inferences about these properties in Experiment 2). To investigate whether children would make more property inferences when the properties were novel, Experiment 7 used our displays and procedure to test 4-year-old children's gender- and race-based third-person inferences about novel activity preferences.

EXPERIMENT 7: THIRD-PERSON NOVEL SOCIAL ACTIVITY INFERENCES (4-YEAR-OLD CHILDREN)

Experiment 7 tested children's use of race and gender categories to make inferences about novel activity preferences. A new group of 4-year-old participants viewed the displays from Experiments 3 through 6 and were told that the target child enjoyed an unfamiliar social activity named by a nonsense word and depicted by a novel visual pattern (on a Velcro sticker). Then children were tested for generalization of these properties from targets to members of test pairs of children who differed in gender or race.

Method

Thirty-two White 4-year-old children participated in Experiment 7 (16 females; $M_{age} = 4;3$; range = 4;0–4;6). The method was similar to Experiments 4 and 6, with the following exceptions: Before each trial, children were presented with a sticker containing a picture of an unfamiliar toy, game, food, or drink that could be shared with others. The experimenter labeled the picture and provided a verbal description of the item in a social context (e.g., ''This is a picture of blicket. Blicket is a brand new game for 4 people. It's really fun, and people can play it at school with their friends.''). Then, children were introduced to the target and were told that the target liked the unfamiliar activity (e.g., ''This is Michael. He loves playing blicket. It's his favorite thing.''). As in Experiments 4 and 6, participants were then asked which of the test children liked the activity. Each trial featured a unique unfamiliar object and activity, along with a unique nonsense word and detailed description.

Results

Children performed at chance on trials testing inferences by gender, $M = 56\%$, $t(31) = 1.35$, $p = ns$, and race, $M = 54\%$, $t(31) < 1$ (Figure 5). A 2 (trial type: gender vs. race trials) × 2 (participant gender) × 2 (order: gender vs. race trials first) ANOVA revealed no significant effects.

When both judgments (i.e., in response to the first query and the repeat query for each trial) are considered, the results show a small but significant effect of gender: Although participants performed at chance on race trials, $M = 55\%$, $t(31) = 1.16$, $p = ns$, they performed above chance on gender trials, $M = 60\%$, $t(31) = 2.60$, $p < .05$. Nevertheless, performance on gender trials was not significantly different from performance on race trials, $t < 1$. At best, therefore, children showed a weak tendency to choose test children who matched the gender of the target child.

A 2 (experiment: 6 vs. 7) × 2 (trial type: gender vs. race) ANOVA comparing performance in Experiments 6 and 7 indicated no significant effects, either when performed on children's initial responses to each test display or when performed on all of children's responses, suggesting that children treated familiar and novel activity preferences similarly.

A final ANOVA examined performance in Experiments 5 and 7 with experiment as a between-subjects factor and trial type (gender vs. race) as a within-subjects factor. This analysis revealed only a significant effect of experiment, $F(1,62) = 7.08$, $p < .01$, with higher overall performance in the friendship inference task (Experiment 5) than in the novel properties inference task (Experiment 7). Children's performance advantage on the friendship inference task was maintained in an analysis of all the data collected (i.e., responses to two queries per trial), $F(1,62) = 4.89$, $p < .05$.

Discussion

Participants in Experiment 7 did not generalize properties from one individual to another based on information about race or gender. Four-year-old children's failure to generalize properties by gender or race in Experiment 6 therefore cannot be explained by the use of common or familiar properties, as Experiment 7 focused on generalization of novel properties. The difference between children's performance in Experiments 5 and 7 provides further support for the hypothesis that

children in the population we tested focus on gender and race when making inferences about the social relationships, but not the common attributes, of other children.

A final set of analyses, conducted on the data from Experiments 5 through 7, probed whether children approached the third-person reasoning tasks presented in those studies from a first-person perspective. Although all three experiments were designed to tap third-person social reasoning, it is possible that children adopted a first-person perspective in these tasks: When the target child matched the participant in race or gender, children might have reasoned about their own preferences rather than those of the target. To examine this possibility, we pooled results from Experiments 5 through 7 and tested whether children's performance on gender (or race) trials differed depending on the gender (or race) of the target in relation to the child. For gender trials, a paired-samples *t*-test revealed that participants were more likely to pick the member of the test pair who matched the target's gender on trials where the target's gender matched that of the child ($M = 73\%$) than on trials where it did not ($M = 41\%$), $t(95) = 6.94$, $p < .001$. For race trials, however, a paired-samples *t*-test revealed that participants' tendency to pick the member of the test pair who matched the target's race did not differ according to whether the target matched the participant's race ($M = 59\%$) or not ($M = 59\%$), $t < 1$. Thus, participants' own gender influenced their gender-based inferences about other children, whereas participants' race did not.

GENERAL DISCUSSION

Taken together, the findings from the present experiments suggest three primary conclusions about the development of children's social categories. Below we discuss each conclusion in turn and raise a host of questions for future research.

Gender Versus Race

The first conclusion is that gender may be a more potent social distinction than race for young White U.S. children. Three-year-old children in Experiments 1 and 2 used gender- but not race-based similarity between themselves and others to guide their own friendship decisions and reasoning about shared activity preferences. Moreover, 4-year-old children were influenced by their own gender, but not their own race, in our third-person experiments. These findings accord with previous evidence that gender is a particularly strong guide to young children's social preferences and reasoning (Abel & Sahinkaya, 1962; Rhodes & Gelman, 2009; Shutts et al., 2010).

All the participants in present studies were White and came from predominantly White social environments in the United States. Studies of children from other racial groups, backgrounds, cultures, and nationalities are critical for understanding the generality of finding that gender is a more potent category than race for children. More generally, studying children from diverse backgrounds and social experiences can shed light on the universal and variable aspects of children's social categories, as well as the factors that produce and modulate these categories.

As one example of the utility of cross-cultural approaches to studying social categories, we recently used a method very similar to that of Experiment 1 to examine the gender- and race-based social preferences of children living in South Africa (Shutts, Kinzler, Katz, Tredoux, & Spelke, 2011). Like the United States, South Africa is a country where Whites historically held power over Blacks and where White citizens still hold a disproportionate amount of the

country's wealth. Unlike the United States, however, Blacks are the racial majority group and Whites are a minority racial group in South Africa. Moreover, both South African leaders and the South African constitution and government promote national unity over racial or ethnic divisions. Studies of South African children can therefore shed light on the role of majority/minority group membership, familiarity, cultural messages, and social status in guiding children's social group attitudes.

Three main findings emerged from our research in South Africa. First, like White 3-year-old participants in our U.S. sample (Experiment 1), Black, White, and multiracial school-aged children in our South African samples showed stronger gender-based social preferences than race-based preferences. Though studies of children in other communities, cultures, and nations are warranted, this finding suggests that gender may be a universally important social distinction for children. Second, unlike most White school-aged children in the United States (Aboud, 1988; Baron & Banaji, 2006), Black children in South Africa did not favor children of their own race over children from other racial groups. This finding suggests that membership in a highly familiar statistical majority group does not necessarily lead to in-group favoritism. Third, regardless of their own racial group membership, South African children tended to favor people from high-status racial groups (e.g., Whites) rather than children from the majority race (i.e., Blacks). This last finding points to an important role for social status in guiding children's early social attitudes.

Why was gender a more robust social distinction for children in the present research as well as for children in our South African research? One possibility is that children are predisposed to consider gender information when evaluating people but do not possess dedicated mechanisms for evaluating others in accord with their race (Cosmides, Tooby, & Kurzban, 2003). Evolutionary psychologists have noted that gender distinctions have been available to human groups throughout our evolutionary history, whereas distinctions of race became available only in modern times, when long-distance trade and immigration brought different racial groups into contact (Cosmides et al., 2003). Thus, sensitivity to gender plausibly was shaped by natural selection, but sensitivity to race almost certainly was not.

A second possibility is that children may attend to gender categories before racial categories because adults tend to label and emphasize gender (but not race) information in their speech to young children (e.g., Arthur, Bigler, Liben, Gelman, & Ruble, 2008; Gelman, Taylor, & Nguyen, 2004; Pahlke, 2009). Societies also tend to mark gender distinctions with distinctive kinds of dress, toys, and activities. Studies of infants (who have minimal exposure to social messages about different social categories) could shed light on the role of social experience in guiding the emergence and development of children's social categories.

A third possibility is that young children do not reliably attend to race information, whereas they are more sensitive to information for gender. In support of this latter possibility, past research suggests that 3-year-old children find it difficult to identify and classify images by race (Renninger & Williams, 1966; Stevenson & Stewart, 1958; Williams & Morland, 1976). Nevertheless, many of the tasks traditionally used to probe children's racial identification and classification may be too demanding for 3-year-old children; indeed, some reveal no sensitivity to gender (e.g., Stevenson & Stewart, 1958), in contrast to the present findings.

A fourth possibility is that the particular photographs we used in our experiments did not effectively convey race information to participants. We believe it is unlikely that children failed to detect the race differences in the present photographs for several reasons. First, even infants detect race differences in photographic stimuli like those we used, and they do so in studies

using a wide range of stimuli (Anzures, Quinn, Pascalis, Slater, & Lee, 2010; Bar-Haim et al., 2006; Kelly et al., 2005). Second, the race (and gender) manipulations were clear to adult raters of the present photographs. Third, and most importantly, a significant effect on race trials emerged in Experiment 5 with 4-year-old children. Nevertheless, we have no independent evidence concerning the ability of the 3-year-old children in our sample to distinguish our photographs by race. For this reason, we cannot exclude this last explanation.

In future research, investigators might work to develop measures that are appropriate for measuring racial and gender identification and classification abilities in very young children. Using classification measures in concert with measures of children's preferences would be especially helpful in understanding the relation between children's detection of social category information and the emergence of category-based social preferences. Though social category-based preferences depend on an ability to detect information relevant for categorization, previous research provides evidence that categorization does not always engender social preferences. For example, Black children in our South African research (Shutts et al., 2011) were better at identifying photographs of ethnic in-group members (i.e., Xhosa people) when Xhosa faces were contrasted with faces from other racial groups (e.g., Whites) than when Xhosa faces were contrasted with non-South African faces from the same racial group (i.e., foreign Blacks). Nevertheless, children in our sample showed more robust in-group favoritism when Xhosa faces were contrasted with foreign Black faces than when Xhosa faces were contrasted with faces from other racial groups (e.g., Whites). For these children, at least, the most easily detectable category distinctions were not the most socially powerful ones. Nevertheless, problems of category discrimination and identification may contribute to the absence of race-based social preferences in the youngest U.S. children whom we studied.

First- Versus Third-Person Reasoning

The second primary conclusion from the present experiments is that children's early social categories appear to be initially marshaled for purposes of first-person reasoning, only later becoming robust enough to support third-person inferences about other children. Whether this pattern would hold across children from other racial groups, nationalities, and testing contexts awaits further research. However, 3-year-old children in our sample used gender to make first-person inferences about relationships and shared attributes but did not reliably use gender or race to reason about others' relationships or shared attributes. Moreover, 4-year-old children in our sample reasoned more consistently that other children would participate in same-gender social relationships and show same-gender social activity preferences when their own gender matched the target's gender. These findings suggest that first-person reasoning about social categories, relationships, and properties may develop in advance of third-party reasoning about the same categories and domains.

It is important to note that limitations in attention, memory, and perspective-taking abilities may explain why 3-year-old children performed above chance when asked to engage in first-person reasoning (i.e., gender trials of Experiments 1 and 2) but tended to perform at chance on third-person reasoning tasks (Experiments 3 and 4). To address this point, future research might examine correlations between children's performance on third-person social reasoning tasks and measures of other cognitive skills (e.g., perspective taking). Additionally, researchers

might work to develop less demanding methods for assessing children's third-person reasoning about social categories (e.g., looking-time procedures).

Social Relationships Versus Attributes

A final conclusion suggested by the present research is that third-person social category-based inferences about interpersonal relationships (friendships) appear to develop before, or more robustly than, third-person inferences about intrapersonal attributes (activity preferences). When asked to reason about others' friendships, 4-year-old children in Experiment 5 attended to both gender and race. In contrast, when asked to reason about others' activity preferences, 4-year-old children in Experiments 6 and 7 did not make reliable social category-based inferences. Early in development, White U.S. children living in predominantly White communities appear to use social categories as guides to social relationships more than as guides to shared personal properties.

Why did children in Experiment 5 infer that people from the same social category were more likely to be friends with one another than with someone from a different social category? One possibility is that young children learn about patterns of social relationships from their early social environments. To test the role of environments in guiding children's social category-based relationship inferences, it would be interesting to conduct similar studies with children from other racial groups, backgrounds, and communities. The participants in our sample likely had few experiences engaging in or observing cross-race friendships. Studies of elementary school-age children provide evidence that the racial and ethnic composition of children's social environments influences children's reasoning about the potential for cross-race relationships (Margie, Killen, Sinno, & McGlothlin, 2005; McGlothlin & Killen, 2006; McGlothlin, Killen, & Edmonds, 2005). In future work, it will be important to determine how young children's experiences in more diverse versus less diverse racial environments influence their reasoning about race and social relationships.

Why didn't children in Experiments 6 and 7 infer that people from the same social category would have activity preferences in common? Children in Experiment 6 may have (reasonably) inferred that children from all gender and racial groups would enjoy activities such as birthday parties. Yet children in Experiment 7 also failed to generalize novel activity preferences by gender or race. One possible explanation is that children in our sample have learned a general rule that children of all genders and race like the same sorts of activities. The literature on children's gender stereotyping casts some doubt on this as an explanation for children's performance on gender trials in Experiment 7, however. Preschool-age children hold robust stereotypes about the kinds of activities boys versus girls enjoy (e.g., playing with trucks vs. playing with dolls; Kuhn, Nash, & Brucken, 1978). Thus, it does not seem to be the case that children never observe or learn that people of different genders like different activities. Rather, we suggest that when young children are thinking about social categories, they are not focused on shared attributes as much as they are focused on shared patterns of social relationships.

It is important to note that children's performance in Experiments 6 and 7 contrasts with previous findings that 4-year-old children use gender and race to guide category-based inferences about others' biological (Gelman et al., 1986) and psychological properties (Waxman, 2010). Two factors may account for these differing findings. First, to test 4- to 6-year-old children's gender category-based reasoning, Gelman et al. (1986) used a procedure in which children heard noun labels (i.e., ''girl'' vs. ''boy'') to describe targets, and this labeling may have encouraged

children to make inferences about hidden properties shared by category members (Waxman & Lidz, 2006). In a direct test of this idea, Waxman (2010) gave children a social category-based property induction task with target individuals presented either with or without unfamiliar noun labels (e.g., ''This one is a Wayshan.''). Children then heard a novel property (e.g., ''likes to go glaving'') applied to a target person and were asked whether various other people also possessed that property. Four-year-old children who heard noun labels were much more likely to extend novel properties preferentially to other members of the target's gender or race category compared with children who did not hear noun labels. In fact, children did not engage in race-based property induction at all when the target was not described with a noun label.

Second, in most social category-based inductive inferences studies—including those conducted by Gelman et al. (1986) and Waxman (2010)—children see and hear about the relevant social category contrast prior to generalizing properties. For example, before teaching children about a target's preference (e.g., ''likes eating naggles''), Waxman (2010) introduced the relevant target (e.g., a picture of a woman) by saying, ''Look at this one.'' Then, she showed a contrasting photograph (e.g., a picture of a man) and said, ''Now look at this one. They are not the same. They are different.'' Such a contrasting procedure might highlight social category information for children and invite them to think about differences between members of different categories. Participants in the present studies, however, only saw one target in isolation before they were asked to make inferences about other category members' properties.

Further Considerations

One general question raised by the present experiments concerns the role of perceptual information in guiding children's choices in first- and third-person tasks. Of course, children had to be sensitive to some sort of perceptual information (e.g., skin color, hair length) to perform above chance in any of these studies because the experimenter did not provide labels for the social categories. Nevertheless, 4-year-old children's performance cannot be due to a general perceptual-matching strategy (e.g., ''point to the person in the pair who looks like the target''), because children in Experiments 5 through 7 saw the same trial displays but only performed above chance in Experiment 5. Still, children's performance in Experiment 5 could be driven by an assumption that people who look alike are likely to be friends (rather than an assumption that people from the same social category are likely to be friends). In future research, it would be interesting to test whether gender and race are unique in guiding children's inferences about shared social relationships or whether perceptual similarity of any sort (e.g., face shape, T-shirt color) invites children to make inferences about social relationships.

Three-year-old children do not seem to use perceptual similarity (either alone or as an indication of shared social category membership) to make third-person inferences. However, it is possible that 3-year-old children's choice of same-gender others in Experiments 1 and 2 was driven by perceptual similarity between themselves and the people pictured in the photographs rather than by reasoning about social category membership per se. Future studies might ask whether 3-year-old children's first-person preferences are sensitive to physical resemblance. If 3-year-old children are more concerned with who looks most similar to them than they are with others' gender, then a brunette wearing an orange shirt might prefer to be friends with a brown-haired boy wearing an orange shirt rather than with a blond girl wearing a green shirt.

One additional direction for future research concerns the scope of children's reasoning about social relationships within and between members of different social categories. The experiments in the present article focused on friendship, but humans participate in many other kinds of relationships and exchanges, including those involving dominance, antagonism, respect, and helping. Do children focus on social category membership to guide inferences about these kinds of social relationships and interactions? Given past findings that children focus on deontic properties when reasoning about social categories (Kalish & Lawson, 2008), it would be particularly interesting to use the methods of the present article to investigate whether and how young children use social categories to guide inferences about deontic social relationships and interactions (e.g., ''Mary is supposed to stand up when Person X comes in the room,'' or ''Mary is supposed to help Person X.'' ''Who else is supposed to do this?'').

More generally, we suggest that studies of young children's early social categories and reasoning may illuminate core features of human social categories throughout development. Older children and adults with extensive experiences in the social world clearly hold beliefs that members of social categories share properties such as preferences, traits, and patterns of behavior (e.g., Fiske, 1998). However, findings from the present experiments suggest that early in development, humans' social categories might be more focused on patterns of interpersonal relationships than on personal attributes. When it comes to reasoning about social categories, children—and perhaps even adults—might be better characterized as naïve sociologists than as naïve psychologists.

ACKNOWLEDGMENTS

The research was supported by National Institutes of Health Grant HD23103 to Elizabeth S. Spelke.

The authors thank Margaret Arnold, Marissa Martinez, and Caitlin McKee for assistance with data collection.

REFERENCES

Abel, H., & Sahinkaya, R. (1962). Emergence of sex and race friendship preferences. *Child Development*, *33*, 939–943.

Aboud, F. (1988). *Children and prejudice*. Cambridge, MA: Basil Blackwell.

Aboud, F., & Skerry, F. E. (1984). The development of ethnic attitudes: A critical review. *Journal of Cross-Cultural Psychology*, *15*, 3–34.

Albert, A. A., & Porter, J. R. (1983). Age patterns in the development of children's gender-role stereotypes. *Sex Roles*, *9*, 59–67.

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., & Lee, K. (2010). Categorization, categorical perception, and asymmetry in infants' representation of face race. *Developmental Science*, *13*, 553–564.

Arthur, A. E., Bigler, R. S., Liben, L. S., Gelman, S. A., & Ruble, D. N. (2008). Gender stereotyping and prejudice in young children: A developmental intergroup perspective. In S. Levy & M. Killen (Eds.), *Intergroup relations: An integrative developmental and social psychological perspective* (pp. 66–86). Philadelphia, PA: Psychology Press.

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*, 1263–1275.

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

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Bigler, R. S., & Liben, L. S. (1993). A cognitive-developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*, *64*, 1507–1518.

Birnbaum, D., Deeb, I., Segall, G., Ben-Eliyahu, A., & Diesendruck, G. (2010). The development of social essentialism: The case of Israeli children's inferences about Jews and Arabs. *Child Development*, *81*, 757–777.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, *7*, 173–178.

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

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 357–411). Boston, MA: McGraw-Hill.

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

Gelman, S. A., & Heyman, G. D. (1999). Carrot eaters and creature believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, *10*, 489–493.

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother–child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, *69*(1, Serial No. 275).

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

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., . . . Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, 31–36.

Kircher, M., & Furby, L. (1971). Racial preferences in young children. *Child Development*, *42*, 2076–2078.

Kuhn, D., Nash, S. C., & Brucken, L. (1978). Sex role concepts of 2- and 3-year-olds. *Child Development*, *49*, 445–451.

La Freniere, P., Strayer, F. F., & Gauthier, R. (1984). The emergence of same-sex affiliative preferences among preschool peers: A developmental/ethological perspective. *Child Development*, *55*, 1958–1965.

Levy, S., & Killen, M. (2008). *Intergroup relations: An integrative developmental and social psychological perspective*. Philadelphia, PA: Psychology Press.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. In E. H. Reese (Ed.), *Advances in child development and behavior* (Vol. 20, pp. 239–287). New York, NY: Academic Press.

Margie, N., Killen, M., Sinno, S., & McGlothlin, H. (2005). Minority children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology*, *23*, 251–269.

Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, *25*, 80–88.

Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's gender-based reasoning about toys. *Child Development*, *66*, 1453–1471.

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, *37*, 431–446.

Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children's beliefs about playing with girls versus boys and their relations to sex-segregated play. *Journal of Social and Personal Relationships*, *16*, 751–771.

McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children attending ethnically homogenous schools. *Child Development*, *77*, 1375–1386.

McGlothlin, H., Killen, M., & Edmonds, C. (2005). European American children's intergroup attitudes about peer relationships. *British Journal of Developmental Psychology*, *23*, 227–249.

Pahlke, E. E. (2009). *European American racial socialization: The influence of mothers' behavior and beliefs on young children's racial attitudes (Unpublished doctoral dissertation)*. Austin, TX: University of Texas at Austin.

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

Powlishta, K. K., Serbin, L. A., & Moller, L. C. (1993). The stability of individual differences in gender typing: Implications for understanding gender segregation. *Sex Roles*, *29*, 723–737.

Quintana, S. M., & McKown, C. (2008). *Handbook of race, racism, and the developing child*. Hoboken, NJ: John Wiley & Sons.

Ramsey, J. L., Langlois, J. H., & Marti, N. C. (2005). Infant categorization of faces: Ladies first. *Developmental Review*, *25*, 212–246.

Renninger, C. A., & Williams, J. E. (1966). Black–white color connotations and racial awareness in preschool children. *Perceptual and Motor Skills*, *22*, 771–785.

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

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 858–931). New York, NY: Wiley.

Samal, A., Subramani, V., & Marx, D. (2007). An analysis of sexual dimorphism in the human face. *Journal of Visual Communication and Image Representation*, *18*, 453–463.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science*, *13*, 599–610.

Shutts, K., Kinzler, K. D., Katz, R. C., Tredoux, C., & Spelke, E. S. (2011). Race preferences in children: Insights from South Africa. *Developmental Science*, *14*, 1283–1291.

Stevenson, H. W., & Stewart, E. C. (1958). A developmental study of racial awareness in young children. *Child Development*, *29*, 399–409.

Taylor, M. G., & Gelman, S. A. (1993). Children's gender- and age-based categorization in similarity and induction tasks. *Social Development*, *2*, 104–121.

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.

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.

Waxman, S. R., & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology* (6th ed., Vol. 2, pp. 299–335). Hoboken, NJ: Wiley.

Williams, J. E., & Morland, J. K. (1976). *Race, color, and the young child*. Chapel Hill, NC: University of North Carolina Press.

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology*, *33*, 183–196.