Beyond Discrete Categories:

Young Children Fail to Privilege Categories when Shared Preferences Compete

Bianca B. Li

Advised by Yarrow Dunham, Assistant Professor of Psychology, and Ashley E. Jordan, Ph.D. Candidate in Psychology.

Submitted to the faculty of Cognitive Science, in partial fulfillment of the requirements of the Bachelor of Science degree.

Yale University

April 22, 2019

TABLE OF CONTENTS

Abstract

Children apprehend the social world by dividing it into discrete categories. They also derive inferences about others' relationships based on shared preference information. The present work attempts to discern whether children, across two age groups (3–4 and 7–9 years old), privilege information about category membership over shared preferences when inferring friendship, intergroup obligation, and intragroup harm among third-parties. By pitting category labels against preferences, this study revealed that younger children did not privilege categorical information over shared preference information. Older children privileged categorical information when the two dimensions were directly pitted against one another; however, the strength of their inferences did not differ from a no-information baseline in either the category or shared preference direction. These findings confirm earlier research conducted on the explanatory power of social categories and shared preferences.

1. Introduction

"If you're a singer and you're Black, you're an R&B artist. Period."

"When I first released music and no one knew what I looked like, I would read comments like, 'I've never heard anything like this before, it's not in a genre.' And then my picture came out six months later – now she's an R&B singer."

These have been the experiences of Black musicians Frank Ocean and FKA Twigs, respectively, whose works span genres and resist classification within a singular musical style. Still, music critics and listeners alike quickly pigeonholed Ocean, FKA Twigs, and other Black musicians into the category of "R&B musician" because of their race, sometimes without careful consideration of their musical style (Younger, 2017). People eagerly divide the social world into discrete categories, which often leads them to draw inferences about individuals based on the groups to which they belong. Sometimes these inferences are based on social categories, like race and gender, and other times they are based on mental states, like preferences and interests. But how do these two types of social information become incorporated into children's developing social sense? How do they emerge? The present work sheds light on this question by empirically assessing how children derive social inferences from these cues early on, and how their inferences change over the course of development.

1.1 Social Categories

Children use various kinds of social information to learn about others. Social category membership, such as language, ethnicity, gender, and race, has emerged as a primary kind of information that children use to make decisions about how a person will generally behave (Gelman & Markman, 1986). Even abstract social category labels, like novel groups with fictional names, can serve as powerful cues for children as they learn new social information

(Baron & Dunham, 2015; Chalik, Rivera, & Rhodes, 2014; Dunham, 2018; Kalish, 2012; Rhodes & Chalik, 2013).

There is ample evidence suggesting that children can reason about category information when making social inferences. For example, Shutts and colleagues found that children tended to select individuals of the same gender more often than individuals of the same race in their decisions about who would be likely to form friendships (Shutts, Pemberton, & Spelke, 2013). This suggests that gender is a more potent category than race, or, at minimum, that the salience of race in children's social reasoning emerges later in childhood. There is also evidence that even novel categories, that is, fictional categories lacking in real-life significance, guide children's reasoning about who is socially obligated to whom, with children predicting that characters who belong to the same novel category will be more likely to help, and avoid harming, one another (Rhodes & Chalik, 2013).

1.2 Psychological Essentialism

Psychological essentialism—the notion that entities are the way that they are because of deep, unobservable properties—is one way in which social categories acquire psychological salience. Reliance on psychological essentialism explains, in part, why children privilege categorical information under some circumstances. In their review paper, Rhodes and Mandalaywala (2017) hypothesized how social essentialism emerges in children, suggesting that these mechanisms arise from an innate desire to make sense of one's environment. That is, they suggest that the way people reason about social categories arises out of the way people make observations about the kinds of distinctions found in nature; categories, like types of plants or animals, are seen as natural kinds, with stable and intrinsic properties (Rhodes et al., 2014; Roberts & Gelman, 2015; Hirschfeld & Gelman, 1997).

There is evidence suggesting that inherent, essentialized properties hold a great deal of explanatory power. Diesendruck and Eldror (2011), for example, investigated how 4–6-year-old children reason about internal properties (e.g., biological or psychological traits) and external properties (e.g., physical or behavioral traits). Using a between-subjects design, the authors presented children with novel social groups with one internal and one external trait (either internal biological traits that are connected to external physical traits, or internal psychological triats that are connected to behavioral traits.) They told half of the children that the internal trait *caused* the external trait, and they told the other half of the children that the two traits were merely correlated. The children were then instructed to choose a new exemplar of a member of this novel category, between a character that had only the internal property, and a character that had only the external property. The researchers found that children chose the character with the (internal) psychological property in both the causal and correlational conditions, but they chose the character with the (internal) biological property in the causal condition only. This suggested that when considering internal properties, children readily reason about psychological traits in an essentialized manner; however, they require more evidence to determine that biological properties can give rise to physical traits. Category labels may be especially informative because children infer that belonging to a category is what *causes* certain behaviors, and that these categories are essentialized properties. Indeed, there is research suggesting that this is the case: children will explain category-typical properties (e.g., why girls like tea sets) with specific reference to the category itself (e.g., "because she is a girl") (Taylor, Rhodes, and Gelman, 2009).

Similarly, in a study by Giffin and colleagues, researchers gave participants descriptions of people who displayed a morally questionable behavioral tendency, and manipulated the

explanation for this behavior through either: 1) a category label indicating that the behavior is caused by a *labeled* disease, or 2) simply a tendency that a person has (Giffin, Wilkenfeld, & Lombrozo, 2017). The researchers found that participants in the category label condition considered the individuals in the vignettes significantly less blameworthy for their actions. This suggests that people make causal inferences about category labels – in this study, it could have been the case that the mere presence of a labeled disease caused participants to reason that there was something about this hypothetical person's behavior that could be traced to the disease and its inherent properties. That certain properties exist simply by virtue of being in certain categories is the hallmark of psychological essentialism.

1.3 Shared Preferences

Another line of research has delved into the explanatory power of another kind of social information: mental states. Evidence suggests that children use the mental states of individuals, over and above their category membership, to predict how individuals will behave: for example, children who were presented with two characters who disliked each other, yet belonged to the same novel category, predicted that the two characters would direct harm toward each other, despite their shared category membership (Chalik et al., 2014). A specific subtype of mental state, namely shared preferences, has received less attention. But there is research indicating that infants use shared food tastes to infer relationship quality (Liberman, Kinzler, & Woodward, 2014), that young children use shared tastes in clothing and toys to guide friendship preferences (Fawcett & Markson, 2010), and that children will allocate fewer resources to recipients who dislike their interests (Sparks, Schinkel, & Moore, 2017).

Given these two types of social information, category labels and mental states, several other studies have attempted to discern whether children privilege one kind of information over

another when making social decisions. For instance, Diesendruck and haLevi (2006) pitted social category against personality trait and asked children and adults to assess the inductive potential of these two kinds of information. Adults and children were presented with two "anchor" characters—each with a specified social category and a personality trait, and with a different preferred hobby. They were subsequently presented a novel character that shared a social category with one of the anchor characters and a personality trait with the other character. The critical test question was which anchor character the novel character would share a hobby preference with: the anchor character with whom they shared a social category, or the anchor character with whom they shared a personality trait. The researchers found that children tended to weigh social categories more heavily in their inferences, while adults tended to weigh personality traits more heavily. These findings illustrated a developmental shift in reasoning, whereby personality traits became a more powerful predictor of behavior and affiliation with age.

1.4 Foundations for the Present Work

A set of studies conducted by Jordan and Dunham served as the basis for the present work. These studies attempted to investigate whether children privilege information about social categories or shared preferences in their reasoning about group structure. Critically, these studies utilized a between-subjects design in which children were assigned either to a condition that focused solely on categorical information, or to a condition that focused solely on shared preferences. In their first study (hereafter "Study 1"), the researchers assigned children to a condition that highlighted either social category membership or shared food preference. To minimize contextual confounds, these categories and foods were given novel names (i.e., "Zertles" and "Lapes"). Children in three age groups (3–4-, 5–6-, and 7–9-year-olds) were asked

questions about who they expected would be friends with, share an activity preference with, and harm, a "target" character – either another individual who had a category label (Category condition) or food preference (Similarity condition) in common with the target, or another individual who did not share the target's category or food preference. Indications of category membership and food preference were marked by differing T-shirt colors (i.e., red and blue), with the target character wearing the same T-shirt color as the anchor character who matched either their category membership or their preference. Based on earlier research suggesting that children tend to weigh category information quite heavily in their decisions about group membership, the researchers hypothesized that children in the Category condition would tend to select the category-biased anchor character more often than children in the Similarity condition would select the preference-biased character.

Interestingly, among the three age groups, and across trial types, the researchers did not discover significant differences between children's tendency to draw inferences based on category labels and shared preferences, although children reliably used both types of information to infer others' preferences and relationships. That is, they tended to select the category-biased and shared preference-biased characters at rates that were significantly above chance, and their performance did not significantly differ between conditions.

A subsequent study (hereafter "Study 2") tapped into a different understanding of how children see social groups: that they exist to mark which individuals are obligated to one another (Rhodes & Chalik, 2013). This study featured the same basic design, but critically, it asked which type of information children privilege when deciding whether third-parties are morally obligated to one another. Specifically, the study assessed children's judgments about shared norms, responsibilities, coalitional defense, and harm. Furthermore, the researchers defined

similarity as a shared toy preference in addition to a shared food preference, as earlier studies have suggested that food preferences share a stronger degree of similarity than toy preferences (Liberman, Woodward, Sullivan, & Kinzler, 2016). The researchers discovered, again, that across all conditions, children generally did not differentiate between category- and preference-based verbal cues, and still selected the category-biased and preference-biased anchor characters at above-chance rates.

One potential concern about the methods used in Studies 1 and 2, however, is that low-level visual similarity cues like T-shirt color may have affected children's performance, leading them to respond without factoring in information about category membership or preference. This was unlikely, as children selected the anchor character that did *not* share the target's category membership or food preference (and thus, wore a different T-shirt color) at above-chance rates on the harm trials, suggesting that the social information provided to children on each trial did sway their decisions.

Still, to address this concern, Jordan and Dunham conducted a Baseline condition (hereafter "Study 3"), where they presented children with the same target and anchor characters, but did not provide them with any sort of social information. Children ages 7–9 years old were tested in this condition because they provided the greatest rates of generalizing in Study 1. In this condition, the researchers simply stated, "Look at this kid," while pointing to each of the anchor and target characters. The researchers reasoned that if children were merely using low-level visual cues to guide their decisions, then they should perform similarly in this baseline condition as compared to Study 1.

This was not what they found. Instead, they found that children in the both the category and similarity conditions in Study 1 selected the predicted anchor character at significantly

higher rates than children in the baseline condition. This indicates that children were guided by the social information provided by the researchers in Studies 1 and 2 over and above visual cues like clothing color and spatial proximity (Jordan & Dunham, under review).

1.5 Overview of the Present Study

The present study is distinct from the previous work in several ways: First, it uses a within-subjects design, by presenting children with shared category and preference information, rather than allocating some children to a category-only, and others to a preference-only, condition. We reasoned that this design would provide a more direct test of the extent to which children privilege one type of information over the other, since a within-subjects design directly pits these two types of information against each other. Second, it features redesigned stimuli, which allow children to easily differentiate between the category and preference dimensions (each signaled by T-shirt color in the previous studies). Third, we communicate shared tastes via a food preference only, because food preferences served as a robust test of shared preferences (over toy preferences) in Jordan and Dunham's earlier studies. Finally, we created a more in-depth training phase featuring more comprehension checks. Because children were required to track novel category labels and shared preferences at once, we added a set of comprehension checks to the training phase that served as exclusion criteria. We used the triad task implemented in Studies 1–3, wherein children were asked to predict which of two "anchor" individuals a "target" individual would befriend, defend, take responsibility for, and harm. Critically, the within-subjects design altered the triad task such that for each of the two anchors, we highlighted their category label *and* preferred food, while the target was described as having the category label of one anchor and the food preference of the other.

Based on earlier work suggesting the explanatory power of social category information for young children, and based on Jordan and Dunham's earlier findings, we predicted that 3–4-year-olds would select characters based on categorical information more than shared preference information. We also predicted a developmental shift, such that 7–9-year-old children would fail to distinguish between the information types, placing equal value on social categories and shared preferences. This is due, in part, to additional evidence that adults were more swayed by shared preferences in their reasoning about group membership, and evidence that as children's capability for theory of mind increases, they tend to rely more on mental states to guide their inferences (Diesendruck & haLevi, 2006; Chalik et al., 2014).

2. Method

2.1 Participants

The participants were 51 children (n = 23 female) from 2 age groups: 3–4- (n = 25) and 7–9- (n = 26) year-olds. For the 3–4-year-olds, the mean age was 3.92 years, and age range was from 3.12 to 4.83 years; for the 7–9-year-olds, the mean age was 7.88 years, and the age range was from 7.21 to 9.87 years. No gender non-binary children were tested. In contrast to Studies 1–3, the intermediate age group (5–6-year-olds) was not tested because they performed similarly to the oldest group of children in prior studies (Jordan & Dunham, under review), and assessing the developmental trajectory of this kind of social reasoning, we reasoned, was equally possible and valid when we tested the youngest and oldest kids. We tested an additional 15 children who were excluded from analyses due to 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 = 25), or an empty classroom at the participant's school (n = 10).

Participants for this study were recruited from the New England region of the United States. We did not collect information about the participants' races, ethnicities, or family incomes, but based on the demographic profiles of the testing sites, we believe that most of the participants are White and from middle-class families. Prior to beginning the study, all parents or legal guardians provided written, informed consent on behalf of their child, and each child provided verbal assent.

2.2 Design & Materials

Stimuli for these studies resembled those used by Jordan and Dunham, but were altered in several critical ways: One goal was to signal category membership and shared preferences in ways that differed from each other, yet were relatively similar in their signaling strength. In contrast to the previous studies, which used T-shirt color to signal both category and preference information, we used colored flags to signal category information, and randomly-drawn shapes inlaid on hand-drawn lunchboxes to signal food preference information. The category and food preference stimuli were created using Keynote, and the characters were the same as those used in the Jordan and Dunham studies. No character was presented more than once during the study. Each character displayed a positive facial expression and matched the participant's gender (as identified by their parent).

All children were assigned to the Pit condition. The study consisted of a series of 4 trials of the following types: "Friend", "Defense," "Responsibility," and "Harm." We wanted to ask specifically about harm because of evidence suggesting that children reason about groups by considering who is obligated to not harm whom (Chalik & Rhodes, 2018). Because harming in-group members is something that is recognized as impermissible early on, we wanted to test the

extent to which younger and older children would reason that intergroup harm is more likely to occur than intragroup harm.

The task was constructed and presented in Keynote, and the experimenter ran all participants in the study on a laptop computer. We counterbalanced the colors of the flags and novel foods (either red and blue or green and orange), the verbal labels of the categories and novel foods (either "*Zertles*" and "*Lapes*", or "*Hoopas*" and "*Flurps*"), and the order of the trial types. We also counterbalanced the order in which the critical information was presented in the training and test phases (either category or preference first); this was to avoid inducing priming for one type of information over the other (Murdock Jr., 1962).

2.3 Procedure & Scoring

The experimenter told each participant that he or she would be "learning about some kids from a storybook," and to "pay really close attention to who each kid is, and what they like to eat". The task then proceeded to a training phase, wherein the experimenter displayed two sets of flags or foods on either the left or right side of the screen (see Figure 1). For example, she may have said of one set 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 foods, the experimenter would present the items again, and ask the participant, for example, "Now can you tell me who these flags are for?" This question served as the first comprehension check.

The experimenter then introduced the child to two sets of 4 introduction characters, one on each side of the screen, who either held flags representing their category membership or lunchboxes with food representing their preference (see Figure 1). 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 of kids, she would say, “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? This question served as the second comprehension check.

For both of these training stages, if a participant answered our comprehension checks incorrectly, the experimenter would correct the participant by pointing out the correct names for each of the categories or foods. These training stages were then repeated for the other information type.

After these two stages of training were completed for each information type, the experimenter presented characters displaying both a flag and a lunchbox, indicating their category membership and food preference, respectively (see Figure 2). She said, for example, of one set of characters while pointing, “See these kids? These kids are all called *Hoopas*, and they all like to eat a food called *Zertles*.” And of the other, “And see these kids? These kids are all called *Flurps*, and they all like to eat a food called *Lapes*.” We counterbalanced whether the experimenter present the foods or the categories first in this stage of training.

The third set of comprehension checks followed these training phases wherein the experimenter presented the participant with a pair of laminated cards containing pictures of either the flags or the foods (see Figure 3). She presented the two sets of characters with *only one* dimension of the critical training information displayed, and asked, for example, “Using these cards, can you show me what the kids like to eat?” and instructed the participant to match up the food cards to the characters. She then repeated this step with the other information dimension (presentation order counterbalanced.) If participants were not able to successfully complete this

matching comprehension task, their data were subsequently excluded from all analyses. We used this comprehension check as our exclusion criteria, as we wanted to ensure that participants understood the category and food preference pairings for each set of characters prior to the test phase. If children failed to retain this information, we reasoned, they may not be basing their decisions on either dimension of social information that we provided to them.

Each test trial began with the experimenter directing the participant's attention to an anchor character on the left side of the screen. She reminded the participant of that anchor character's group label and food preference (Figure 4). The experimenter then presented a second anchor character on the right side of the screen, and reminded the participant of that character's group label and food preference in the same way. While presenting the anchor characters, the experimenter would say of each character, for instance, "See this kid? This kid is called a *Zertle,* and s/he likes to eat *Flurps.*" After presenting the two anchor characters, the experimenter displayed a child with their attributes concealed by a gray block marked with a question mark. The experimenter said while pointing to this target, "Now see this mystery kid?" She would then reveal the group label and food preference for this target, highlighting the fact that the target shared one dimension of similarity with each of the anchor characters. For instance, she would say, "This kid is called a *Zertle* like him (while pointing to the left anchor character), and likes to eat *Hoopas* like him (while pointing to the right anchor character.)" The trial block determined which type of test question the experimenter presented: in the friend trials, she asked which of the two anchor characters the target would be friends with; in the defense trials, she asked which of the two anchor kids would protect the target from a harmful action (i.e., who would stop someone from breaking the target's favorite toy); in the responsibility trials, she asked which of the two anchor characters would apologize on behalf of the target

when the target committed a harmful act (i.e., hitting someone); in the harm trials, she asked who, of the two anchor characters, the target would likely harm (i.e., hit.) For each trial, the participant was instructed to point to the anchor character whom they believed most appropriately answered the test question. If the participant did not respond, or failed to choose just one anchor character, the experimenter prompted him or her to answer up to two more times (see Appendix A for full trial script).

We coded our data as follows: for the Friend, Defense, and Responsibility trials, a score of "1" indicated that a participant selected the anchor character who shared the target's category label. A score of "-1" indicated that the participant selected the anchor character who shared the target's food preference. This was reversed for the harm trials, where a score of "-1" indicated a category label match, and a score of "1" indicated a food preference match. We reverse-coded the Harm trials because we reasoned that if children were weighing one type of social information over another, they would expect the target to harm the anchor character that was *not* similar along that social information dimension. We calculated an average bias score for each trial block by taking the mean of scores for each block, and calculated an aggregate average bias score by taking the mean of each of these means.

3. Results

3.1 Main Analyses

All of these main analyses were preregistered. We used R Studio to analyze our data and create our plots. We used one-sample *t*-tests to assess whether children performed at above-chance levels (chance = 0). By comparing each trial type to chance, we used the Bonferroni correction for multiple comparisons, resulting in an adjusted alpha of 0.0125. These *t*-tests revealed that the 3–4-year-old children did not select the category-biased or preference-biased

anchor character at above-chance rates for any of the 4 trial types (all *p*s > 0.0125). We collapsed across the trial types and discovered that younger children's performance across the types was not significantly biased in either the category or preference direction, (M = -0.055, SD = 0.385, $t(26)$ = -0.71, p = 0.482) (Figure 5). However, turning to the 7–9-year-old children, one-sample *t*-tests revealed that for the Defense trials, older children chose the category-biased character at above-chance levels (M = 0.40, SD = 0.63, $t(26)$ = 3.25, p = 0.003.) Collapsing across all trial types, we found older children selected the category-biased anchor character at above-chance rates (M = 0.33, SD = 0.53, $t(26)$ = 3.12, p = 0.005.)

We conducted a 2 (Age: 3–4-year-olds vs. 7–9-year-olds) x 4 (Trial type: harm vs. friend vs. responsibility vs. defense) Analysis of Variance (ANOVA), and observed a main effect of age group (F = 15.48, p = 0.001.) Older children were more likely than younger children to select anchor characters that were category-biased. We did not observe a main effect of trial type or an interaction between the factors (*p*s > 0.05).

3.2 Analysis of Comprehension Check Passers

One interpretation of these results, particularly when examining the younger children, is that younger children were not capable of understanding the task. This is a plausible explanation, given that the task required children to track two pairs of novel category labels and food preferences, and appreciate that the target characters shared only one dimension of similarity with each of the anchor characters. We included the comprehension checks that involved matching characters' category labels with their food preferences, using laminated cards, as a way to screen out children for whom the task may have been too confusing (Figure 3). We excluded all children who failed the first set of matching comprehension checks (during the training phase), but we did not exclude children who failed only the final set of matching comprehension

checks. We reasoned that if these children were able to correctly match the category labels and food preferences in the training phase, this should be sufficient to demonstrate that they comprehended the nature of the task. Furthermore, being presented with target characters that did not conform to the anchor characters presented in the training phase may have confused children and subsequently caused them to mismatch the category labels and food preferences in the final comprehension checks. Supposing that this was the case, we performed the same statistical tests described above with only the subset of children who passed *both* the initial and final comprehension checks. We reasoned that this would eliminate the children who may have been even slightly confused by the nature of the task. The following analyses are exploratory, and should be interpreted accordingly.

Seven 3–4-year-olds and one 9-year-old were excluded on this basis. We again used one-sample *t*-tests to assess whether children performed at above-chance levels (chance = 0.) By comparing each trial type to chance, we used the Bonferroni correction for multiple comparisons, resulting in an adjusted alpha of 0.0125. Again, these *t*-tests revealed that the 3–4-year-old children selected the category- and preference-biased anchor characters at chance rates on each of the 4 trial types (all *p*s > 0.05). Compared to the above analyses, however, this subset of younger children chose the preference-biased anchor character slightly more often, although their performance did not reach significance ($p = 0.1223$.) (Figure 6). We collapsed across the trial types and discovered that younger children's performance was not significantly biased in either the category or preference direction ($M = -0.13$, $SD = 0.34$, $t(18) = -1.62$.) Turning to the 7–9-year-old children, one-sample *t*-tests revealed that, again, for the Defense trial type, older children chose the category-biased character at levels significantly above chance ($M = 0.38$, $SD = 0.63$, $t(25) = 3.00$, $p = 0.006$). Again, collapsing across trial types, we found that older

children's performance was significantly biased in the category direction ($M = 0.30$, $SD = 0.53$, $t(25) = 2.84$, $p = 0.009$.) One-way ANOVAs for both the younger and older children did not reveal significant differences in their performances between trial types (ps > 0.05).

We, again, conducted a 2 (Age: 3–4-year-olds vs. 7–9-year-olds) x 4 (Trial type: harm vs. friend vs. responsibility vs. defense) Analysis of Variance (ANOVA), and observed a main effect of age group ($F = 15.87$, $p = 0.001$.) Older children, again, were more likely than younger children to select anchor characters that were category-biased. We, again, did not observe a main effect of trial type or an interaction between the factors (ps > 0.05).

3.3 Comparison to Baseline

From our primary analyses, we discovered that older children, but not younger children, significantly privileged information about social categories over information about shared preferences. But to what extent? To answer this question, we compared older children's performance to a baseline condition to investigate how much these older children were guided *exclusively* by the social information we provided to them about the category labels and food preferences assigned to the target and anchor characters, *over and above* simple visual cues (i.e., the colors of the flags and foods). The baseline condition employed by Jordan and Dunham (Study 3) proved to be useful for comparison in this respect. We do acknowledge that the stimuli differed slightly between the baseline condition and the present study (the characters in the baseline condition wore colored T-shirts, while the characters in the present study held flags and lunchboxes with foods.) Still, many of the factors between two studies remained the same: the types of cartoon characters used, the triads in which these characters were set up, and the sequences in which we presented the characters. Since the 7–9-year-old children in the present study seemed to show a category bias, we were interested to know if this bias was still significant

when compared to 7–9-year-olds' performance in the baseline condition. The following analyses are also exploratory, and should be interpreted accordingly.

Since Jordan and Dunham's studies used a different coding method than we did, we transformed our data to match theirs to facilitate comparison. Jordan and Dunham assigned a "1" for selection of the predicted anchor character in their studies, and a "0" for selecting the other character. They then took the sum of predicted-test-character matches for each block. We transformed our data similarly for the present study: we assigned a "1" for selecting a category-biased character, and assigned a "0" for selecting a preference-biased character (and reverse-coding for the Harm trial type.) Instead of taking the averages, we took the sum of the category-biased selections for each block. Thus, the minimum score a child could receive for each block was 0, and the maximum score was 4. A score of 4 indicated that the child selected the category-biased character on each trial, while a score of 0 indicated that the child selected the preference-biased character on each trial. Chance performance was a score of 2. Since there were no significant differences in children's performances across trial types, we collapsed across them all.

We compared the average number of category, preference (similarity), and baseline matches (from Jordan and Dunham's Studies 1 and 3) to the average number of category-biased matches from the present study. All 7–9-year-old children performed at above-chance rates across each condition: that is, they chose the predicted anchor characters in both conditions of Jordan and Dunham's Study 1 and their baseline condition, and they chose the category-biased character at above-chance rates in the present study.

We then ran a Welch's two-sample *t*-test to examine whether the difference in means between the baseline condition and the present study (Pit condition) was significant. The results

of this test demonstrated that the difference in means was not significant (M(baseline) = 2.41, M(pit) = 2.65), t = -1.06, p = 0.292) (Figure 7). This suggests that older children may not have privileged category information, over and above low-level perceptual/visual similarity cues.

4. Discussion

4.1. General Discussion

This study featured a within-subjects design that directly pitted information about social categories and information about shared preferences against each other. This, on its own, was the most rigorous test of whether children were more swayed by one kind of social information over another, since it asked children to reason about both kinds of information at once. Critically, it differs from Jordan and Dunham's earlier studies in this respect: while children in the previous between-subjects studies were asked only to reason about one kind of social information at a time, the present work introduced competition between the two cue types allowing for a more direct assessment of their relative strength.

On their face, the results we gathered seem to refute Jordan and Dunham's earlier findings, given that older children in the present study privileged information about social categories over information about shared preferences, while younger children did not privilege either information type (though they did show a slight, though not significant, bias towards shared preferences). By comparison, Jordan and Dunham's earlier studies showed that children robustly used information about category membership and shared preference in their judgments, and did not significantly discern between the two types of information in their inferences.

What could explain the apparent discrepancies between the present study and Jordan and Dunham's prior studies?

We hypothesized that younger children would be biased in the category direction, yet this was not what we found. This could have been due to the strength of the competing cue that we selected, namely shared preferences. Indeed, there is ample evidence that even young children can use information about both social categories and mental states to inform their inferences about group membership and intergroup obligation (e.g., Sparks et al., 2017; Liberman et al., 2014; Hamlin, Mahajan, Liberman, & Wynn, 2013). Given these accounts, it is plausible that our results confirmed that children, at even 3–4 years of age, can robustly use both kinds of information to inform their reasoning, which could explain why they performed at chance in the present study.

One potential reason why older children privileged social category information here could be that category information was more perceptually salient than shared preference information. We can potentially rule this out, however, given that we counterbalanced the order in which these two types of information were presented in our study, and ensured that the visual cues for social categories were not more salient than the visual cues for shared preferences.

Why, then, could older children have exhibited a bias toward social category information? Chalik and colleagues (2014) found that when children were presented with characters who belonged to different groups and had different individual mental states, children's capability for theory of mind (ToM) reasoning was positively correlated with their likelihood to rely on individuals' mental states *over* their group membership in deciding how these characters would behave toward each other. Since ToM reasoning is a capability that emerges early in childhood and becomes a skill that older children commonly utilize in their everyday thoughts and decisions (Wellman, 1992), we might have expected older children to rely more on shared preference information, as shared preferences are, fundamentally, a kind of mental state. But it

did not seem to be the case that older children in our study were as reliant on information about shared preferences.

One possible explanation for why social categories held so much weight for older children is that the explanatory power of a category label might be particularly strong. As Giffin and colleagues (2017) found, adults were more likely to excuse a morally questionable behavior if they were told that the behavior was due to a labeled mental/physical condition, compared to when they were told that the behavior was simply due to a "tendency." That is, adults were more likely to ascribe causality to category labels. Similarly, it could be the case that the category labels we presented to children in this study could, on their own, have been explanatorily more powerful than the shared food preference information. The older children may have inferred that category labels were more meaningful, or explained something intrinsic about the anchor and target characters, and were therefore more likely to be swayed by this kind of information. Given prior research that suggests that children readily do this (e.g., Baron & Dunham, 2015; Dunham, 2018; Rhodes & Chalik, 2013), this is a plausible explanation for our pattern of results as well.

4.2. Comparison to Baseline

One way to refine our finding that older children seemed to display a category bias is by examining just how strong their category bias was. When we compared 7-9-year-olds' performance to the baseline condition, where children were given visual cues but no other information, we found that 7–9-year-olds in our study did not select the category-biased character at significantly higher rates than the rates at which 7–9-year-olds selected the predicted anchor character in the baseline study. This suggests that while older children in the present study may have privileged category information more than shared preference information, they did not do so to the extent that literature in this area seems to suggest. That is, findings in the

literature that seem to suggest that children rely heavily on information about social categories, and that they are some of the first to emerge in the developmental trajectory of social cognition (e.g., Gelman & Markman, 1986). We might expect, therefore, that children would robustly rely on social categories, over most other kinds of social information, to guide their inferences about inter-/intragroup interactions. And we did find this, to some extent, for the 7-9-year-olds here; importantly, however, the extent to which they privileged social category information did not prevail over comparable kinds of low-level visual and perceptual similarities (e.g., different colors to signal categories and shared preferences; spatial proximality of characters, etc.) that children might have reflexively relied upon.

4.3 Limitations

While we did not require that participants report their racial/ethnic background or their socioeconomic status, we believe that the majority of our participants were white and came from middle-class families. Given that our study asked children to reason about social categories (of which race, and to an extent, social class, are types), and to reason about target characters that differed from the category and preference pairs exhibited by the anchor characters, we must be cautious about the generalizability of our findings (Henrich, Heine, & Norenzayan, 2010).

Additionally, while Jordan and Dunham's Study 3 served as a useful baseline for comparison for our study, it would have been helpful to include a baseline condition for the present study that used the exact same visual stimuli. This would have served as the strongest test to determine whether children relied on the social category information that we provided over and beyond their reliance on other visual cues.

4.4 Future Directions

We purposely used novel category names and food types and names in our study to minimize the contextual biases that children might bring from their everyday lives had we decided to use real-world social categories like race or gender and familiar food preferences. Certainly, using novel descriptions provides the strongest test of how children reason about these types of social information, as it allows us to examine if and how they do so, *absent* the kinds of social information that they are already familiar with. But it may be additionally illuminating to incorporate real-life social categories and preferences into a study that has the same design as this one to see how children respond. Additionally, it would be useful to include participants from a wide range of cultural backgrounds which would allow us to generalize our conclusions beyond the sample tested here (Henrich et al., 2010).

Our study assessed children's third-person evaluations of fictional others. To that end, it may be additionally interesting to examine how children, across ages, form social groups from information communicated in their day-to-day lives. If older children are more affected by labeled category information, it might be the case that the people with whom they most often affiliate are people who belong to their labeled social categories. On the other hand, if younger children do not tend to privilege one kind of information over another, we might not observe networks that form around social categories or shared preferences. Studies (e.g., Eagle, Pentland, & Lazer, 2009) have examined how networks of groups form within a larger group (say, a classroom of students), and it may be fascinating to construct a similar model with groups of children in different age groups, to examine whether groups tend form around any type of social information. It might be the case that, in a classroom of children, those who share a social category share more connections (i.e., friendships) with each other. It could also be possible that those who share a preference share more connections with each other.

4.5 Concluding Remarks

In an era where attention to, and awareness of, social groups is increasing, and where increased migration and globalization has facilitated the interaction and integration of people across different social groups, it is important to assess the emergence of the understanding of social groups. While we know that people can be understood as belonging to different groups, and as being similar and different from others based on various dimensions of grouping, investigating how people are guided by different kinds of social information, and how this capability shifts over time, can inform how we talk with children about sociality, guide them to interact with each other in classroom settings, and facilitate adults coexisting peacefully within a society.

Acknowledgements

A world of thanks to my advisors, Ashley E. Jordan and Dr. Yarrow Dunham, whose earlier studies laid the groundwork for this thesis, and whose guidance and feedback throughout this process has been consistent, patient, and empathetic. I am additionally thankful for the Cognitive Science department and the Social Cognitive Development Lab at Yale, without whom I would not have been able to complete these eight months of research. And finally, I am grateful for my family and friends, for their unbridled love and support.

Author Contributions

Thanks to my advisor, Ashley E. Jordan, for creating Figures 3 and 4, and for helping me to streamline my discussion section in response to peer reviewers finding section 4.2 confusing. Thank you to my peer reviewers for pointing out this confusion.

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. http://doi.org/10.1080/15248372.2014.1000459

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

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

Diesendruck, G., & Eldror, E. (2011). What children infer from social categories. *Cognitive Development*, *26*(2), 118–126. https://doi.org/10.1016/j.cogdev.2010.11.001

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in children's social category based induction. *Child Development*, 77, 539-553. http://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., & Emory, J. (2014). Of affect and ambiguity: The emergence of preference for arbitrary ingroups. *Journal of Social Issues*, *70*(1), 81-98. http://doi.org/10.1111/josi.12048

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, *106*(36), 15274-15278. https://doi.org/10.1073/pnas.0900282106

Fawcett, C. A., & Markson, L. (2010). Similarity predicts liking in 3-year-old children. *Journal of experimental child psychology*, *105*(4), 345-358. http://doi.org/10.1016/j.jecp.2009.12.002

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

Giffin, C., Wilkenfeld, D., & Lombrozo, T. (2017). The explanatory effect of a label: Explanations with named categories are more satisfying. *Cognition*, 168, 357–369. https://doi.org/10.1016/j.cognition.2017.07.011

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. http://doi.org/10.1177/0956797612457785

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature,* 466(7302), 29. https://doi.org/10.1038/466029a

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development,* 12(2), 213-238. https://doi.org/10.1016/S0885-2014(97)90014-9

Jordan, A.E., & Dunham, Y. (under review) Are Categories Primary? Children Use Similarities to Reason About Social Groups.

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

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. http://dx.doi.org/10.1037/a0034481.

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. http://doi.org/10.1073/pnas.1605456113

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482. http://dx.doi.org/10.1037/h0045106

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

Rhodes, M., & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), e1437. https://doi.org/10.1002/wcs.1437

Rhodes, M., Gelman, S. A., & Karuza, J. C. (2014). Preschool ontology: The role of beliefs about category boundaries in early categorization. *Journal of Cognition and Development,* 15(1), 78-93. https://doi.org/10.1080/15248372.2012.713875

Roberts, S. O., & Gelman, S. A. (2015). Do children see in Black and White? Children's and adults' categorizations of multiracial individuals. *Child Development,* 86(6), 1830-1847. https://doi.org/10.1111/cdev.12410

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. http://doi.org/10.1080/15248372.2011.638686

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. http://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

Wellman, H. M. (1992). *The Child's Theory of Mind.* The MIT Press.

Younger, B. (2017, September 28). Black Musicians on Being Boxed in by R&B and Rap Expectations: "We Fit in So Many Things". Retrieved from https://pitchfork.com/thepitch/black-musicians-on-being-boxed-in-by-randb-and-rap-expectations-we-fit-in-so-many-things/

Figures

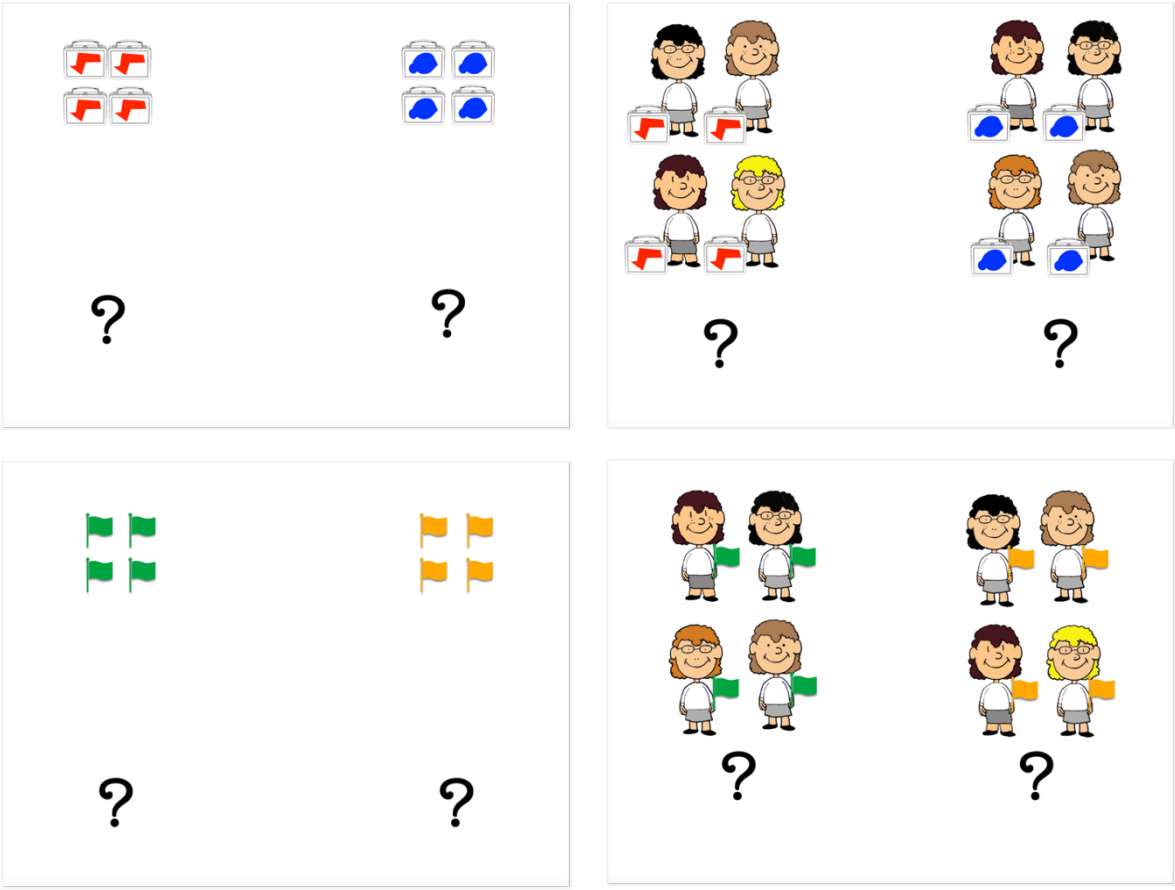

Figure 1. Example cue introduction displays. The experimenter said, while pointing to the corresponding locations: (*top-left*) "See these foods, they're called Hoopas. Can you say Hoopas? See these foods, they're called Flurps. Can you say Flurps? (*top-right*) These kids all like to eat a food called Hoopas, and these kids all like to eat a food called Flurps. (*bottom-left*) Now, see these flags, they're for kids who are called Zertles. Can you say Zertles? And see these flags, they're for kids who are called Lapes. Can you say Lapes? *(bottom-right*) These kids are all called the Zertles, and these kids are all called the Lapes."

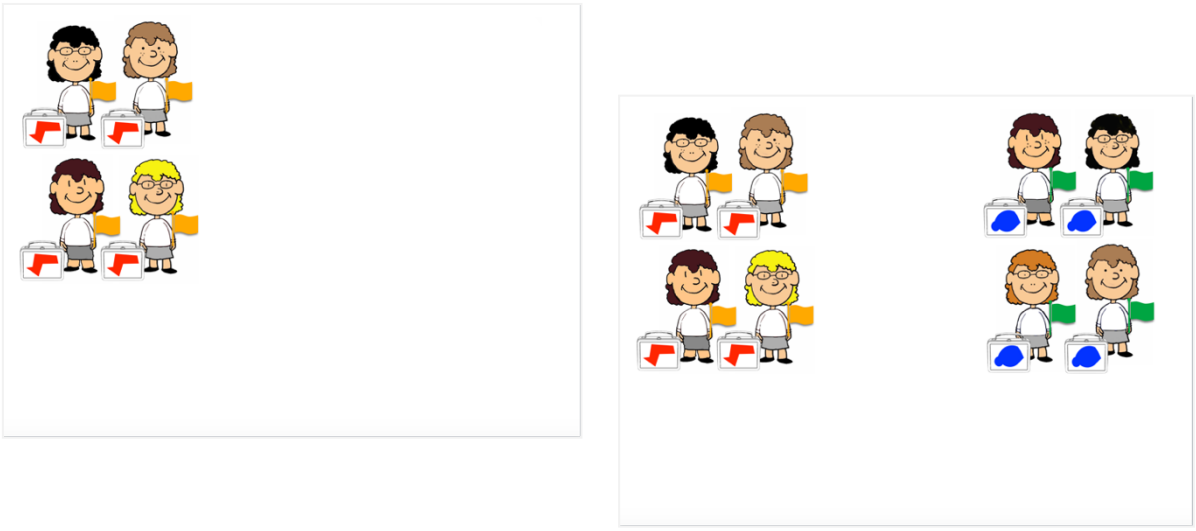

Figure 2. Example displays combining the category and food preference cues. The experimenter said: **“**See these kids? These kids all like to eat Hoopas, and they’re called the Lapes *(left panel).* And see these kids? These kids all like to eat Flurps, and they’re called the Zertles *(right panel)*.

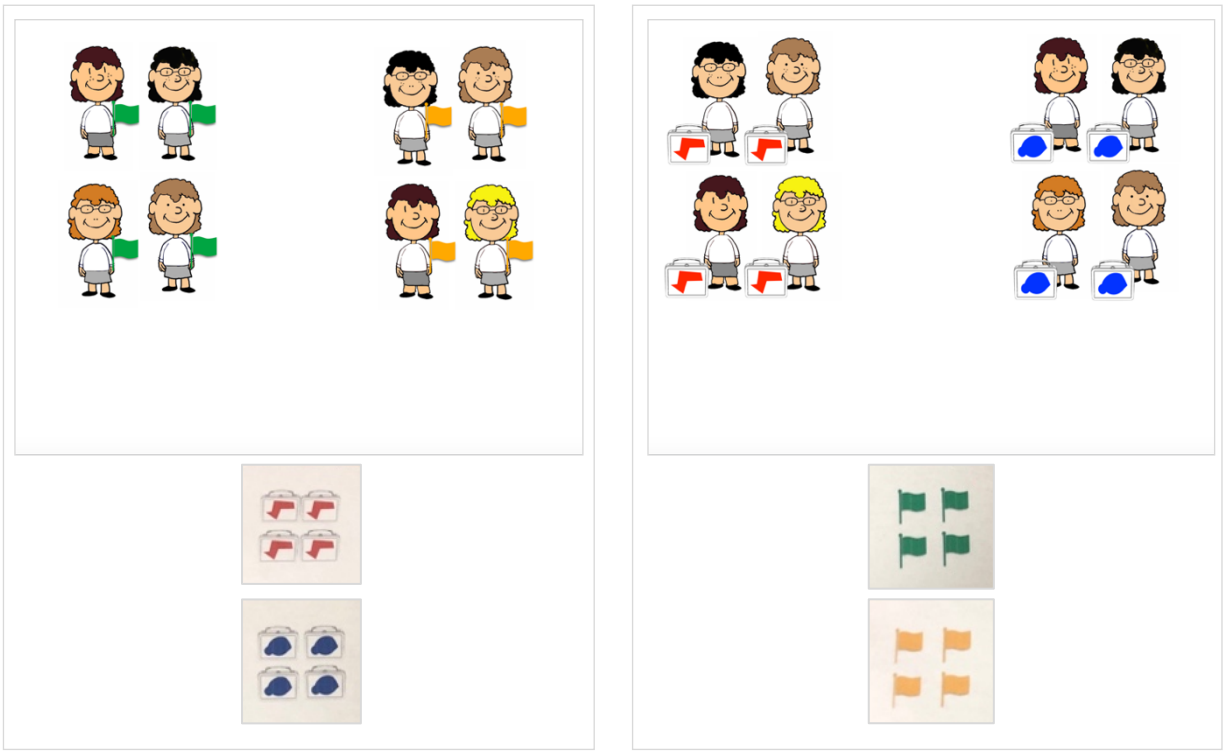

Figure 3. Example matching comprehension check displays. The experimenter said: "Using these pictures [*she placed two laminated cards in front of the participant*] can you show me which foods these kids like (*left panel*) / group these kids belong to (*right panel*)? Can you match them up?" The participant placed the cards to the left or right side beneath the group of choice.

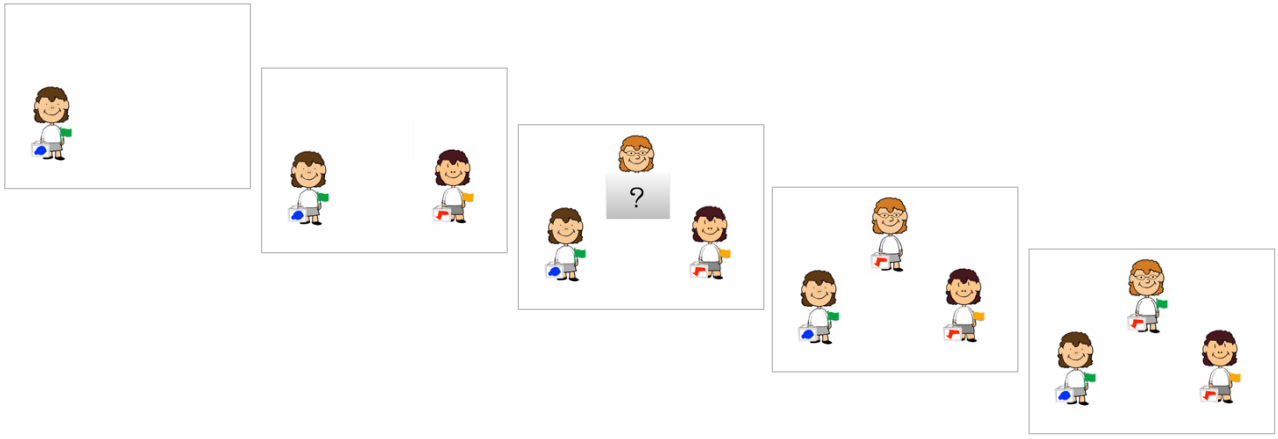

Figure 4. Example test trial displays. The experimenter said: **“**See this kid [left anchor]? This kid likes to eat flurps, and she’s called a Zertle. And see this kid [right anchor]? This kid likes to eat hoopas, and she’s called a Lape. Now see this mystery kid? She likes to eat Hoopas like her [*points to right anchor*], and is called a Zertle like her [*points to left anchor*]”.

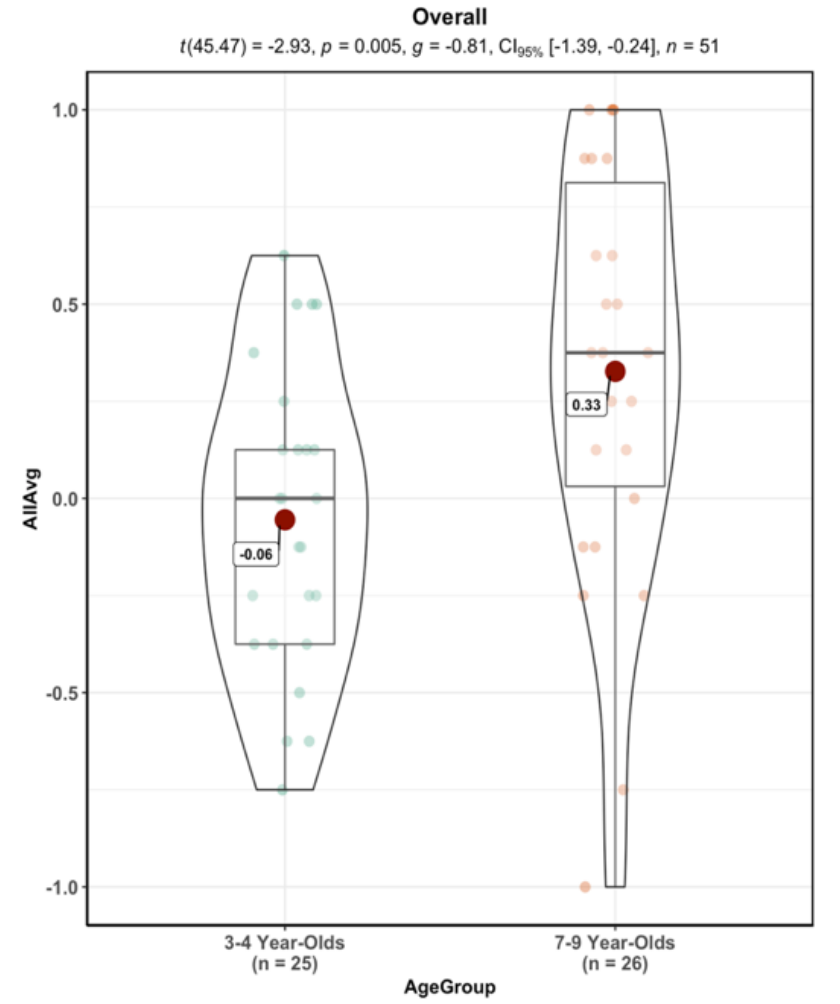

Figure 5. Average bias scores, collapsed across trial types, for each age group. Higher values indicate category bias, and lower values indicate preference bias. Box and whisker plots wherein the box represents the interquartile range; each vertical line extending from the box points to the highest and lowest data points; the larger red dots represent the means for each age group; the smaller dots represent each data point; and the curved lines represent the distribution of the data.

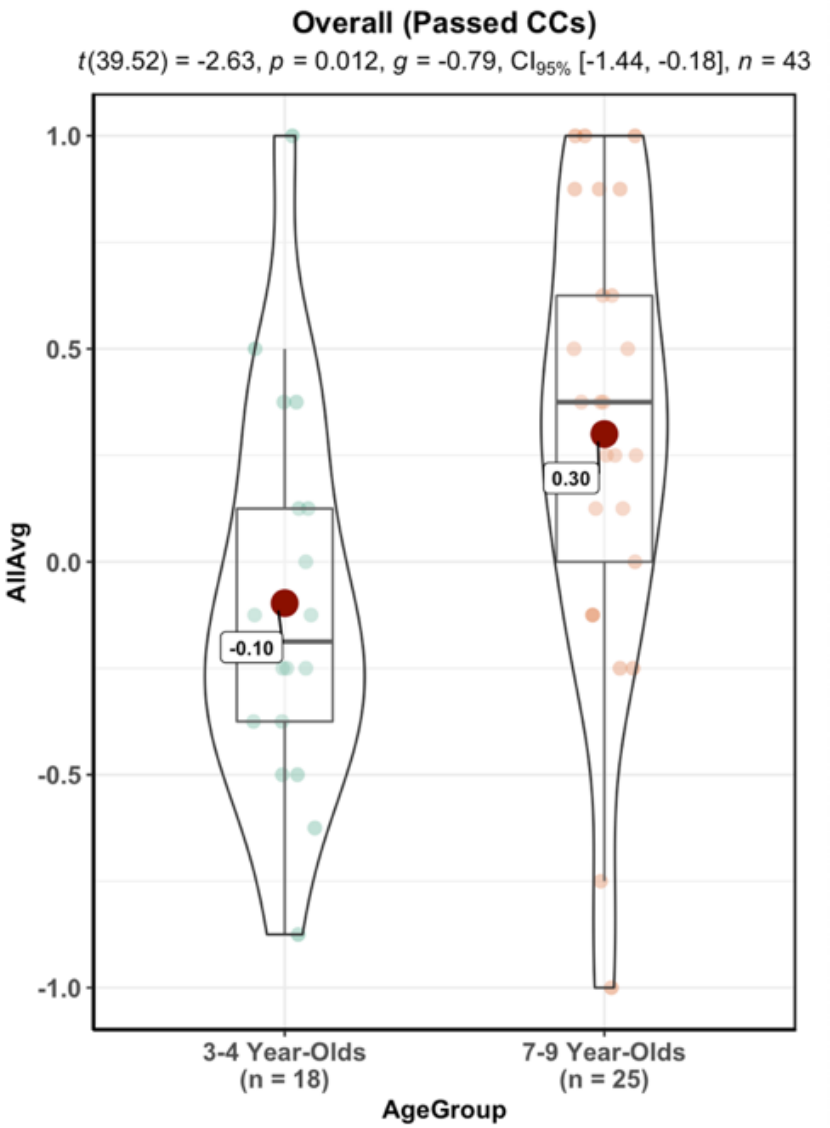

Figure 6. For the subset of children who passed all comprehension check sets, average bias scores, collapsed across trial types, for each age group. Higher values indicate category bias, and lower values indicate preference bias. Box and whisker plots wherein the box represents the interquartile range; each vertical line extending from the box points to the highest and lowest data points; the larger red dots represent the means for each age group; the smaller dots represent each data point; and the curved lines represent the distribution of the data.

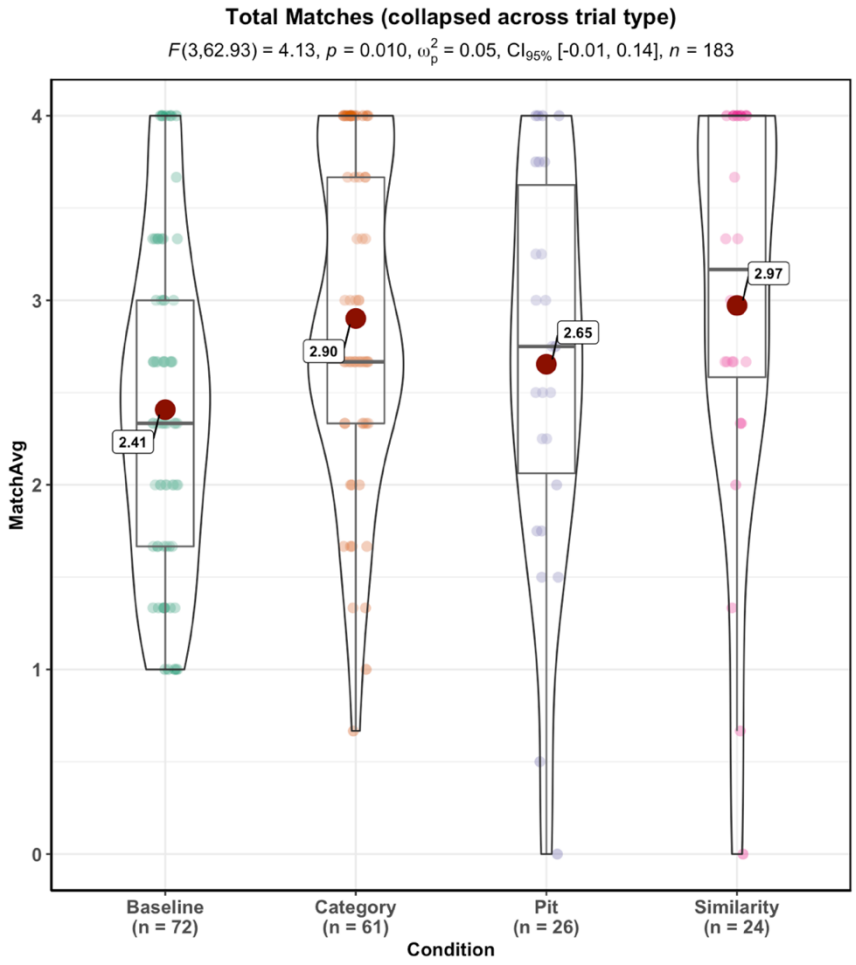

Figure 7. Average total matches for the baseline, category, pit, and preference (similarity) conditions, collapsed across trial types, for 7–9-year-olds. For the pit condition, higher values indicate greater category bias, and lower values indicate greater preference bias. For the other three conditions, higher values indicate more predicted anchor character matches. Box and whisker plots: box represents interquartile range; each vertical line extending from the box points to the highest and lowest data points; the larger red dots represent the means for each age group; the smaller dots represent each data point; curved lines represent the distribution of the data.

Appendix A

Trial Types

Friend: Which one of these two kids do you think she wants to be friends with?

Defense: One day, someone tried to break one of her favorite toys. Which one of these two kids made them stop doing that to her?

Responsibility: One day, she hit someone really hard and didn't say sorry. Which one of these two kids will say sorry for her?

Harm: One day, she hit one of these two kids. Which one of these kids do you think she hit?