Received: 11 June 2019 | Revised: 19 March 2020 | Accepted: 16 June 2020

DOI: 10.1111/desc.13013

PAPER

Developmental Science WILEY

Are category labels primary? Children use similarities to reason about social groups

Ashley Jordan | Yarrow Dunham

Department of Psychology, Yale University, New Haven, CT, USA

Correspondence
Ashley Jordan, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA.
Email: ashley.jordan@yale.edu

Funding information
The John Templeton Foundation, Grant/Award Number: 56036

Abstract

While interpersonal similarities impact young children's peer judgments, little work has assessed whether they also guide group-based reasoning. A common assumption has been that category labels rather than 'mere' similarities are unique constituents of such reasoning; the present work challenges this. Children (ages 3–9) viewed groups defined by category labels or shared preferences, and their social inferences were assessed. By age 5, children used both types of information to licence predictions about preferences (Study 1, $n = 129$) and richer forms of coalitional structure (Study 2, $n = 205$). Low-level explanations could not account for this pattern (Study 3, $n = 72$). Finally, older but not younger children privileged labelled categories when they were pitted against similarity (Study 4, $n = 51$). These studies show that young children use shared preferences to reason about relationships and coalitional structure, suggesting that similarities are central to the emergence of group representations.

KEYWORDS

inductive reasoning, similarity, social categories

1 | INTRODUCTION

From early in development, children are faced with the crucial task of deciding which types of social information to attend to when learning about others. Understanding the ways in which people are connected is important because it can help children predict how individuals will behave and interact (Shutts, Roben, & Spelke, 2013), thereby anchoring children's inferences about the social world. Social category membership has emerged as one of the most important types of input children use to make these determinations. A number of studies have shown that children use group distinctions such as language, ethnicity, gender and race to support their generalizations about others' behaviours and preferences (Diesendruck & haLevi, 2006; Liberman, Woodward, & Kinzler, 2017; Shutts et al., 2013; Taylor, Rhodes, & Gelman, 2009). For example, Shutts et al. (2013) found that by age 4 young children use shared race and gender category membership to infer friendship among third parties. Children also rely on abstractly defined social categories, such as noun-labelled, novel groups, to support their inferences about social relationships (Baron & Dunham, 2015; Chalik, Rivera, & Rhodes, 2014; Dunham, 2018; Kalish, 2012; Rhodes & Chalik, 2013). Taken together, these studies show that categorical information supports children's reasoning about the social world. Specifically, it indicates the closeness of individuals within a group and to what extent properties are likely to generalize across group members.

In addition to helping children grasp interpersonal relationships, some have argued that social categories serve a more powerful function, guiding children's inferences about coalitional structure (Chalik & Rhodes, 2018; Cimpian, 2016; Kurzban, Tooby, & Cosmides, 2001; Rhodes & Chalik, 2013). This makes intuitive sense given that categories facilitate young children's concept acquisition in a number of domains, such as natural kinds (Cimpian & Erickson, 2012; Gelman, Collman, & Maccoby, 1986; Gelman & Markman, 1986), artefacts (Dewar & Xu, 2009; Mandler & McDonough, 1996; Xu, 2002) and word-learning (Soja, Carey, & Spelke, 1991). In the social domain

specifically, several studies have revealed that categories, denoted by labels, carry an advantage over other features such as perceptual similarity (Baron, Dunham, Banaji, & Carey, 2014; Diesendruck & haLevi, 2006).

If categorical information, conveyed via labels, most powerfully informs children's acquisition of social knowledge, too, one should expect children to privilege this form of information over other relevant types of input. For example, it should support social inferences to a greater extent than other dimensions of similarity such as shared interests. Furthermore, one would expect that this inferential advantage for category labels would be at its strongest in cases where children are asked to reason about the structure of social relationships, because these have been considered inferences that uniquely follow from the assumed presence of social categories and the social coalitions they imply.

However, despite evidence suggesting that children rely most heavily on social categories to apprehend those in their environment, they clearly also rely on input about individuals to inform such evaluations. For example, Chalik et al. (2014) introduced 3- and 4-year-old children to a pair of noun-labelled groups and provided them with information about the mental states of individuals from these groups. Afterward, they asked children to predict how new exemplars would behave towards one another. Critically, they wanted to know whether children would privilege information about individuals' mental states or group membership when deciding to whom these individuals would direct a negative action. The authors found that as children's theory of mind ability increased, they were increasingly likely to rely on mental states as compared to group membership to predict others' actions. More recent work also suggests that shared interests impact children's affiliation judgments and resource allocation decisions to a similar extent as categorical information (Sparks, Schinkel, & Moore, 2017).

Even young infants use abstract mental states like emotions, goals, opinions and tastes to predict others' dispositions and likely future behaviours (Hamlin, Mahajan, Liberman, & Wynn, 2013; Kuhlmeier, Wynn, & Bloom, 2003; Liberman, Kinzler, & Woodward, 2014). Indeed, shared tastes are a particularly salient cue that infants rely on when reasoning about social partners (Liberman et al., 2014; Mahajan & Wynn, 2012). For example, a study by Liberman et al. (2014) showed that infants use shared tastes to guide their expectations about the quality of others' relationships. Moreover, shared interests in toys and clothing predict young children's preferences for peers (Fawcett & Markson, 2010). These findings suggest that interpersonal similarities are a fundamental part of children's developing social sense.

Thus, an alternative perspective, which these findings might be taken to support, holds that children's category-based induction operates via a bottom-up system for tracking similarity (Landau, Smith, & Jones, 1988; Sloutsky, 2003; Sloutsky & Fisher, 2004; Sloutsky, Kloos, & Fisher, 2007; Sloutsky, Lo, & Fisher, 2001; Smith, Jones, & Landau, 1996). Under this account, rather than serving as symbolic markers of group membership, category labels are features which distinct entities within a group share. If they have any privileged role it simply stems from their ease of acquisition, perceptual salience and status as common knowledge in a linguistic community. From this perspective, despite evidence to the contrary (Baron et al., 2014; Cimpian & Erickson, 2012; Dewar & Xu, 2009; Diesendruck & haLevi, 2006; Gelman et al., 1986; Gelman & Markman, 1986; Mandler & McDonough, 1996; Soja et al., 1991; Taylor et al., 2009; Xu, 2002), children will not *necessarily* demonstrate a category advantage in their social reasoning. It is important to note, however, that the bulk of these studies have generally focused on children's reasoning about natural and artefact kinds, making it difficult to draw conclusions about children's developing *social* cognition. For this reason, it is important to specifically assess the relative strength of category labels and shared preference information in the social domain. Moreover, it is important to assess this using a task that holds other aspects of similarity, such as lower-level perceptual cues, constant across conditions.

Here we provide the first direct assessment of a central assumption in the developmental literature on social categories, namely, that the patterns of reasoning supported by category labels are uniquely due to *category* relationships as opposed to other types of similarity relationships. To this end we tested the strength of children's social inferences after they received information about either category membership or shared preferences in a matched design. We selected an age range of 3–9 years as this is similar to that of other work assessing children's social category reasoning (e.g. Rhodes & Chalik, 2013). As in most past work, we signalled category membership via the use of novel noun labels to denote the social categories that other children belonged to. We predicted that children would rely on labelled category membership (e.g. ''These kids are the Zertles'.) more than common interest (e.g. 'These kids like to eat zertles'.) to infer others' preferences and behaviours. Since much of the prior work on social inferences focuses on familiar categories like gender, race and ethnicity (e.g. Diesendruck & haLevi, 2006; Shutts et al., 2013), we used a novel groups paradigm to isolate the influence of children's abstract categorical reasoning from information they may have acquired outside of the experimental context.

Research Highlights

- Children (ages 3–9) viewed groups defined by labelled categories or shared preferences, and their social inferences were assessed.
- By age 5, children used both types of information to licence predictions about preferences and richer forms of coalitional structure.
- Low-level explanations could not account for this pattern.
- Older but not younger children privileged category labels when they were pitted directly against shared preference.

This method is commonly used with developmental populations to circumvent their potentially confounding knowledge of familiar groups (Baron et al., 2014; Chalik et al., 2014; Dunham, Baron, & Carey, 2011; Dunham & Emory, 2014; Patterson & Bigler, 2006; Rhodes & Chalik, 2013).

1.1 | Overview of studies

In Study 1, we tested whether category membership or shared preferences better supports children's predictions about others' activity and social preferences. Study 2 assessed whether this distinction had an impact on children's predictions about deontic relationship and coalitional expectations, that is, whom children believe to be obligated to one another. Prior work suggests that these types of predictions are uniquely supported by social category membership over other forms of similarity (Rhodes & Chalik, 2013). In Study 3, we assessed children's baseline performance in the absence of either social category labels or shared preference information but in the presence of lower-level perceptual commonalities, in order to rule out low-level explanations for the results we obtained in Studies 1 and 2. Finally, in Study 4, we directly pitted category cues against similarity cues, providing the most stringent test of the relative strength of each type of information. To forecast our results, which were contrary to our predictions, in nearly all cases the patterns of inferences generated by shared interests were indistinguishable from those generated by shared social category membership, particularly among the youngest children we tested. Based on these results, we argue that shared interests are central to children's understanding of social groups, and that the inductive advantage frequently attributed to category labels may not be so secure.

2 | STUDY 1

Prior work has assessed children's reliance on category labels when reasoning about social preferences and shared interests (Diesendruck & haLevi, 2006; Shutts et al., 2013). However, these studies have focused on familiar social distinctions such as gender, race and ethnic or religious groups. Study 1 employed a novel groups paradigm to examine whether children use cues to category membership (Category condition), as compared to shared food preference (Similarity condition), when predicting others' relationship patterns. Following previous work, we used a triad task in which we asked children to predict which of two individuals a 'target' individual would befriend, commit a harmful action against, or share an activity preference with, in cases in which one individual shared and the other did not share a particular feature (i.e. a category label or a common preference, depending on the condition) (Rhodes & Chalik, 2013; Shutts et al., 2013). Although in the case of activity preference, we are asking children to infer a new shared preference from information about prior shared preferences, we felt it important to include these trials to allow for comparison to prior literature (e.g. Shutts et al., 2013).

2.1 | Method

2.1.1 | Participants

The participants were 129 children (n = 65 female, M_{age} = 6.30 years; range = 3.11–9.97 years) from three age groups: 3–4- (n = 43), 5–6- (n = 36) and 7–9 (n = 50) years old, with this sample size selected based on past studies of children's category-based inferences (e.g. Taylor et al., 2009). We tested 12 additional children who we excluded from analyses due to the following: experimenter error (n = 4: similarity condition, n = 2), failure to complete the task (n = 7: similarity condition, n = 4) or lack of written informed consent from a parent or legal guardian (n = 1, similarity condition). Data collection took place from late fall of 2015 to the summer of 2016. The study took place in either a university laboratory (n = 27), a children's museum (n = 52) or an empty classroom at the participant's school (n = 50).

For each study reported here: we recruited participants from the New England region of the United States; we did not collect information about participant race or family income, however, given the demographic profiles of our testing sites, we believe most participants were white and from middle-income families; all parents or legal guardians provided written informed consent on their child's behalf, and each child provided verbal assent before beginning the study.

2.1.2 | Design and materials

We randomly assigned participants to either the category (n = 68) or similarity (n = 61) condition. The task consisted of three test trial blocks, each with four trials of the following types: 'harm', 'friend' and 'activity'. The experimenter conducted the study in PowerPoint on a laptop computer. We used Photoshop to generate 88 unique characters marked by T-shirt colour for use in the experiment (16 introduction characters and 72 test trial characters). Each character displayed a positive facial expression, and the gender of the participant matched that of the characters he or she viewed. We counterbalanced the order of the trial blocks, the lateral position of the characters during the introduction and test phases, and the pairing of verbal label to T-shirt colour.

2.1.3 | Procedure and scoring

The experimenter explained to each participant that he or she would 'look at kids from a storybook and learn things about them'. She told participants in the category condition that they would learn about 'who each kid is' and participants in the similarity condition that they would learn about 'what each kid likes'. Next, the experimenter displayed two sets of four introduction characters, one on each side of

the screen. One set of characters wore blue T-shirts, and the other set wore red T-shirts. In the category condition, the experimenter told participants that each set belonged to a novel-labelled group. She said of one set while pointing to the characters: 'See these kids? They are all called the "*Zertles*"'. And she said of the other set while pointing: 'See these kids? They are all called the "*Lapes*"' (Figure 1a). In the similarity condition, she told participants that each character in the set liked to eat the same noun-labelled food. She said of one set while pointing to the characters: 'See these kids? Each of them likes to eat a food called "*zertles*"'. And she said of the other set while pointing: 'See these kids? Each of them likes to eat a food called "*lapes*"'.

The experimenter asked each participant to recall which set of introduction characters was associated with each group or food label saying either: 'Do you remember what these kids are called?' Or: 'Do you remember which food these kids like to eat?' If a participant responded incorrectly, the experimenter corrected him or her by saying, for example: 'Actually, I think *these* kids are called the Lapes, and *these* kids are called the Zertles', while pointing to the correct set of characters. Once the experimenter ascertained that the participant understood this information, she told him or her that they would answer questions about new kids that they had not yet seen.

Each test trial began with the experimenter directing the participant's attention to a target character that appeared in the upper-middle portion of the screen. First, she reminded participants of the target's group label or food preference (e.g. in the category condition: 'This kid is a Zertle'; in the similarity condition: 'This kid likes to eat a food called zertles'). Next, the experimenter provided a statement about the target, which described him or her in relation to one of two test characters that appeared on the lower-left and -right sides of the screen (Figure 1b). The trial block determined which type of target statement and test question the experimenter presented: friend trials assessed which of the two characters participants thought the target was friends with; harm trials assessed which character participants thought the target would direct a harmful action towards; activity trials assessed which character participants thought the target shared a preference for an activity with (see Table S1). To indicate their response on each trial, participants had the opportunity to point to a character who either wore the target's same T-shirt colour or the other T-shirt colour. If the participant failed initially to provide a response, the experimenter prompted him or her to answer up to two additional times.

A score of '1' indicated that a participant selected the predicted test character, and a score of '0' indicated that he or she did not. For friend and activity trials, this was the character that wore the same T-shirt colour as the target; however, for harm trials, this was the character that wore the other T-shirt colour. We calculated a 'match

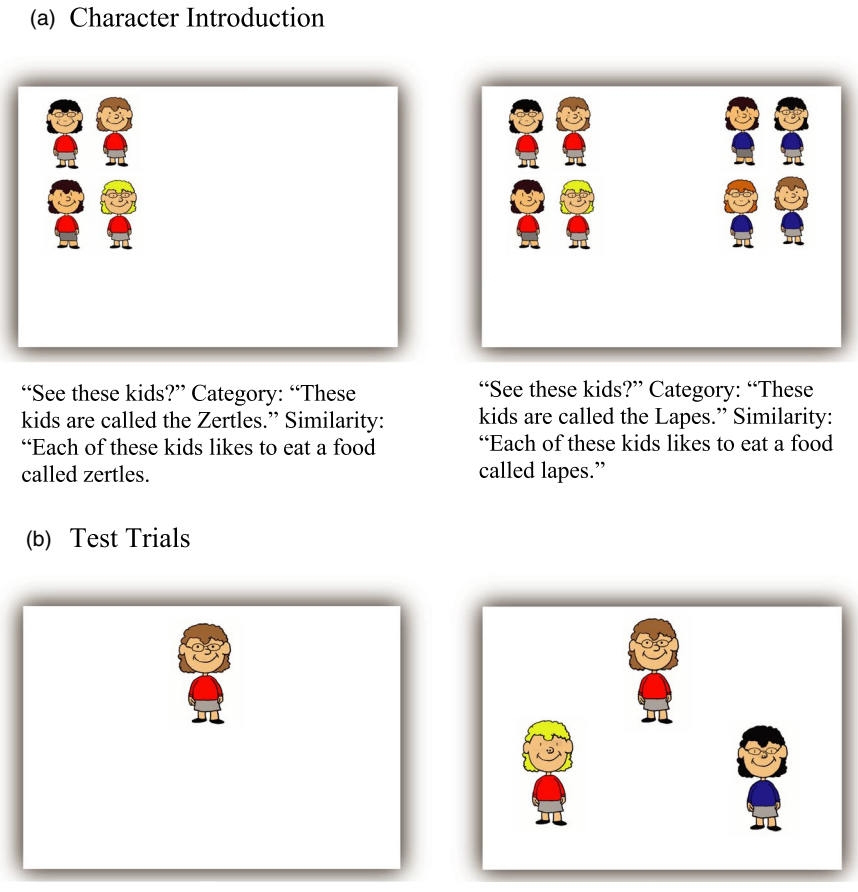

FIGURE 1 (a) Example displays from the character introduction phase in category and similarity conditions of Study 1. The experimenter pointed to each set of characters as she described them. (b) Example displays from the test trials

score' for each participant within each trial block. The minimum score a child could receive was 0 and the maximum score was 4. A score of 4 indicated that the child selected the predicted test character on each trial, while a score of 0 indicated that the child failed to select the predicted test character on each trial.

2.2 | Results

A score of 2 indicated chance performance. Since we compared each trial type to chance, we used the Bonferroni correction for multiple comparisons, which resulted in an adjusted alpha of 0.008. Children selected the predicted test character at above-chance rates for each of the three trial types in both conditions (all *p*s < .008).

We collapsed across the trial types to examine whether children's performance differed by condition. One-sample *t* tests revealed the youngest children, 3–4 years old, performed above chance in the category ($M = 2.64$), $t(83) = 4.58$, $p < .001$ and similarity ($M = 2.43$), $t(62) = 2.32$, $p = .024$ conditions, and their performance did not differ by condition, $t(145) = 0.94$, $p = .349$ (Figure 2). Children ages 5–6 years old performed above chance in the category ($M = 2.83$), $t(47) = 3.87$, $p < .001$ and similarity conditions ($M = 2.92$), $t(47) = 4.05$, $p < .001$, and their performance did not differ by condition, $t(94) = 0.27$, $p = .788$. Likewise, 7–9-year-old children performed above chance in the category ($M = 2.97$), $t(71) = 5.92$, $p < .001$ and similarity ($M = 3.31$), $t(71) = 8.85$, $p < .001$ conditions, and their performance did not differ by condition, $t(142) = 1.51$, $p = .133$.

We conducted a 2 (condition: category vs. similarity) × 3 (trial type: friend vs. harm vs. activity) x 3 (age group: 3–4 vs. 5–6 vs. 7–9 years) Analysis of Variance (ANOVA), and observed a main effect of age group, $F(2,123) = 5.53$, $p = .005$. Children in the oldest age group (7–9 years old) were more likely than children in the youngest age group (3–4 years old) to select test characters in the predicted direction, $t(289) = 3.72$, $p < .001$. Children in the oldest two age groups (5–6- and 7–9 years old) did not perform differently from each other, $t(238) = 1.42$, $p = .157$. There was also a

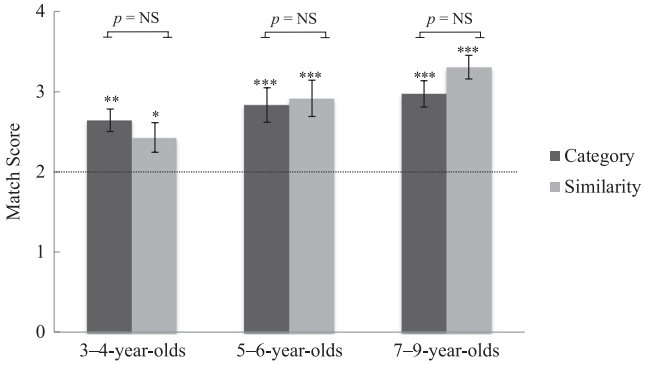

FIGURE 2 Mean match scores of 3–4-, 5–6-, and 7–9-year-olds in the category and similarity conditions of Study 1. Scores could range from 0 to 4, and the dotted line represents chance (2). Bars represent 1 *SEM* in either direction, and asterisks indicate that the means differ significantly from chance (***$p < .001$, **$p < .01$, *$p < .05$)

significant main effect of trial type, $F(1,123) = 8.41$, $p = .004$. On harm trials ($M = 2.54$) children were less likely to select test characters in the predicted direction than on friend trials ($M = 3.06$), $t(128) = 2.78$, $p = .006$, or activity trials ($M = 2.95$), $t(128) = 2.16$, $p = .033$. Surprisingly, children's performance did not differ based on the type of verbal information we provided: category ($M = 2.80$) or similarity ($M = 2.90$), $F(1,123) = 0.18$, $p = .674$, and their condition assignment did not interact with trial type, $F(1,123) = 0.85$, $p = .359$ or age, $F(2,123) = 1.14$, $p = .324$.

We observed a two-way interaction between age and trial type, $F(2,123) = 5.04$, $p = .008$. Children in the middle age group (5–6 years old) were more likely to select predicted test characters on friend trials ($M = 3.34$), $t(31) = 3.52$, $p = .001$ and activity trials ($M = 3.38$), $t(31) = 3.80$, $p < .001$ than on harm trials ($M = 1.91$). Additionally, children in the oldest age group (7–9 years old) selected predicted characters more often on friend trials ($M = 3.38$) than on activity trials ($M = 3.00$), $t(47) = 2.28$, $p = .027$, although their performance was above chance on all trial types (all *p*s < 0.001). We did not observe a three-way interaction between the factors, $F(2,123) = 1.92$, $p = .151$.

2.3 | Discussion

In Study 1 we predicted that children would make stronger social inferences in the category condition than in the similarity condition (Diesendruck & haLevi, 2006; Rhodes & Chalik, 2013; Taylor et al., 2009). However, we did not find evidence that children differentiated the two conditions. Instead, children across our age range relied on both category labels and shared preferences to make social inferences about interests and preferences in third-party cases, and did so to a statistically indistinguishable degree.

Although we observed a difference between the trial types among 5–6-year-old children, such that they selected the predicted character at higher rates on friend and activity trials than on harm trials, this may be due to the fact that selecting the predicted character required children to choose a test character that wore a different colour T-shirt than the target, and younger children may have been more inclined to select the 'colour match' instead. Overall, Study 1 demonstrates that when using standard intergroup measures children rely on information about shared interests and category labels to a similar extent when reasoning about others' relationships. The lack of category advantage is striking given that the measures we chose are highly social in nature and have previously been considered central to social category-based inference. However, examining a wider range of measures would bolster this conclusion, and we turn to this in Study 2.

3 | STUDY 2

In Study 2, we examined participants' inferences about relationships involving social and coalitional obligation, which we expected

to provide a stronger test of the impact of category labels on children's inferences (Kalish & Lawson, 2008; Rhodes & Chalik, 2013). That is, while our results from Study 1 are surprising, one could argue that our measures failed to tap into a more fundamental aspect of children's intergroup reasoning, specifically, that social groups serve the unique purpose of marking (a) which individuals are obligated to one another (Rhodes & Chalik, 2013), and (b) what members of a common social group are obliged to do (Kalish & Lawson, 2008). To address this concern, in Study 2 we examined whether children use cues to similarity as strongly as they use cues to category membership when reasoning about scenarios that are explicitly coalitional, in a broad sense, relating to patterns of intragroup obligation.

This study differed from Study 1 in two critical respects: First, turning our focus from social preferences, we tested social obligations, such as taking responsibility for another's actions or defending one against negative outcomes, rather than merely assessing friendship relations or additional shared interests. Second, we elected to define similarity as a shared toy preference (Similarity-toy condition) in addition to a shared food preference (Similarity-food condition), since work on the early emergence of food selection reasoning demonstrates that, from infancy, group membership constrains generalizations about others' food preferences (Liberman, Woodward, Sullivan, & Kinzler, 2016). This work suggests that cues to food preference carry more weight in the social domain, which may have inflated our results in the similarity condition of Study 1.

3.1 | Method

3.1.1 | Participants

The participants were a new group of 205 3–9-year-old children (97 females; M_{age} = 6.49 years; range = 3.07–9.98 years) from three age groups: 3–4 (n = 62), 5–6 (n = 58) and 7–9 (n = 85) years of age. We tested an additional nine children who were excluded from analyses because they failed to complete the task (n = 8: similarity-food condition, n = 2; similarity-toy condition, n = 2; category condition, n = 4) or because their parent or legal guardian did not provide written informed consent (n = 1, similarity-toy condition). Data collection took place from spring to late fall of 2016. The study took place in either a university laboratory (n = 47), a children's museum (n = 80) or an empty classroom at the participant's school (n = 78).

3.1.2 | Design and materials

The design and materials differed from Study 1 only in the following respects: We randomly assigned participants to either the category (n = 73), similarity-food (n = 66) or similarity-toy (n = 66) condition. The task consisted of four test trial blocks, each with four trials of the following types: 'obligation', 'responsibility', 'defence' and 'harm'. We elected to test harm trials again to allow for a direct comparison between Studies 1 and 2 and because this has been a central measure in exploring children's category-based coalitional reasoning (Chalik et al., 2014; Rhodes & Chalik, 2013). Additionally, because the prediction for harm trials involves selecting the *non-matching* character, they also provide a check on the possibility that children are merely making colour matches based on the characters' appearance. Finally, we used Photoshop to generate 24 additional, unique test characters; this was due to the new requirements created by the addition of a test trial block.

3.1.3 | Procedure and scoring

The procedure for this study differed from Study 1 only in the following respects: In the similarity-toy condition, the experimenter told participants that each character in the set liked to play with the same novel-labelled toy. She said of one set while pointing to the characters: 'See these kids? Each of them likes to play with a toy called *zertles'*. And she said of the other set while pointing: 'See these kids? Each of them likes to play with a toy called *lapes'*. Each test trial in the similarity-toy condition began with the experimenter reminding the participant which of the two toys the target preferred based on their T-shirt colour (e.g. 'See this kid? This kid likes to play with a toy called zertles'). The experimenter then presented one of the following trials: Obligation trials assessed which character participants thought was obligated to complete the same action as the target; responsibility trials assessed which character participants thought would take responsibility for the target's negative action; defence trials assessed which character participants thought would stop another individual from committing a harmful action against the target; harm trials did not differ from Study 1 (see Table S1).

We scored participants' performance in the same manner as in Study 1, with children receiving a score of '1' for selecting the test character in the same coloured T-shirt as the target on obligation, responsibility and defence trials, and the test character in the other coloured T-shirt on harm trials; participants received a score of '0' for providing the opposite responses respectively. As in Study 1, a participant's match score could range from 0 to 4.

3.2 | Results

We used one-sample t tests to assess whether children performed at above-chance levels (chance = 2), and the Bonferroni correction for multiple comparisons, which resulted in an adjusted alpha of 0.004. With the exception of responsibility trials in the similarity-toy condition (M = 2.39), $t(65)$ = 2.02, p = .048, children selected the predicted test character at above-chance levels for each of the four trial types in each condition (all $ps < .004$).

We collapsed across the trial types to examine whether children's performance differed by condition. One-sample t tests revealed that the youngest children, 3–4 years old, performed above chance in the category condition only (M = 2.42), $t(95)$ = 2.94,

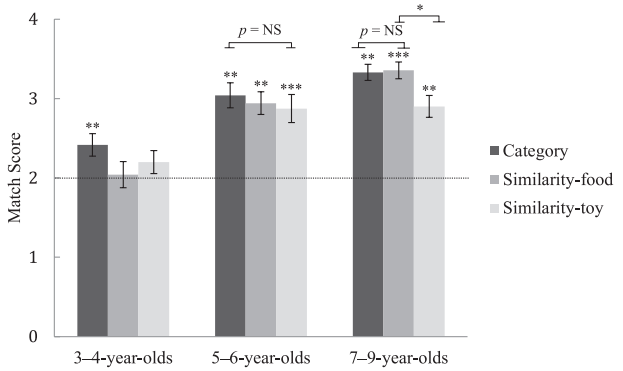

FIGURE 3 Mean match scores of 3–4-, 5–6-, and 7–9-year-olds in the category, similarity-food, and similarity-toy conditions of Study 2. Scores could range from 0 to 4, the dotted line represents chance (2), and bars represent 1 *SEM* in either direction. Asterisks indicate that 3–4-year-olds scored above chance in the category condition (**$p < .01$), and the older age groups did so in all conditions (***$p < .001$); 7–9-year-olds scored higher in the similarity-food than similarity-toy condition (*$p < .05$), and their performance did not differ between the category and similarity-food conditions (p = NS)

$p = .004$ (Figure 3). Their performance did not differ from chance in the similarity-food ($M = 2.04$), $t(71) = 0.25$, $p = .801$ or similarity-toy ($M = 2.20$), $t(79) = 1.38$, $p = .172$ conditions. Despite these differences in significance, their performance did not differ by condition ($ps > .05$). Children ages 5–6 years old performed above chance in the category ($M = 3.04$), $t(71) = 6.61$, $p < .001$, similarity-food ($M = 2.94$), $t(87) = 6.59$, $p < .001$, and similarity-toy ($M = 2.88$), $t(71) = 4.95$, $p < .001$ conditions, and their performance did not differ by condition ($ps > .05$). The oldest children, 7–9 years old, performed above chance in the category ($M = 3.33$), $t(123) = 13.10$, $p < .001$, similarity-food ($M = 3.36$), $t(103) = 12.82$, $p < .001$, and similarity-toy ($M = 2.92$), $t(107) = 6.47$, $p < .001$ conditions. Moreover, they scored higher in the similarity-food condition than in the similarity-toy condition, $t(210) = 2.47$, $p = .014$. As in Study 1, the oldest children's performance did not differ between the category and similarity-food conditions, $t(226) = 0.17$, $p = .865$.

We conducted a 3 (condition: category vs. similarity-food vs. similarity-toy) × 4 (trial type: obligation vs. responsibility vs. defence vs. harm) × 3 (age group: 3–4 vs. 5–6 vs. 7–9 years) ANOVA. We observed a main effect of age group, $F(2,196) = 25.69$, $p < .001$. Children in the oldest age group (7–9 years old) ($M = 3.20$) selected test characters at higher rates than children in the middle (5–6 years old) ($M = 2.95$), $t(478) = 5.68$, $p < .001$ and youngest (3–4 years old) ($M = 2.24$), $t(586) = 8.84$, $p < .001$ age groups. We did not obtain main effects of trial type, $F(1,196) = 3.14$, $p = .078$ or condition, $F(2,196) = 1.80$, $p = .168$, or an interaction between any of the factors (all $ps > .05$).

3.3 | Discussion

The results of Study 2 again suggest that children overall did not differentiate between the two types of verbal information we provided them with—category labels and similar preferences. While there was a hint that the youngest age group may have been more impacted by category cues, any such differences were not supported by direct comparisons across conditions. Thus, by at least 5 years of age, if not before, children use both social categories and shared preferences to infer interpersonal obligation and coalitional structure.

We observed two slight differences in children's patterns of performance worth noting. First, the oldest children in the sample showed sensitivity to the manner in which we defined similarity. Specifically, when we defined similarity as a shared food preference it carried the same weight as category labels, however, when we defined it as a toy preference it carried somewhat less weight. Second, the youngest children performed above chance in the category condition, but their performance did not differ from children assigned to either of the similarity conditions. It is important to note, however, that these results did not emerge from the larger model, suggesting the need for caution in interpreting differences across trial types and between conditions.

It is striking that across most of our age range children made at best weak distinctions between the three conditions. That is to say, children treated information about shared tastes (for both food and toys) as central to understanding social alliances, using it as the basis for inferences about richly coalitional actions such as harming others and providing assistance. While their scores are more extreme, the oldest children's patterns of performance look quite similar to the middle group. They relied on each type of information to predict the deontic social relationships of others. However, not all shared tastes are equally useful; rather, some commonalities like food preferences may signal coalitional structure more powerfully than others.

4 | STUDY 3

Study 3 addressed one potentially deflationary worry about the method used in Studies 1 and 2, namely the effect of our visual stimuli on performance. That is, in both the category and shared interest conditions in those studies, spatial proximity and shared T-shirt colour provided an additional cue beyond the verbal information that constituted the difference between conditions. If these cues were powerful enough, they could have driven children's performance in Studies 1 and 2, potentially drowning out what we considered our primary manipulation of information type. Of course, those studies already contain data that speak against this possibility, specifically performance on the harm trials, in which children generally selected the *non-matching* individual as the one the target directed a harmful action towards. This demonstrates that children are not *merely* selecting the matching character. Still, perhaps in the presence of visual similarities the verbal information about category membership or preferences simply did not contribute to children's performance; that would represent a quite different interpretation of the results of Studies 1 and 2 that we felt the need to rule out (or in). To explore this, we reasoned as follows: If children relied merely on these visual cues to licence their socially rich judgments about peer and activity

preferences (Study 1) and deontic relationships (Study 2), then we would expect to see them perform similarly in a baseline condition that excludes information about category membership and shared interests but retains the same visual and spatial cues. To test this, we recruited a new group of 7–9-year-old children to respond to the dependent measures we tested in Study 1. This age group provides the best test because it provided the strongest rates of generalizing in Studies 1 and 2.

4.1 | Method

4.1.1 | Participants

The participants were a new group of 72 7–9-year-old children (36 females; *M* age = 8.39 years; range = 7.02–9.95 years). Data collection took place from late winter to spring of 2017. The study took place in either a university laboratory ($n = 8$), a children's museum ($n = 63$) or an empty classroom at the participant's afterschool program ($n = 1$).

4.1.2 | Design and materials

The design and materials were identical to those of Study 1 with the exception that we assigned each child to the Baseline condition. The task consisted of three test trial blocks, each with four trials of the following types: friend, activity and harm, as in Study 1.

4.1.3 | Procedure and scoring

Using the procedure from Study 1 we assessed baseline performance when neither type of verbal information—category membership or shared preferences—described the character sets. This allowed us to rule out low-level explanations for the results we obtained in Studies 1 and 2, such as a mere reliance on clothing colour. Instead of telling children that the introduction characters were members of a labelled social group, or that they all shared a common preference, an experimenter pointed to the stimuli and simply said: 'Do you see these kids?' Afterward, she asked children to respond to the dependent measures from Study 1, except instead of noting the target's group membership or food preference first, she simply said, 'See this kid?' prior to revealing the two test characters on the bottom half of the screen.

We scored performance in the same manner as in Studies 1 and 2 with a participant's match score ranging from 0 to 4.

4.2 | Results

We compared performance on each trial type to chance (chance = 2) using the Bonferroni correction for multiple comparisons, which resulted in an adjusted alpha of 0.01. One-sample *t* tests revealed that children performed above chance on friend trials only ($M = 2.46$), $t(71) = 2.68$, $p = .009$. On harm ($M = 2.39$) and activity ($M = 2.38$) trials children's performance did not differ from chance ($ps > .01$). A one-way ANOVA on the three trial types revealed no difference in performance, $F(1,71) = 0.08$, $p = .784$.

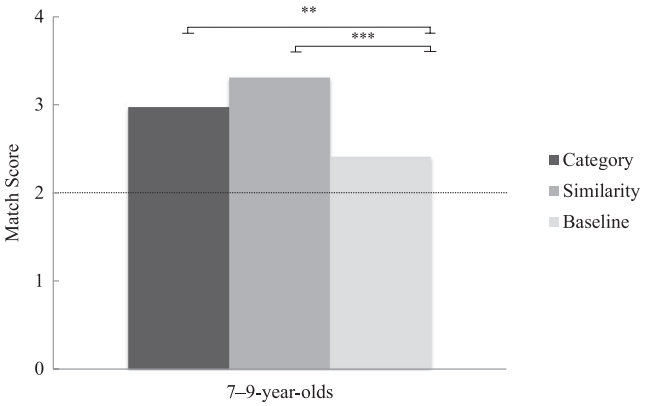

FIGURE 4 Mean match scores of 7–9-year-olds in the baseline (Study 3), category, and similarity (Study 1) conditions. Scores could range from 0 to 4, and the dotted line represents chance (2). Bars represent 1 *SEM* in either direction, and asterisks indicate that 7–9-year-olds selected predicted test characters at higher rates in the category (**$p < .01$) and similarity (***$p < .001$) conditions than in the baseline condition

To determine the extent to which the verbal information we provided drove children's performance in the first two studies, we compared children's baseline performance to the same age group's performance in the category and similarity conditions of Study 1. A 3 (condition: category vs. similarity vs. baseline) × 3 (trial type: friend vs. activity vs. harm) ANOVA revealed a main effect of condition, $F(2,117) = 10.20$, $p < .001$ (Figure 4). Children in the category condition ($M = 2.97$) selected the predicted test character at significantly higher rates than children in the baseline condition ($M = 2.41$), $t(286) = 2.95$, $p = .003$. Likewise, children in the similarity condition ($M = 3.31$) selected the predicted test character at significantly higher rates than children in the baseline condition, $t(286) = 4.81$, $p < .001$. There was no main effect of trial type, $F(1,117) = 1.31$, $p = .255$, and no interaction between the factors, $F(2,117) = 0.64$, $p = .531$.

4.3 | Discussion

The results of Study 3 establish a baseline level of performance in our task. Children responded to each of three dependent measures from Study 1 after viewing identical displays, but children were not given either type of verbal information to base their judgments on. Only on friend trials did children perform above chance, and it is important to note that their performance did not differ significantly between friend trials and the other trial types. This pattern of results suggests that the spatial and visual information present in our study was at best a weak cue to the presence of commonalities among the characters.

The central goal of this study was to rule out performance explanations in the previous studies based solely on low-level cues. And here the results were clear: A comparison of children's performance in Study 1 to these results revealed that the verbal information we provided had a large and consistent impact on children's judgments over and above the visual cues. Thus, a low-level reliance on those cues cannot explain the powerful impact of information about either shared interests or category membership. However, it is still possible that children only relied on shared interest cues in the category and similarity conditions because we presented them in isolation. To address this concern, in Study 4 we directly pit the two types of cue against one another in order to provide the most stringent test of their relative strength.

5 | STUDY 4

Following the approach of Diesendruck & haLevi, 2006, we pitted our two cues of interest against one another in a preregistered test of their relative efficacy. To assess a developmental trajectory, we chose to test the youngest and oldest age groups from Studies 1 and 2, because the middle age group patterned closely to the oldest children in the previous studies. We predicted that if children privileged category labels over shared interests they would perform above chance in the pitted case (preregistration: http://aspredicted.org/blind.php?x=6ti3xg). Moreover, based on our data from Studies 1 and 2, in which young children's performance in the category condition was slightly higher, and older children's performance in the similarity condition was slightly higher, we predicted that in this more direct comparison young children would show a category bias, and older children would show a similarity bias. Thus, we asked children to indicate which of two test characters matched a target character on a combination of social preference and coalitional dependent measures from the previous studies.

5.1 | Method

5.1.1 | Participants

The participants were 51 children (n = 23 female) from two age groups: 3–4- (n = 25; M age = 3.92; range = 3.12–4.83) and 7–9- (n = 26; M age = 7.88; range = 7.21–9.87) years old. We tested 15 additional children who we excluded from analyses due to the following: experimenter error (n = 3), failure to complete the task (n = 3) or failure to pass the comprehension checks (n = 9). Data collection took place from early fall to mid-winter of 2018. The study took place in either a university laboratory (n = 13), a children's museum (n = 28), or an empty classroom at the participant's school (n = 10).

5.1.2 | Design and materials

The task consisted of four test trial blocks, each with four trials of the following types: harm, friend, responsibility and defence. The design and materials differed from the previous studies in several key respects: We signalled category membership and shared preferences such that the cues were distinct and of relatively equal strength. In doing so, we represented category labels with differently coloured flags and food preference with differently coloured lunch boxes; the two cues were approximately equal in size as we aimed for roughly equal signalling strength in terms of lower-level visual salience. Because T-shirt colour no longer served as a cue to category membership or shared preference, we depicted each character in a white T-shirt.

We counterbalanced the colours of the flags and novel foods (either red and blue or green and orange), the verbal labels assigned to each category and novel food (either *zertles* and *lapes*, or '*hoopas*' and '*flurps*'), the order of the trial types, and the order in which we presented category and shared preference information in the training and test phases.

5.1.3 | Procedure and scoring

The procedure and scoring differed from the previous studies in the following ways: In the introduction phase, the experimenter guided the participant through a thorough training in which they: (a) learned the labels associated with each flag or lunchbox; (b) met the group of kids associated with each label and visual cue; and (c) answered a set of comprehension questions followed by corrective feedback. For example, she said of one set of items while pointing: 'See these flags? These flags are for kids called *Hoopas'*. She then pointed to the other set and said: 'And see these flags? These flags are for kids called *Flurps'*. After presenting the flags or the lunchboxes, the experimenter presented the items again, and asks the participant, for example: 'Now, can you tell me who these flags are for?'

Next, the experimenter introduced the child to two sets of four introduction characters, one on each side of the screen, who either held differently coloured flags representing their category membership or lunchboxes with pictures of food representing their preference. She said of one set of characters while pointing, for example: 'See these kids? These kids are all called *Hoopas'*. And of the other set: 'And see these kids? These kids are all called *Flurps'*. After presenting each set of characters with their respective category membership or food preference, the experimenter presented the same characters again, and asked the participant: 'Now, can you tell me what these kids are called?'

After learning about both cue types, children viewed each set of characters with their flags held in one hand and their lunchboxes in the other, and the experimenter stated each group's label and food preference once more. Next, the experimenter presented the participant with a pair of laminated cards containing pictures of either the flags or the foods. She then presented the two sets of characters with the *other* dimension depicted, and asked, for example: 'Using these cards, can you show me what the kids like to eat?' and instructed the participant to match the cards to the characters. She then repeated this step with the other dimension. If participants

failed to match correctly in both cases, their data were subsequently excluded from additional analyses, as we wanted to ensure that participants understood the category and preference pairings for each character set prior to the test phase.

Each test trial began with the experimenter directing the participant's attention to the test characters, in turn, to remind the participant of the characters' group labels and food preferences (see Figure 5). Then the experimenter presented a 'mystery kid' partially occluded by a grey box with a question mark on it. She revealed the kid's flag and lunchbox and said, for example, while pointing: 'See this mystery kid? This kid is called a *Zertle* like her (while pointing to the corresponding test character), and likes to eat a food called *hoopas* like her (while pointing the other test character)'. Then, the experimenter asked one of the four test questions. After completing all of the test questions children repeated the comprehension matching procedure from the introduction phase.

We coded the data as follows: for friend, defence, and responsibility trials, a score of '1' indicated that a participant selected the test character who shared the target's category label, and a score of '−1' indicated that the participant selected the test character who shared the target's food preference. As in the previous studies, we reverse scored harm trials. We calculated an average bias score for participants within each trial block, and scores could range from 4 to −4.

5.2 | Results

We compared children's average scores within each trial block to chance (chance = 0) using the Bonferroni correction for multiple comparisons, which resulted in an adjusted alpha of 0.0125. We preregistered the following analyses: Two-tailed *t* tests revealed that 3–4-year-old children did not select the category- or preference-biased test character at above-chance rates for any of the four trial types: harm ($M = -0.12$), friend ($M = -0.10$), responsibility ($M = 0.02$) and defence ($M = -0.02$) (all $ps > .265$). However, 7–9-year-old children selected the category-biased character at above-chance rates on defence ($M = 0.40$) trials, $t(25) = 3.25$, $p = .003$, and marginally so on friend ($M = 0.37$) trials, $t(25) = 2.56$, $p = .017$; their performance on harm and responsibility ($Ms = 0.27$) trials did not differ significantly from chance ($ps = .115$, and .119, respectively).

We conducted a 2 (Age: 3–4 years old vs. 7–9 years old) × 4 (Trial type: harm vs. friend vs. responsibility vs. defence) ANOVA, and observed a main effect of age, $F(1,196) = 15.48$, $p < .001$. Older children ($M = 0.33$) were more likely than younger children ($M = -0.06$) to select category-biased test characters. We did not observe a main effect of trial type or an interaction between the factors ($ps = .863$ and .873 respectively) (Figure 6).

Seven 3–4 years old and one 9 years old did not pass the final comprehension matching task. Although we did not initially plan to exclude participants who failed the final comprehension check from our main analyses, we decided to explore the possibility that younger children may have failed to demonstrate a bias in either direction due to their failure to retain which cues we paired with which throughout the study. However, we did not find evidence for this possibility; older children ($M = 0.30$) were still more likely than younger children ($M = -0.10$) to select category-biased characters, $F(1, 164) = 15.87$, $p = .001$, and the younger children's performance did not differ from chance even when excluding those who failed the final comprehension check ($p = .122$).

5.3 | Discussion

In Study 4 we conducted an even stronger test of whether labelled categories are privileged in children's group reasoning relative to other types of similarity, such as shared interest. We directly pitted cues to category membership, a shared labelled and distinct flag, against cues to common interest, a shared food preference and distinct lunchbox. This revealed a striking developmental shift: Young children did not preferentially base their inferences on categories more than shared interests. This finding is in line with our conclusions from Studies 1 and 2 that young children fail to distinguish between the two cues in a matched design that holds perceptual similarity constant. And although older children did privilege categorical cues to infer group structure, particularly coalitional defence, even here the difference was not dramatic.

One possible reason that older children preferentially selected category-biased characters is because, by at least 5 years of age, children assume that social category labels carry rich explanatory power (Taylor et al., 2009). This essentialized notion of category

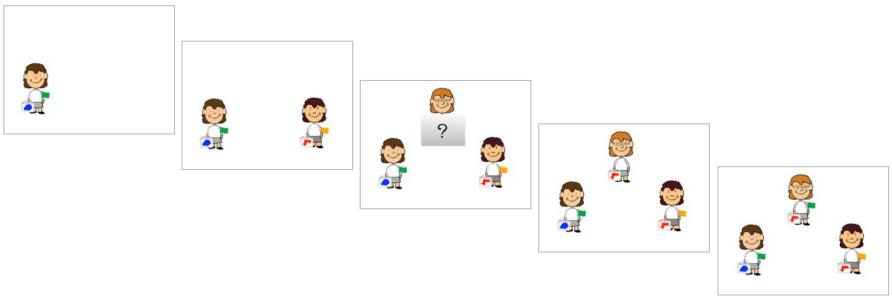

FIGURE 5 Example displays from the test trials of Study 4. The experimenter said: "This kid likes to eat flurps and is called a Zertle [left], and this kid likes to eat hoopas and is called a Lape [right]. Now, see this mystery kid [center]? She likes to eat hoopas like her [points right], and is called a Zertle like her [points left]." Each test question followed this type of prompt

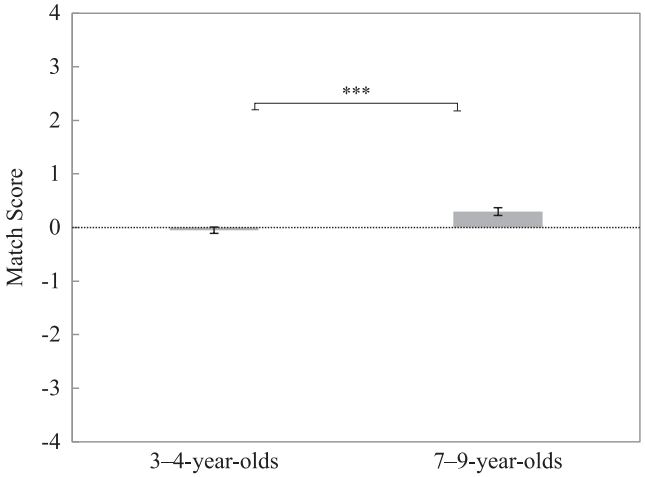

FIGURE 6 Mean match scores of 3–4- and 7–9-year-olds in the pit condition of Study 4. Scores could range from −4 to 4, positive values indicate category bias, negative values indicate similarity bias, and 0 = chance. Bars represent 1 *SEM* in either direction, and asterisks indicate that 7–9-year-olds selected category-biased test characters more often than 3–4-year-olds (***$p < .001$)

labels may have supported children's inferences to a greater extent in the direction of category bias. This may also explain why children made the strongest inferences on trials that required them to predict which character would defend the target individual, a hallmark of rich coalitional reasoning.

6 | GENERAL DISCUSSION

The studies presented here used a novel groups paradigm to test children's third-person social inferences based on two types of verbal information, labelled categories and shared preferences. Based on both past theorizing and past empirical work, we expected to observe a robust advantage for labelled categories over shared interests. However, to our surprise, these studies indicate that children generally rely on both types of information to make predictions about others' activity and peer preferences (Study 1) and deontic relationship expectations (Study 2). Critically, these effects cannot be explained solely via low-level visual cues to group membership, such as spatial proximity and T-shirt colour (Study 3). Only when we directly pitted these cues against one another did we observe a category bias, but only a modest such bias, and only among older children (Study 4).

In general, children who received information about sharing interests made social inferences at comparable rates to children who received information about belonging to a labelled category. This suggests that (a) cues to interpersonal similarities carry more weight than past literature proposes, and (b) category labels are sufficient, but not necessary to support children's categorical inferences. These findings also dovetail with recent work showing that shared preferences increase children's friendship judgments and resource allocations at comparable rates to shared membership in a minimally assigned group (Sparks et al., 2017).

Previous studies argue that noun-labelled social categories uniquely cue children into the complex structure of the social world. However, the studies here paint a more nuanced picture. Our data suggest that children are equally sensitive to other types of socially meaningful information, such as shared interests; this is true even in cases where children are asked to make inferences about coalitional structure, something previously considered the hallmark of category-based reasoning about social kinds. Perhaps children consider shared category membership one of several important types of interpersonal similarity that signal coalitional relationships. If so, this suggests that prior work, which argues for a category advantage, fails to provide a comprehensive account of children's inductive reasoning in the social domain. Here we propose that personal preferences are more central to children's representations of social groups than prior work suggests.

It is worth noting that although Study 2 revealed that interpersonal similarities, regardless of type, support children's group-based inferences by age 5, our data show that by age 7 children are sensitive to the type of shared preference cue. As other work suggests (Liberman et al., 2014, 2016), children may have prioritized information about food preferences over object preferences due to the social potency of food selection information. However, it is striking that we only observed this difference among the oldest children in our sample (7–9 years old). One possibility is that children come to value shared food preference more as they gain opportunities to select and prepare foods that they prefer. Another possibility is that they become increasingly aware of the cultural salience of food preference and its link to richer forms of social and cultural identity.

We acknowledge that in some instances children showed a slight category advantage. For example, the youngest children in Study 2 performed at above-chance levels when they received group labels, but not when they received shared preference information. Moreover, the oldest children in Study 2 made stronger inferences in the category condition as compared to the similarity-toy condition (although this difference did not emerge when we compared it to the similarity-food condition). Thus, while there was some scattered evidence for a category advantage at specific ages, and under specific conditions, any such differences were small and, in most cases, did not emerge in a combined statistical model. Thus, on balance we argue that the overall pattern of data here is more consistent with rough equality between category and shared preference cues, particularly among younger children, although of course, we interpret these null effects cautiously. At the very least we can safely conclude that in paradigms like these, labelled categories and shared preferences induce patterns of judgment that are highly similar in magnitude and that are not likely to be distinguished by typical developmental sample sizes.

These results mark the first direct test comparing children's reliance on labelled categories and interpersonal similarities when reasoning specifically in the social domain. Prior work on familiar social groups, as well as natural and artefact kinds, points to a category advantage, but this conclusion may be unwarranted with regard to children's social reasoning for several reasons. First, children's

performance in such tasks may be inflated by the salience of and the specific social knowledge children have about familiar social groups (e.g. specific content knowledge that children in Israel have about the importance of religion and/or ethnicity; Diesendruck & haLevi, 2006). Second, studies of natural and artefact kinds have generally supported a category advantage hypothesis, but those results may have been too hastily extended to children's reasoning about social partners.

Finally, to the best of our knowledge, none of the studies to date directly compared category and shared interest information in a matched design, nor have they pitted the two cues against one another. It was this approach that here allowed us to assess the relative strength of each dimension. When we presented these two critical types of information in isolation, children failed to privilege one type over the other. And only when we directly pitted these cues against each other did we observe a modest category bias among older children. These findings challenge classic notions of a category label advantage by suggesting that interpersonal similarities are equally central to children's emerging understanding of social groups.

Most critically, shared interests not only support children's inferences about other likely commonalities, such as shared activity preferences, they also support deeper coalitional inferences, like predicting who will take responsibility for another's antisocial actions. This raises further questions about the functional role of shared interests in children's abstract reasoning about social categories. One possibility is that young children regard interpersonal similarities in and of themselves as important indicators of social connectedness. Yet another possibility is that children come to see those interpersonal similarities as a means by which social groups are formed. This second possibility is reasonable given vast literature on minimal groups, which shows that children can formulate rich social judgments on the basis of modest cues to group membership (e.g. Baron & Dunham, 2015; Dunham et al., 2011; Dunham & Emory, 2014). Indeed, our similarity manipulation was even more pronounced. This interpretation suggests that children in our similarity conditions used the presence of shared interests (in addition to the visual cues) to infer the presence of social categories themselves. However, even if this is right, the fact that shared interest cues then drive children's inferences so strongly is striking, and it speaks against the view that labels provide a privileged link to social categorization in children. Furthermore, the data from the shared preference conditions speak to the robustness of children's minimal-group inferences in the third party. Nonetheless, future work should examine whether these interpretations are in fact distinct, and if so, which more accurately reflects children's understanding of interpersonal similarities.

Of course, future work, for example that pits shared category membership against shared interests when other visual cues are absent, or that explores other dimensions of similarity, might reveal a category advantage. Still, the present findings support that non-categorical information is a strikingly powerful elicitor of social inference in children, and thus such a result could not undercut the power of shared preferences that we observe here. This conclusion is especially relevant to the many contexts in which similarity information is present, but category labels are not. Future theorizing will need to integrate the powerful role that other non-categorical shared properties may have in guiding children's social inferences.

ACKNOWLEDGEMENT

The authors thank members of the Yale Social Cognitive Development lab, including Bianca Li for stimuli creation, data collection and data entry, Carleen Liu, Jaiqi Liu and Maria Maier for their assistance with data collection, and Allison Bradshaw, Salima Clark, Shirley Duong and Karen Yao for their assistance with data entry. We thank the schools and museums that served as testing sites, including Alcott School, Alphabet Academy, The Connecticut Science Center, Long Ridge School, St. Rita School, and the Yale Peabody Museum of Natural History, along with the children, parents, teachers and staff who participated. Finally, we thank two anonymous reviewers for their useful feedback. This work was generously supported by the John Templeton Foundation, Grant No. 56036.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on the Open Science Framework at https://osf.io/zwfym/.

REFERENCES

Baron, A. S., & Dunham, Y. (2015). Representing 'Us' and 'Them': Building blocks of intergroup cognition. *Journal of Cognition and Development*, *16*(5), 780–801. https://doi.org/10.1080/15248372.2014.1000459

Baron, A. S., Dunham, Y., Banaji, M., & Carey, S. (2014). Constraints on the acquisition of social category concepts. *Journal of Cognition and Development*, *15*(2), 238–268. https://doi.org/10.1080/15248372.2012.742902

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

Chalik, L., Rivera, C., & Rhodes, M. (2014). Children's use of categories and mental states to predict social behavior. *Developmental Psychology*, *50*(10), 2360. https://doi.org/10.1037/a0037729

Cimpian, A. (2016). The privileged status of category representations in early development. *Child Development Perspectives*, *10*(2), 99–104. https://doi.org/10.1111/cdep.12166

Cimpian, A., & Erickson, L. C. (2012). Remembering kinds: New evidence that categories are privileged in children's thinking. *Cognitive Psychology*, *64*(3), 161–185. https://doi.org/10.1016/j.cogpsych.2011.11.002

Dewar, K., & Xu, F. (2009). Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, *20*(2), 252–257. https://doi.org/10.1111/j.1467-9280.2009.02278.x

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, *77*, 539–553. https://doi.org/10.1111/j.1467-8624.2006.00889.x

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780–793. https://doi.org/10.1016/j.tics.2018.06.004

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, *82*(3), 793–811. https://doi.org/10.1111/j.1467-8624.2011.01577.x

Dunham, Y., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, *70*(1), 81–98. https://doi.org/10.1111/josi.12048

Fawcett, C. A., & Markson, L. (2010). Similarity predicts liking in 3-year-old children. *Journal of Experimental Child Psychology*, *105*(4), 345–358. https://doi.org/10.1016/j.jecp.2009.12.002

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *396–404*, https://doi.org/10.2307/1130595

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. https://doi.org/10.1016/0010-0277(86)90034-X

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science*, *24*(4), 589–594. https://doi.org/10.1177/0956797612457785

Kalish, C. W. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology*, *48*(4), 1133. https://doi.org/10.1037/a0026344

Kalish, C., & Lawson, C. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development*, *79*, 577–593. https://doi.org/10.1111/j.1467-8624.2008.01144.x

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by12-month-olds. *Psychological Science*, *14*, 402–408. https://doi.org/10.1111/1467-9280.01454

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–15392. https://doi.org/10.1073/pnas.251541498

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2014). Friends or foes: Infants use shared evaluations to infer others' social relationships. *Journal of Experimental Psychology: General*, *143*(3), 966–971. https://doi.org/10.1037/a0034481

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). Preverbal infants infer third-party social relationships based on language. *Cognitive Science*, *41*(S3), 622–634. https://doi.org/10.1111/cogs.12403

Liberman, Z., Woodward, A. L., Sullivan, K. R., & Kinzler, K. D. (2016). Early emerging system for reasoning about the social nature of food. *Proceedings of the National Academy of Sciences*, *113*(34), 9480–9485. https://doi.org/10.1073/pnas.1605456113

Mahajan, N., & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, *124*(2), 227–233. https://doi.org/10.1016/j.cognition.2012.05.003

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition*, *59*(3), 307–335. https://doi.org/10.1016/0010-0277(95)00696-6

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*(4), 847–860. https://doi.org/10.1111/j.1467-8624.2006.00906.x

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*(6), 999–1006. https://doi.org/10.1177/0956797612466267

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*(1), 35–62. https://doi.org/10.1080/15248372.2011.638686

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246–251. https://doi.org/10.1016/S1364-6613(03)00109-8

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166. https://doi.org/10.1037/0096-3445.133.2.166

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*(2), 179–185. https://doi.org/10.1111/j.1467-9280.2007.01869.x

Sloutsky, V. M., Lo, Y. F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development*, *72*(6), 1695–1709. https://doi.org/10.1111/1467-8624.00373

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*(2), 143–171. https://doi.org/10.1016/0010-0277(96)00709-3

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, *38*(2), 179–211. https://doi.org/10.1016/0010-0277(91)90051-5

Sparks, E., Schinkel, M. G., & Moore, C. (2017). Affiliation affects generosity in young children: The roles of minimal group membership and shared interests. *Journal of Experimental Child Psychology*, *159*, 242–262. https://doi.org/10.1016/j.jecp.2017.02.007

Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys; cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development*, *80*(2), 461–481. https://doi.org/10.1111/j.1467-8624.2009.01272.x

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250. https://doi.org/10.1016/S0010-0277(02)00109-9

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Jordan A, Dunham Y. Are category labels primary? Children use similarities to reason about social groups. *Dev Sci*. 2021;24:e13013. https://doi.org/10.1111/desc.13013