Development of category-based reasoning in 4- to 7-year-old children: The influence of label co-occurrence and kinship knowledge

Karrie E. Godwin *, Bryan J. Matlen, Anna V. Fisher *

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Article info

Article history:
Received 7 March 2012
Revised 1 November 2012
Available online 29 January 2013

Keywords:
Induction
Category-based reasoning
Linguistic labels
Cognitive development
Kinship knowledge
Semantic development

Abstract

Category-based reasoning is central to mature cognition; however, the developmental course of this ability remains contested. One strong indicator of category-based reasoning is the propensity to make inferences based on semantically similar labels. Recent evidence indicates that in preschool-age children the effects of semantically similar labels are limited to a small subset of labels that co-occur in child-directed speech, suggesting that performance with these labels may reflect lexical priming rather than category-based reasoning. However, most co-occurring labels used in prior research refer to offspring–parent relationships (e.g., puppy–dog). Thus, it is possible that children in previous research performed induction by relying on kinship rather than co-occurrence information. To address this possibility, the current studies examined the role of kinship knowledge and label co-occurrence in induction in 4- to 7-year-old children and adults. The results point to a gradual age-related increase in the ability to spontaneously rely on kinship knowledge when making inferences.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Category-based reasoning is central to mature cognition and underlies much of our learning and functioning in the world (e.g., Heit & Rubinstein, 1994; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993; Yamauchi & Markman, 2000). For example, on learning that English Setters have 39 pairs of chromosomes, we may conclude (without explicitly being told) that Dalmatians also have...
39 pairs of chromosomes because English Setters and Dalmatians are the same kind of animal (i.e., both are a kind of dog). However, the developmental course of this fundamental ability remains contested.

A considerable body of prior research suggests that young children, including infants, are capable of spontaneous category-based reasoning and that category labels promote this type of reasoning (Gelman, 1988; Gelman & Coley, 1990; Gelman & Markman, 1986; Jaswal, 2004; Jaswal & Markman, 2007; Welder & Graham, 2001). The strongest evidence in support of this possibility comes from a study demonstrating that preschool-age children make inferences based not only on identical category labels but also on semantically similar labels (henceforth referred to as synonyms for brevity) (Gelman & Markman, 1986). In that study, children were presented with a triad of objects and provided with respective category labels. For example, children could be presented with a rabbit (target item), a squirrel (Test Item 1) that was designed to look similar to the target, and another rabbit (Test Item 2) that was designed to look dissimilar from the target. Children were told about the properties of each test item (e.g., the rabbit eats grass and the squirrel eats bugs). Children were then asked to generalize one of these properties to the target item. Importantly, similarity in category membership was conveyed either by identical labels (e.g., rabbit–rabbit) or by synonymous labels (e.g., bunny–rabbit). The results indicated that the rate of category-based inferences was above chance in both conditions (i.e., 67% with identical labels and 63% with synonymous labels), suggesting that children use category labels to guide their inductive inferences.

However, recent findings (Fisher, 2010; Fisher, Matlen, & Godwin, 2011) suggest that preschoolers’ ability to make inferences using synonyms is limited to a small set of words that not only share meaning but also co-occur in child-directed speech. For example, Fisher et al. (2011) revisited the question of whether young children engage in category-based reasoning with synonymous labels by analyzing children’s responses separately for co-occurring synonyms (e.g., puppy–dog, kitty–cat) and non-co-occurring synonyms (e.g., crocodile–alligator, mouse–rat). Consistent with the possibility that children’s performance in earlier studies stemmed from label co-occurrence rather than semantic similarity, the results of that study indicated that most 4-year-olds performed at above chance in making category-based inferences with co-occurring synonyms, but these same children did not systematically select category choices when presented with non-co-occurring synonyms.

Fisher et al. (2011) hypothesized that these effects stemmed from lexical priming. Specifically, prior research suggests that co-occurrence plays an important role in the formation of lexical associations (Brown & Berko, 1960; McKoon & Ratcliff, 1992; Spence & Owens, 1990). Strong lexical associations between co-occurring labels may facilitate inductive generalization via priming. For example, when children are asked whether a “bunny” shares a property with a “rabbit” or a “squirrel,” they may select the category match (i.e., “bunny”) due to lexical priming rather than category-based reasoning. Based on current evidence, the co-occurrence hypothesis offers a plausible explanation for the observed pattern of results.

However, there is an alternative explanation consistent with the notion that young children are capable of engaging in category-based reasoning with semantically similar labels. Prior studies have identified only a few semantically similar labels that not only are familiar to preschool-age children but also co-occur in child-directed speech in the English language according to the CHILDES database (MacWhinney, 2000). Incidentally, these words can be construed as referring to offspring–parent relationships (e.g., puppy–dog, kitty–cat, bunny–rabbit). Therefore, it is possible that children’s induction with these labels is driven by the knowledge of kinship information rather than label co-occurrence. In other words, children may spontaneously engage in category-based reasoning when they are

---

1 Although kitty and bunny are not labels exclusively used for baby animals, these labels are often used to refer to the young of these species. For instance, the Merriam–Webster dictionary defines bunny as a “rabbit; especially young rabbit” and kitty as a “cat; especially kitten.” Common use of these words seems consistent with the dictionary definitions; a Google picture search using the term “bunny” yielded 74 animal images of which 62% depicted young rabbits, and a search using the term “kitty” yielded 79 animal images of which 49% depicted kittens. Conversely, a Google picture search using the term “rabbit” yielded 151 animal images of which 80% depicted mature rabbits, and a search using the term “cat” yielded 214 animal images of which 74% depicted mature cats. Therefore, it is reasonable that children may interpret bunny and kitty as words that refer to the young of the species.
presented with semantically similar labels, but they do so only when these labels refer to kinship relationships.

There is evidence suggesting that children may be sensitive to kinship information and that children can use this information during the course of inductive generalization (Bullock & Opfer, 2009; Opfer & Bullock, 2007; Springer, 1992; Springer & Keil, 1991). For example, Springer (1992) showed that preschool-age children generalize properties such as skin color on the basis of perceptual similarity in the absence of kinship information, yet when kinship information was provided children preferred to generalize on the basis of kinship. Moreover, in Springer’s study children considered kinship to be a more important factor in biological inheritance than friendship. These and other related findings suggest that kinship information can aid young children’s inductive generalization.

The current studies were designed to examine whether children’s performance in prior research (Fisher, 2010; Fisher et al., 2011) was driven by co-occurrence probability of semantically similar labels or kinship information conveyed by these labels. Toward this goal, we presented participants with an induction task with semantically similar labels pointing to kinship relations. Some of these labels were likely to co-occur in child-directed speech (e.g., puppy–dog), whereas other labels were unlikely to co-occur (e.g., chick–hen). Importantly, co-occurrence probability of labels was manipulated within participants, with co-occurring and non-co-occurring trials intermixed (rather than blocked, as in Fisher et al., 2011). This design was advantageous because (a) any observed differences could be attributed to the experimental manipulation rather than to differences in generalization strategies between participants and (b) it allowed a more straightforward comparison with the seminal study by Gelman and Markman (1986) in which co-occurring and non-co-occurring label pairs were intermixed. In Study 1, we examined whether adults capitalize on kinship information conveyed by co-occurring and non-co-occurring label pairs when performing induction. In Study 2, we investigated the developmental trajectory of category-based induction with co-occurring and non-co-occurring kinship labels in 4-, 5-, and 7-year-old children.

Study 1

Methods

Participants

Participants were undergraduate students from a local university (N = 20, mean age = 19.93 years, SD = 1.03) who received partial course credit for participation.

Design

The current experiment used co-occurrence probability as the within-participant factor; participants performed induction with both co-occurring and non-co-occurring labels. The use of a within-participant design is a strength of this study because some of the previous work manipulated co-occurrence as a between-participant factor (Fisher, 2010). In Fisher et al. (2011), co-occurrence was manipulated within participants; however, in that study co-occurrence was blocked such that some participants first performed inferences with co-occurring label pairs and then with non-co-occurring label pairs, and the rest of the participants performed the tasks in reverse order. In the current study, co-occurring and non-co-occurring labels were intermixed, thereby reducing any possible linguistic cohort effects.

Materials

Verbal stimuli consisted of eight label triads. Each triad consisted of a target word, a category choice, and an unrelated lure. The properties that participants were asked to generalize during the induction task consisted of two-syllable blank predicates. A full list of linguistic stimuli is provided in Table 1.

Visual stimuli consisted of sets of three identical doors. Participants were told that the objects were hiding behind each of the doors. The objects themselves were not depicted. This procedure was used to eliminate conflict between labels and appearances and, thus, to encourage reliance on category
information conveyed by the labels. The presence of conflict among stimuli dimensions is known to impose considerable performance costs for children as well as adults (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Diamond, 2006; Fan, McCandliss, Thomas, & Posner, 2003; Rueda et al., 2004); therefore, eliminating task conflict should facilitate reliance on semantic similarity because it is the only source of information available in the task. These same visual stimuli and procedures have been successfully used in prior research (Fisher et al., 2011).

An additional set of eight pictures was used for the kinship knowledge task, which assessed knowledge of biological inheritance. A detailed description of these tasks is provided in the “Procedure” section below.

Label selection. Label selection was based on a separate calibration study. Participants were a sample of 4- and 5-year-old children (N = 16, mean age = 5.20 years, SD = 0.50, 8 girls and 8 boys) who participated in a picture identification task similar to the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). Prior research has established that preschool-age children are highly familiar with the co-occurring label pairs used in the current study (Fisher, 2010; Fisher et al., 2011); therefore, co-occurring labels were not included in the calibration study. Verbal stimuli consisted of 22 non-co-occurring label pairs that referred to kin relations. Visual stimuli consisted of pictures presented on a computer screen. Children were asked to select the animal labeled by the experimenter from four pictorial response options (see Fig. 1). The picture identification task served to assess children’s familiarity with the labels. Only non-co-occurring word pairs for which children exhibited high levels of accuracy were selected for the experiment proper (M = .94, SD = .07, range = .75–1.00 for the selected labels; see Fig. 2). The final list of stimuli consisted of three co-occurring semantically similar word pairs referring to kinship relations (bunny–rabbit, puppy–dog, and kitty–cat) and five non-co-occurring semantically similar word pairs referring to kinship relations (caterpillar–butterfly, chick-hen, tadpole–frog, lamb–sheep, and lion–cub).

The asymmetry in the number of co-occurring and non-co-occurring trials was a result of natural language limitations. As mentioned previously, prior research has identified only a few semantically similar label pairs that are familiar to children and co-occur in child-directed speech. The added constraint of identifying highly familiar co-occurring word pairs that also refer to kinship relations further restricted the pool of available natural language stimuli. Consequently, there were only three co-occurring semantically similar label pairs in the co-occurring condition. However, it should be noted that presenting a small number of co-occurring label pairs among a larger number of non-co-occurring label pairs could dilute any possible effects of label co-occurrence on inductive generalization. Therefore, observing the effect of label co-occurrence on inductive inferences of preschoolers under these conditions would be particularly notable.

Similar to prior studies (Fisher, 2010; Fisher et al., 2011), co-occurrence probability was analyzed using CHILDES, a corpus of child speech and child-directed speech (MacWhinney, 2000).² Five databases in the corpus were analyzed: the Bates, Brown, Gleason, HSLLD, and Wells databases. Labels were considered to be co-occurring when one of the words in a pair immediately preceded the other word in

<table>
<thead>
<tr>
<th>Target word</th>
<th>Category choice</th>
<th>Lure</th>
<th>To-be-generalized property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bunny</td>
<td>Rabbit</td>
<td>Squirrel</td>
<td>Creighan</td>
</tr>
<tr>
<td>Kitty</td>
<td>Cat</td>
<td>Fox</td>
<td>Manchin</td>
</tr>
<tr>
<td>Puppy</td>
<td>Dog</td>
<td>Bear</td>
<td>Erwin</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>Butterfly</td>
<td>Ladybug</td>
<td>Higa</td>
</tr>
<tr>
<td>Lion</td>
<td>Cub</td>
<td>Pig</td>
<td>Matlen</td>
</tr>
<tr>
<td>Lamb</td>
<td>Sheep</td>
<td>Cow</td>
<td>Koski</td>
</tr>
<tr>
<td>Chick</td>
<td>Hen</td>
<td>Mouse</td>
<td>Troxel</td>
</tr>
<tr>
<td>Tadpole</td>
<td>Frog</td>
<td>Fish</td>
<td>Omat</td>
</tr>
</tbody>
</table>

² Data were retrieved from CHILDES between May and July 2012.
that pair within the same utterance. Raw co-occurrence frequencies were normalized using the Jaccard index (Van Eck & Waltman, 2009); the number of raw co-occurrences was divided by the sum of each word occurring individually minus the number of times the two words co-occurred. Based on the analysis of the CHILDES corpus, the mean co-occurrence probability of labels in the co-occurring condition was .04 and in the non-co-occurring condition was .00; this difference was statistically significant, independent-samples $t(6) = 3.16, p = .02$.

A separate calibration study, with a sample of 20 undergraduate students, was conducted to validate the semantic similarity of the labels selected for the experiment proper. For each label triad, participants were asked to rate the degree to which the words shared meaning. Participants used a 7-point Likert scale to rate semantic similarity, where 7 indicated that the words have identical or nearly identical meaning and may be used interchangeably (e.g., pants–trousers) and 1 indicated that
the words are very dissimilar in meaning (e.g., wall–zebra). For each label triad, participants were asked to rate the semantic similarity among the target object, category match, and lure (e.g., lamb–sheep, lamb–cow, and sheep–cow). The label pairs were presented in a pseudo-random order, with a constraint that no word could occur on two consecutive trials. Results of the calibration study confirmed that related test items (i.e., the target and category match) were more semantically similar ($M = 5.71, SD = 0.83$) than unrelated test items (i.e., the target/category match and lure) ($M = 2.96, SD = 1.16$), paired-samples $t(19) = 14.97, p < .0001$. The same pattern of results was obtained when the analysis was conducted separately for each co-occurrence condition (non-co-occurring labels: related test items $M = 5.07, SD = 1.14$, unrelated test items $M = 2.86, SD = 1.13$, $t(19) = 13.05, p < .0001$; co-occurring labels: related test items $M = 6.34, SD = 0.77$, unrelated test items $M = 3.06, SD = 1.21$, $t(19) = 12.52, p < .0001$). It is worth noting that the related items in the co-occurring condition were judged to be more semantically similar than the related items in the non-co-occurring condition, paired-samples $t(19) = 5.61, p < .001$. We return to this issue in the General Discussion.

**Procedure**

Adult participants were tested in a laboratory on campus. Visual stimuli were presented on a computer, and instructions and labels were presented verbally by experimenters who were blind to the hypothesis.

**Induction task.** The induction task consisted of eight trials, with each trial consisting of a target item, a category choice, and an unrelated lure; the target was the baby animal (e.g., puppy), the category-choice was the parent animal (e.g., dog), and the lure was an unrelated animal (e.g., bear). Participants were told that the target had a novel property and were asked to generalize the property to one of the test items. Induction trials were presented in one of two pseudo-random orders; all label pairs were intermixed for presentation in Order 1 (with the stipulation that the three co-occurring word pairs could not occur in succession), and the order of the stimuli was reversed in Order 2.

Participants were told that they would play a game in which objects were hiding behind doors. The experimenter told participants what was hiding behind each door and then asked them a question. On every trial, the target object was hidden behind the top-most door. The locations of the response options were randomized across trials (e.g., to the right or left of the target). On every trial, the experimenter pointed to the top-most door first and told participants what was hiding behind the door (e.g., “There is a puppy hiding behind this door”). Then, the experimenter disclosed what was hiding behind the remaining two doors (e.g., “There is a bear hiding behind this door” and “There is a dog hiding behind this door”). The presentation order of the category choice first or lure first was randomized across trials. Subsequently, the experimenter asked participants to infer which object (category choice or lure) shared the same property with the target object (e.g., “This puppy has erwin inside; do you think the dog behind this door or the bear behind this door also has erwin inside like this puppy?”). A schematic description of the induction task is presented in Fig. 3.

**Kinship knowledge task.** A kinship knowledge task was administered immediately after the experiment proper to ensure that participants had the prerequisite knowledge of biological inheritance to perform category-based induction. Participants were told that the baby animal was hiding behind a rock. The baby animal was never depicted to prevent participants from selecting a response based on perceptual similarity rather than kinship knowledge. For each baby animal, participants were asked to select the corresponding “mother” from two pictorial response options (i.e., the category choice and lure item from the induction task). Fig. 4 presents a schematic description of the kinship knowledge task.

---

3 This was true for all label pairs except the lion–cub pair, in which lion was the target and cub was the category choice. This was done because (a) the analysis of the CHILDES database proceeded in parallel with data collection and (b) we hypothesized that if these two labels co-occurred, the direction of co-occurrence would be lion–cub rather than cub–lion. However, the results of the CHILDES analysis indicated that the words lion and cub did not co-occur in either order.
Fig. 3. Schematic depiction of the induction task. All instructions were given verbally by the experimenter.
Results and discussion

Induction accuracy

Proportions of category-based responses were compared with chance level (.50) using single-sample *t* tests. Adults’ induction performance was above chance regardless of co-occurrence probability; adults averaged 92% and 93% of category-based responses in the co-occurring and non-co-occurring conditions, respectively (all *ts* > 8.75, all *ps* < .0001). There was no significant difference in the proportion of category-based responses in the co-occurring and non-co-occurring conditions, paired-samples *t*(19) = 0.53, *p* = .60.

Kinship knowledge accuracy

Adults obtained mean scores significantly above chance (.50) in both co-occurrence conditions (all single-sample *ts* > 14.78, all *ps* < .0001). Adults obtained statistically equivalent scores on the kinship knowledge task in both the co-occurring and non-co-occurring conditions (100% and 96%, respectively), paired-samples *t*(19) = 1.28, *p* = .21.

These results suggest that adults spontaneously use kinship information conveyed by semantically similar labels. Importantly, adult participants can do so with both co-occurring and non-co-occurring labels, as evidenced by their near-ceiling selections of the category match in both co-occurrence conditions. Although the calibration results indicated that semantic similarity ratings were stronger for co-occurring label pairs than for non-co-occurring label pairs, adults exhibited identical generalization patterns across both conditions, suggesting that in this task mature induction is not significantly affected by the differences in semantic similarity ratings between the two conditions. In other words, even though adults judged co-occurring synonyms to be somewhat more semantically similar than non-co-occurring synonyms, the latter were sufficiently semantically similar to support inductive generalizations from the target to the related test item (e.g., from *chick* to *hen*) rather than to the lure (e.g., *mouse*) at a near-ceiling level. Study 2 explored whether 4- to 7-year-old children also spontaneously use kinship information conveyed by linguistic labels during the course of inductive reasoning.

Study 2

Methods

Participants

Participants were 20 4-year-olds (mean age = 4.48 years, *SD* = 0.25, 8 girls and 12 boys), 20 5-year-olds (mean age = 5.28 years, *SD* = 0.21, 13 girls and 7 boys), and 24 7-year-olds.
(mean age = 7.06 years, SD = 0.38, 11 girls and 13 boys) recruited from local schools and preschools.

**Design and procedure**

The experiment had a 2 (Co-occurrence Probability: non-co-occurring vs. co-occurring labels) × 3 (Age: 4-year-olds vs. 5-year-olds vs. 7-year-olds) mixed design. Co-occurrence probability was a within-participant factor. As in Study 1, co-occurring and non-co-occurring test items were intermixed and presented in one of the two pseudo-random orders.

Children were tested individually in a quiet room adjacent to their classroom. Visual stimuli were presented on a computer, and instructions and labels were presented verbally by experimenters who were blind to the hypothesis. All stimuli and procedures were identical to those in Study 1.

**Results and discussion**

**Induction accuracy**

Proportions of category-based responses were analyzed in a two-way mixed analysis of variance (ANOVA), with age as the between-participant factor and co-occurrence condition as a within-participant factor. This analysis revealed a significant effect of co-occurrence condition, $F(1,61) = 16.18$, $p < .0001$, $\eta^2_p = .21$. The effect of age, $F(1,61) = 2.92$, $p = .06$, $\eta^2_p = .087$, and the interaction between age and co-occurrence condition, $F(2,61) = 2.58$, $p = .084$, $\eta^2_p = .078$, approached significance. We further explored these findings through planned comparisons due to their nontrivial effect sizes (according to Cohen, 1988, $\eta^2_p$ values between .059 and .138 correspond to a medium effect size).

Proportions of category-based responses were compared with chance level (.50) using single-sample $t$ tests. Similar to adults, induction performance of 5- and 7-year-olds was above chance regardless of co-occurrence probability: the 5-year-olds averaged 80% and 70% of category-based responses in the co-occurring and non-co-occurring conditions, respectively (all $t$s > 3.50, all $p$s < .001), and the 7-year-olds averaged 87% and 81% of category-based responses in the co-occurring and non-co-occurring conditions, respectively (all $t$s > 7.25, all $p$s < .0001). In contrast, the 4-year-olds’ induction performance in the non-co-occurring condition (57%) was not significantly different from chance, single-sample $t(19) = 1.32$, $p = .10$. At the same time, the rate of category-based responses in 4-year-olds in the co-occurring condition (82%) was above chance, single-sample $t(19) = 5.59$, $p < .0001$. These findings are displayed in Fig. 5.

Induction performance as a function of condition (non-co-occurring labels vs. co-occurring labels) was analyzed using paired-samples $t$ tests. For both 7- and 5-year-olds, there was no significant

---

**Fig. 5.** Proportions of category-based responses by age group and co-occurrence condition in the induction task. The line indicates chance performance. Error bars represent standard errors of the mean.
difference in mean induction scores for co-occurring labels and non-co-occurring labels (7-year-olds: 87% and 81%, respectively; 5-year-olds: 80% and 70%, respectively; all paired-sample ts < 1.62, all ps > .12). In contrast, among 4-year-olds, the rate of category-based responses was significantly higher in the co-occurring condition (82%) than in the non-co-occurring condition (57%), paired-samples t(19) = 3.85, p < .001, Cohen’s d = 1.01.

Kinship knowledge accuracy

Mean scores on the kinship knowledge task are displayed in Fig. 6. Children in all age groups obtained mean scores that were significantly above chance (.50) in both co-occurrence conditions (all single-sample ts > 4.75, all ps < .0001). Similar to adults, 7- and 5-year-olds were equally accurate in both the co-occurring and non-co-occurring conditions (7-year-olds: 99% and 97%, respectively; 5-year-olds: 90% and 87%, respectively; all paired-sample ts < 0.85, all ps > .40). However, 4-year-olds exhibited higher accuracy on the kinship knowledge task in the co-occurring condition (95%) than in the non-co-occurring condition (74%), t(19) = 4.25, p < .0001.

Could the difference in induction performance among 4-year-olds with co-occurring and non-co-occurring labels stem from the differences in children's kinship knowledge? Below, we describe three follow-up analyses that considered the possibility that induction performance in 4-year-olds was a function of their knowledge of the kinship terms. A detailed summary of children's performance on the induction and kinship knowledge tasks for each label pair is presented in Table 2.

In the first follow-up analysis, we submitted 4-year-olds' induction scores to a repeated-measures analysis of covariance (ANCOVA) with co-occurrence condition as a within-participant factor and kinship knowledge scores in each condition as covariates. This analysis indicated a main effect of co-occurrence condition, F(1,17) = 4.75, p < .05, η²p = .218. However, the effects of kinship knowledge within each condition and the kinship knowledge by co-occurrence condition interaction were not significant (all Fs < 1.70, all ps > .21).

To address the possibility that the pictorial stimuli contained in the kinship knowledge task inadvertently supported children's performance, resulting in an overestimation of children's kinship knowledge, we collected data from 14 children who did not participate in Experiment 2 (mean age = 4.66 years, SD = 0.32) on a modified version of the kinship knowledge task. In the modified version of the kinship knowledge task, only labels were provided and none of the animals was depicted (we refer to this version as label-only). The procedure was analogous to the induction task; children were told that the animals were all hiding behind rocks (e.g., "Now I'm going to tell you what is hiding behind the rocks: lamb, sheep, cow"). Then, the children were asked to identify the baby animal's mother (e.g., "Do you think the sheep behind this rock or the cow behind this rock is the lamb's mother?"). Children's performances on both versions of the kinship knowledge task were compared. In the original version of the task 4-year-olds' mean accuracy was 95% on the co-occurring word pairs, and in the label-only version children exhibited 88% accuracy, t(32) = 1.07, p = .29. For the non-co-occurring word pairs, 4-year-olds' accuracy was 74% in the original version and 80% in the label-only version, t(32) = 0.75, p = .46. The lack of significant differences in children's performance on the original and label-only versions of the kinship knowledge task suggests that presenting pictures in this task did not facilitate performance in Experiment 2.
In the second follow-up analysis, we reanalyzed the induction data correcting for kinship knowledge. In particular, for every participant, we removed the induction score for any triad that included labels that children misidentified on the kinship knowledge task. For instance, if a child did not correctly identify that a frog was a tadpole’s mother, the induction data for this trial were removed from the child’s aggregate induction score. In other words, a child’s overall induction score was not penalized if the child did not know the kinship relation for a particular label pair. This correction resulted in mean induction scores that were very close to those displayed in Fig. 5; after correcting for knowledge of kinship relations, the rate of category-based induction was 81% in the co-occurring condition (compared with the uncorrected mean of 82%) and 55.83% in the non-co-occurring condition (compared with the uncorrected mean of 57%). Results of all statistical analyses of induction performance remained unchanged after correcting for children’s knowledge of kinship relations.

Finally, in the non-co-occurring condition, we identified three label pairs that elicited performance on the kinship knowledge task similar to that in the co-occurring condition in 4-year-olds. Specifically, 4-year-olds were highly accurate in identifying kinship relations with the following non-co-occurring label pairs:

- chick–hen
- lamb–sheep
- caterpillar–butterfly

(henceforth, we refer to this subset as the top 3). The average rate of correct responses on the kinship knowledge task with these label pairs was 90%, comparable to that in the co-occurring condition (95%), paired-samples t(19) = 1.37, p = .19. For this subset of labels, differences on the induction task can hardly be attributed to children’s superior knowledge of kinship relations, the rate of category-based induction was 81% in the co-occurring condition (compared with the uncorrected mean of 82%) and 55.83% in the non-co-occurring condition (compared with the uncorrected mean of 57%). Results of all statistical analyses of induction performance remained unchanged after correcting for children’s knowledge of kinship relations.

Individual response patterns

To investigate individual patterns of responses, participants were classified as either category-based or non-category-based responders. For comparison purposes, the analysis of individual response patterns includes adult data from Study 1. To mitigate concerns about possible kinship knowledge effects for non-co-occurring labels, we limited this analysis to the top 3 non-co-occurring condition trials—the trials on which 4-year-olds exhibited near-ceiling accuracy in the kinship knowledge task. Thus, analysis of the individual patterns of responses involved three trials in each co-occurrence condition.

A category-based responder was defined as a participant who provided category-based responses on all three trials within each co-occurrence condition. Data were analyzed using McNemar’s chi-square test, which is a chi-square test for within-participant designs. Results of this analysis

<table>
<thead>
<tr>
<th>Label pair</th>
<th>Proportion of category-based responses on the induction task and binomial probability</th>
<th>Proportion of correct responses on the kinship knowledge task and binomial probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-occurring labels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kitty–Cat</td>
<td>.80 (16 of 20 children), p &lt; .01</td>
<td>.95 (19 of 20 children), p &lt; .001</td>
</tr>
<tr>
<td>Bunny–Rabbit</td>
<td>.75 (15 of 20 children), p &lt; .05</td>
<td>1.00 (20 of 20 children), p &lt; .001</td>
</tr>
<tr>
<td>Non-co-occurring labels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamb–Sheep</td>
<td>.50 (10 of 20 children), p &gt; .58</td>
<td>1.00 (20 of 20 children), p &lt; .001</td>
</tr>
<tr>
<td>Caterpillar–Butterfly</td>
<td>.60 (12 of 20 children), p &gt; .25</td>
<td>.75 (15 of 20 children), p &lt; .001</td>
</tr>
<tr>
<td>Chick–Hen</td>
<td>.60 (12 of 20 children), p &gt; .25</td>
<td>.95 (19 of 20 children), p &lt; .001</td>
</tr>
<tr>
<td>Lion–Cub</td>
<td>.50 (10 of 20 children), p &gt; .58</td>
<td>.65 (13 of 20 children), p &gt; .13</td>
</tr>
</tbody>
</table>

Note: The shaded portion of this table describes children’s performance on the top 3 non-co-occurring label pairs (i.e., the label pairs that elicited comparable performance to the co-occurring label pairs on the kinship knowledge task).

* Induction data are uncorrected for kinship knowledge.
(displayed in Fig. 7) mirrored group data. For adults, the majority of participants were classified as category-based responders regardless of co-occurrence condition; fully 85% (17/20) of adult participants were classified as category-based responders in the co-occurring condition and 90% (18/20) in the non-co-occurring condition, McNemar’s $\chi^2 = 1.00, p = 1.00$. Among 7-year-olds, 71% (17/24) of participants were classified as category-based responders in the co-occurring condition and 54% (13/24) in the non-co-occurring condition, McNemar’s $\chi^2 = 2.78, p = .18$. Among 5-year-olds, 65% (13/20) of participants were classified as category-based responders in the co-occurring condition and 45% (9/20) in the non-co-occurring condition, McNemar’s $\chi^2 = 2.00, p = .29$. In contrast to older participants, the majority of 4-year-olds were classified as category-based responders in the co-occurring condition (60%, 12/20), but only a small percentage of 4-year-olds were classified as category-based responders in the non-co-occurring condition (20%, 4/20), McNemar’s $\chi^2 = 6.40, p = .02$.

The final analysis examined the developmental trajectory of category-based responding on the inductive reasoning task with kinship terms. This analysis was limited to the top 3 non-co-occurring label pairs in order to examine category-based reasoning uncontaminated by the (a) effects of label co-occurrence and (b) possible differences in kinship knowledge among the non-co-occurring label pairs in older and younger children. First, children’s responses on the top 3 non-co-occurring label pairs (for which kinship knowledge was approaching ceiling level even in the youngest age group) were entered into a one-way ANOVA with age as a between-participant factor. The results of this analysis pointed to a significant effect of age, $F(2, 61) = 3.21, p < .05, \eta^2_p = .095$. Post hoc Tukey’s tests suggested that this effect was driven by different rates of category-based responding in 4- and 7-year-olds (57% and 79% of category-based responses, respectively, $p < .05$). Proportions of category-based responses in 5-year-olds (67%) were not different from the other two age groups (both $p > .34$). The group analyses were consistent with the individual patterns of responses. Specifically, 13 of 24 children (54%) were classified as category-based responders in the 7-year-old group, 9 of 20 (45%) in the 5-year-old group, and 4 of 20 (20%) in the 4-year-old group. These differences were significant for the comparison between 4- and 7-year-olds, $\chi^2 = 5.37, p = .02$, and only marginally significant for the comparison between 4- and 5-year-olds, $\chi^2 = 2.85, p = .09$. Overall, results of Study 2 point to a gradual increase in children’s ability to spontaneously use kinship knowledge during the course of category-based reasoning. Unlike older children and adults, the majority of 4-year-olds consistently generalized properties to items referred by kinship terms only when the kinship terms co-occurred in natural language. This finding emerged even when care was taken to equate older and younger participants on their knowledge of kinship relations for the items used in the study.

Fig. 7. Numbers of participants classified as category-based responders and non-category-based responders by age and co-occurrence condition. These data reflect performance only on the top 3 non-co-occurring label pairs (i.e., the label pairs that elicited comparable performance to the co-occurring label pairs on the kinship knowledge task). NC, non-co-occurring condition; Co, co-occurring condition.
Role of semantic similarity of labels

One possible interpretation of the current findings is that the observed pattern of results may be driven by higher semantic similarity of label pairs in the co-occurring condition than in the non-co-occurring condition. This possibility appears unlikely for several reasons. First, in Fisher et al. (2011) study, co-occurring and non-co-occurring semantically similar labels were well matched for the level of semantic similarity, yet children exhibited superior performance with co-occurring labels. Furthermore, for the 24 different label pairs used in Fisher and colleagues’ study, there was no relationship between the semantic similarity ratings of label pairs and probability of category-based responding in 4-year-olds ($r = -.07$); in contrast, there was a positive relationship between label co-occurrence and category-based responding ($r = .56$).

Recall that the current study was designed specifically to control for the possibility left open by the prior studies, namely that children relied on kinship information rather than label co-occurrence. Toward this goal, we carefully selected several kinship terms that do not co-occur in child-directed speech but are familiar to young children. Even though these non-co-occurring kinship terms were judged to be somewhat less semantically similar to each other than the co-occurring kinship terms, the non-co-occurring kinship terms were considerably more semantically similar to each other than to the unrelated lures (e.g., chick–hen vs. chick–mouse or hen–mouse). Most older children and nearly all adults were able to use semantic similarity information provided by the kinship terms during the course of inductive reasoning. However, very few 4-year-olds (20%) were able to do so even when we controlled for the differences in kinship knowledge between the co-occurrence conditions. For instance, in the current study, the label pairs puppy–dog and lamb–sheep were well matched for both semantic similarity and kinship knowledge in 4-year-olds: puppy–dog, semantic similarity score = 6.0, 90% correct on the kinship knowledge task; lamb–sheep, semantic similarity score = 5.75, 100% correct on the kinship knowledge task. Despite a good match on the dimensions of semantic similarity and kinship knowledge, 4-year-olds generated on average 90% of category-based responses with the co-occurring label pair (puppy–dog) but only 50% of category-based responses with the non-co-occurring label pair (lamb–sheep). Overall, it appears doubtful that the performance of younger children was due to the (relatively small) differences in semantic similarity ratings between co-occurring and non-co-occurring labels, particularly when the current findings are considered together with those reported by Fisher and colleagues (2011).

Role of individual label frequencies

Another possible interpretation of the current findings is that children’s performance was influenced by individual label frequencies. As can be seen in Table 3, word frequency was higher in the co-occurring condition than in the non-co-occurring condition for targets, category choices, and lures. To address the possibility that children’s responses on the induction task were related to differences in label frequency rather than co-occurrence, we conducted a stepwise linear regression analysis with probability of category-based responding as a dependent variable and frequency of target, category match, and lure labels as predictor variables. The regression model explained more than 75% of the variance in children’s responses, adjusted $R^2 = .759$, $F(1,6) = 23.07$, $p < .005$. The only significant predictor was co-occurrence condition ($\beta = .89$, $p = .003$); individual word frequencies were not found to be significant predictors ($-.015 \leq \beta \leq .30$, all $p$s $\geq .13$).

It could be argued that the number of label triads in the current study was small and, therefore, the outcome of the above analysis is not conclusive. However, it should be noted that the above finding is consistent with the outcomes of prior research in which we investigated the relationship among co-occurrence, label frequency, and induction performance with a substantially larger set of stimuli ($N = 24$; Fisher et al., 2011). Specifically, our prior findings indicated that category-based responding in preschool-age children was related to co-occurrence but not to label frequency. Scrutinizing individual trials in Table 3 (rather than condition means) supports these conclusions. For example, in the trial puppy–dog–bear, the target label (puppy) had the lowest frequency among the co-occurring label targets (157 vs. 617 for bunny and 879 for kitty); nevertheless, category-based responding was higher on this item (90%) than on any other item. Furthermore, in the trial puppy–dog–bear, the frequency of the category match (dog: 2052) was lower than the frequency of the lure (bear: 2156); nevertheless children overwhelmingly selected the lower frequency category match on this trial. At the
same time, in the trial caterpillar–butterfly–ladybug, the category match (butterfly: 704) was more frequent than the lure (ladybug: 5); however, the higher frequency of the correct response did not appear to significantly facilitate category-based responding (60%, \( p > .25 \)). Overall, if word frequency had an influence on children’s performance on the induction task, it appears unlikely that the observed pattern of results can be explained solely by the differences in word frequency.

**General discussion**

The results reported in this article replicate prior findings and point to several novel findings. First, similar to the results reported by Fisher et al. (2011), adults and most children over 5 years of age spontaneously engaged in category-based reasoning when there was no conflict between semantic similarity and perceptual information. The current studies extend this finding to a case in which semantic similarity is communicated by kinship terms. Second, adults and most children over 5 years of age were equally successful in engaging in category-based reasoning with both co-occurring and non-co-occurring kinship terms; however, unlike older participants, preschool-age children successfully generalized properties only when semantic similarity was communicated by co-occurring kinship terms.

One methodological improvement of the current studies over prior studies is that in the current research co-occurrence probability of labels was manipulated within participants such that co-occurring and non-co-occurring trials were intermixed. Therefore, the current results provide strong evidence that overall above chance responding with semantically similar labels reported by Gelman and Markman (1986) could indeed stem from intermixing a relatively small number of trials in which semantically similar labels co-occur (and in which children are likely to select the category match at above chance level) with a larger number of trials in which synonyms do not co-occur (and in which children are likely to respond at chance level). In the current studies, when responses of 4-year-olds were aggregated across all eight trials (i.e., three co-occurring and five non-co-occurring trials), the rate of choosing category matches was 66% (above chance, single-sample \( t(19) = 3.72, \ p = .001 \) and very close to that reported by Gelman and Markman in a comparable condition (i.e., 63%).

At the same time, the reported results provide further support to the co-occurrence hypothesis. As discussed earlier, one mechanism by which co-occurrence of labels may influence children’s performance on reasoning tasks is lexical priming. Co-occurrence may result in strong lexical associations between word pairs (Brown & Berko, 1960; McKoon & Ratcliff, 1992; Spence & Owens, 1990). Consequently, children may select the semantically similar response option not because they are engaging in category-based reasoning but rather because of spreading activation from the target to the synonymous word (e.g., from bunny to rabbit) and not to the lure (e.g., squirrel).

**Table 3**

Frequency of linguistic stimuli.

<table>
<thead>
<tr>
<th>Target word</th>
<th>Frequency</th>
<th>Category choice</th>
<th>Frequency</th>
<th>Lure</th>
<th>Frequency</th>
<th>Induction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Co-occurring labels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bunny</td>
<td>617</td>
<td>Rabbit</td>
<td>460</td>
<td>Squirrel</td>
<td>67</td>
<td>.75</td>
</tr>
<tr>
<td>Kitty</td>
<td>879</td>
<td>Cat</td>
<td>2565</td>
<td>Fox</td>
<td>63</td>
<td>.80</td>
</tr>
<tr>
<td>Puppy</td>
<td>157</td>
<td>Dog</td>
<td>2052</td>
<td>Bear</td>
<td>2156</td>
<td>.90</td>
</tr>
<tr>
<td>Mean frequency for co-occurring labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>551.00</td>
<td>Category choice</td>
<td>1692.33</td>
<td>Lure</td>
<td>762.00</td>
<td></td>
</tr>
<tr>
<td><strong>Non-co-occurring labels</strong></td>
<td>Co-occurring labels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cat</td>
<td>532.80</td>
<td>Category choice</td>
<td>201.40</td>
<td>Lure</td>
<td>387.40</td>
<td></td>
</tr>
</tbody>
</table>

A second pathway by which co-occurrence may aid induction is by simplifying the mapping task between labels and their referents. Consistent with this idea, one difference between co-occurring and non-co-occurring labels used in the current studies is that co-occurring labels can often be used interchangeably between offspring and parent items (e.g., bunny–rabbit, kitty–cat), whereas non-co-occurring label pairs often refer to animals at distinct stages in their development (e.g., caterpillar–butterfly). Co-occurring labels, therefore, be used concurrently to refer to the same object (e.g., “Look at the puppy–dog”) in contrast to the non-co-occurring labels. Concurrent use of semantically similar labels in reference to the same object may reinforce the relationship between semantically similar terms and may in turn lead to enhanced category-based responding. Future research is required to evaluate these two possibilities.

Broader theoretical implications

The reported results also have implications for different developmental accounts of inductive generalization. According to one account, often referred to as a naive theory approach, from early infancy induction is driven by knowledge about language and the world. When “trying to determine whether to draw an inference from Object A to Object B, a child would not simply calculate the similarity between the two objects. Rather the child would determine whether A and B belong to members of the same natural kind category that encompasses both A and B” (Gelman & Coley, 1991, p. 185). According to this account, labels communicate category membership to infants even before they utter their first words (e.g., Ferry, Hespos, & Waxman, 2010; Waxman & Gelman, 2009). Furthermore, this account suggests that labels are more central to the category membership of objects than appearances (e.g., Gelman & Coley, 1991; Jaswal, 2004; Keil, Smith, Simons, & Levin, 1998). To summarize, according to the naive theory approach, from early infancy (a) induction is category-based and (b) labels are proxies for category membership.

According to the alternative approach, early in development induction is an automatic generalization process based on the overall similarity of compared entities. This approach suggests that labels start out as object features and become proxies for categories during the course of development (e.g., Deng & Sloutsky, 2012; Sloutsky & Fisher, 2004, 2012; Sloutsky, Lo, & Fisher, 2001). According to this approach, the similarity of compared entities is computed over both visual and auditory attributes, including labels. Therefore, early in development, inductive generalization can be label-based without necessarily being category-based; that is, children may rely on labels during the course of inductive generalization without the knowledge that labels are proxies for categories as long as the labels are identical or phonologically similar (Sloutsky & Fisher, 2004, 2011).

The finding that preschool-age children rely not only on identical labels but also on semantically similar labels to perform induction (Gelman & Markman, 1986) has been traditionally interpreted as strong evidence for the naive theory account of inductive generalization. However, the current findings cast doubt on this interpretation and are more consistent with the possibility that early in development inductive generalization is driven primarily by low-level factors such as visual similarity of objects (e.g., Sloutsky & Fisher, 2004), phonological similarity of labels (Sloutsky & Fisher, 2011), and label co-occurrence (Fisher, 2010; Fisher et al., 2011; see also current studies).

It remains an open question whether young children can use labels as proxies for categories in task settings and paradigms other than a label generalization task (Fisher, 2010) and an inductive inference task (Fisher et al., 2011; see also current studies). At the same time, the current findings contribute to a growing body of evidence suggesting that the ability to make spontaneous category-based inferences follows a protracted developmental course. The precise age at which children are credited with category-based reasoning differs based on the specific methodologies and stimuli used in various studies; however, a number of findings point to the emergence of this ability at approximately 6 or 7 years of age (Badger & Shapiro, 2012; Fisher, 2010; Fisher & Sloutsky, 2005; Fisher et al., 2011; Sloutsky, Kloos, & Fisher, 2007). It is conceivable that different tasks and different materials could scaffold children’s induction performance and improve category-based responding; it is also likely that children are sensitive to kinship information in other task contexts (e.g., Bullock & Opfer, 2009; Opfer & Bullock, 2007). Nonetheless, the current findings contribute to the body of evidence suggesting that there are clear
developmental differences in the propensity toward category-based reasoning among preschoolers, older children, and adults.

Acknowledgments

We thank Malika Sinha, Alyssa Montanaro, Yael Engel, Stephanie Brown, and Siyan Zhao for their help in collecting data and retrieving data from the CHILDES databases. We are grateful to Erik Thiessen and David Rakison for their comments on earlier versions of the manuscript. We also offer our thanks to the children and undergraduate students who participated in this study as well as the parents, teachers, and administrators who made this project possible. This work was supported in part by Graduate Training Grants awarded to Carnegie Mellon University by the Department of Education (R305B090023 and R305B040063).

References


