Inductions from Novel Categories: The Role of Language and Conceptual Structure

Natalie S. Davidson
Susan A. Gelman
University of Michigan

Young children expect members of a familiar category to share nonobvious prop-
erties. The present experiments were designed to determine whether children
expect unfamiliar objects with the same novel label to share nonobvious proper-
ties. In three experiments, 4- and 5-year-old children were taught properties about
unfamiliar objects (e.g., a gnu-like animal) and then were tested on their gener-
alization of each property to four test objects. The test objects were either similar
or dissimilar in appearance to the target and received either the same or a different
novel label from that of the target. When category labels and appearances were
completely independent of one another, children drew inferences based on ap-
pearances rather than novel category labels. However, when the same pictures
received familiar labels (e.g., a cow, a deer) instead of novel labels (e.g., a fep, a
wug), children drew inferences based on category labels as well as appearances.
Furthermore, children did draw inferences on the basis of novel labels when
perceptual cues were not completely discrepant from the labels. These results
suggest that novel labels help guide children's inferences from unfamiliar catego-
ries, but only when the conceptual basis of the names is clear.

An important function of categories is to extend knowledge beyond what is
known from direct experience or observation. That is, categories promote induc-
tive inferences concerning nonobvious properties. For example, once we dis-
cover that one bird has a right aortic arch, we can reasonably infer that other
birds also have right aortic arches. Inferences such as these at times go beyond

appearances, as when we infer that a penguin and a robin share certain underlying similarities because both are birds, despite their surface dissimilarities. In a recent set of experiments, Gelman and her colleagues (Gelman, 1988; Gelman & Markman, 1986, 1987; Gelman & O'Reilly, 1988) found that children grasp this function of categories by the age of 3 years. That is, young children realize that familiar categories often share important properties that are not perceptually obvious.

In order to determine whether children realize that category membership can be more important than perceptual cues in their inductive inferences, Gelman and Markman conducted several studies in which, for several different categories, the perceptual similarity of the objects was pitted against their category membership. The task involved teaching children new facts and then charting how they generalized the new information. For example, on one item, children were shown a triad of pictures consisting of a flamingo, a bat, and a blackbird. The experimenter told the child, "This bird gives its babies mashed-up food" (pointing to the flamingo), and "This bat gives its babies milk" (pointing to the bat). Then the child was asked, "Does this bird (the blackbird) give its babies mashed-up food like this bird (the flamingo) or milk like this bat?" In this and other items, the test picture was labeled with the same category label as one of the first two pictures (e.g., both were called *birds*), but was perceptually dissimilar (e.g., the blackbird looked different from the flamingo). The test picture was similar in appearance to the other one of the first two pictures (e.g., the blackbird resembled the bat) yet received a different label from it. The results showed that children consistently based their generalizations on the category membership of the objects and not on their perceptual similarities. For example, in this case, children typically reported that the blackbird gives its babies mashed-up food like the flamingo, despite the striking similarity between the blackbird and the bat. Control conditions were included to rule out the possibility that children were basing their inductions on the identity of category labels rather than on the category membership of the objects. For example, children continued to draw inferences from one category member to another even when category members received synonymous rather than identical labels.

Gelman and Markman's finding is surprising in light of traditional developmental theories which would predict that children attend only to obvious perceptual qualities of things (Wellman & Gelman, 1988). It also raises the question of what the basis is of children's ability to look beyond the obvious on this task. How is it that children at such an early age, with little or no scientific training, can appreciate the importance of category membership for inductions? There are at least two distinct possibilities, representing two ends of a continuum. At one extreme, children's beliefs in the rich structure of categories may be individually built up for each category. That is, children may need experience with and knowledge about a category before assuming that it promotes inductions. Chil-

dren may reason something like this: "I know that members of this category have many obvious properties in common, so they probably have many nonobvious properties in common, too." For example, children rapidly learn that all or most birds have feathers, lay eggs, build nests, fly, have wings, make chirping sounds, and so forth. On the basis of these observed similarities, children may extrapolate that birds are also likely to share less overt features (e.g., internal parts, parenting behaviors, diet). In other words, through their experience with specific categories, children may come to expect members of these categories to share deep similarities. If this is the case, then children will base their inductions on category membership only for familiar categories.

At the other extreme, children may hold a general expectation about language that is independent of their specific experiences with a given category. Children may assume that objects with the same name share important, nonobvious properties. This expectation could apply to novel objects and novel labels not previously encountered. This latter possibility would be at least consistent with recent research, which has uncovered several early expectations that children have about the structure of natural language categories. For example, Macnamara (1982) claims that children as young as 18 to 24 months assume that a common noun refers to an object as a whole rather than to one of its properties. Moreover, children assume that a common noun picks out an entire category of objects whereas a proper noun refers to only an individual object (Gelman & Taylor, 1984; Katz, Baker, & Macnamara, 1974). Children also have expectations about how categories encoded in language are organized. By 3 years of age, children assume that nouns refer to objects that are taxonomically related (e.g., a dog and a cat) even though in simple classification tasks in which no label is used they typically group objects on the basis of thematic relations (e.g., a dog and a bone) (Markman & Hutchinson, 1984; see Waxman & R. Gelman, 1986, for a related finding).

The purpose of the present studies is to investigate which of the above two possibilities better characterizes why children draw category-based inferences. In particular, do children need extended experience with a given category before treating it as the basis of inductions, or do children hold a general expectation that objects with the same name share deep similarities, even for novel objects with which children have little specific experience? To examine this question, we looked at the role that language plays in children's inductive inferences concerning unfamiliar objects. In the first experiment, we examined whether children make use of a novel label of an unfamiliar object in determining when to generalize one of its properties. In the second experiment, we examined the effect that children's experience with a category had on their inductions. In the third experiment, we varied the structure of the category involved to determine whether children are more willing to draw inferences within some novel categories than within others.
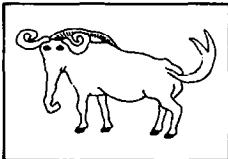
EXPERIMENT 1

In this experiment, we investigated the importance of language for children's inductions from unfamiliar categories using a task similar to that of Gelman and Markman (1986, 1987). However, rather than using familiar objects and familiar category labels (e.g., a *bird*), we used unfamiliar objects (e.g., a gnu-like animal) and labeled them with unfamiliar terms (e.g., a *zav*). The task involved teaching children facts about a target object and then charting the extent of their generalizations of the fact to four test objects. The test objects varied in their relatedness to the target on two dimensions, category label and perceptual similarity. That is, two of the test objects were perceptually similar to the target while the other two looked quite different from the target. Moreover, the category labels of the two objects at each level of perceptual similarity were varied. One object at each level received the same label as the target and the other received a different label. In this manner, we were able to determine whether children base their inductive inferences on the labels of objects with which they have no prior experience.

Method

Subjects.   Fifty-three children (27 girls and 26 boys) participated, recruited from a list of interested parents. The mean age of the children was 4;9, ranging in age from 4;3 to 5;3. In addition, 18 adults participated in a rating task to preselect the pictures (see the following).

Materials.   Eight sets of pictures of unfamiliar objects were used, each set containing five objects from the same superordinate category. Four of the sets included natural kinds (mammals, reptiles, insects, and vegetables); the other four included artifacts (furniture, machines, measurement devices, and vehicles). Each set of pictures consisted of one target and four test pictures. Two of the test pictures looked very similar to the target, while the other two looked very different from it. The pictures were drawn in accord with several specific criteria. That is, the similar pictures differed from the target only in shade of color, position, and one salient feature (e.g., a different tail than the target's). The dissimilar pictures differed from the target in several respects: color, position, size, coat or material (e.g., wood versus metal), and two to three other salient features. One of the picture sets is shown in Figure 1.

Adult Ratings.   It was important to ensure that, for each pair of pictures, both pictures at a given level of similarity were equally similar to the target picture. For this reason, we asked 18 adult volunteers to rate the perceptual similarity of each test picture to the target, for 16 sets of pictures that had been drawn according to the criteria described above. The rating scale ranged from 1 ("not similar at all") to 7 ("extremely similar"). The final eight sets that were used in the study were chosen from these original 16 because they met the

	Conditions		
	Novel Label	Familiar Label	Unrelated Objects[a]
Target Picture			
	a zav	*a cow*	*a zav*
Test Pictures			
	a zav	*a cow*	*a zav*
	a traw	*a deer*	*a traw*
	a zav	*a cow*	*a zav*
	a pume	*a giraffe*	*a pume*

[a] A different target picture was used in this condition (i.e., a measuring instrument).

Figure 1.   The structure of each picture set for Experiments 1 and 2.

following requirements: (a) The similar items were perceived as very similar to the target (i.e., the mean similarity rating of each picture was 5 or greater, with a mean total of 5.9), (b) the dissimilar items were perceived as very different from the target (i.e., the mean similarity rating of each picture was 3.1 or lower, with a mean total of 2.5), (c) the two similar items per set were equally similar to the

target (i.e., the mean difference between their similarity ratings was 0.45), and
(d) the two dissimilar items per set were equally different from the target (i.e.,
the mean difference between their similarity ratings was 0.40).

All of the properties taught to children in the study were simple attributes that
the children could easily understand (see Table 1). They were modified from a
collection of such attributes generated by adult subjects as ones typical of in-
stances of these superordinate categories (Gelman, 1988, footnote 5, p. 73).
Because the categories being tested were unfamiliar, and because the attributes
could not be determined from the appearances of the pictures alone, children
would be unable to use prior knowledge of these attributes in answering the
questions.

Procedure.   All children were seen individually. Children were told that they
would be shown some pictures of unfamiliar things and would be asked to help
answer some questions about them. There were three conditions: the Novel Label
condition ($n = 20$), the No Label condition ($n = 16$), and the Unrelated Objects
condition ($n = 17$).

Novel Label Condition.   For each set, the experimenter showed the child the
target picture and taught him or her a label for the object (e.g., "This is a zav").
The labels used were all nonsense words, including *biv, bof, cheem, durn, fant,
jop, kif, lobe, lorse, prote, pume, rom, shap, sim, skub, thorp, traw, turp, vaw,
vit, woj, wug, yub,* and *zav*. These labels were randomly assigned to the pictures
in the eight sets. The child also learned a property of the target (e.g., "This zav
has four stomachs"). Then, the child was shown each of the four test objects, one
at a time. For each of the four test pictures, the child learned a label for the object
and was then asked whether he or she thought that the property learned about the
target was also true of the test object [e.g., "This is a traw (test picture). Do you

Table 1.   Categories and Properties Used
in Experiment 1

Category[a]	Property
Mammals	has four stomachs
Reptiles	lives in trees
Insects	eats bananas
Vegetables	is good for your fingernails when you eat it
Vehicles	uses electricity
Furniture	is used to hang shirts on
Machines	is used to make tables
Measuring instruments	is used to find lost pennies

[a]We did not use these category labels in the experiment. They
are listed here to indicate that the pictures were intended to belong
to these categories.

think this traw has four stomachs like this zav (target)?"]. One of the similar objects was labeled with the same label as the target (e.g., *zav*), and the other similar object received a different label (e.g., *traw*). Likewise, one of the dissimilar objects received the same label as the target (e.g., *zav*), and the other received a different label (e.g., *pume*). (A sample item with the novel labels is shown in Figure 1.) This procedure was repeated for each of the four test pictures in all eight sets. The order of presentation of the sets and the order of test pictures within each set was randomized for each subject. Also, the assignment of same vs. different labels to each member of a pair of pictures at the same level of similarity was counterbalanced across subjects.

    No Label Condition.   The purpose of this control condition was to determine the separate effect of appearances on children's inferences. Children were shown the same pictures as in the Novel Label condition, but were not taught any labels for them [e.g., "See this? Do you think this (test picture) has four stomachs like this (target)?"].

    Unrelated Objects Condition.   The purpose of this condition was to ensure that any effects we might find for labeling would not be due to a response bias (i.e., to answer "yes" whenever the target and test pictures had the same label and "no" whenever they were different), without any consideration of how plausible the inferences were for a given set of test pictures. In this condition, the procedure was the same as in the Novel Label condition, but the picture sets were designed so that the target was completely unrelated to the test pictures. This was accomplished by pairing a target picture from a natural kind category with a set of test pictures from an artifact category, and vice versa. For example, one set consisted of a mammal as the target object (such as the one in Figure 1) and four measuring instruments as the test objects. Thus, on one trial children heard, "See this zav (pointing to one of the measuring instruments)? Do you think this zav has four stomachs like this zav (pointing to a mammal)?"

Results and Discussion
If children expect objects that receive the same label to share important novel properties, then we would expect them to show a tendency to draw inferences on the basis of the category labels in our task.[1] That is, they should more often draw inferences when the label of a test object matches that of the target than when the labels do not match. For example, children should draw more inferences when asked, "Do you think this zav has four stomachs like this zav?" than when they are asked, "Do you think this traw has four stomachs like this zav?" On the other hand, if children base their inferences strictly on the appearances of objects, they

---

[1] It is not clear from the current studies whether this tendency would be due to a bias that children have to interpret category labels in this way or to a constraint on the interpretations that they consider (see Keil, 1981; Nelson, 1988).

should draw an equal number of inferences to both of the test pictures at each level of perceptual similarity to the target.

Each child scored a "1" every time he or she drew an inference from the target to the test picture, that is, every time he or she answered "yes." Since preliminary analyses showed no significant sex differences, sex was not included as a factor in subsequent analyses. A four-way ANOVA was performed on the total scores for each test picture, calculated across all eight sets. Condition (Novel Label vs. No Label vs. Unrelated Objects) was the between-subjects variable, with repeated measures on category type (natural kind vs. artifact), appearance (similar to target vs. dissimilar), and label (same as target vs. different). (Note that the appearance factor does not apply to the Unrelated Objects condition since all pictures in that condition were dissimilar from the target in appearance.) Within each cell of this design, a subject's score could range from 0 to 4. Results are shown in Figure 2. There were five significant results.

First, there was a main effect for condition, $F(2,50) = 8.69$, $p < .001$. Children drew significantly fewer inferences in the Unrelated Objects control condition ($M = 20\%$) than they did in either the Novel Label condition ($M =$
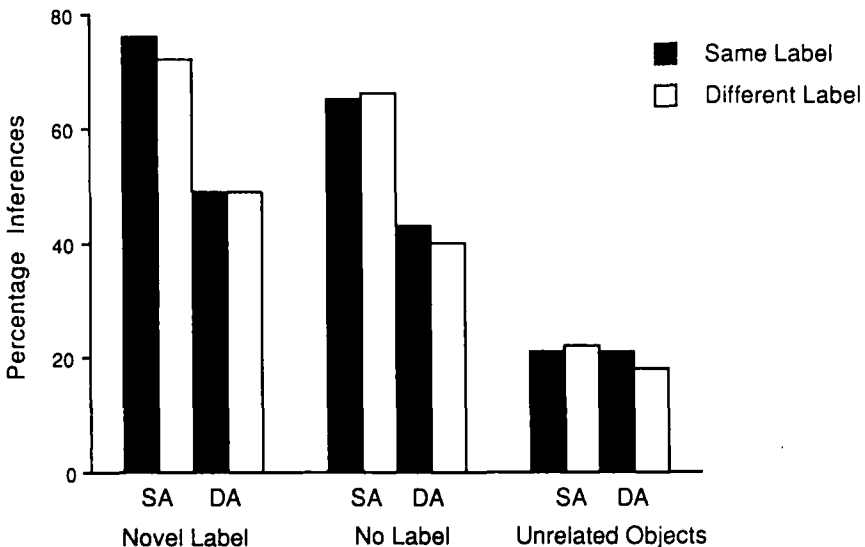
Figure 2.   Mean percentage of inferences as a function of picture and condition: Experiment 1. [Pictures did not receive labels in the No Label condition, so appearance is the only relevant dimension for that condition. The target picture in the Unrelated Objects condition was from a different category (e.g., a target picture of a mammal was paired with four test pictures of measuring instruments), so all of the test pictures in that condition were dissimilar in appearance from the target. The bar labels "SA" and "DA" refer to "Similar Appearance" and "Dissimilar Appearance," respectively.]

62%), $t(35) = 4.12, p < .002$, or the No Label condition ($M = 54\%$), $t(31) = 2.92, p < .01$. This result indicates that children rarely made cross-category inferences even when the labels matched exactly. Thus, children are sensitive to how plausible the inference would be in making their responses. In an additional t test it was found that there was no significant difference between the Novel label and No Label conditions.

The second significant result was that children drew more inferences to the test pictures that were perceptually similar to the target ($M = 55\%$) than to those that were perceptually dissimilar from the target ($M = 37\%$), as shown by a main effect for appearance, $F(1,50) = 44.12, p < .0001$. There was also an Appearance $\times$ Condition interaction, $F(2,50) = 7.90, p < .01$. This result was due to the absence of an appearance effect for the Unrelated Objects condition, as predicted.

Finally, category type was found to have two significant effects on children's inductions. First, children drew significantly more inferences to artifacts ($M = 49\%$) than to natural kinds ($M = 43\%$), $F(1,50) = 13.72, p < .001$.[2] However, this result must be interpreted in light of an Appearance $\times$ Category Type interaction, $F(1,50) = 4.60, p < .05$. Appearance had a larger effect on the inductions that children made for the natural kinds than it did for the artifacts. The mean difference between the number of inferences made to pictures of similar appearance to the target and the number made to pictures of different appearance from the target was 21% for the natural kinds and only 14% for the artifacts.

The type of label that a picture received did not affect the number of inferences children drew. In the Novel Label condition, for pictures similar in appearance to the target, there was no significant difference between the number of inferences drawn to the test objects with the same label as the target ($M = 76\%$) and those with a different label ($M = 72\%$). The same was true of pictures dissimilar in appearance from the target ($M = 49\%$ for both kinds of pictures). Thus, children did not use labels as the basis of their inferences for these novel objects. Rather, children appeared to be using appearances exclusively.

There are at least two possible explanations for the previous result. First, children might require experience with a category before they set up expectations about the inferences they can draw from it. All of the pictures and labels in the Novel Label condition were ones with which children had no experience.

---

[2] This result is somewhat surprising, given recent findings that, for *familiar* categories, natural kinds generally promote more inferences than do artifact categories (Gelman, 1988; Gelman & O'Reilly, 1988). Perhaps the natural kind—artifact differences found previously were partly a result of subjects' experience with, and knowledge about, actual natural kinds and artifacts. However, it is also important to realize that in the present study neither the properties taught nor the complexity of the objects included were equated across category type. Thus, the category type differences obtained here should be viewed with caution, and may not generalize to a different set of items.

However, there is another possible explanation. To explain, we first point out that the Novel Label condition provided a very strong test of children's understanding that labels imply deep similarities. For every set, the target and one test picture differed substantially in their appearance yet received the *same* label, while the target and another test picture were highly similar in appearance yet received *different* labels. Children may have been unable to understand how an object that looked like the target could be a member of a different category, while an object that appeared different from the target was in the same category. Some minimum amount of similarity (perceptual or otherwise) may be necessary before children accept a common noun label. This could be a particularly serious problem with novel labels, about which children have minimal information. Thus, the labels in Experiment 1 may not have made sense to our subjects because they violated children's expectation that things with the same name are deeply alike. Thus, the names could have been ignored. But if objects were labeled in a way that made sense to children on a conceptual level, with labels that were more consistent with their expectations, perhaps children would use the labels as clues to induction.

In the following experiments, we examined the preceding possible explanations for our results. In Experiment 2, we examined the effect that children's experience with a category has on their inductions. In Experiment 3, we labeled the pictures more closely in accord with children's nonlinguistic understanding.

EXPERIMENT 2

In the second experiment, we investigated more explicitly the effect that experience with a category has on children's inductions. Rather than include only unfamiliar categories (as in Experiment 1), we compared children's inferences within familiar and unfamiliar categories directly. More specifically, we varied how familiar the category labels were to the children by using either familiar labels (e.g., a *cow*) or unfamiliar labels (e.g., a *zav*) for the same set of pictures. We then charted the extent of children's inductions to the test objects in order to determine whether children would draw more inferences within categories that were familiar than within categories that were unfamiliar.

Method

Subjects.   Thirty-three children (18 girls and 15 boys) participated in the main experiment. The mean age of the children was 4;8, ranging in age from 4;1 to 5;4. In addition, 12 children (4;1 to 5;2, mean age: 4;8) participated in the pretest (see the following). None of these children had participated in Experiment 1.

Materials.   The same pictures were used in this experiment as in the previous one. All of the properties taught about the targets were simple attributes that children could easily understand.

Procedure.   All children were seen individually. Children were told that they would be shown some pictures of unfamiliar things and would be asked to help answer some questions about them. There were two conditions: the Novel Label condition ($n = 16$) and the Familiar Label condition ($n = 17$).

Novel Label Condition.   The task for this condition was the same as that in the Novel Label condition of Experiment 1. For each picture set, the experimenter showed the child the target picture and taught him or her a label for the object. The same nonsense words were used as in Experiment 1. However, in this study, since the labels for the Familiar Label condition could not sensibly be assigned to the pictures randomly, the labels for the Novel Label condition, rather than being randomly assigned, were fixed for each picture as well. Only the initial assignment of labels to each set was random. Within each set, the type of label used for a picture (i.e., either the same or different from the target's label) was held constant across the two conditions.

Each child learned a property of each target picture (e.g., "This zav has four stomachs inside"). He or she was then asked, for each of the four test pictures, whether the property learned about the target was also true of the test picture [e.g., "This is a traw (test picture). Do you think this traw has four stomachs inside like this zav (target)?"]. This procedure was repeated for each of the four test pictures in all eight sets. The order of presentation of the sets and of the test pictures within each set was randomized for each subject.

Familiar Label Condition.   The only difference between this condition and the Novel Label condition was the use of real words that children would know for the labels in the Familiar Label condition. These included *cow, deer, giraffe, dinosaur, lizard, frog, spider, grasshopper, caterpillar, onion, corn, pineapple, spaceship, submarine, car, hanger, television antenna, chair, coffeemaker, saw, mixer, clock, camera,* and *lamp.* The words were assigned to the pictures so that they were all unusual but plausible members of these familiar categories. That is, the primary characteristics of each picture were consistent with those of its corresponding familiar object. All objects were drawn to have the correct material (e.g., the lizards had scales; the saws appeared to be made of metal; the deer had fur), to have parts that were consistent with those of their familiar counterparts (e.g., the cows had hooves; the car had tires; the submarine had a periscope), and to have approximately the same shape as their familiar counterparts (e.g., the pineapple was short and squat; the giraffe was tall and lean; the camera was a square box sitting on a tripod).

For example, one of the target pictures was of a gnu-like animal with an elephant's trunk and was labeled a *cow* ("See this cow? This cow has four stomachs inside."). Just as in the Novel Label condition, each child was asked whether the property taught about the target was also true of the test picture [e.g., "This is a deer (test picture). Do you think this deer has four stomachs inside like this cow (target)?"].

Pretest of Properties.   It was important to ensure that children did not al-

ready know the properties when asked about the familiar categories. For this reason, it was necessary to change several of the properties that would otherwise be too familiar to children. We then pretested all of the properties to make certain that the final set included ones that children did not already know. We showed 11 children all of the test pictures, one picture at a time, and labeled each with the category label it would receive in the Familiar Label condition. When each picture was shown, the child was asked whether or not it had the particular property of interest (e.g., "This is a coffeemaker. Do you think this coffeemaker has batteries inside?"). The final set of eight properties was chosen to include ones that children did not already know. For these properties, there was no significant difference between the number of "yes" responses to the test objects that would be labeled the same as the target ($M = 32\%$) and the number to those that would be labeled differently from it ($M = 31\%$). The final set of properties is shown in Table 2.

Results

A child scored a "1" every time he or she drew an inference from the target to the test picture. Since preliminary analyses showed no significant sex differences, sex was not included as a factor in subsequent analyses. A four-way ANOVA was performed on the total scores for each test picture, calculated across all eight sets. Condition (Novel Label vs. Familiar Label) was the between-subjects variable, with repeated measures on category type (natural kind vs. artifact), appearance (similar to target vs. dissimilar), and label (same as target vs. different). Results are shown in Figure 3. There were four significant results.

Table 2.   Categories and Properties Used in Experiment 2

Category[a]	Familiar Labels[b]	Property
Mammals	cow, deer, giraffe	has four stomachs inside
Reptiles	lizard, dinosaur, frog	lives in sand
Insects	grasshopper, spider, cater-pillar	eats raisins
Vegetables	onion, corn, pineapple	is good for your fingernails when you eat it
Vehicles	spaceship, submarine, car	is made of steel
Furniture	hanger, TV antenna, chair	has wires inside
Machines	saw, coffeemaker, mixer	has batteries inside
Measuring instruments	clock, camera, lamp	has a motor inside

[a]We did not use these category labels in the experiment. They are listed here to indicate that the pictures were intended to belong to these categories.
    [b]These are the labels used in the Familiar Label condition. The labels used in the Novel Label condition are listed in the text under the *Novel Label Condition* heading in Experiment 1.
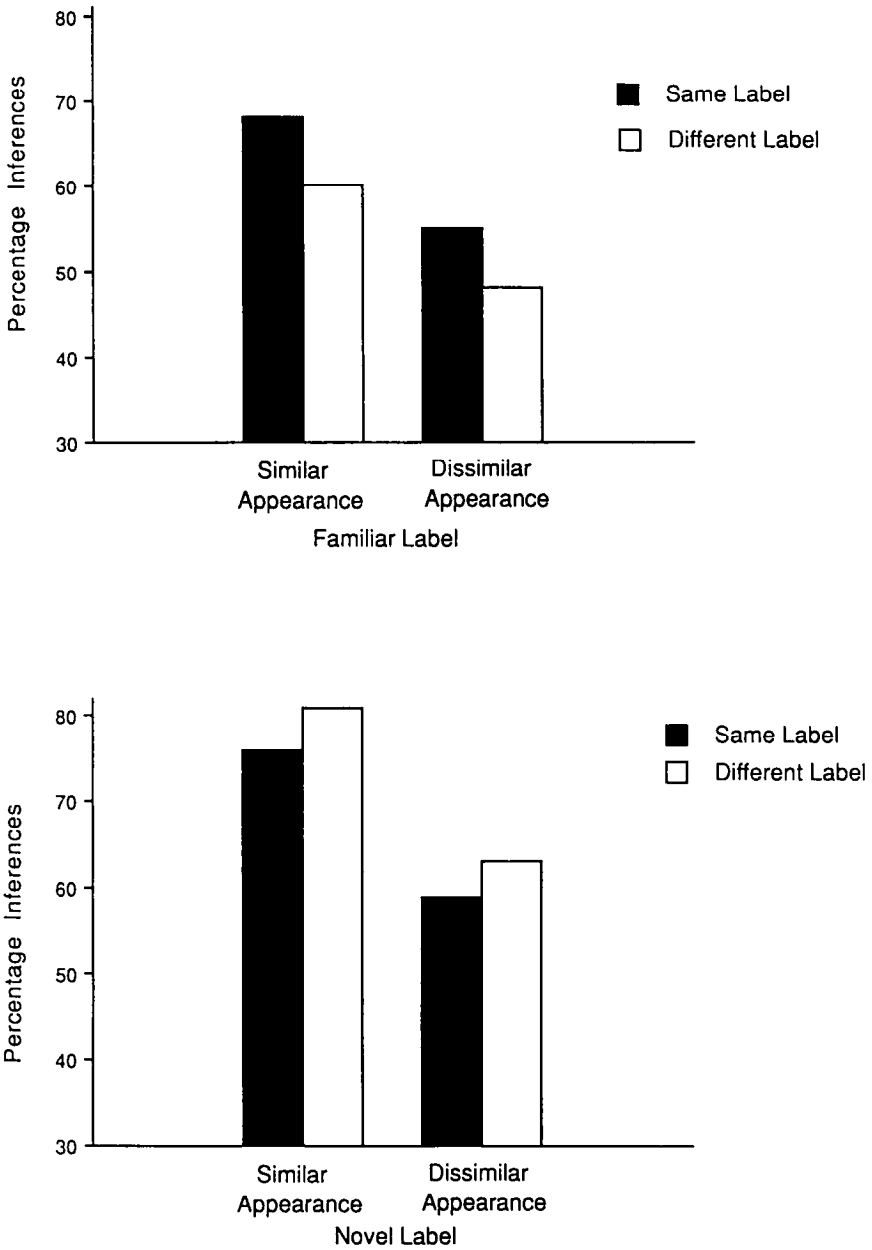
Figure 3.   **Mean percentage of inferences as a function of picture and condition: Experiment 2.**

First, as in Experiment 1, there was a main effect for appearance, $F(1,31) =$ 32.94, $p < .0001$. Children drew significantly more inferences to test pictures that were perceptually similar to the target picture ($M = 71\%$) than to those that were perceptually dissimilar from the target ($M = 56\%$).

Second, there was a Label × Condition interaction, $F(1,31) = 4.48, p < .05$. In the Familiar Label condition, children showed a nonsignificant tendency to draw more inferences to objects labeled the same as the target ($M = 62\%$) than to those labeled differently ($M = 54\%$). In contrast, in the Novel Label condition, children showed a nonsignificant tendency to draw fewer inferences to objects labeled the same as the target ($M = 68\%$) than to those labeled differently ($M = 72\%$). Using the Novel Label condition as a baseline, this result suggests that children were more likely to use a familiar-word category label for their inductions than an unfamiliar one. As in Experiment 1, children did not assume that unfamiliar labels were informative with respect to induction of novel properties.

Third, children made more inductions to artifacts ($M = 69\%$) than to natural kinds ($M = 59\%$), as indicated by a main effect for category type, $F(1,31) =$ 14.66, $p < .001$. Since we did not control for the types of properties used across the two category types, it is possible that the properties of the artifacts and the natural kinds were not equally generalizable at the outset of the experiment. This possibility is supported by the fact that this difference between artifacts and natural kinds was also found in the pretest of the properties. Children responded "yes" more often when asked whether a particular artifact had the property of interest ($M = 39\%$) than when asked the same of a natural kind ($M = 24\%$). Accordingly, this difference probably does not indicate anything very general about children's beliefs about differences in the inductive power of natural kind versus artifact categories.

The final significant result was a Label × Category Type interaction, $F(1,31)$ $= 5.86, p < .05$. For the natural kind categories, children tended to draw more inferences to objects labeled the same as the target ($M = 63\%$) than to those labeled differently ($M = 55\%$). In contrast, for the artifact categories, children tended to draw fewer inferences to objects labeled the same as the target ($M = 66\%$) than to those labeled differently ($M = 71\%$). However, neither of these pairwise comparisons was significant.

Discussion

In the Novel Label condition of Experiment 2, as in Experiment 1, we found that children did not base their inductive inferences from unfamiliar objects on the novel labels (e.g., *zav, traw*) that the objects received. However, when the very same unfamiliar objects were given familiar names (e.g., *cow, deer*; Familiar Label condition of Experiment 2), children showed more of a tendency to use the labels as the basis of induction. Although there was no significant difference between the number of inferences to pictures with the same versus different labels from the target in either condition, the Label × Condition interaction

suggests that knowledge of a category label increases its power in guiding induc-tive inferences. However, simply hearing a label for an object is not always sufficient to signal to a child whether or not the category is richly structured and promoting of inductions. Children may in fact require a certain amount of experi-ence with that label; they may need to observe that objects with that name have many other similarities in common, before the label reaches the status of guiding inductions. In more concrete terms, just knowing that two animals are both called *zavs* does not imply that they have any further commonalities above and beyond what one would assume based on their appearances. But knowing that zavs eat the same food, live in the same habitat, are pursued by the same enemies, react similarly to fear, etc., could lead a child to expect that any two zavs will share many further properties as well.

EXPERIMENT 3

The results of Experiments 1 and 2 demonstrate that children do not hold a completely general assumption about language, in which objects with the same name necessarily share deep similarities whereas objects with different names have different underlying natures. However, this finding does not necessarily imply that children's understanding of the power of language for induction is built up for every category individually. That is, children may not always require direct or extended experience with a word in order to use it as the basis of induction.

We suggest that children may hold an expectation about language that is more directly related to the structure of the category being named. This expectation has two parts: (a) When children are unable to see the conceptual basis of a novel label—that is, when they fail to see what objects with the same new name have in common—they will not use the label to guide their inductions. This is what we found in the Novel Label condition of Experiments 1 and 2. In those experi-ments, the novel labels were too discrepant with how children divided up the objects based on appearances, and the label was not used as the basis for induc-tions. (b) In contrast, when children are able to detect the conceptual basis of a novel label—when they see that objects with the same new name share important similarities—they will use the label to guide their inductions. That is, if the novel label is fairly consistent with children's nonlinguistic partitioning of the objects under consideration (e.g., based on their appearances), then children may believe that the label maps onto a richly structured category and may tend to draw inferences based on the label.

In the present experiment, we tested the preceding notion indirectly by provid-ing children with the same pictures as in Experiments 1 and 2, but showing each child only three of the pictures per set. In particular, one of the "conflict" pictures in each set was omitted (either the one of similar appearance but a different label from the target, or the one of dissimilar appearance but the same

label). In this way we hoped to reduce the contradiction between the labels and
the child's nonlinguistic (perceptual) partitioning of the pictures. Note that label-
ing still provides information beyond that of appearances (i.e., there is still a
conflict between appearances and labeling). However, children should find it
easier to grasp the conceptual basis of a set with one anomalous instance, than a
set with two anomalous instances that run counter to one another.

Method

Subjects.   Seventy-four children (39 girls and 35 boys) participated in the
experiment. The mean age of the children was 4;10, ranging in age from 4;0 to
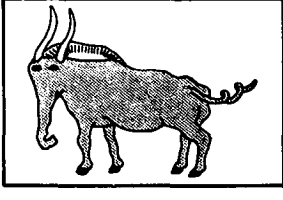5;9. None of these children had participated in Experiments 1 or 2.

Materials.   The same pictures were used in this experiment as in the two
previous ones. However, in this experiment children saw only three test pictures
from each set rather than four, as in Experiments 1 and 2. Two separate configu-
rations of three test pictures per set were each presented to two groups of
children, so that children's patterns of inferences to all four test pictures from
Experiments 1 and 2 could be determined.

The design of the two configurations was identical, except for the nature of
one of the pictures in each set. Both configurations included the following two
kinds of pictures: (1) Similar Appearance-Same Label (SASL) pictures. (These
pictures were similar in appearance to their respective targets, and also received
the same label as the target.) (2) Dissimilar Appearance-Different Label (DADL)
pictures. (These pictures were markedly different in appearance from the target,
and received a different label from that of the target.) Finally, the third picture in
each set posed a conflict between appearances and labeling. In one configura-
tion, the third picture was similar in appearance to the target but received a
different label from that of the target. [This configuration will be referred to as
the Similar Appearance-Different Label (SADL) set; an example of this configu-
ration is shown in Figure 4.] In the other configuration, the third picture was very
different in appearance from the target but received the same label. [This config-
uration will be referred to as the Dissimilar Appearance-Same Label (DASL) set.
An example of this configuration is shown in Figure 5.]

The purpose of this design was to determine the number of inferences that
children would draw to the two conflict pictures individually relative to the
number they draw to the two pictures that pose no conflict. In each of the
configurations—the Similar Appearance-Different Label configuration and the
Dissimilar Appearance-Same Label configuration—comparison of the conflict
picture with one of the no-conflict pictures will show whether labels affect the
number of inferences children draw, while comparison with the other no-conflict
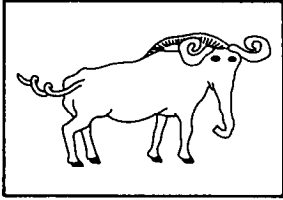picture will show whether there is an appearance effect.
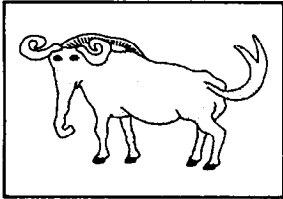
There were no individual properties taught about the targets in this experi-
ment. Instead, the same general question was asked about the test objects in all
the categories, namely, "Do you think this (test object) and this (target object)

	Conditions	
	Novel Label	Familiar Label
Target Picture		
	a zav	a cow
Test Pictures		
	a zav	a cow
	a traw	a deer
	a pume	a giraffe

Figure 4. The structure of each picture set for the Similar Appearance-Different Label configuration: Experiment 3.

have the same kinds of stuff inside?" This change from the properties used in Experiment 2 was included in an attempt to reduce the role of individual variability in children's knowledge about particular properties, and to shorten the task. Previous research (Gelman & O'Reilly, 1988) has found that preschool children tend to attribute identity of internal parts primarily to members of the

	Conditions	
	Novel Label	Familiar Label

Target Picture

| | a zav | a cow |

Test Pictures

| | a zav | a cow |

| | a zav | a cow |

| | a pume | a giraffe |

Figure 5. The structure of each picture set for the Dissimilar Appearance-Same Label configuration: Experiment 3.

same basic-level category (e.g., all dogs are thought to have "the same kind of stuff inside," whereas dogs and horses are less often thought to do so).

   Procedure.   All children were seen individually. Children were introduced to a stuffed bear and told that they would see pictures of objects from the bear's

homeland far away. They were asked to help the experimenter answer some questions about them.

There were three conditions: the Novel Label condition ($n = 25$), the Familiar Label condition ($n = 25$), and the Novel Label with Information condition ($n = 24$). Each condition was completed with an approximately equal number of children who saw the Similar Appearance-Different Label picture configuration and the Dissimilar Appearance-Same Label configuration.

Novel Label Condition.   The task for this condition was similar to that in the Novel Label condition of Experiment 2. For each picture set, the experimenter showed the child the target picture and taught him or her a label for the object. The same nonsense words were used as in Experiments 1 and 2. The child was then asked, for each of the three test objects in that set, whether the test object and the target have the same kind of stuff inside [e.g., "This (test object) is a traw. Do you think this traw and this zav (target) have the same kind of stuff inside?"]. This procedure was repeated for each of the eight picture sets. The order of presentation of the sets and the order of test pictures within each set was randomized for each subject.

Familiar Label Condition.   As in Experiment 2, the only difference between the Novel Label and the Familiar Label condition was the use of real words that children would know in the latter condition. The same words were used as in Experiment 2. Just as in the Novel Label condition, each child was asked whether the test object and the target have the same kind of stuff inside [e.g., "This (test object) is a deer. Do you think this deer and this cow (target) have the same kind of stuff inside?"].

Novel Label with Information Condition.   The only difference between the Novel Label condition and the Novel Label with Information condition was that children were taught a property of the target category in the latter condition [e.g., "Here's a zav (target animal). Zavs eat bugs," or "Here's a durn (target artifact). Durns are used to carry people underwater."]. The purpose of the extra sentence, in this case "Zavs eat bugs" or "Durns are used to carry people underwater," was to provide information about the category as a whole, beyond just the label. We hypothesized that the extra information might help children realize the structure of the category and therefore help them to draw inferences on the basis of the novel label. The new information always concerned some important characteristic of the category as a whole (e.g., habitat, diet, form of locomotion). The presentation of the test pictures and the subsequent questions were the same as in the Novel Label condition [e.g., "This (test object) is a traw. Do you think this traw and this zav (target) have the same kind of stuff inside?"].

Results
Each child scored a "1" every time he or she drew an inference from the target to the test picture, that is, every time he or she answered "yes." Analyses were performed separately for the Similar Appearance-Different Label set and the Dissimilar Appearance-Same Label set.
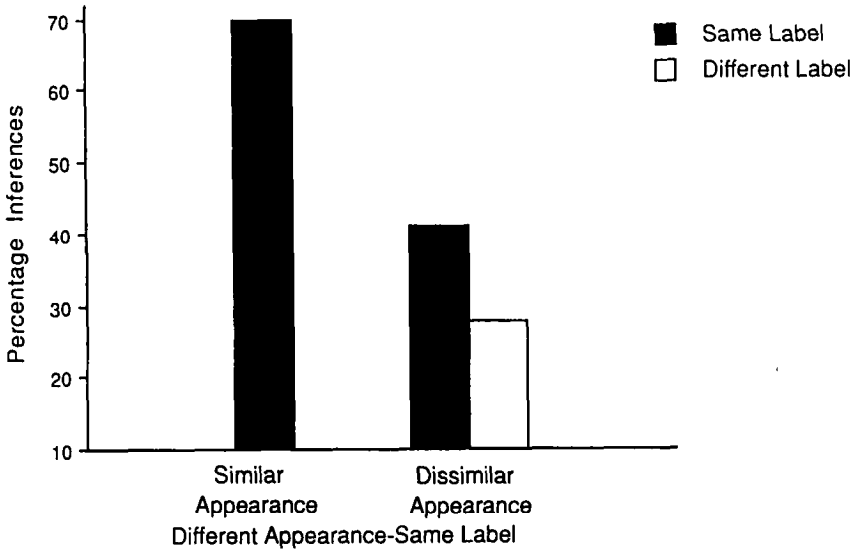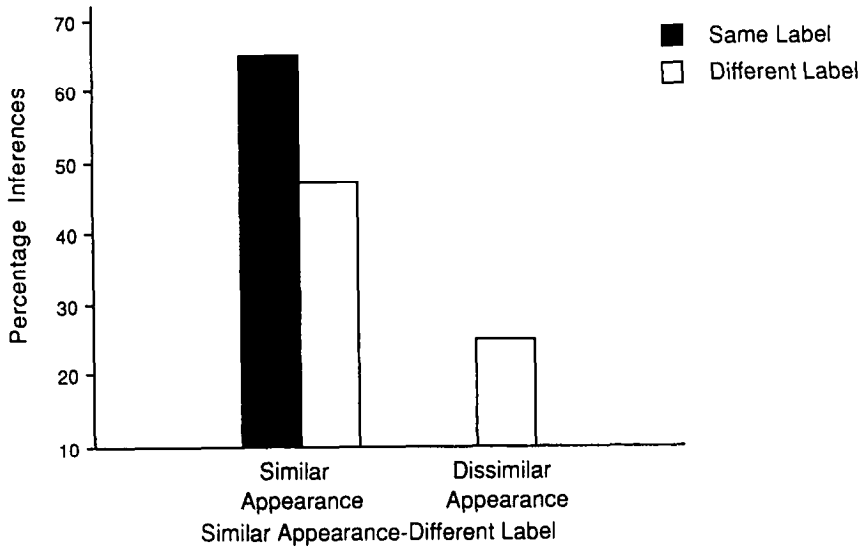
Figure 6.　Mean percentage of inferences as a function of picture and configuration: Experiment 3.

Similar Appearance-Different Label Set. A three-way ANOVA was performed on the total scores of each test picture, calculated across all eight sets. Condition (Novel Label vs. Familiar Label vs. Novel Label with Information) was the between-subjects variable, with repeated measures on category type (natural kind vs. artifact) and picture type [Similar Appearance-Same Label (SASL) vs. Similar Appearance-Different Label (SADL) vs. Dissimilar Appearance-Different Label (DADL)]. Results are shown in Figure 6.

There were no main effects or interactions due to labeling condition. Children's inferences were roughly equal across the three conditions (41%, 44%, and 52% overall in the Novel Label, Familiar Label, and Novel Label with Information conditions). Furthermore, category type had no effects on the number of inferences children drew (48% overall to natural kinds, 44% overall to artifacts).

The only significant result was a main effect for picture type, $F(2,68) = 31.71$, $p < .0001$. In particular, there were significant effects of both labels and appearances on the number of inferences that children drew. When appearance was held constant, children drew significantly more inferences to test objects with the same label as the target (Picture SASL: $M = 65\%$) than to those with a different label (Picture SADL: $M = 47\%$), $p < .02$. Furthermore, when labeling was held constant, children drew significantly more inferences to test objects that were perceptually similar to the target (Picture SADL: $M = 47\%$) than to those that were perceptually dissimilar from the target (Picture DADL: $M = 25\%$), $p < .01$.

Dissimilar Appearance-Same Label Set. A separate three-way ANOVA was performed on the total scores of the children who were shown this set of pictures. Condition (Novel Label vs. Familiar Label vs. Novel Label with Information) was the between-subjects variable, with repeated measures on category type (natural kind vs. artifact) and picture type [Similar Appearance-Same Label (SASL) vs. Dissimilar Appearance-Same Label (DASL) vs. Dissimilar Appearance-Different Label (DADL)]. Results are shown in Figure 6.

The results were similar to those of the Similar Appearance-Different Label set. There were no significant main effects or interactions due to condition. The number of inferences drawn in the Novel Label, Familiar Label, and Novel Label with Information conditions were 43%, 50%, and 47%, respectively. As before, category type did not affect the number of inferences children drew (46% overall to natural kinds, 46% overall to artifacts).

There was a significant main effect for picture type, $F(2,68) = 48.92$, $p < .0001$. Again, there were significant effects of appearance and of labeling. When labels were held constant, children drew significantly more inferences to test objects that were perceptually similar to the target (Picture SASL: $M = 70\%$) than to those that were perceptually dissimilar from the target (Picture DASL: $M = 41\%$), $p < .01$. Furthermore, when appearances were held constant, children drew significantly more inferences to test objects with the same label as the target (Picture DASL: $M = 41\%$) than to those with a different label (Picture DADL: $M = 28\%$), $p < .05$.

Discussion

In Experiment 3, children drew inferences based on novel labels as well as familiar labels. This result contrasts with the results of Experiment 2, in which children showed a tendency to draw inferences based only on familiar labels.

What could account for the difference in results between Experiments 2 and 3? We suggest that the primary difference was in the structure of the picture sets that were presented. Recall that in Experiment 2, names were totally orthogonal to appearances for the two conflict pictures. That is, the name provided no clue to an object's appearance. For this reason, we propose that in Experiment 2 children were unable to see the conceptual basis of the novel labels. Without any further information about the categories being named, children may have been unable to understand how one object in a set could be similar in appearance to the target but have a different label, while another could be dissimilar in appearance but have the same label. However, in Experiment 3, there was only one such anomalous picture per set in which the label and appearance were discrepant. Therefore, children were able to understand how the labels mapped onto the pictures, because the structure conveyed by the labels was not too discrepant with children's own nonlinguistic, perceptual partitioning of the objects.

It is also possible that the different results across experiments could be due partly to the different properties taught across the different experiments. In Experiments 1 and 2, children were taught a specific property of one member of a category, and their inductions of that property to other objects were then tested [e.g., "Do you think this traw (test picture) has four stomachs inside like this zav (target)?"]. In Experiment 3, however, children's inductions of a general property about the target were tested (e.g., "Do you think this traw and this zav have the same kind of stuff inside?"). Children could be more willing to infer that two members of a category have the same internal structure than that they share the more specific properties of Experiments 1 and 2.

However, it is unlikely that the preceding difference in procedure could account entirely for the different results. Gelman and O'Reilly (1988) have also tested children's inductions of specific properties in one experiment, and their beliefs about internal structure in the other ("Do these have the same kinds of stuff inside?"), and have generally found the same patterns of results in both experiments, with a high level of inferences drawn to members of familiar categories. Furthermore, the specific properties used in Experiment 2 are, in principle, capable of promoting inductions, since they did evoke category-based inferences in the Familiar Label condition. Similarly, many of the specific properties used in Experiments 1 and 2 are strictly analogous to properties that were used in earlier studies that also evoked category-based inductions from familiar categories. For example, "eats raisins" is analogous to "eats alfalfa" (Gelman, 1988), "eats grass" (Gelman & Markman, 1987), and "eats plants" (Gelman & Markman, 1986). Thus, there appears to be nothing about the properties per se that would limit children's inferences in Experiments 1 and 2. Rather, the pri-

mary difference among the experiments appears to be the structure of the picture sets that children saw.

GENERAL DISCUSSION

Young children consider category membership to be at least as important as perceptual cues in guiding their inductive inferences within familiar categories (Gelman & Markman, 1986, 1987). The purpose of the current studies was to consider two possible hypotheses in order to explain the basis of children's unexpectedly early appreciation of the role of categories. The first hypothesis we examined is that children may need specific experience with a category label before they expect members of that category to share deep, nonobvious properties. Their knowledge of the many, more obvious properties that category members share (e.g., that birds have beaks and feathers, and typically fly) might lead children to assume that category members also share other less obvious properties (e.g., that birds have the same kind of heart, bone structure, and diet). In contrast, the second hypothesis is that children may not require specific experience with a category before they assume that it promotes inductions. Instead, they may hold a general assumption about language, such that all objects with the same name are assumed to share important, nonobvious properties.

We tested these two hypotheses in a series of three experiments. The major difference between the current experiments and previous research (Gelman & Markman, 1986, 1987) was in the familiarity of the objects used. Gelman and Markman (1986, 1987) used familiar objects with familiar labels (e.g., a rabbit, snake, pearl, etc.), while the current studies used novel objects (e.g., a gnu-like animal with an elephant's trunk) with novel labels (e.g., a *zav*). The use of novel objects allowed us to separate the effect of labeling from the effect of children's familiarity with the objects being labeled. This design provided a better test of whether the category-based inferences children demonstrated in Gelman and Markman's (1986, 1987) studies were due to children's early appreciation of a general principle of language, or to their knowledge of other similarities that members of familiar categories share.

There are two primary findings from the current studies. First, children do at times draw inferences on the basis of novel labels for unfamiliar objects, even when such labels conflict with perceptual appearances. Specifically, in Experiment 3, children who saw pictures of novel animals and objects drew significantly more inferences when labels matched than when they differed, regardless of whether the labels were familiar or novel. In this case, the labels provided information to children above and beyond the perceptual information from the pictures alone. Thus, children have at least a somewhat general assumption about language, namely that objects with the same common noun label share deeper underlying similarities.

On the other hand, the second finding shows that children do not always draw inferences on the basis of a novel label (Experiments 1 and 2). This is actually an important finding in its own right, because it helps rule out the possibility that children in earlier studies (e.g., Gelman & Markman, 1986, 1987) were simply answering on the basis of category labels per se. Indeed, Experiments 1 and 2 clearly show that children do not blindly draw inferences from one object to another just because they have the same name.

When were labels important? In the present studies, labels gained importance in either of two ways: (a) Past experience (in the form of a familiar basic-level label) could convey that the category in question was one that promotes inductions. This was suggested in Experiment 2, in which the very same pictures were labeled with either familiar (e.g., a *cow*) or novel (e.g., a *zav*) labels. In that experiment, children drew inferences partly on the basis of the labels when the pictures received familiar labels, but on the basis of only appearances when they received novel labels. The familiar labels, therefore, seemed to provide information about the structure of the category in a way that the novel labels did not. (b) The structure of the category itself was also important. When labels did not completely contradict appearances, as in Experiment 3, children were willing to draw inferences based on the novel label.

Altogether, then, children do not have a completely general assumption about the power that language has for conveying the deeper identity of novel objects. Instead, it appears that language guides children's inductions in only a limited set of cases. In particular, children seem to base their inductions on the category only when they are able to understand how (or why) the category labels map onto the pictures in the way that they do. Experiments 1 and 3 used exactly the same pictures, but it was only in Experiment 3 that children drew inferences based on the category labels. Apparently, children were evaluating the entire *set* of objects that were named when deciding when to draw inferences. It may be that children were computing a value of "conceptual coherence" (Medin & Wattenmaker, 1987) among objects that received the same name, which was then used to determine whether a category promotes inductions. However, at this point the notion of coherence is entirely unspecified, and more research is needed to determine what goes into creating conceptual coherence. Developmental research may be particularly helpful in illuminating the notion, because children may have special difficulty learning incoherent categories. These difficulties could then help specify what makes a category coherent or not. Regardless of whether a satisfactory account of coherence can be constructed, our main point here is that the effectiveness of language interacts with other factors such as the structure of the category being named.

Our findings and the preceding interpretation fit well with a range of related data. To begin with, Billman (1988) and Shipley (1988) have also recently found that labeling can influence the inductions that people make. Billman gave adults an induction task with unfamiliar categories and found that subjects assumed that

novel labels provided important information. Importantly, in her study, appearances did not conflict with category labels; rather, appearances were neutral with respect to category membership. Similarly, Shipley gave preschool children an induction task with categories which were somewhat unfamiliar, and also found labeling effects. Significantly, the structure of her items was like the Similar Appearance-Different Label sets in our Experiment 3. That is, there was only moderate discrepancy between labels and appearances.

The suggestion that conceptual structure and not language alone is important to induction also helps make sense of recent findings by Gelman and O'Reilly (1988). In their study of familiar categories, language had no effect on children's inductive inferences. Preschoolers and second graders drew just as many basic-level inferences without basic-level labels [e.g., "Do you think this (a soft, stuffed armchair) has lignin all through it like this (a straight-back wooden chair)?"] as they did when the basic-level labels were explicitly mentioned [e.g., "Do you think this chair (armchair) has lignin all through it like this chair (straight-back chair)?"]. Moreover, at the superordinate level, preschoolers were not able to make the appropriate within-category inductions even with the help of the labels [e.g., "Do you think this piece of furniture (a bed) has lignin all through it like this piece of furniture (a straight-back chair)?"]. Older children drew more superordinate-level inductions regardless of how the pictures were named. Thus, these results suggest that children were able to invoke the appropriate basic-level categories for familiar objects on their own. However, for categories that were conceptually more difficult, such as superordinates, children did not benefit from hearing the categories explicitly mentioned. Apparently, some conceptual understanding of the category was necessary, above and beyond linguistic information, for children to draw category-based inductions.

Similarly, Mervis (1984) has suggested that children will not reorganize a category based on language alone if the label does not have a conceptual basis that the child can understand. She has found that, even if a parent repeatedly corrects a child's naming error (e.g., pointing out to a child that a goose is called a goose, not a duck), the child continues to make the naming error until he grasps the conceptual basis of the name. Her findings on word learning with very young children (approximately 14 months of age) are fully consistent with our findings on induction with children who are a few years older.

In sum, our results and those of other investigators suggest that children have neither an entirely general assumption about language, nor do they build up their expectations about categories one at a time. Rather, they have a moderately general assumption that gets overridden when the structure of a category is inconsistent with how it is named. In future work it will be important to determine what the experiences and underlying mechanisms are that transform a word from being unfamiliar and able to promote inductions in only limited cases (e.g., a *zav*), to being familiar and able to promote inductions across a broader range of circumstances (e.g., a *cow*).

REFERENCES

Billman, D.O. (1988, November). *Effects of novel category membership and labeling on induction of new properties.* Paper presented at the meeting of the Psychonomic Society, Chicago, IL.

Gelman, S.A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20,* 65–95.

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., & O'Reilly, A.W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development, 59,* 876–887.

Gelman, S.A., & Taylor, M. (1984). How 2-year-old children interpret proper and common names for unfamiliar objects. *Child Development, 55,* 1535–1540.

Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development, 45,* 469–473.

Keil, F. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88,* 197–227.

Macnamara, J. (1982). *Names for things: A study of human learning.* Cambridge, MA: MIT Press.

Markman, E.M., & Hutchinson, J.E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology, 16,* 1–27.

Medin, D.L., & Wattenmaker, W.D. (1987). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 25–62). Cambridge, England: Cambridge University Press.

Mervis, C.B. (1984). Early lexical development: The contributions of mother and child. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 339–370). Hillsdale, NJ: Erlbaum.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3,* 221–246.

Shipley, E.F. (1988, June). *Categories, hierarchies, and induction.* Paper presented at the annual meeting of the Jean Piaget Society, Philadelphia, PA.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1,* 139–156.

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.