Cognition 118 (2011) 432–438

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Semantic similarity of labels and inductive generalization: Taking a second look

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

Department of Psychology, Carnegie Mellon University, United States

ARTICLE INFO

Article history:
Received 12 July 2010
Revised 13 December 2010
Accepted 15 December 2010
Available online 11 January 2011

Keywords:
Inductive reasoning
Semantic similarity
Categories
Labels
Cognitive development

ABSTRACT

Prior research suggests that preschoolers can generalize object properties based on category information conveyed by semantically-similar labels. However, previous research did not control for co-occurrence probability of labels in natural speech. The current studies re-assessed children's generalization with semantically-similar labels. Experiment 1 indicated that adults made category-based inferences regardless of co-occurrence probability; however, 4-year-olds generalized with semantically-similar labels that co-occurred in child-directed speech (e.g., *bunny–rabbit*) but not with non-co-occurring labels (e.g., *crocodile–alligator*). Experiment 2 indicated that generalization with semantically-similar labels increased gradually between 4- and 6-years of age. These results are discussed in relation to theories of early learning.

1. Introduction

It is well documented that adults readily utilize category information conveyed by labels to make inductive inferences. For example, if a sheepdog has a certain property, adults may infer that a bulldog may also have this property because they are both 'dogs'. However, the developmental course of this ability remains contested. Some research has suggested that children rely on category information conveyed by labels from early in life (Gelman & Coley, 1990; Jaswal & Markman, 2007), and in some cases as early as 13 months of age (Welder & Graham, 2001). For example, Gelman and Coley (1990) found that when 2-year-old children were presented with items that shared category but not appearance similarity (e.g. a bluebird and a dodo, both labeled 'birds') and items that shared appearance but not categorical similarity (e.g. a bluebird and a pterodactyl, labeled 'bird' and 'dinosaur'), children made generalizations to objects that shared labels rather than appearances.

An alternative interpretation of these findings is that early in development children may treat labels as perceptual attributes of objects rather than category markers (Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001). Under this view, when two objects share a label, children may engage in *label-based* induction rather than in *category-based* induction: children may rely on shared labels not because they understand that labels refer to categories but because identical labels increase overall perceived similarity among presented objects.

It can be difficult to disentangle the label-based and category-based accounts because both predict that children should rely on identical labels during induction. One way to tease apart these perspectives is to convey category membership via semantically-similar labels. Gelman and Markman (1986) hypothesized that if children treat labels as proxies for kinds, then children should generalize based on identical as well as semantically-similar labels. They presented 4-year-olds with a property induction task in which category information was communicated by either identical or synonymous labels. Findings indicated that children made inferences to categorically-related items at

* Corresponding author. Address: Carnegie Mellon University, Department of Psychology, 5000 Forbes Ave., 335-I Baker Hall, Pittsburgh PA 15213, United States. Tel.: +1 412 268 8656; fax: +1 412 268 2798.
E-mail address: fisher49@andrew.cmu.edu (A.V. Fisher).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.12.008

above chance level. Notably, synonyms and identical labels yielded similar levels of performance (63% and 67% of category-based responses, respectively).

However, Fisher (2010) recently suggested that some label-pairs used in Gelman and Markman's (1986) Synonyms condition were not only semantically-similar but also co-occurred in child-directed speech (e.g., *bunny–rabbit*) according to the CHILDES database (MacWhinney, 2000). Co-occurrence has been argued to give rise to strong lexical associations (Brown & Berko, 1960; McKoon & Ratcliff, 1992). Therefore, it is possible that children's responses to some items were based on lexical priming rather than category-based reasoning. For example, when children were asked whether a *bunny* shared a property with a *rabbit* or a *squirrel*, children's responses could be based not on the reasoning that bunnies and rabbits are the same kind of animal, but on the fact that the word *bunny* primed the word *rabbit* but not the word *squirrel*. The present studies were designed to revisit the question of whether preschool-age children use semantic similarity of labels in property induction.

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were 33 4-year-olds ($M = 4.52$, $SD = .40$ years; 17 females) recruited from local preschools and 30 undergraduate students.

2.1.2. Design and materials

The experiment had a 2 (labels: semantically-similar vs. identical) by 2 (co-occurrence: non-co-occurring vs. co-occurring) by 2 (age: 4-year-olds vs. adults) mixed design, with age and labeling condition as between-subject factors and co-occurrence as a within-subject factor.

There were eight label-triads, with each triad comprised of a target, a related test item, and an unrelated test item. Related items were conveyed either by identical or semantically-similar labels; unrelated items consisted of labels that a separate group of adult participants judged to be unrelated to the target; to-be-generalized properties consisted of two-syllable blank predicates (see Table 1).

Visual stimuli consisted of three identical doors (Fig. 1). Participants were told that objects were hiding behind each door, but the objects were never depicted. This procedure was used to eliminate conflict between labels and appearances, and to encourage reliance on semantic information, as it was the only available source of information.

2.1.2.1. Label selection. Assignment of label-pairs to co-occurrence conditions was based on the analysis of five databases in the CHILDES corpus (for details see Fisher, 2010). The average co-occurrence probability of semantically-similar labels was .03 in the co-occurring condition and .00 in the non-co-occurring condition.

A separate group of 22 adults rated semantic similarity of the targets-to-related and targets-to-unrelated test items on a scale of 1 (lowest) to 7 (highest). Results confirmed that targets were more semantically-similar to related items ($M = 6.3$; e.g., *mouse–rat*) than to unrelated items ($M = 2.8$; e.g., *mouse–duck* and *rat–duck*), $t(14) = 11.43$, $p < .001$. No differences were found when analyses were separated by the co-occurrence condition, $F(1, 15) < 1$, *ns*.

Table 1
Labels used in Experiment 1 and co-occurrence probabilities of semantically-similar labels.

Target items	Related test items	Unrelated test items	Blank predicates	Co-occurrence probabilities
Non-co-occurring condition				
Dolphin	Whale	Seal	Omat	.000
Alligator	Crocodile	Hippo	Matlen	.000
Toad	Frog	Bird	Koski	.000
Mouse	Rat	Duck	Lignin	.000
Co-occurring condition				
Puppy	Doe	Cow	Erwin	.010
Kitty	Cat	Pig	Manchin	.040
Bunny	Rabbit	Squirrel	Creighan	.070
Pony	Horse	Fox	Troxel	.01

2.1.3. Procedure

Children were tested individually at their daycares and adults were tested individually in the laboratory. Adult data were collected by hypothesis-blind experimenters. Half of the children's data was collected by hypothesis-blind experimenters and the other half by the second author; no differences were found between the data collected by hypothesis-blind experimenters and the second author, $t(31) < 1$, *ns*.

Visual stimuli were presented on a computer and labels were provided verbally by experimenters. In the identical labels condition the target and related items were referred to by identical labels (e.g., *mouse–mouse* for the half of participants and *rat–rat* for the other half). Participants were randomly assigned to the order of co-occurrence conditions (co-occurring first or non-co-occurring first) and to one of the two random orders of trials within each condition.

Following the induction task children participated in a Picture Identification (Pic-Id) task similar to the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). On each trial children were presented with four pictures and asked to point to the target picture (e.g., "*can you find the whale?*"). The Pic-Id task included all target, related, and unrelated labels used in Experiment 1. Knowledge of semantically-similar labels was tested using *identical* pictures on separate trials (see Appendix A), with location of correct response counterbalanced across trials. The rate of correct identifications was 99% in each co-occurrence condition, indicating that children were familiar with the words used in the experiment proper and could readily apply semantically-similar labels to the *same* objects.

2.1.4. Results and discussion

Preliminary analyses revealed no effects of block order (all $ps > .20$); therefore the data were collapsed across orders for the reported analyses. Proportions of category-based responses (i.e., choices of identical or semantically-similar labels) were analyzed in a three-way mixed ANOVA, with label condition and age as between-subject

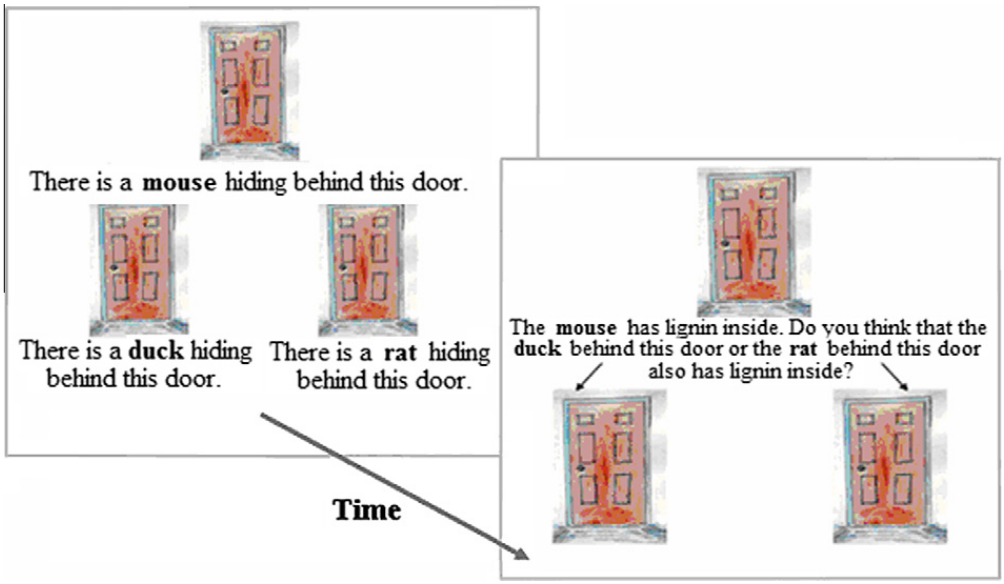


Fig. 1. Schematic depiction of the task.

factors and co-occurrence condition as a within-subject factor. There was a significant effect of age, $F(1, 58) = 29.57$, $p < .001$, $\eta_p^2 = .45$; a significant interaction between co-occurrence and age $F(1, 59) = 5.58$, $p < .05$, $\eta_p^2 = .09$; and a significant three-way interaction, $F(1, 59) = 4.41$, $p < .05$, $\eta_p^2 = .07$.

There were no differences among conditions for adults (all means above .97, all $ps > .63$). For children there was a significant interaction between label type and co-occurrence condition, $F(1, 31) = 6.07$, $p < .05$, $\eta_p^2 = .16$. Further tests revealed a reliable difference in performance between the non-co-occurring semantically-similar labels ($M = .51$) and co-occurring semantically-similar labels conditions ($M = .74$), paired-sample $t(16) = 3.45$, $p < .005$, Cohen's $d = .85$. Within the non-co-occurring condition, there was a reliable difference between the semantically-similar and identical label conditions, independent-sample $t(31) = 2.41$, $p < .05$ ($M = .51$ and .75, respectively), Cohen's $d = .86$. Furthermore, children's performance in the non-co-occurring semantically-similar labels condition did not exceed chance, one-sample $t(17) = .20$, *ns*., whereas performance in all other conditions was above chance (all one-sample $ts > 2.54$, $ps < .05$). There were no differences in children's performance with identical co-occurring and identical non-co-occurring labels ($M = .75$ and .70, respectively), paired-samples $t(15) = .53$, *ns* (see Fig. 2).

To analyze individual response patterns participants were classified into category-based and non-category-based responders. Category-based responding was defined as choosing semantically-similar labels on at least three of four trials (75%) per co-occurrence condition. All adult participants in all conditions were classified as category-based responders. In the co-occurring condition most 4-year-olds were classified as category-based responders: 11 out of 17 (65%) in the identical and 10 out of 16 (63%) in the semantically-similar labels condition. In the non-co-occurring/identical labels condition most children were classified as category-based responders: 12 out of 16 (75%). However, in the non-co-occurring/semantically-similar labels condition only 6 out of 17 children (35%) were classified as category-based responders. The association between condition and responder type in the non-co-occurring/semantically-similar labels and non-co-occurring/identical labels condition was significant, Fisher's exact $p < .05$ (see Fig. 3).

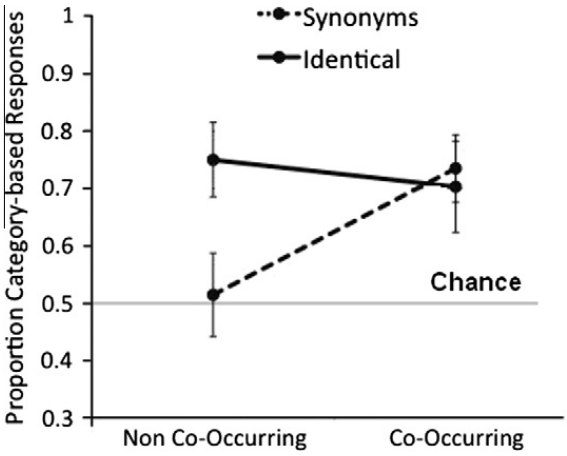


Fig. 2. Proportion of category-based responses in 4-year-old children in Experiment 1.

Overall, in contrast to adults, 4-year-old children found it challenging to use semantically-similar labels during induction, unless these labels co-occurred in child-directed speech.[1] However, Experiment 1 was limited in that very few label-pairs were used due to the paucity of co-occurring

[1] It could be argued that this finding may stem from better memory for co-occurring than for non-co-occurring label pairs. To address this possibility, 4-year-old children were asked to recall which objects were hiding behind the doors at the conclusion of each induction trial. Children's responses indicated that there was no difference in memory performance in the co-occurring and non-co-occurring conditions (86% and 85% of correct memory responses, respectively), paired-sample $t(32) < 1$, *ns*. A linear regression performed on children's memory scores and induction scores revealed no significant relationship in the semantically-similar labels condition, $r^2(134) = .055$, *ns*, or the identical labels condition, $r^2(126) = .019$, *ns*.

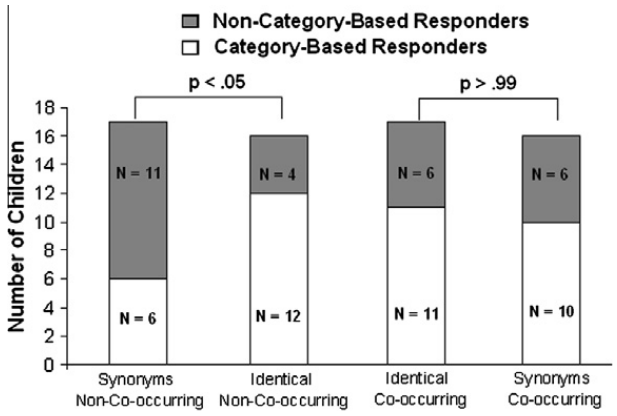


Fig. 3. Individual response patterns in Experiment 1.

semantically-similar labels. Experiment 2 was designed to explore the development of category-based induction using a larger and more diverse set of non-co-occurring semantically-similar labels with 4-, 5-, and 6-year-old children.

3. Experiment 2

3.1. Method

3.1.1. Participants

Participants were 18 4-year-olds (M = 4.61, SD = .43 years; nine females), 21 5-year-olds (M = 5.40, SD = .31 years; 16 females), and 21 6-year-olds (M = 6.60, SD = .32 years; nine females), none of whom had participated in Experiment 1. Three 4-year-old children were not available for all testing sessions and were excluded from the analyses.

3.1.2. Design and materials

The experiment had a 3 (age: 4-, 5-, and 6-year-olds) by 2 (ontological kind: natural kinds vs. artifacts) mixed design, with age as a between-subject factor and ontological kind as a within-subject factor. There were 16 label-triads referring to eight natural kind and eight artifact sets (see Table 2); visual stimuli were identical to those in Experiment 1.

3.1.2.1. Label selection. A separate group of 21 adults rated semantic similarity of labels using the same procedure as in Experiment 1. Results indicated that semantic similarity of the target-to-related test items (M = 5.5; e.g., *shoe–boot*) was higher than semantic similarity of the target-to-unrelated test items (M = 1.41; e.g., *shoe–sponge*), paired-sample $t(15) = 18.07$, $p < .0001$. Overall, the ratings were lower in Experiment 2 than in Experiment 1 (5.5 vs. 6.3 for targets-to-related items, and 1.41 vs. 2.8 for targets-to-unrelated items; both independent-samples ts > 2.56, ps < .05). We examine below whether this could potentially account for the observed pattern of results.

A separate group of 18 4-year-olds participated in a Pic-Id task. The procedure was similar to the one described in Experiment 1 with one important difference: six control trials were added to mitigate the concern that Pic-Id results could be influenced by children's reluctance to tell the experimenter they do not see the object they are being asked to identify. In the control trials children were asked to point to objects that were not pictured (e.g., 'egg' when the depicted objects were 'chair', 'bear', 'cat', and 'skateboard'). The results indicated that children were comfortable telling the experimenter 'I don't see it' when they thought the label supplied by the experimenter did not apply to any of the depicted objects – the rate of correct rejections was 81%. On the experimental trials, the rate of correct identifications was 94%; therefore, children were familiar with the semantically-similar labels used in Experiment 2 and could readily apply these labels to the *same* object (see Appendix A). CHILDES analyses of the labels used in Experiment 2 confirmed that none of the label-pairs were likely to co-occur in natural speech to children (co-occurrence probability = .00).

3.1.3. Procedure

Target items in the natural kinds condition were said to have a particular property inside. However, artifacts are inherently more variable than natural kinds in the materials they can be made from; therefore, in the artifacts condition half of the target items were said to have "X inside" and the other half were said to be "made of X" (see Table 2). This manipulation allowed us to examine whether children's induction is influenced by these subtle differences in phrasing.

The induction task was similar to that used in Experiment 1. However, in order to avoid fatigue due to the increased number of trials, Experiment 2 consisted of two separate test sessions that took place 1 week apart. Children were randomly assigned to the starting order: natural kinds condition first or artifacts condition first.

3.1.4. Results and discussion

Preliminary analyses revealed no effects of condition order (p > .58). We thus conducted a 2 (ontological kind: natural kinds vs. artifacts) × 3 (age: 4-, 5-, and 6-year-olds) mixed ANOVA with age as a between-subject factor and ontological kind as a within-subject factor. This analysis revealed a significant effect of age, $F(2, 54) = 54.51$, $p < .001$, $\eta_p^2 = .38$. No other effects or interactions were significant (all Fs < 1.40, all ps > .24; see Fig. 4).

Post-hoc analyses revealed that all age contrasts were significant: 5-year-olds gave significantly more category-based responses than 4-year-olds (Tukey's $p < .05$), and 6-year-olds gave significantly more category-based responses than 5-year-olds (Tukey's $p < .01$). Four-year-old children performed at chance level in both the natural kinds (M = .58) and artifacts conditions (M = .50), both one-sample ts < 1.42, ps > .17. In contrast, 5-year-olds exhibited above-chance performance in both the natural kinds (M = .71) and artifacts conditions (M = .70), as did 6-year-olds (M = .90 and .87, respectively), all one-sample ts > 3.50, ps < .005. Comparisons of children's performance in the artifacts condition based on property phrasing (e.g. "*is made of X*" vs. "*has X inside*") evidenced no significant differences within any age group (all ps > .49). Similarly, comparisons of children's performance with artifacts and

Table 2
List of stimuli in Experiment 2.

Natural kinds				Artifacts			
Target items	Related test items	Unrelated test items	Blank predicates	Target items	Related test items	Unrelated test items	Blank predicates
Dolphin	Whale	Sand	has Matlen inside	Axe	Hammer	Bed	is made of Timmon
Rock	Stone	Cloud	has Omat inside	Mug	Cup	Rope	is made of Ramesh
Mountain	Hill	Wood	has Koski inside	Bowl	Plate	Pen	is made of Doyle
Mouse	Hamster	Lake	has Creighan inside	Glove	Mitten	Road	is made of Pilter
Tree	Bush	Bread	has Pollick inside	Cap	Hat	Rug	has Quittle inside
Fly	Bee	Sun	has Troxel inside	Couch	Sofa	Ball	has Zelter inside
Duck	Chicken	Grass	has Nemoy inside	Bus	Truck	Spoon	has Viston inside
Lemon	Orange	Rain	has Lignin inside	Shoe	Boot	Sponge	has Mufton inside

natural kinds did not yield any significant differences for any age group (all *ps* > .10).

To explore individual response patterns we classified children into category-based responders and non-category-based responders. We collapsed children's responses across ontological conditions and defined category-based responding as providing at least 12 category-based responses on 16 trials (75%), the same proportional criterion used in Experiment 1 (see Fig. 5).

Among 4-year-olds, only 2 out of 15 children (13%) reliably selected semantically-similar labels over unrelated labels. Reliance on semantically-similar labels increased gradually with age: 9 out of 21 children (43%) were classified as category-based responders in the 5-year-old group, and 18 out of 21 children (86%) in the 6-year-old group. A chi-square test revealed a significant association between responder type and age, $\chi^2(2, 57) = 19.20$, $p < .001$. Follow-up chi-square tests revealed that the proportion of category-based responders was higher in the 6-year-old group than in both younger groups, both $ps \leqslant .01$; no other contrasts were significant.

It is possible that 4-year-old children in Experiment 2 failed to perform category-based induction because the label-pairs used were less semantically-similar than those in Experiment 1. Second, it could be argued that certain label-pairs yielded higher performance in 4-year-old children in Experiment 1 not because these labels co-occurred, but because they were highly frequent. To examine these possibilities we examined the relationship among induction performance, word frequency,[2] semantic similarity, and co-occurrence condition.

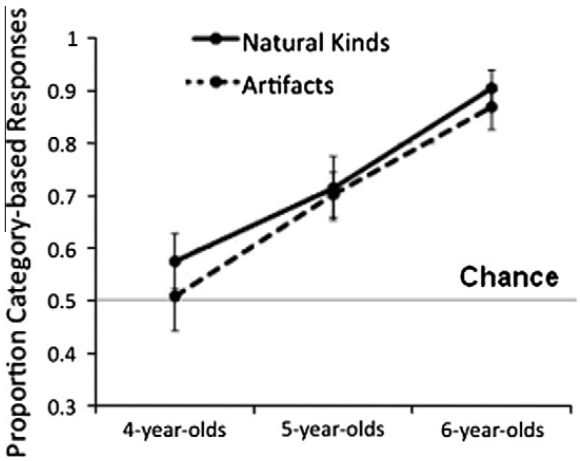


Fig. 4. Proportions of category-based responses in Experiment 2.

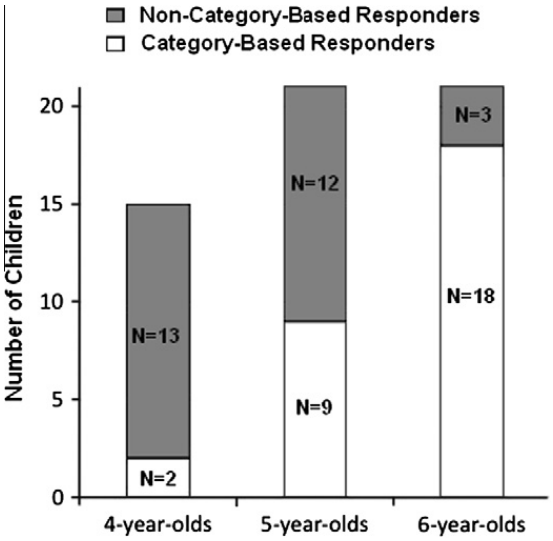


Fig. 5. Individual response patterns in Experiment 2.

The analyses revealed that co-occurrence condition was positively correlated with both word frequency ($r = .56$, $p < .01$) and semantic similarity ($r = .46$, $p < .05$). However, neither word frequency nor semantic similarity was correlated with induction performance ($r = .11$ and $r = -.07$, respectively, both *ps* > .58), whereas co-occurrence condition was, $r = .56$, $p < .005$; see Table 3). Therefore, differences in semantic similarity and word frequency are unlikely to account for the differences in induction performance between the co-occurring and non-co-occurring labels.

[2] Analysis reported below used combined frequency of both words within each pair of semantically-similar labels. Results of all analyses remain unchanged if average frequency of the labels in a pair or individual frequency of each label is used instead of combined frequency.

Table 3
Label frequency and performance on semantically-similar label-pairs in 4-year-old children in Experiments 1 and 2.

Label–fair	Proportion of category-based responses	Binomial probability	Combined frequency in CHILDES
Experiment 1			
Co-occurring labels			
Bunny–Rabbit	.76 (13 out of 17 children)	$p = .0245$	854
Kitty–Cat	.59 (10 out of 17 children)	$p = .314$	3166
Pony–Horse	.76 (13 out of 17 children)	$p = .0245$	473
Puppy–Dog	.82 (14 out of 17 children)	$p = .006$	1824
Non-co-occurring labels			
Alligator–Crocodile	.53 (9 out of 17 children)	$p = .500$	80
Dolphin–Whale	.47 (8 out of 17 children)	$p = .685$	134
Mouse–Rat	.59 (10 out of 17 children)	$p = .314$	369
Toad–Frog	.47 (8 out of 17 children)	$p = .685$	94
Experiment 2			
Natural kinds			
Dolphin–Whale	.60 (9 out of 15 children)	$p = .304$	134
Duck–Chicken	.60 (9 out of 15 children)	$p = .304$	851
Fly–Bee	.67 (10 out of 15 children)	$p = .151$	200
Lemon–Orange	.53 (8 out of 15 children)	$p = .500$	714
Mountain–Hill	.33 (5 out of 15 children)	$p = .941$	209
Mouse–Hamster	.67 (10 out of 15 children)	$p = .151$	317
Rock–Stone	.53 (8 out of 15 children)	$p = .500$	290
Tree–Bush	.60 (9 out of 15 children)	$p = .304$	682
Artifacts			
Axe–Hammer	.60 (9 out of 15 children)	$p = .304$	149
Bowl–Plate	.60 (9 out of 15 children)	$p = .304$	898
Bus–Truck	.67 (10 out of 15 children)	$p = .151$	1510
Cap–Hat	.46 (7 out of 15 children)	$p = .696$	944
Couch–Sofa	.60 (9 out of 15 children)	$p = .304$	113
Glove–Mitten	.53 (8 out of 15 children)	$p = .500$	138
Mug–Cup	.20 (3 out of 15 children)	$p = .996$	1062
Shoe–Boot	.40 (6 out of 15 children)	$p = .349$	1402

4. General discussion

The results presented here point to several novel findings. Contrary to prior research, Experiment 1 revealed that reasoning with semantically-similar labels is more challenging than reasoning with identical labels: while most 4-year-old children readily relied on identical labels in property induction, few children relied on semantically-similar labels unless these labels not only shared meaning but also co-occurred in child-directed speech. Experiment 2 replicated the results with non-co-occurring labels, and extended the finding beyond natural kinds to artifact categories. Furthermore, Experiment 2 provided developmental evidence that although approximately half of 5-year-old children spontaneously relied on semantically-similar labels during induction, the majority of children did not do so until 6-years of age. This trend suggests a gradual increase in the use of category information during induction and provides converging evidence to other related findings in the literature (Fisher & Sloutsky, 2005; Fisher, 2010).

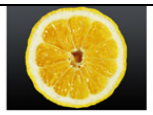
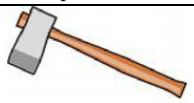
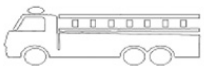
These results are inconsistent with prior research on children's use of semantically-similar labels during induction (Gelman & Markman, 1986). It is possible that these discrepant findings stem from averaging across two different distributions of responses in prior research: non-co-occurring semantically-similar labels (e.g., *rock–stone*) may have resulted in chance performance, whereas co-occurring semantically-similar labels (e.g., *puppy–dog*) may have resulted in above-chance performance. Aggregating over a bimodal distribution of responses could result in a mean that exceeds chance level. In support of this hypothesis we found that by aggregating across both co-occurrence conditions in Experiment 1, the rate of category-based responding was 63%, above chance (single-sample $t(16) = 2.17$, $p < .05$) – which is identical to that reported by Gelman and Markman.

It is possible that results reported in this study were driven by factors other than co-occurrence and that 4-year-old children can rely on some semantically-similar label-pairs that were not utilized in this research. Nonetheless, the results reported here provide support for the notion that children's ability to rely on semantic information conveyed by linguistic labels increases gradually with development (Sloutsky & Fisher, 2004; Fisher, 2010). Furthermore, the results suggest that semantic similarity alone is insufficient to promote category-based induction in 4-year-old children. Overall, the present findings suggest that category-based reasoning follows a more protracted developmental course than previously believed.

Acknowledgments

We thank Malika Sinha, Janelle Higa, Jamie Troxel, Tina Li, Sarah Shade, and Stephanie Brown for help in data collection and CHILDES analyses. We thank three anonymous reviewers, Vladimir Sloutsky, David Rakison, and Yevdokiya Yermolayeva for helpful comments on the earlier versions of this manuscript. Finally, we thank the children, parents, teachers, and administrators for their participation. This research was supported by the Institute of Education Sciences, US Department of Education through Grant R305B040063 to Carnegie Mellon University.

Appendix A

Experiment 1			
 Alligator - Crocodile	 Mouse-Rat	 Toad-Frog	 Dolphin-Whale
 Kitty-Cat	 Puppy-Dog	 Bunny-Rabbit	 Pony-Horse
Experiment 2			
 Rock-Stone	 Mountain-Hill	 Tree-Bush	 Lemon-Orange
 Dolphin-Whale	 Mouse-Hamster	 Fly-Bee	 Duck-Chicken
 Axe-Hammer	 Bowl-Plate	 Shoe-Boot	 Bus-Truck
 Glove-Mitten	 Cup-Mug	 Couch-Sofa	 Cap-Hat

References

Brown, R., & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development, 31*, 1–14.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test: Manual* (3rd ed.). Circle Pines, MN: American Guidance Services.

Fisher, A. V. (2010). What's in a name? How rocks and stones are different from dogs and puppies. *Journal of Experimental Child Psychology, 105*, 198–212.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development, 76*, 583–597.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796–804.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development, 8*, 93–111.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155–1172.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188.

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

Welder, A. N., & Graham, S. A. (2001). The influences of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72*, 1653–1673.