Category coherence in children's inductive inferences with cross-classified entities

Simone P. Nguyen*, Tess Chevalier

University of North Carolina Wilmington, Department of Psychology, United States

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 27 April 2015
Accepted 4 May 2015

Keywords:
Category coherence
Inductive inferences
Cross-classification
Categorization

ABSTRACT

Items rarely belong to a single category, but rather can be cross-classified into many categories, each serving a very different basis for induction. Presently, little is known about how children determine which category to use for induction in cross-classification situations. This research examined the role of category coherence in children’s and adults’ (N = 329) inductive reasoning about cross-classified entities. In Study 1 and 2, participants were presented with a person who could be cross-classified into two categories, one coherent and the other incoherent. Category coherence is the extent to which members of a category and/or the properties of a category make sense together, given one’s background knowledge. The results showed that by age 5 years, children are systematic in their use of coherence, tending to select the coherent category to inform their inference about the person. There is weaker evidence for 4-year-olds. Study 3 and 4 revealed that 4-year-olds have an appreciation for coherence in tasks of categorization. The results of Study 5 demonstrated 5-year-olds’ and adults’ tendency to use coherence when reasoning about cross-classified entities across different induction tasks. Overall, these results contribute to our emerging understanding of how and why children select some categories versus others during induction with cross-classified items.

© 2015 Elsevier Inc. All rights reserved.

* Corresponding author at: University of North Carolina Wilmington, Department of Psychology, 601 South College Road, Wilmington, NC 28403-5612, United States.
E-mail address: nguyens@uncw.edu (S.P. Nguyen).

http://dx.doi.org/10.1016/j.cogdev.2015.05.001
0885-2014 © 2015 Elsevier Inc. All rights reserved.
Category-based induction is a key way through which we use categories to broaden our knowledge, using a premise category to make inferences about the properties of unfamiliar instances of a category (see Heit, 2000, 2007, for a review). There is a tremendous amount of research showing that young children are capable of inductive reasoning (see Gelman & Davidson, 2013; Hayes, 2007; Murphy, 2002, for a review) with much of this research focusing exclusively on children’s ability to use single category membership as a basis for induction (see Gelman, 2003; Krackow & Gordon, 1998; Nguyen, 2008 for discussion). However, items rarely belong to a single category, but can be cross-classified into many categories, each serving a very different basis for induction (Murphy & Ross, 1999; Nguyen & Murphy, 2003; Nguyen, 2012; Ross & Murphy, 1999). This is especially the case for categories pertaining to people, where a person can potentially belong to several categories with different bases (e.g., age, gender, occupation, hobbies) (see Baron, Dunham, Banaji, & Carey, 2014; Shutts, Pemberton Roben, & Spelke, 2013 for a review of children’s acquisition and use of social group membership). Each of these bases can lead to very different inferences about that person, and subsequently affect interactions with that person.

Therefore, cross-classification raises a number of questions about children’s inductive reasoning. When children are faced with an entity that is a member of several categories, how do they decide which category to use for induction? Do children use only a single category or all of the categories? For example, if a person can be cross-classified into different categories based on their occupation and hobby, do children use their knowledge of the former or latter category to inform their inferences?

There is small body of research that is beginning to document how and why children select some categories over others when making inferences in cross-classification contexts. This research has found that children’s category selection is influenced by context, specifically the relevance of the property to the inductive inference. That is, with cross-classified concepts, children tend to select the category that is most relevant to the property that is being generalized in the induction task (Heit & Rubinstein, 1994; Kalish & Gelman, 1992; Nguyen & Murphy, 2003; Nguyen, 2008, 2012; Shafto, Coley, & Vitkin, 2007; see also Medin & Waxman, 2007 for a discussion regarding the relevance of targets in induction). For example, Nguyen (2012) has found that by 4 years of age, children know that taxonomic and script categories are informative of different kinds of properties, and if one category is more relevant in a given moment, then children will use it as a basis for induction. Specifically, on one trial, when children are asked to make a biochemical inference about an item that can be cross-classified into both categories (e.g., a clown is a person and a part of the circus), children know to use the item’s taxonomic category membership, which is defined by shared common properties, as a basis for induction (e.g., inferring that a clown and person have the “same stuff inside”). On a different trial, when children are asked to make a situational inference about the same cross-classified item, children also know to use the item’s script category membership, which is defined by common roles in a routine or event, as a basis for induction (e.g., inferring that the clown and circus tent belong to the “same special time”).

Although this body of research shows that property relevance helps children select the appropriate category to use on different trials, it does not address what children do when more than one category remains relevant for induction at the same time during the same trial. Research from the adult concepts literature has begun to address this issue by investigating how the structure of the category, particularly category coherence, might also play a role in the induction of cross-classified concepts (e.g., Hayes, Kurniawan, & Newell, 2011; Patalano, Chin-Parker, & Ross, 2006; Patalano, Wengrovitz, & Sharpes, 2009). Category coherence is the degree to which members of a category and/or the properties of a category hang together and make sense within the context of one’s greater knowledge about their underlying origins or purpose. In other words, category coherence is the extent to which category members and their features go together in light of an individual’s prior knowledge (Medin, 1989; Murphy & Medin, 1985).

For example, using cross-classified job and hobby categories, Patalano et al. (2006) operationalized coherence as the extent to which category members share causal features that give rise to surface level features, such as uniformity in behaviors. These researchers found that adults tend to use the high coherence category more often than the low coherence category to make inferences about people who could be cross-classified into both categories. For instance, in this study, adults were told that the majority of feminists (high coherence) prefer Coke to Pepsi, whereas the majority of waiters (low coherence category) prefer the opposite. When asked to predict the drink preferences of someone
belonging to both categories, a feminist supporter who is a waiter, adults made predictions consistent with the high coherence category (predicting a preference for Coke).

To date, research from the conceptual development literature has yet to investigate how category coherence might also play a role in the induction of cross-classified concepts. Despite this dearth of research, category coherence has been found to support category learning with singly classified items, suggesting that coherence is relevant to children’s concepts. Past research has shown that children learn coherent category structures with more ease and accuracy than less coherent ones (e.g., Booth, 2014; Krascum & Andrews, 1998; Murphy, 2002; Nguyen, McCullough, & Noble, 2011). For example, the coherent category structure of basic level categories helps to explain why these categories show an advantage in early category learning (e.g., see Murphy, 2002, for a review). Young children are more apt to learn new labels for experimenter-created, novel objects when the labels reference conceptually coherent categories (Booth, 2014). In addition, coherent explanations such as how the features of fictitious animals are related to their fighting or hiding behaviors have also been found to have a facilitative effect on children’s learning of a novel category (Krascum & Andrews, 1998, see also Hayes & Taplin, 1992). Research has also demonstrated that providing children with coherent information that is organized and linked through causal explanations promotes children’s ability to accurately classify items into real-world categories such as healthy foods (e.g., “healthy foods give your body what it needs because they have vitamins inside,” Nguyen et al., 2011). While this research has shown that coherence influences children’s learning of singly classified items, its influence on children’s induction with cross-classified items remains largely unknown. Results showing that children use coherence would help to elucidate how children narrow down the categories of a cross-classified entity.

Thus, in this research, we explore the influence of coherence on children’s induction with cross-classified entities. That is, when considering cross-classified entities, are children more inclined to make inferences involving highly coherent category memberships compared to less coherent ones? In a series of five studies, 4- and 5-year-old children as well as adults were asked to make predictions about entities that could be cross-classified into coherent and incoherent categories. The task that we devised to test children’s use of coherence was based on Nelson and Miller (1995). These researchers presented adults with the hypothetical situation of buying a gift for a person belonging to two different categories. Since Nelson and Miller (1995) were particularly interested in category distinctiveness (as opposed to coherence), which is the relative number of members that each category possesses, these researchers tested adults on entities belonging to both a distinct category (e.g., a skydive a and nondistinct category (e.g., a tennis player). Adults in this study were asked to decide which gift that person would like the most (e.g., a book about skydiving or a book about playing tennis). The results of Nelson and Miller (1995) showed that participants chose the distinctive choices significantly more often than the nondistinct choices, indicating that the participants utilized the distinctive categories when making their inferences.

In the present research we adapted Nelson and Miller’s (1995) gift giving task to test 4- and 5-year-olds’ use of coherence for induction with cross-classified entities. This age range was selected because of previous findings demonstrating that children are capable of cross-classification at this point in development (Nguyen, 2007; Nguyen & Murphy, 2003). This ability is imperative because the first step in induction is being able to categorize an item so that this category knowledge can then be used to make an inductive inference. Without this initial ability, it would be difficult to expect children to engage in induction. Thus, if children use coherence, then they should tend to select the coherent choices more often than the incoherent ones, which may increase with development.

1. Study 1

1.1. Method

1.1.1. Participants

The participants were 70 children: 30 four-year-olds ($M = 4.70$, range $= 4.43–4.99$, 21 females, 9 males) and 40 five-year-olds ($M = 5.50$, range $= 5.02–6.34$, 22 females, 18 males). Forty adults ($M = 20.8$, range $= 18.58–35.20$, 26 females, 14 males) also participated as a developmental comparison.
Table 1
Coherent and incoherent categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Coherent</th>
<th>Incoherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coherence ratings (M, SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseball player (4.92, 1.73)</td>
<td>Board game player (3.31, 1.68)</td>
<td></td>
</tr>
<tr>
<td>Basketball player (5.43, 1.50)</td>
<td>Movie watcher (3.00, 1.83)</td>
<td></td>
</tr>
<tr>
<td>Sailor (4.93, 1.73)</td>
<td>Toy collector (3.31, 1.38)</td>
<td></td>
</tr>
<tr>
<td>Librarian (5.21, 1.42)</td>
<td>Secretary (4.00, 1.66)</td>
<td></td>
</tr>
<tr>
<td>Ballet dancer (4.71, 1.89)</td>
<td>Sticker collector (3.43, 1.28)</td>
<td></td>
</tr>
<tr>
<td>Distinctiveness ratings (M, SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseball player (2.85, 2.07)</td>
<td>Board game player (3.57, 2.40)</td>
<td></td>
</tr>
<tr>
<td>Basketball player (1.71, 1.13)</td>
<td>Movie watcher (1.07, 2.67)</td>
<td></td>
</tr>
<tr>
<td>Sailor (4.71, 1.68)</td>
<td>Toy collector (5.78, 1.71)</td>
<td></td>
</tr>
<tr>
<td>Librarian (3.21, 1.67)</td>
<td>Secretary (3.14, 1.51)</td>
<td></td>
</tr>
<tr>
<td>Ballet dancer (4.92, 1.63)</td>
<td>Sticker collector (5.71, 1.63)</td>
<td></td>
</tr>
</tbody>
</table>

A separate group of 44 adults (M = 19.58, range = 18.21–34.17, 25 females, 19 males) and 12 children (M = 4.87, range = 4.38–5.53, 4 females, 8 males) also assisted with stimuli preparation and selection. These adults and children did not participate in the actual study.

All of the children and adults were recruited from a predominantly middle-class European-American community located in the Southeastern United States. Children were recruited from local preschools and adults were recruited from a local university.

1.1.2. Materials

The materials included a task with five trials that measured the use of category coherence when making inferences about cross-classifiable entities. Each trial was comprised of a description of a person who could be cross-classified into a coherent (e.g., baseball player) and an incoherent (e.g., board game player) category. The 10 categories were related to jobs and hobbies, inspired by Nelson and Miller (1995) and Patalano et al. (2006). The descriptions were accompanied by black and white drawings of gender neutral faces that were assigned gender neutral names as well drawings representing the coherent and incoherent categories (e.g., a bat for baseball player, dice for a board game player). These drawings were printed on an 8½ × 11 sheet of paper and placed in a 3-ring-binder. Thus, each trial included a category pair. See Table 1 for a complete list.

These pairs were carefully selected based on adult ratings of coherence and distinctiveness. In particular, it was important to ensure that the coherent categories were indeed more coherent than the incoherent categories. It was also important to ensure that distinctiveness, the relative number of members that each category possesses, was not a confound by equating the coherent and incoherent categories on this dimension (see Patalano et al.’s, 2006 criticism of Nelson & Miller, 1995). Specifically, 14 adults provided two types of ratings for each of the categories within the pairs: coherence and distinctiveness. A coherent category was defined to adults as one in which the members are relatively uniform. A distinctive category was defined to adults as one in which members are infrequently encountered. The scales ranged from 1 (not at all coherent/distinctive) to 7 (very coherent/distinctive). The coherence ratings results for each pair in Table 1 indicated that the coherent categories were significantly more coherent than the incoherent categories, t’s(13) > 2, p’s = .001, .000, .006, .015, .030, d’s = 0.96, 1.44, 1.05, 0.78, 0.79, respectively. The distinctiveness ratings results for each pair in Table 1 indicated that the coherent and the incoherent categories did not differ significantly in their distinctiveness, t’s(13) < 2, p’s = .355, .070, .082, .909, .136, respectively.

We also pre-tested a separate group of 30 adults on these pairs by equalizing their property prevalence (see Patalano et al., 2006, Experiment 1). We presented adults with 5 trials, one at a time. For each trial, adults were told that 80% of people who belong to a coherent category and 80% of people who belong to an incoherent category have different book preferences. Then adults were asked which book a person who belongs to both categories would prefer as a gift. For example, adults heard, “80% of baseball players prefer books about baseball to books about board games. Eighty percent of board
game players prefer books about board games to books about baseball. This is Pat. Pat is a baseball player (coherent category) and a board game player (incoherent category). What book does Pat prefer, a book about baseball or a book about board games?” The results showed that adults selected the coherent category 69% of the time, which was significantly above chance, $t(29) = 7.02$, $p < .001$, $d = 1.28$. These results provided us with confidence that any effects that were to be found in the actual study could withstand or survive once property prevalence was equated across categories.

In any test of induction it is also crucial to ascertain that children have knowledge of the categories. To establish that children in the age range tested in this study would be knowledgeable of the categories that were selected from the adult ratings, a group of children participated in a category identification task using the drawings representing these categories. In this task, children were asked to identify which of two choices is associated with a coherent/incoherent category target, a categorically related or unrelated choice (e.g., “This is a bat. This is a set of dice. Point to the picture that goes with a baseball player?”). The results indicated that the children were indeed knowledgeable of the categories, selecting the categorically related choice 87% of the time, which is significantly above chance (50%), $t(11) = 8.16$, $p < .001$, $d = 2.35$.

1.1.3. Procedure

Participants were tested individually for approximately 10 min. Children were tested in a quiet area of their preschool, and adults were tested in a campus laboratory.

Participants were initially told that they were going to play a game about people. Then, participants were presented with five test trials, one at a time. For each trial, participants were presented with a drawing of a person and asked what type of gift they would give to a person if they wanted to be certain the person would like the gift, one based on the coherent or incoherent category. Participants could indicate their response by pointing at one of the two drawings representing the two categories. For example, participants heard, “This is Pat. Pat is a baseball player (coherent category) and a board game player (incoherent category). If you were going to give Pat a gift that you were very sure Pat would like, would you give Pat a book about playing baseball or a book about playing board games?” Thus, each trial, offered participants an opportunity to choose either the coherent or incoherent category choice. If participants have the ability to use category cohesion to make inferences about cross-classifiable items, then they should tend to select the coherent choices. No feedback was given to the participants on any of the trials.

The trials were presented in one of two random orders, and the ordering of the coherent and incoherent choices within each trial was counterbalanced.

1.2. Results and discussion

The data were scored by assigning a 1 when the coherent category was chosen and a 0 when the incoherent category was chosen by the participant. The scores from the five trials were then collapsed into a summary variable and converted into a percentage. A univariate analysis of variance (ANOVA) was conducted with age group as the independent variable and coherent category selection as the dependent variable. Results revealed that coherent category selection varied by age group, $F(2,107) = 10.31$, $p < .001$, $\eta^2_p = .162$. Four-year-olds, 5-year-olds, and adults selected the coherent choice 55%, 61%, and 81% of the time ($SDs = .27%, .24%, .25%$, respectively. Tukey post-hoc comparisons indicated that there was only a significant difference between adults and 4-year-olds, $p < .001$, and between adults and 5-year-olds, $p = .002$.

Each age group’s performance was also compared to chance (50%). Only 5-year-olds and adults performed significantly above chance, $t’s(39) = 2.85, 7.78, p = .007, p < .001, d’s = 0.45, 1.23$, respectively.

Please see Fig. 1. The results of the subsequent studies are also depicted in this figure.

Although the analyses above are informative of performance averaging across participants of each age group, it is important to determine if the results reflect individual participants. To explore individual differences in the data, we also calculated the percentage of each age group showing consistent coherent category responses (i.e., selecting the coherent category choice the majority of the time on three or more of the five trials). The distribution of consistent category responses was relatively similar across the age groups. There was not a significant difference between 4-year-olds (60%) and
5-year-olds (70%) nor between 4-year-olds and adults (85%), Fisher’s exact tests, $p^s = .450, .027$. There was also not a significant difference between 5-year-olds and adults, Fisher’s exact test, $p = .180$.

Bonferroni-adjusted significance level of .0167 was calculated for the Tukey post-hocs, chance comparisons, and Fisher’s exact tests in Study 1 as well as the subsequent studies.

Taken together, the results from the analyses averaging across participants and the analyses examining individual differences provide some evidence that 5-year-olds are systematic in their use of coherence. That evidence appears weaker for 4-year-olds. The individual differences data revealed that there are some individual 4-year-olds who are capable of consistently using coherence in induction, however, when averaging across participants in this age group, their performance as a whole was not above chance.

2. Study 2

Study 2 was designed to address the concern of whether the use of non-neutral properties in Study 1 might be masking 4-year-olds’ abilities. Given that Study 1 used non-neutral properties, the concern is that young children’s inferences may have been guided by their prior knowledge or expectations that link both the coherent and incoherent categories to these properties (cf. Rehder & Hastie, 2004). Specifically, it may be that young children were willing to select the incoherent choice some of the time because they expect that people would also enjoy books related to their less coherent category memberships, not just their coherent ones. For example, even if board game players are a less coherent category than baseball players, they are still united by their love of board games. Therefore, it would be surprising if board game players did not like books about board games. Thus, we used neutral properties to address this potential concern in Study 2.

2.1. Method

2.1.1. Participants

The participants were 18 four-year-olds ($M = 4.18$, range $= 3.66–4.53$, 9 females, 9 males), 18 five-year-olds ($M = 5.40$, range $= 5.00–6.00$, 10 females, 8 males), and 18 adults ($M = 19.24$, range $= 18.26–22.97$, 9 females, 9 males). A separate group of 27 adults ($M = 19.10$, range $= 18.12–22.67$, 13 females, 14 males) who did not participate in the actual study also assisted with stimuli preparation. All of the children and adults were recruited from the same population as Study 1. However, none of these individuals had previously participated in Study 1.
2.1.2. Materials
The materials were similar to Study 1, except that the properties were not books related to category membership. Rather the properties were neutral, and were related to food preferences.

As in Study 1, we pre-tested adults on these materials by equalizing the proportion of people in a given coherent/incoherent category who possess a property. For example, adults heard, “80% of baseball players prefer apples to bananas. Eighty percent of board game players prefer bananas to apples. This is Pat. Pat is a baseball player (coherent category) and a board game player (incoherent category). What fruit does Pat prefer, an apple or banana?” The results showed that adults selected the coherent choices 67% of the time, which was significantly above chance, $t(26) = 4.50, p < .001, d = 0.870$.

2.1.3. Procedure
The procedure was also similar to the one in Study 1 with the exception of neutral properties. For each trial, participants were presented with a drawing of a person and asked what type of food they would prefer, one based on the coherent or incoherent category membership. For example, participants heard, “Baseball players like apples and board game players like bananas. This is Pat. Pat is a baseball player (coherent category) and a board game player (incoherent category). Do you think Pat would like an apple or a banana more?” The food pairs for the other 4 trials were pie-cake, hamburger–hot dog, juice–milk, and waffles–pancakes.

2.2. Results and discussion
The same analytic strategy from Study 1 was applied to Study 2, and revealed a similar pattern of results. As in Study 1, coherent category selection varied by age group, $F(2,51) = 11.56, p < .001$, $\eta^2_p = .312$. Four-year-olds, 5-year-olds, and adults selected the coherent choice 52%, 60%, and 78% of the time ($SD's = 22\%, 14\%, 12\%$), respectively. Tukey post-hoc comparisons indicated that there was only a significant difference between adults and 4-year-olds, $p < .001$, and adults and 5-year-olds, $p = .005$.

Only 5-year-olds and adults, but not 4-year-olds, performed significantly above chance (50%), $t'(17) = 3.09, 10.10, p = .007, p < .001$, $d's = 0.73, 2.38$, respectively.

As in Study 1 we also examined the distribution of consistent category responding (i.e., selecting the coherent category choice on three or more of the five trials). There was not a significant difference between 4-year-olds (56%) and 5-year-olds (78%), Fisher's exact test, $p = .489$. There was, however, a significant difference between 4-year-olds and adults (100%), Fisher's exact test, $p = .003$. There was not a significant difference between 5-year-olds and adults, Fisher’s exact tests, $p = .045$.

Overall, these results based on neutral properties are generally consistent with Study 1, offering some evidence that 5-year-olds are systematic in their use of category coherence to make inferences about cross-classified items. This evidence continues to be weaker for 4-year-olds.

While 4-year-olds' limited performance in Study 1 and 2 may be due to young children's limited ability to use category coherence as a basis for induction with cross-classified items, it is critical to rule out an alternative explanation. As opposed to an inability to use category coherence per se, these results could be due to young children's difficulty with spontaneously invoking coherence information on their own. This difficulty may simply be a consequence of age differences in knowledge and understanding of categories and their coherence. Recall that the induction task in Study 1 and Study 2 required children to engage in a two-step-process (cf. Liu, Gelman, & Wellman, 2007). First, children had to spontaneously cross-classify the person into two categories on their own based on coherence since they were never explicitly told which category was coherent versus incoherent. Then, children had to subsequently use that information to make an inference about the person. If children failed to complete the first step with cross-classification, then it would be unreasonable to expect them to succeed at the second step with induction. Based on the results so far it is difficult to determine with any certainty whether children failed at step 1 or 2.

3. Study 3
Study 3 was designed to reexamine 4-year-olds' ability to use category coherence to make inferences about cross-classified items. To address the issue of young children's potential difficulty with
spontaneously invoking coherence information, Study 3 used a modified version of the induction task from Study 1 and 2. The first modification was simplifying the task into a single step. Children were given explicit category information up front, and were required to only make an inference based on this category information. This way, children did not have to spontaneously cross-classify before engaging in induction. Thus, this modified task could more directly test young children’s ability to use category coherence to reason about cross-classified items.

The second modification involved making the task friendlier for young children. Recall that the gift-giving scenario in Study 1 and 2 was based on Nelson and Miller (1995), who conducted their research on adults. To create a more fun, child-friendly scenario, in Study 3, children were told that they would be playing a game about aliens. This game-like approach with aliens was informed by past cognitive development research that has used descriptions of aliens as a part of their methodology (e.g., Livingston & Andrews, 2005; Notaro, Gelman, & Zimmerman, 2001). Thus, this version of the induction task was situated in a game-like context that was age-appropriate for young children.

If 4-year-olds are indeed capable of using coherence when making inferences about cross-classified items, then this modified version of the induction task should begin to reveal this ability in young children.

3.1. Method

3.1.1. Participants

Participants consisted of 25 four-year-olds ($M = 4.62$, range = 4.13–4.95, 11 females, 14 males), 25 five-year-olds ($M = 5.32$, range = 5.03–5.75, 13 females, 12 males), and 25 adults ($M = 19.76$, range = 18.10–26.89, 1 female, 24 males). Participants were recruited from the same community as Study 1 and 2, but none of these individuals had previously been tested in these studies.

3.1.2. Materials

The materials were the same as in Study 1, except that antennae were added to the gender-neutral faces. In addition, there was a drawing of an alien planet, named “Zog.”

3.1.3. Procedure

The procedure was similar to Study 1, but with the two modifications to the task. Specifically, participants were initially told that they would be playing a game about aliens living on a planet called “Zog.” Participants were also shown a drawing of this alien planet. Next, participants were presented with five test trials, one at a time. For each trial, children were asked what type of job an alien should do if they want to be the same as the other aliens, a coherent or incoherent job. In each trial, the coherent and incoherent categories were explicitly identified for participants using child-friendly language. For example, participants heard, “Pat is a baseball player (coherent category). Aliens who are baseball players are all the same. Pat is also a board game player (incoherent category). Aliens who are board game players are not the same. Today, Pat wants to be the same as the other aliens. So, should Pat be a baseball player or a board game player?”

As in Study 1, the trials were presented in one of two random orders, and the ordering of the coherent and incoherent choices within each trial was counterbalanced.

3.2. Results and discussion

The data were scored and analyzed as in the previous studies. The results of the univariate analysis of variance (ANOVA) indicated that coherent category selection varied by age group, $F(2,72) = 5.38$, $p = .007$, $η^2_\text{p} = .130$. Four-year-olds, 5-year-olds, and adults selected the coherent choice 80%, 83%, and 98% of the time (SD’s = 27%, 23%, 5%), respectively. Only 4-year-olds and adults differed significantly from each other, Tukey post-hoc, $p = .01$.

Each age group’s performance was also compared to chance (50%). All of the age groups, 4-year-olds, 5-year-olds, and adults, performed significantly above chance, $t$’s(24) = 5.75, 7.26, 43.70, $p$’s < .001, $d$’s = 1.15, 1.45, 8.73, respectively.
An independent sample t-test was also conducted to compare the performance of the 4-year-olds' in Studies 1 and 3. The results showed that 4-year-olds in Study 3 (M = 80%, SD = 27%) performed significantly better than 4-year-olds in Study 1 (M = 55%, SD = 27%), t(53) = 3.48, p = .001, d = .945. This increased level of performance was also apparent in our exploration of individual differences in Study 3. There was a 28% increase in the number of 4-year-olds who were consistently selecting the coherent category choice on three or more of the five trials in Study 3 (88%) compared to Study 1 (60%). Overall, there was not a significant difference in the distribution of consistent responding between 4-year-olds and 5-year-olds (92%), 4-year-olds and adults (100%), and 5-year-olds and adults, Fisher's exact tests, p's = 1.00, .235, .490, respectively.

Thus, these results begin to suggest that 4-year-olds' performance in Study 1 may be due to young children's difficulty with spontaneously invoking category coherence information, as opposed to an inability to use category coherence for induction. Yet, an alternative explanation for the results could be that children were not engaging in induction at all, but were rather just matching up the word “same” when it was used both in the prompt and in the category description (cf. Gelman & Markman, 1986). This matching process could lead to the identical results seen here in this study. If this is the case, then coherence may not have been guiding children's choices at all. On the one hand, this strategy seems highly unlikely since the word “same” appeared in both the coherent and incoherent descriptions. If the presence of this word affected children's responding, then children would have been equally likely to select the coherent and incoherent responses, which was not the case.

On the other hand, some of the children may have detected the negation of the word “same” in the description of the incoherent category, which may have also lead to higher matching with the coherent response. Study 4 was designed to rule out this possibility by testing children on a different version of the test question that did not use the word “same,” but still carried the same meaning.

4. Study 4

The aim of Study 4 was to address the concern that children in Study 3 were merely matching up the word “same” in the prompt and in the category description. Thus, in this study, the word “same” was replaced with “just like” in the test question to eliminate the possibility of children engaging in a matching processing, as opposed to induction. If children are capable of using coherence, then they should tend to select the coherent choices during testing, even with this modified wording.

4.1. Method

4.1.1. Participants

Participants consisted of 18 four-year-olds (M = 4.35, range = 3.87–4.99, 10 females, 8 males) and 18 five-year-olds (M = 5.65, range = 5.02–6.28, 11 females, 7 males). Children were recruited from the same community as in the prior studies, but had not participated in any of these studies.

4.1.2. Materials

The materials were the same as in Study 3.

4.1.3. Procedure

The overall procedure was identical to Study 3, except that the word “same” was replaced with “just like” in the test question. For example, participants heard, “Pat is a baseball player (coherent category). Aliens who are baseball players are all the same. Pat is also a board game player (incoherent category). Aliens who are board game players are not the same. Today, Pat wants to be just like other aliens. So, should Pat be a baseball player or a board game player?”

4.2. Results and discussion

The results showed that 4-year-olds and 5-year-olds selected the coherent choice 86% and 89% of the time (SD’s = 21%, 18%), respectively. There was not a significant difference in 4- and 5-year-olds'
selection of the coherent choice, as indicated by an independent sample t-test, $t(34) = 0.50, p = .621$. Both 4-year-olds’ and 5-year-olds’ level of performance was significantly above chance, $t's(17) = 7.02, 8.95, p's < .001, d's = 1.65, 2.10$, respectively.

Individual differences in the data were also explored using the same criteria as in the previous studies. The distribution of consistent responding did not vary between 4-year-olds (89%) and 5-year-olds (94%), Fisher’s exact test, $p = 1.00$.

In all, these results suggest that the children in Study 3 were not merely matching the word “same” in the category description and test question. When the word “same” was replaced with “just like,” children still tended to select the coherent choices.

Of course we cannot isolate what precisely helped 4-year-olds in Study 3 and 4 since the modifications to these studies entailed two components, including a simplified task and child-friendly language. It may be that one, or both these changes facilitated young children’s performance. It could also be argued that these changes consequently altered the task such that it was no longer a test of induction per se, but rather a test of categorization. Typically, studies of induction involve a premise category, a potential property, and then some conclusion category or individual, which participants must make an inference about (see Gelman & Davidson, 2013; Hayes, 2007; Murphy, 2002, for a review). In Study 3 and 4, children were presented with two categories that differed in their coherence, and were asked which category an individual should belong to if they wanted to be a part of the coherent category. Thus, this methodology arguably resembles more a categorization task rather than an inductive inference task.

5. Study 5

In our final study we address the concern that Study 3 and 4 may not necessarily test induction, but rather categorization. Therefore, in Study 5 we attempted to provide a clearer test of whether children use coherence for inductive inference by employing a task closer to Study 1, but with the added information about coherence from Study 3.

5.1. Method

5.1.1. Participants

The participants were 18 four-year-olds ($M = 4.41$, range = 4.00–4.97, 8 females, 10 males), 18 five-year-olds ($M = 5.28$, range = 5.00–6.15, 9 females, 9 males), and 18 adults ($M = 20.73$, range = 18.79–27.59, 10 female, 8 males).

5.1.2. Materials

The materials were the same as those in Study 3.

5.1.3. Procedure

The procedure was based on aspects of the procedures in Study 1 and Study 3. In particular, participants were introduced to five aliens, one at a time, and were asked to make an inference about the type of book that the alien would most prefer as a gift. For example, participants heard, “This is Pat. Pat is a baseball player (coherent category). Aliens who are baseball players are all the same. Pat is also a board game player (incoherent category). Aliens who are board game players are not the same. If you were going to give Pat a gift that you were very sure Pat would like, would you give Pat a book about playing baseball or a book about playing board games?”

5.2. Results and discussion

The data were scored and analyzed as in the previous studies. The results of a one-way ANOVA indicated that coherent category selection varied by age group, $F(2, 51) = 11.39, p < .001, \eta^2_g = .309$. Four-year-olds, 5-year-olds, and adults selected the coherent choice 52%, 67%, and 89% of the time ($SD's = 23%, 27%, 21%$), respectively. The only significant age difference was between adults and 4-year-olds, Tukey post-hoc, $p < .001$. 
Each age group's performance was also compared to chance (50%). Only 5-year-olds and adults performed significantly above chance, $t$'s(17) = 2.75, 7.92, $p = .014, .001$, $d's = 0.65, 1.86$, respectively.

To explore individual differences in the data, as in the previous studies, we examined the percentage of each age group showing consistent coherent category responses (i.e., selecting the coherent category choice the majority of the time on three or more of the five trials). The results indicated that there was not a significant difference in the distribution of consistent coherent category responding between the 4-year-olds (56%) and 5-year-olds (61%), 4-year-olds and adults (89%), and 5-year-olds and adults, Fisher’s exact tests, $p's = 1.00, .235, .490$, respectively.

The decreased performance of 4-year-olds in Study 5 confirms the concern that Study 3 and 4 may have been more of a test of categorization, rather than induction. In Study 5, four-year-olds' performance declined to a similar level as in Study 1 and 2 (55%, 52%), despite having above chance performance in Study 3 and 4 (80%, 86%). The different task demands in each study shed light on the change in 4-year-olds’ performance. The tasks in Study 3 and 4 appear to have evoked only a categorization judgment (i.e., given that two categories differ in coherence, which category is a new individual most likely to belong to?). The task in Study 5, however, involved an extra inductive step that required children to use coherence to infer a preferred gift. Although the relative coherence of the two categories was stated in Study 5, as in Study 3 and 4, this information may not have been enough to assist 4-year-olds in taking this extra inductive step.

6. General discussion

The aim of this research was to investigate how category coherence influences children’s induction with cross-classified entities. We predicted that if children use category coherence to guide their inductive inferences with cross-classified entities, then they should favor the coherent choices over the incoherent choices in the induction task. Overall, the findings of the present investigation offer evidence that 5-year-olds and adults are systematic in their use of coherence as a basis for induction with cross-classified entities. The evidence is weaker for 4-year-olds. Specifically, Study 1, 2, and 5 found that five-year-olds and adults were reliably above chance whereas 4-year-olds were at chance in their coherent category selection. However, it should be pointed out that there was generally not a difference in the mean level of performance between 4- and 5-year-olds in Study 1, 2, and 5. This was also case when individual differences were examined in Study 1, 2 and 5. It appears that at least half of the 4-year-olds were consistently selecting the coherent category choices in these studies. This pattern of responding suggests that some 4-year-olds do understand the implications that their knowledge of coherence (as seen in Study 3 and 4) has for making predictions about category members.

Taken together, these novel results make a major contribution to our emerging understanding of how and why children select certain categories versus others during induction with cross-classified items. To our knowledge, the research reported here is the first of its kind to demonstrate the valuable role that coherence plays in children’s inductive reasoning with cross-classified concepts. The results of this research extend previous developmental research that has found that coherence supports children’s learning of single categories (e.g., Booth, 2014; Krascum & Andrews, 1998; Murphy, 2002; Nguyen et al., 2011). Our findings are consistent with the adult concepts literature, and add a developmental aspect to this body of research, lending further support for the claim that coherence affects beliefs about what makes the categories of a cross-classified concept informative for induction (Hayes et al., 2011; Patalano et al., 2006, 2009).

Thus far, we have interpreted the findings of the current investigation as suggesting that coherence influences children’s inductive reasoning with cross-classified entities. However, it is important to pause and consider alternative explanations. Arguably, other confounding factors could account for the results found in this investigation. First, it could be argued that similarity, the resemblance of an entity’s properties to a category, but not coherence, was actually driving the results. This argument was raised by Patalano et al. (2006) as an alternative explanation for their results with adults. This argument is applicable to our developmental data as well, since the principle of homogeneity does make the prediction that children will be more inclined to generalize to the same person rather than to a different person (Gelman, 1988). When it comes to inductive reasoning with single categories, children are in fact more willing to make generalizations within groups that are high on similarity, compared to
groups that are low on this dimension. That is, more inferences are made within homogenous groups with similar cases as opposed to dissimilar ones (e.g., Gelman, 1988; Lawson & Kalish, 2006; Rhodes & Brickman, 2010). A number of studies have shown that that all things being equal, young children tend to overestimate category homogeneity, expecting that category members will be very similar in many respects (e.g., Rhodes & Brickman, 2010; Rhodes, Brickman, & Gelman, 2008; Rhodes, Gelman, & Brickman, 2008).

The principle of homogeneity may very well provide a partial explanation for some of the results, as coherence is strongly associated with this factor (see Patalano et al., 2006). If the principle of homogeneity was operating in the present investigation, then we would have expected young children to have a default expectation that both coherent and incoherent category choices have high coherence, resulting in chance level performance. This was indeed the case for 4-year-olds. However, a strong argument against this principle as providing a full explanation of the present results is based on differences in the developmental trajectories. Rhodes and Brickman (2010), for example, have found that children only show a reduction in the homogeneity assumption, and begin to use category diversity in induction when primed at approximately 7 years of age. In the present research, however, there is evidence of reliable use of coherence at 5 years.

With that said, it is crucial to emphasize that this explanation does not discount our findings since judgments about similarity are not just influenced by common properties alone, but rather are also strongly influenced by common relations among members of a category and/or the properties of a category, in other words, coherence. While the job and hobby categories in the present research allowed us an opportunity to study meaningful, real world categories that are rich in cross-classification potential, we unfortunately could not manipulate their category structure. In future research, to help tease apart similarity and coherence, it may be helpful to use artificial, experimenter-generated categories that allow for increased control and manipulation of these factors (see Patalano et al., 2006).

Secondly, it could also be argued that the results of the current investigation reflect a bias against incoherent categories, as opposed to participants favoring coherent categories for their strong inductive potential. Similar to the last argument, this one was raised by Patalano et al. (2009) as an alternative explanation for their results with adults, which again applies to our developmental data. In contrast with coherent categories, incoherent ones include members of a category and/or the properties of a category that do not go together or make sense in light of one’s existing knowledge. Given the disjointed nature of incoherent categories, it could be argued that participants in the current research were merely selecting the coherent category choices in an effort to avoid the incoherent ones. In other words, arguably, the findings presented here are not necessarily unique to induction with cross-classified entities given that participants could have selected the coherent choices under any circumstance, regardless if they were making an inference. This possibility seems highly unlikely because a priori reasons exist for why participants might be drawn to incoherent category choices. For example, participants’ attention may be directed to incoherent categories because of the salience of their disjointed category structure. Also, participants may perceive the breadth of category members and properties that fall within the scope of an incoherent category as functional for induction. Moreover, there may be a preference for incoherent category choices over coherent ones in some situations where the property to be generalized is most relevant to the incoherent category (see Patalano et al., 2009). In future research, it would be useful to ask participants to provide explanations for their category selections. These explanations would shed light on the degree to which their category selections are affected by coherent category structure.

7. Conclusions

Cross-classified concepts are ubiquitous in our everyday lives with most entities sharing membership in more than one category as in the case of a person who may potentially belong to many categories based on their hobbies, jobs, gender, age, etc. Cross-classification raises central questions about children’s inductive reasoning, especially how and why children select some categories, but not others for induction with cross-classified entities. The present research offers new insight into this poorly understood and understudied issue within the conceptual development literature. Here, we
document for the first time, to our knowledge, the contribution of category coherence to children’s inductive reasoning with cross-classified entities.

It is important to note that our findings do not assume that coherence solely influences category-based induction, but rather suggest that coherence contributes to this complex process (see also Patalano et al., 2009). Thus, these findings do not intend to diminish the research on property relevance or even singly classified items. In this investigation we were primarily interested in examining whether children have any capacity at all to use coherence as a basis for induction in cross-classification situations, as opposed to making any direct comparisons between property relevance or singly classified items. To have found that children do have this capacity is critical as it speaks to what influences children’s inductive reasoning in cross-classification situations. We believe that these results provide a strong groundwork for future studies to continue exploring varying factors that might also impact children’s inductive reasoning within cross-classification contexts (e.g., attributional cues such as consensus, see Kelley, 1972, 1973; Schuster, Ruble, & Weinert, 1998; generalization of social norms versus psychological states, see Kalish, 2012; Kalish & Lawson, 2008; Kalish & Shiverick, 2004; sample size and sample diversity, see Lawson, 2014; Lawson & Fisher, 2011).

Acknowledgements

We thank Cameron Gordon and Brett Hayes for their discussions and encouragement regarding this research. We are also thankful to Jennifer Huff, Bethany Kratoska, Julie Robinson, Bethany Wentworth, Scott Ferris, Tonya Vandenbrink, and Catherine McDermott for their research assistance, as well as all of the participants, schools, and families for their invaluable help.

References


