

## The Importance of Knowing a Dodo Is a Bird: Categories and Inferences in 2-Year-Old Children

Susan A. Gelman and John D. Coley  
University of Michigan

This study examined 2-year-old children's ability to make category-based inferences. Subjects were asked a series of questions that they could answer based on category membership, appearances, or both. In one condition, all pictures were named; in a second condition, none were named. Children performed well on prototypical pictures regardless of whether they were named; on atypical pictures, they performed better when category labels were provided. A control study demonstrated that children ignored the label when it named a transient property rather than a stable category. Contrary to standard views of young children, these results indicate an early-emerging capacity to overlook salient appearances. However, one important development still to take place is the ability to use subtle perceptual cues to determine category membership in the absence of language.

Are young children's categories based on appearances alone, or are children aware that categories can reflect deeper commonalities? Many years of developmental research and theory have suggested that children cannot look beyond the obvious (Wellman & Gelman, 1988). Children are highly attentive to perceptual similarities on a range of important tasks (e.g., Fenson, Cameron, & Kennedy, 1988; Gentner, 1988; L. B. Smith, 1989). At the same time, growing evidence suggests that surface similarity alone cannot adequately describe adult concepts and that theories play an important role in the organization of knowledge for adults (Murphy & Medin, 1985). Accordingly, a number of researchers have suggested that conceptual development reflects a shift from categories based on perceptual similarity to categories based on theories that reflect deeper underlying commonalities among category members (e.g., Neisser, 1987, p. 6).

Recently, the notion of a developmental shift has been challenged by studies demonstrating that preschool children can overlook misleading perceptual information when reasoning about categories (S. A. Gelman & Markman, 1986). The impetus for this work was the assumption that categories function to extend knowledge beyond what is obvious or already known. If we learn certain facts about one category member, we are likely to infer that they are true of other category members as well. For example, if we learn that one dog has leukocytes inside it, we are likely to infer that other dogs also have leukocytes inside

them. Although children often have difficulty constructing certain kinds of categories (Inhelder & Piaget, 1964), S. A. Gelman and Markman hypothesized that once children are told the category membership of an object, they might be able to use the category as the basis of further inferences.

To test this idea, S. A. Gelman and Markman (1986) gave preschool children a task in which category membership was put into conflict with superficial appearances and children were asked to make a series of inductive inferences. For example, on one item children saw a brontosaurus, a rhinoceros, and a triceratops, which were labeled as "dinosaur," "rhinoceros," and "dinosaur," respectively. Category labels and outward appearances conflicted: The brontosaurus and triceratops are members of the same category, whereas the rhinoceros and triceratops look more alike outwardly. Then children learned a new property of the brontosaurus and the rhinoceros (that they had cold blood and warm blood, respectively) and were asked which property was true of the triceratops. Children reported that the triceratops has cold blood like the brontosaurus, even though it more closely resembled the rhinoceros. The results of this and other related experiments (S. A. Gelman, Collman, & Maccoby, 1986; S. A. Gelman & Markman, 1987) indicated that by 3½ years of age, children chose to base inferences on category membership despite conflicting surface appearances.

Control studies showed that these effects were not simply a response bias that was due to hearing the same word for the two category members. First, when category membership was conveyed by synonymous rather than identical labels (e.g., "rock" and "stone" rather than "rock" and "rock"), children still drew inferences from one category member to another (S. A. Gelman & Markman, 1986). Second, when children learned properties that should not generalize (e.g., accidental properties such as an animal's age), they did not generalize from one category member to another (S. A. Gelman, 1988; S. A. Gelman & Markman, 1986). Thus, children do not draw inferences blindly as a consequence of hearing the category label. Finally, on some items, preschool children drew category-based inferences even in the absence of labels. For example, even when none of the

---

This research was supported by National Institute of Mental Health Grant 1-R29-HD23378 and a Spencer Fellowship to Susan A. Gelman and by a University of Michigan Regent's Fellowship to John D. Coley.

We thank Stacey Watkins and Michelle Weckler for their assistance. We are grateful to the children, parents, and teachers of the following schools for their generous cooperation: Ann Arbor "Y" Child Care Center, Humpty Dumpty Day Care, Jack and Jill Nursery School, the University of Michigan Children's Center, and the University of Michigan Children's Center for Working Families.

Correspondence concerning this article should be addressed to Susan A. Gelman, Department of Psychology, 3433 Mason Hall, University of Michigan, Ann Arbor, Michigan 48109-1027.

pictures were named, children drew more inferences from a green leaf-insect to a black beetle than from a green leaf-insect to a green leaf (S. A. Gelman & Markman, 1987). Follow-up analyses suggest that children were able to determine that the leaf-insect was an insect and not a leaf on the basis of subtle properties of the appearance (e.g., the animal's legs and head). Although misleading appearances often do trick children into thinking that anomalous objects are not what they are, children do not always require the experimenter's label in order to determine category membership and to use it to draw the appropriate inference.

In the present studies, we examined the inferences of 2½-year-old children to determine whether these younger children also make use of categories to infer important, nonobvious properties. It is important to extend past findings to younger children because relatively little is known about the conceptual capabilities of children this age. Few studies have compared the performance of 2-year-olds to that of older children. By studying inferential processes in children between 2 and 3 years old, we can begin to bridge the gap between what is known about the early accomplishments of infancy (cf. Cohen & Younger, 1983) and the developments of later childhood. The present studies help fill this gap and thus expand our knowledge of the skills of a relatively understudied age group.

For example, data from 2-year-olds could help address the question of a possible developmental shift from similarity-based to theory-based categories. If 2-year-olds are restricted to similarity-based categories, then they should be unable to draw inferences that contradict an object's appearances. However, if we find evidence for category-based inferences at age 2, either there is no developmental shift or it must be taking place before this age. There is already evidence that by age 2 children expect labels to refer to coherent categories (S. A. Gelman & Taylor, 1984; Katz, Baker, & Macnamara, 1974; Markman & Hutchinson, 1984). However, it is not known when children realize that category membership can be a better basis than perceptual appearances for drawing certain inferences.

In addition, a study of younger children promises eventually to provide insight concerning the mechanisms that could underlie children's ability to draw category-based inferences. There are vast changes in early preschool in a variety of cognitive domains, including knowledge (Carey, 1985), vocabulary (Anglin, 1977), metacognition (Flavell, 1985), and the kinds of classifications children spontaneously construct (R. Gelman & Baillargeon, 1983; Sugarman, 1982). If children's category-based inferences depend on these skills, then 2-year-olds, because of their relatively limited abilities in these areas, should experience more difficulties than older children. However, if 2-year-olds perform well, then presumably such skills are not critical prerequisites for category-based inferences.

Studying these issues required developing appropriate techniques to uncover younger children's understanding. Our major challenge was to devise a procedure that would not overtax the attentional capacity of 2-year-olds. Previous studies of induction placed relatively heavy information-processing demands on young children, requiring them to focus on two or three pictures at a time and to learn one or two new facts for every picture set. In the present study, we used a simpler procedure to reduce processing demands on the younger subjects. We asked

children to consider only one picture at a time. They were also asked to make inferences concerning familiar properties, given typical and atypical (novel) exemplars of known categories. Thus we imposed fewer memory demands on subjects while still allowing a comparison between category-based and appearance-based inferences to be made.

## Study 1

### *Method*

#### *Subjects*

Twenty-two children (ages 2 years, 1 month–3 years, 1 month,  $M = 2$  years, 8 months) participated in the main experiment. Of these, 11 children ( $M = 2$  years, 8 months) were randomly assigned to the label condition; the remaining 11 ( $M = 2$  years, 8 months) were assigned to the no-label condition (see Procedure). An additional 11 children (2 years, 6 months–3 years, 0 months,  $M = 2$  years, 10 months) participated in a preselection of the materials (described later). Eighteen adults provided several measures as a basis of comparison.

#### *Items*

Each child saw nine picture sets, listed in Table 1, each consisting of one target picture and four test pictures. All were brightly colored drawings.<sup>1</sup> A sample set is provided in Figure 1 (without the details provided by color and shading). In every set, the target picture was a typical instance of a familiar category (e.g., a typical bird). The test pictures varied in their category membership and typicality. Two test pictures were from the same category as the target picture, and two were from a contrasting category. In the example cited, two of the test pictures were birds and two were dinosaurs. Within each category, one of the test pictures was a typical category member and the other was an atypical category member that was likely to be unfamiliar to the child. For example, within the bird category, the test pictures were a bluebird (typical bird) and a dodo bird (atypical bird); within the dinosaur category, the test pictures were a stegosaurus (typical) and a pterodactyl<sup>2</sup> (atypical). Finally, for each picture set, the typical member of the same category as the target strongly resembled the target, and the atypical member of the contrasting category also strongly resembled the target picture. For example, the bluebird test picture and the pterodactyl both looked very much like the target bird. In this way, half of the atypical pictures posed a direct conflict between category identity and appearances.

Although the notion of "similarity" is notoriously difficult to pin

<sup>1</sup> One possible concern with the stimuli was that, for children, pictures would provide relatively meager perceptual information compared with real objects. However, much of the evidence in the literature suggesting that children rely heavily on perceptual similarity is based on two-dimensional pictures such as the present stimuli (e.g., Melkman, Tversky, & Baratz, 1981; Olver & Hornsby, 1966). Moreover, in a recent study of classification in 2-year-olds, Fenson, Vella, and Kennedy (1989) found that children "were able to match pictures with pictures as accurately as they were able to match objects with pictures" (p. 915), which again suggests that pictures are highly salient to young preschoolers.

<sup>2</sup> Although a pterodactyl is not actually a dinosaur, informal pretesting with adults revealed that most adult subjects believed it to be a dinosaur. We included the item because it is consistent with adults' folk classifications, although we note here that it does conflict with current scientific taxonomies.

Table 1  
*Pictures and Properties Used in Study 1*

Target/instance	Category 1	Category 2	Property
Within ontological category			
Worm Typical Atypical	Worm Brown worm Planarian	Snake Black snake Worm-like snake	Eats dirt
Rabbit Typical Atypical	Rabbit Rabbit Short-eared pika	Squirrel Squirrel Long-eared kaibab	Eats carrots
Bird Typical Atypical	Bird Bluebird Dodo bird	Dinosaur Stegosaurus Pterodactyl	Lives in a nest
Dog Typical Atypical	Dog Golden retriever Chihuahua	Lamb White lamb Dog-like lamb	Barks
Bird Typical Atypical	Bird Swallow Kiwi bird	Fish Turquoise fish Flying fish	Says tweet-tweet
Across ontological categories			
Leaf Typical Atypical	Leaf Green leaf Pine needle	Bug Black beetle Green leaf-insect	Grows on a tree
Snake Typical Atypical	Snake Green snake Horned snake	Stick Brown stick Curved, green stick	Moves by itself
Rock Typical Atypical	Rock Brown rock Spotted cube	Fish Trout Lumpfish	Sits on the ground
Flower Typical Atypical	Flower Flower Jack-in-the-pulpit	Fish Gray fish Starfish	Grows in a garden

down precisely (Goodman, 1983; Murphy & Medin, 1985) and there are clearly a range of different kinds of similarity relations (L. B. Smith, 1989), here we consider salient perceptual similarities to include color, size, shape, and prominent parts (e.g., wings). In addition, to make certain that our intuitions about similarity were accurate, we asked 18 adult subjects to rate the similarity of each picture to the target ("how much the two pictures look like one another") on a scale from 1 (*not at all alike*) to 7 (*almost identical*). As predicted, for the typical items, adults rated the pictures that belonged to the same category as the target pictures as much more similar to the target pictures ( $M = 6.39$ ) than they rated the pictures from a different category ( $M = 1.76$ ),  $t(17) = 31.85, p < .001$ . For example, the bluebird appeared to be more similar to the target bird than did the stegosaurus. Conversely, for the atypical items, adults rated the pictures that belonged to a category different from that of the target pictures (which were drawn to resemble the target pictures) as substantially more similar to the targets ( $M = 4.52$ ) than they rated the pictures from the same category as the target ( $M = 2.32$ ),  $t(17) = 8.38, p < .001$ . For example, the pterodactyl appeared to be more similar to the target bird than did the dodo.

We manipulated typicality for two reasons. First, it allowed us to construct a situation in which categories and appearances conflicted with one another. For example, the dodo does not much resemble a typical bird (although it is a bird), whereas the pterodactyl does resemble a bird (although it is not one). Second, by including atypical pic-

tures we ensured that children had not seen these pictures before and so were drawing new inferences beyond what they already knew with certainty.

In five of the sets, all pictures in the set came from within the same ontological category, or "basic category of existence" (Keil, 1979). The bird/dinosaur set described above is such a set; all exemplars are animals. For the remaining items, categories depicted within a set crossed an ontological boundary and so were conceptually more distinct. For example, one set consisted of snakes (which are animals) and sticks (which are inanimate objects). We hypothesized that children would grasp the importance of the category label for the cross-boundary sets (e.g., snakes/sticks) more readily than for the within-boundary sets (e.g., birds/dinosaurs) because the distinction between categories would be more obvious in the former (see Keil, 1979).

#### *Property Preselection*

To ensure that the properties we used were ones that young children would find familiar when applied to typical category members, we pretested a larger set of properties on 11 children. Two subjects were dropped from this preselection task because they answered "yes" to each question, leaving 9 subjects (mean age = 2 years, 10 months). Each child was shown one of the two typical pictures for each of nine picture sets and was asked a series of yes/no questions concerning each pic-

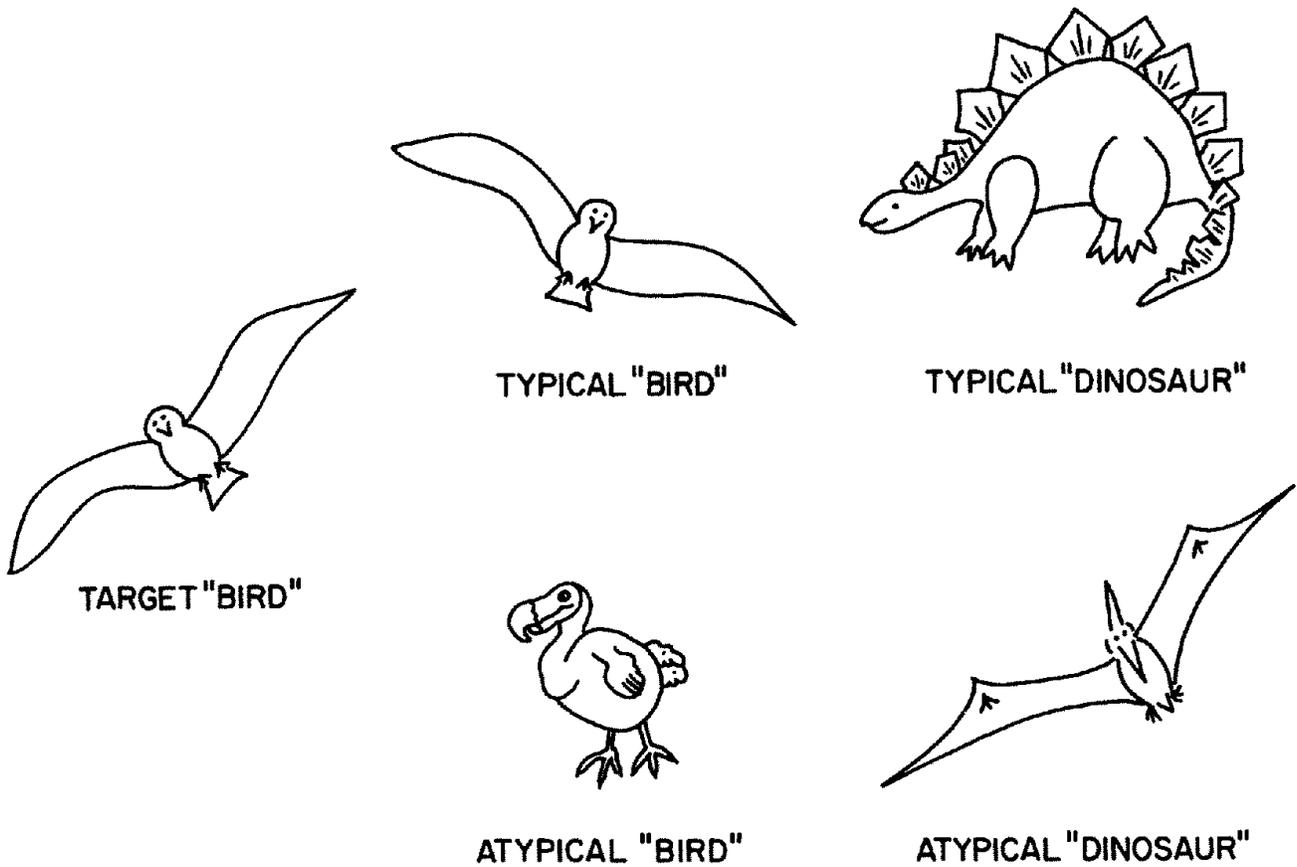


Figure 1. Sample item used in Studies 1 and 2.

ture. The pictures were not labeled. For example, a child was shown either the bluebird or the stegosaurus and was asked each of the following questions: Does it live in a nest? Does it have big teeth? Does it say "tweet tweet"? Does it eat worms? The properties used were chosen from among those that children most often answered correctly for both typical pictures in the set.<sup>3</sup> For example, the 4 children who saw the bluebird each answered that it lived in a nest, and the 5 children who saw the stegosaurus answered that it did not live in a nest. "Lives in a nest" was therefore deemed a familiar property for the typical pictures of this set (bluebird and stegosaurus) and was included in the subsequent studies.

### Procedure

The nine picture sets were presented in randomized order in picture-book format. All of the target pictures were in one book; all of the test pictures were in a different book. For every picture set, each child first saw the target picture and was reminded of one property that was true of that object. For example, they were shown the target picture of a bluebird and were reminded that it lives in a nest. The target picture was left in view, although the experimenter made no reference to the target picture after reminding the child of that initial fact. Children were then shown the four test pictures, one at a time, in random order. For each picture, children were asked whether or not the picture had the property that was given for the target picture. For example, children were shown the bluebird, dodo, stegosaurus, and pterodactyl one at a time and were asked, for each one, "Does this live in a nest?"

There were two conditions in Study 1. In the label condition, each picture was labeled for the child with the category name as it was presented. For example, the experimenter said, "This is a bird" as the picture book was opened to the target bird. The name of the test picture was also supplied when the experimenter asked the test question (e.g., "This is a bird. Does this bird live in a nest?") while pointing to the dodo). In the no-label condition, none of the pictures was labeled.

If children answered on the basis of category membership, they would judge that the bluebird and the dodo each live in a nest and that the dinosaurs do not. If they answered based on overall appearances, they would judge that the bluebird and the pterodactyl each live in a nest and that the stegosaurus and dodo bird do not. Or, children could have shown an inconsistent response (e.g., judging that only the dodo lives in a nest) or a response bias (e.g., saying "yes" on all items).

*Adult tasks.* Adult subjects completed a paper-and-pencil version of the same task that children received. Half of the adults were randomly assigned to the label condition; the remainder were assigned to the no-label condition. For each yes/no question (these were identical to those that children received), subjects circled their response and then rated how confident they were that their answer was correct on a scale from 1 (*not at all confident*) to 7 (*extremely confident*). Then they completed the similarity ratings (see Items, above). Finally, subjects in the no-label condition wrote down what they thought was the name of each picture.

<sup>3</sup> On seven of the properties, pretest subjects were 100% correct; on the remaining two properties, pretest subjects were 89% correct.

Table 2  
*Study 1: Mean Percentage Correct as a Function of Labeling Condition and Typicality*

Group	Typical	Atypical
2-year-olds		
Label	74***	69**
No label	76***	42*
Adults		
Label	98***	75***
No label	94***	76***

\* Below chance,  $p < .05$ . \*\* Above chance,  $p < .01$ . \*\*\* Above chance,  $p < .001$ .

### Results

Children's responses were scored for accuracy, with 1 indicating a correct answer and 0 indicating an incorrect answer. The data were summed across items and entered into a 2 (condition: label vs. no label)  $\times$  2 (picture typicality: typical vs. atypical instance)  $\times$  2 (within- vs. across-ontological boundary) analysis of variance (ANOVA). The results are shown in Table 2. As predicted, children were more accurate on the typical than atypical instances,  $F(1, 20) = 33.39$ ,  $p < .0001$ , and more accurate in the label than the no-label condition,  $F(1, 20) = 6.11$ ,  $p < .05$ . Moreover, the predicted Condition  $\times$  Typicality interaction occurred,  $F(1, 20) = 17.92$ ,  $p < .001$ .<sup>4</sup> Whereas performance on the typical items was unaffected by labeling (75% correct overall), performance on the atypical items was better when the pictures were named (69% correct) than when they were unnamed (42% correct). Thus, on the typical pictures, children correctly inferred the relevant property regardless of whether the pictures were labeled. But on the atypical pictures, the presence or absence of a label significantly affected performance. There were no differences between the within- versus across-ontological-boundary items. No other significant main effects or interactions were found.

We also analyzed the results of individual children (see Table 3). Children were scored as showing a *category-based* pattern if on 26 or more of the 36 trials (binomial  $p < .005$ ) they answered correctly ("yes" to pictures of the same category as the target, "no" to pictures of a category different from that of the target). Children were scored as showing an *appearance-based* pattern if on 26 or more of the 36 trials they answered based on appearances (correct on the typical instances and incorrect on the atypical instances). Children were scored as showing *response bias* if on at least 34 trials out of 36 they answered "yes" to every

Table 3  
*Study 1: Response Patterns for Individual Children*

Group	Category based	Appearance based	Response bias	Inconsistent
Label	6	0	3	2
No label	1	6	2	2

question or "no" to every question. Finally, all other children were scored as having *inconsistent* patterns.

Note that the responses of individual children mirror the results of the group as a whole: Category-based responses are found primarily when the label is provided. When a Fisher exact test was performed on just those children showing either a category-based or an appearance-based pattern, the effect for labeling condition was significant,  $p < .005$ . Also, response bias was fairly common. This result is not surprising, given the difficulty that children below age 3 have with yes/no questions (S. A. Gelman & Taylor, 1984). The response bias suggests that we have reached the lower age limit that can be tested with this procedure.<sup>5</sup>

We also examined individual items and discovered that on eight of the nine items, children performed better in the label than the no-label condition when responses to all four pictures in each set were considered. (On the ninth item children performed equally across conditions.) When we examined responses to the atypical pictures separately, we found that children performed better in the label than the no-label condition on all nine items. Thus, these results are fairly general across a range of different items.

*Adults.* The data from adults are presented in Table 2. Adults performed extremely well regardless of whether they were given the category labels. The yes/no data were scored as for children and entered into a 2 (label)  $\times$  2 (typicality) ANOVA. Adults were correct more often on the typical than the atypical items,  $F(1, 16) = 91.23$ ,  $p < .0001$ . There were no effects for labeling. Similarly, although subjects were highly confident overall that their answers were correct (mean rating of 6.06 on a scale of 1 to 7), they were more confident when judging typical ( $M = 6.40$ ) than atypical ( $M = 5.72$ ) pictures,  $F(1, 16) = 70.62$ ,  $p < .0001$ .

We suspected that one reason adults were able to perform so well in the no-label condition was that they often were able to figure out the category to which the pictures belonged on the basis of the perceptual information alone. For example, based on the animal's face and fins, they could tell that the flying fish was a fish and not a bird, despite its wings and bird-like shape and coloration. To take a closer look at this issue, we examined the labels that adults in the no-label condition provided on the naming task. For each subject, we noted whether the label they gave for each test picture was the same as the label they gave for the target picture for that set. For test pictures of the same category as the target, subjects correctly gave the same label on 78% of all trials (combining typical and atypical items). For test pictures of a category different from that of the target, subjects correctly produced a different label on 94% of all trials. Thus, adults were highly successful in determining category labels on the basis of subtle perceptual features alone.

Furthermore, to determine more directly whether knowledge

<sup>4</sup> All of these findings were reproduced using multivariate statistical tests. Univariate tests are reported here for clarity.

<sup>5</sup> When the 5 children with response bias are removed from the main analyses, we obtain the same pattern of results as when all subjects are included: a main effect for label,  $F(1, 15) = 16.26$ ,  $p < .001$ , a main effect for typicality,  $F(1, 15) = 56.10$ ,  $p < .0001$ , and a Label  $\times$  Typicality interaction,  $F(1, 15) = 29.51$ ,  $p < .0001$ .

of the label predicted adults' ability to draw the appropriate inferences, we computed Spearman nonparametric correlations between the naming scores (as calculated above) and the scores on the inference task on a subject-by-subject basis. The correlation was nonsignificant on the typical items, as would be predicted by subjects' near-ceiling performance on both tasks. However, on the atypical items, the two tasks correlated  $.73$  ( $df = 8$ ),  $p < .05$ . Adults' ability to categorize the atypical items predicts their performance on the inference task.

### Discussion

These results suggest that for children as young as age 2½ years, categories extend beyond salient perceptual similarity. Children show an early attention to the importance of names and an early ability to overlook salient appearances when drawing inferences. In this respect, the 2-year-olds in the present study are similar to 4-year-olds and adults in previous research. On the basis of this finding, we suggest that if there is a developmental shift away from similarity-based categories, it must occur below age 2½.

However, it could be that children were not drawing inferences about important underlying properties in this experiment. An alternative explanation for the results of Study 1 is that children may have had a bias to answer in accord with the experimenter's labels without consideration of the questions being asked. That is, because the target and test pictures received identical labels, children may have assumed that they should give the same answer for both.

We believe it is unlikely that children were simply showing this sort of response bias. They were asked about only one picture at a time, and the experimenter never made explicit comparison to the target after its initial presentation. Both of these factors should minimize any tendency toward response bias. Furthermore, performance on typical items was no better in the label condition than in the no-label condition. More important, performance was not at ceiling in either condition. Thus, if children were simply answering on the basis of words paired with pictures, performance should also have improved when the typical pictures were labeled.

Nonetheless, we wanted to rule out this possibility as an explanation of the findings of Study 1. To this end, we conducted a control study in which the pictures were paired with words that would not promote inferences for adults (e.g., simple adjectives such as "dirty"). If children have a tendency simply to say that two objects paired with the same word share other properties, then they should say that two things that are "dirty" are alike in other ways as well (e.g., they eat the same food). However, if the results of Study 1 reflect not response bias but rather children's beliefs about category structure, then children should not base their inferences on adjective labels such as "dirty" because such adjectives indicate neither category identity nor the associated deeper properties.

## Study 2

### Method

#### Subjects and Items

Eleven children (ages 2 years, 6 months–3 years, 1 month,  $M = 2$  years, 10 months) participated in the study. None of these children had

participated in Study 1. Each child saw the same picture sets used in Study 1.

### Procedure

The procedure was identical to that of the label condition of Study 1, with one exception: Rather than supplying category labels, the experimenter described each picture with a familiar adjective or adjectival phrase that denoted a transient property of the object under consideration. As in Study 1, each target picture was "named" as it was presented and was then followed by the reminder fact. For example, as the experimenter opened the picture book to the target bird, she said, "This is *wide awake*. It *lives in a nest*." Each test picture was also named with an adjective that either matched or contrasted with the adjective given to the target (e.g., "This [pointing to the dodo] is *wide awake*" or "This [pointing to the pterodactyl] is *sleepy*"). As in Study 1, after a picture was named, children were asked whether or not it had the property that was given for the target picture, for example, "This is *wide awake*. Does it live in a nest?" The full list of adjectives used is shown in Table 4.<sup>6</sup>

If children answer based on the adjectives, they will judge that the bluebird ("wide awake") and the dodo ("wide awake") live in nests and that the dinosaurs ("sleepy") do not. If they ignore the adjectives and answer based on overall appearances, they will judge that the bluebird and the pterodactyl live in nests and that the stegosaurus and dodo bird do not.

### Results

Children's responses were scored for accuracy, with 1 indicating a correct (category-based) answer and 0 indicating an incorrect answer. The data were summed across items and entered into a 2 (picture typicality: typical vs. atypical instance)  $\times$  2 (within- vs. across-ontological boundary) ANOVA. As in the no-label condition of Study 1, children performed much better on the typical items ( $M = 74\%$  correct) than on the atypical ( $M = 49\%$  correct) items,  $F(1, 10) = 19.58$ ,  $p < .002$ . Performance on the typical items was significantly above chance,  $t(10) = 4.81$ ,  $p < .002$ , whereas performance on the atypical items did not differ from chance,  $t(10) = 0.377$ ,  $p > .50$ . Children also performed better on items that crossed ontological boundaries (such as the snakes/sticks item;  $M = 68\%$  correct) than they did on items that did not cross ontological boundaries (such as the birds/dinosaurs item;  $M = 57\%$  correct),  $F(1, 10) = 8.52$ ,  $p < .02$ . There were no significant interactions.

Finally, we compared children's performance in Study 2 directly to that in Study 1. Linear contrasts revealed that performance in the adjective control condition did not differ from performance in the no-label condition and that both differed from performance in the label condition, but only on the atypical items (as expected),  $F(1, 30) = 29.14$ ,  $p < .001$ .

In short, children did not use familiar adjectives as the basis

<sup>6</sup> As is apparent in Table 4, most of the adjectives used in this experiment refer to properties that vary over time for a given individual and so clearly do not indicate stable category membership. Other adjectives (e.g., "pregnant," "dead," "smart") would probably function more like nouns, at least for adults, and so were deliberately excluded from the study. See Markman (1989) for evidence suggesting that adjectives and nouns generally differ in the extent to which they refer to stable categories, although counterexamples can also be found.

Table 4  
*Pictures, Adjectives, and Properties Used in Study 2*

Target/instance	Category 1	Category 2	Property
Within ontological category			
Wet	Wet	Dry	Eats dirt
Typical	Brown worm	Black snake	
Atypical	Planarian	Worm-like snake	
Nearby	Nearby	Far away	Eats carrots
Typical	Rabbit	Squirrel	
Atypical	Short-eared pika	Long-eared kaibab	
Wide awake	Wide awake	Sleepy	Lives in a nest
Typical	Bluebird	Stegosaurus	
Atypical	Dodo bird	Pterodactyl	
Dirty	Dirty	Clean	Barks
Typical	Golden retriever	White lamb	
Atypical	Chihuahua	Dog-like lamb	
Full	Full	Hungry	Says tweet-tweet
Typical	Swallow	Turquoise fish	
Atypical	Kiwi bird	Flying fish	
Across ontological categories			
Cold	Cold	Warm	Grows on a tree
Typical	Green leaf	Beetle	
Atypical	Pine needle	Green leaf-insect	
Little	Little	Big	Moves by itself
Typical	Green snake	Brown stick	
Atypical	Horned snake	Curved, green stick	
Pretty	Pretty	Ugly	Sits on the ground
Typical	Brown rock	Trout	
Atypical	Spotted cube	Lumpfish	
Good	Good	Bad	Grows in a garden
Typical	Flower	Gray fish	
Atypical	Jack-in-the-pulpit	Starfish	

of their inferences. Thus, these results rule out the possibility that children in Study 1 simply had a response bias to say “yes” when the experimenter provided matching labels and “no” when different labels were provided. Children were selective in their inferences, making a distinction between category names, which promote inferences (Study 1, label condition), and words labeling transient properties, which do not (Study 2).

### General Discussion

In Study 1, 2½-year-old children relied on language (in particular, category labels) for inferring the identity and further properties of novel objects. They reported, for example, that a pterodactyl does not live in a nest but that a dodo bird does. When the same pictures were labeled with adjectives rather than nouns in Study 2 (e.g., “This is sleepy” rather than “This is a dinosaur”), the results mirrored those from when the pictures were not named. Children did not use temporary-state adjectives as the basis of their inferences; they paid special attention to labels only when they named the object category. (See also Waxman, 1990, for evidence that adjectives do not highlight category relations for young children.)

Thus, by age 2½ years, children expect categories to promote rich inductive inferences concerning important properties, and they can overlook conflicting perceptual appearances in doing so. The children in the present studies were approximately 1 year younger than those previously reported to base inferences on category membership (conveyed by the label) when it conflicts with perceptual similarity (S. A. Gelman & Markman, 1987). Thus, the present results suggest that if there is a developmental shift away from similarity-based categories, it has taken place by age 2½.

The present results provide even stronger evidence that children’s belief in the power of language for their inferences is not dependent on extensive scientific knowledge. That is, young children apparently do not require detailed knowledge of biology in order to use words such as “bird” and “dog” as the basis of their inferences. (See Carey, 1985, for evidence of clear misconceptions about biology throughout the preschool years.) The children in this study were only 2 years old, too young to have been taught the particulars of why pterodactyls are not birds. Rather, children learn the system of language taught by adults and draw inductive inferences in accordance with that system. Thus, children’s inferences are consistent with adult

theories without necessarily being based on them. We suggest that children have an early-emerging expectation that does not itself constitute a theory but that will pave the way for later theory building. In particular, the category name sets up an expectation on the part of the child that there is much more in common to members of the category than meets the eye.

### *Children's Inferences: Induction or Deduction?*

The design of this study differs from previous work (S. A. Gelman, 1988; S. A. Gelman & Markman, 1986, 1987) in that all of the questions concerned properties that children knew to be true or false of prototypical category instances (e.g., "Does this [rabbit] eat carrots?" vs. "Does this [dog] have leukocytes all through it?"). This raises the issue of whether children were actually making inductive or deductive inferences. That is, one might argue that children were not drawing *inductive* inferences (which, by definition, mean that children can use categories to extend their knowledge beyond what they already know; Skyrms, 1975) but rather were drawing *deductive* inferences that follow directly from their prior knowledge. Whether children are making inductions or deductions is of great theoretical importance because it tells us whether children expect categories to capture important new information (as revealed by their inductions) or whether they are simply reporting facts that they have been taught directly (indicating use of deductions). Our position is that children were making inductive inferences. In the following, we explain why.

In order for their inferences to have been *deductively* valid, children in the label condition would have had to start with the premise that, for example, "all rabbits eat carrots." Coupled with the information provided in the experiment that "this [animal  $x$ ] is a rabbit," the premise would lead to the conclusion that animal  $x$  eats carrots. Therefore, in order to argue that children were reasoning *deductively*, one would need to account for how children obtained the premise that "all rabbits eat carrots."

One possibility is that children could have been told explicitly that all rabbits eat carrots. We did not provide this information. We simply reminded children that the property in question was true for the target picture. Furthermore, it seems improbable that parents or other adults would have done so for all of the property–category pairings in the study ("all worms eat dirt," "all snakes move by themselves," "all leaves grow on trees," etc.). But even if they had, preschoolers do not seem to understand the word "all" as a universal quantifier (Inhelder & Piaget, 1964), often confusing "all" with "some" and not understanding that "all" implicates the entire category. Similar misunderstandings arise with "always" (Braine & Romain, 1983). Thus, even if the input consistently included statements with "all" or "always" (which seems unlikely), children would not necessarily know that such statements warrant deductive inferences. Indeed, older preschoolers have difficulty drawing appropriate deductive inferences even when the premises are clearly stated (C. L. Smith, 1979).

A second possibility is that children may have constructed the premise "all rabbits eat carrots" themselves. Notice, however, that this proposition would itself have had to be constructed by means of an *induction*: From a limited number of

exemplars (Rabbit 1 eats carrots, Rabbit 2 eats carrots, etc.), the child could have induced that *all* rabbits eat carrots. It is also important to note here that children could have formed a very different kind of induction based on their prior experience. Considering the rabbit example again, children could have induced that "animals that look like rabbits eat carrots." There is no logical or empirical necessity for children to have preferred the induction "all rabbits eat carrots" over the induction "all animals that look like rabbits eat carrots" because, in the past, children found equally strong evidence for both inductions. Thus, even if children had formed inductions prior to participating in the present experiment, that in itself would not tell us what *kind* of inductions had been formed.

From the present experiment, we cannot determine when children's inductions took place. In other words, the present task could be assessing either the inductions that children made spontaneously outside of the laboratory or inductions made during the experiment itself. Whichever is the case, the results are valuable for telling us which kinds of inductions children find most reasonable or natural. In particular, the results indicate that 2½-year-olds are willing to overlook appearances in the inferences they draw (either spontaneously or in response to the experimental questions).

### *Conclusions*

For 2-year-old children, category membership can be more powerful than surface appearances in guiding inferences about important properties. However, we do not wish to underestimate the role of appearances in children's concepts. Perceptual clues are still the primary means of discovering category membership for unlabeled objects. For example, when shown atypical instances that were not labeled, children performed significantly below chance, suggesting that they were answering on the basis of appearances. An important question we have not addressed is when children begin to sort out which of the many available perceptual cues reliably indicate category membership. Children in the present study apparently had difficulty determining the category membership of the atypical items. In contrast, adults performed well even when pictures were not named; in fact, there were no differences in performance between the two conditions.

This finding suggests that one important development still to take place is the ability to focus on subtle perceptual cues to determine category membership in the absence of language. Which cues are used to determine category membership may be partly a function of the theories one holds. For example, to someone with biological training, the presence of a blowhole is a critical clue that a sea creature is a mammal and not a fish; the biologist knows that a blowhole indicates that the animal breathes air with lungs. But to a naive observer who is unaware of the theoretical link between blowholes and lungs, the same creature may be classed as a fish (cf. the classification of whales before Linnaeus). Thus it is important that even before children can make use of subtle perceptual cues to determine category membership,<sup>7</sup> they readily use category labels as the basis of their inferences (see also S. A. Gelman, Collman, & Maccoby, 1986; Keil, 1986). In sum, for children as young as age 2½,

language conveys important information beyond that which meets the eye.

<sup>7</sup> We do not wish to imply that these children are wholly incapable of attending to subtle cues to category membership, however. In recent pilot data, we found that 2-year-olds could detect the category membership of some novel pictures on the basis of subtle perceptual cues (e.g., a lamb-like dog was recognized as a dog on the basis of its nose, feet, and tail, despite its overall color and fur). However, children clearly gain much sophistication along these lines as they grow older.

### References

- Anglin, J. M. (1977). *Word, object, and conceptual development*. New York: Norton.
- Braine, M. D. S., & Romain, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 263-340). New York: Wiley.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Cohen, L. B., & Younger, B. A. (1983). Perceptual categorization in the infant. In E. F. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory?* (pp. 197-220). Hillsdale, NJ: Erlbaum.
- Fenson, L., Cameron, M. S., & Kennedy, M. (1988). Role of perceptual and conceptual similarity in category matching at age two years. *Child Development, 59*, 897-907.
- Fenson, L., Vella, D., & Kennedy, M. (1989). Children's knowledge of thematic and taxonomic relations at two years of age. *Child Development, 60*, 911-919.
- Flavell, J. H. (1985). *Cognitive development* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 167-230). New York: Wiley.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.
- Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57*, 396-404.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-209.
- Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532-1541.
- Gelman, S. A., & Taylor, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Development, 55*, 1535-1540.
- Gentner, D. (1988). Metaphor as structure-mapping: The relational shift. *Child Development, 59*, 47-59.
- Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: Norton.
- Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? *Child Development, 45*, 469-473.
- Keil, F. C. (1979). *Semantic and conceptual development*. Cambridge, MA: Harvard University Press.
- Keil, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition* (pp. 133-153). Norwood, NJ: Ablex.
- Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
- Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology, 16*, 1-27.
- Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. *Journal of Experimental Child Psychology, 31*, 470-486.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.
- Neisser, U. (1987). Introduction: The ecological and intellectual bases of categorization. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 1-24). New York: Cambridge University Press.
- Olver, R. R., & Hornsby, J. R. (1966). On equivalence. In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 68-85). New York: Wiley.
- Skyrms, B. (1975). *Choice and chance: An introduction to inductive logic* (2nd ed.). Encino, CA: Dickenson.
- Smith, C. L. (1979). Children's understanding of natural language hierarchies. *Journal of Experimental Child Psychology, 27*, 437-458.
- Smith, L. B. (1989). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 146-178). New York: Cambridge University Press.
- Sugarman, S. (1982). Developmental change in early representational intelligence: Evidence from spatial classification strategies and related verbal expressions. *Cognitive Psychology, 14*, 410-449.
- Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5*, 123-150.
- Wellman, H. M., & Gelman, S. A. (1988). Children's understanding of the nonobvious. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 99-135). Hillsdale, NJ: Erlbaum.

Received May 19, 1989

Revision received December 27, 1989

Accepted January 5, 1990 ■