

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/11589981>

# How Much Does a Shared Name Make Things Similar? Linguistic Labels, Similarity, and the Development of Inductive Inference

Article *in* Child Development · November 2001

Impact Factor: 4.72 · DOI: 10.1111/1467-8624.00373 · Source: PubMed

---

CITATIONS

109

---

READS

99

3 authors, including:



[Vladimir Sloutsky](#)

The Ohio State University

142 PUBLICATIONS 2,362 CITATIONS

[SEE PROFILE](#)



[Anna V Fisher](#)

Carnegie Mellon University

56 PUBLICATIONS 854 CITATIONS

[SEE PROFILE](#)

## How Much Does a Shared Name Make Things Similar? Linguistic Labels, Similarity, and the Development of Inductive Inference

*Vladimir M. Sloutsky, Ya-Fen Lo, and Anna V. Fisher*

This article examines the development of inductive generalization, and presents a model of young children's induction and two experiments testing the model. The model specifies contribution of linguistic labels and perceptual similarity to young children's induction and predicts a correspondence between similarity judgment and induction of young children. In Experiment 1, 4- to 5-year-olds, 7- to 8-year-olds, and 11- to 12-year-olds were presented with triads of schematic faces (a Target and two Test stimuli), which varied in perceptual similarity, with one of the Test stimuli sharing a linguistic label with the Target, and another having a different label. Participants were taught an unobservable biological property about the Target and asked to generalize the property to one of the Test stimuli. Although 4- to 5-year-olds' proportions of label-based inductive generalizations varied with the degree of perceptual similarity among the compared stimuli, 11- to 12-year-olds relied exclusively on labels, and 7- to 8-year-olds appeared to be a transitional group. In Experiment 2 these findings were replicated using naturalistic stimuli (i.e., photographs of animals), with perceptual similarity manipulated by "morphing" naturalistic pictures into each other in a fixed number of steps. Overall results support predictions of the model and point to a developmental shift from treating linguistic labels as an attribute contributing to similarity to treating them as markers of a common category—a shift that appears to occur between 8 and 11 years of age.

### INTRODUCTION

Inductive generalization is an important component of human thinking. Furthermore, some believe that induction is the most important component because "inductive inference is the only process . . . by which new knowledge comes into the world" (Fisher, 1935/1951, p. 7). Induction requires at least two stimuli or stimuli sets: the base and the target. Examples of inductive generalizations are (1)  $X_1$  has property  $Y$ , therefore  $X_2$  has property  $Y$ , (2)  $X$ s have property  $Y$ ,  $Z$ s are like  $X$ s, therefore  $Z$ s have property  $Y$ ; or (3)  $X_1$  has property  $Y$ , therefore all  $X$  have property  $Y$ . The first type of induction has been defined as specific induction, whereas the latter two have been defined as general induction (for a discussion, see Osherson, Smith, Wilkie, Lopez, & Shafir, 1990).

There is much evidence that infants and young children are capable of specific inductive generalizations (Gelman & Markman, 1986, 1987; Mandler & McDonough, 1996, 1998); however, mechanisms underlying such generalizations remain unclear. Several proposals attempting to specify these mechanisms have been discussed (for discussions, see Eimas, 1994; Keil, 1989; Mandler & McDonough, 1998). These include a similarity-based approach (i.e., induction from  $X_1$  to  $X_2$  is a function of similarity between  $X_1$  and  $X_2$ ) and a category-based approach (i.e., induction from  $X_1$  to  $X_2$  is a function of  $X_1$  and  $X_2$  belonging to the same category  $X$ ).

According to the similarity-based approach, induction starts out as a special case of the "universal law of

generalization" (Shepard, 1987). The law states that the probability of generalizing a response (e.g., fear) from one stimulus to another stimulus varies with featural overlap between the stimuli. An alternative proposal suggests that for certain categories, such as natural kinds, induction starts out as a category-based process (for reviews and discussions, see Gelman & Coley, 1991; Gelman, Coley, & Gottfried, 1994). In this case, at least for natural kinds, inductive generalization is driven by the category membership rather than by featural overlap. For example, a person is more likely to generalize a property (e.g., the ability to drink) from a bird to another dissimilar-looking bird than from a bird to a similar-looking airplane (Gelman & Coley, 1991; Gelman & Markman, 1986; Mandler & McDonough, 1998).

Both positions have well-recognized weaknesses. Weaknesses of the similarity position were originally pointed out by the philosopher Nelson Goodman (1972/1992). Consider a simple inductive task. A child has two pets—a white bird and a black dog. The bird likes oats, whereas the dog likes meat. Then the child gets another pet—a white dog. Should the child expect the new dog to be a meat lover, like another dog, or to be an oat lover, like another white animal? And what about the age of the new pet: should it be more similar to the other dog, or to the white bird? When

asked both of these questions, adults tend to balk at the second question and to make specific induction based on natural kind membership for the first, that is, *dogs* rather than *white things* (Gelman, 1988; Nisbett, Krantz, Jepson, & Kunda, 1983). The argument against similarity thus states that induction requires more than a mere comparison of features; one needs to know which dimensions and features are important, which requires prior knowledge (for related discussions, see also Medin, Goldstone, & Gentner, 1993; Murphy & Medin, 1985).

There is a cost attached to the category-based position, however. In particular, this position has to assume the existence of complex knowledge that is necessary for performing induction in a category-based manner. First, the children have to have a mental representation of the category in question (or a category placeholder, if an entity is novel). This is necessary because the generalization decision is based on the membership of both the base and the target in the category. Second, children have to have an identification procedure, to identify the base and the target as members of the category. This is necessary because even if they have a mental representation of the category and knowledge of its properties, they also have to assign an instance to a particular category. In addition, children have to have knowledge of the distinction between natural or theoretical kinds that support induction (e.g., *mammals*) and arbitrary groupings or intuitive kinds that do not support induction (e.g., *red things*). Finally, children have to have knowledge of properties that are important for a particular natural kind (e.g., “lays eggs” is a category-important property for birds, whereas “are fun to watch” is not). This knowledge is important because many properties could be induced only within a particular natural kind, but not across (e.g., “lays eggs” could be generalized within the category *bird*, but not to the category *mammal*, *tree*, or *atom*). These assumed abilities are quite complex and, although used to explain induction, require explanation themselves.

Of course, arguments such as those listed above cannot eliminate either theoretical position. For example, the similarity-based position assumes the existence of attentional and perceptual constraints on the process of similarity computation. In particular, some researchers have argued that often people do not have to extract and individuate features from objects; rather they detect features and integrate those features they attend to into a perceived whole (Thompson & Massaro, 1989; Treisman & Gelade, 1980). In this case, computation of similarity does not require knowledge of “important” features and dimensions, because computation of similarity will be performed only over those features that are attended to. Furthermore, high atten-

tional weight of a feature does not have to be accompanied by the knowledge of the importance of this feature. By the same token, the category-based position may argue that identification of the base and the target as members of the same category could be assisted by linguistic labels, and in the absence of such labels induction may be performed along the lines of perceptual similarity. Furthermore, the knowledge of category-important properties does not have to be full blown—children and adults may occasionally generalize properties erroneously. To perform induction in a category-based manner, they just need some crude understanding that (1) individual entities are members of stable categories; (2) members of stable categories share many properties; and (3) some properties can be generalized from one member of a category to other members (e.g., lays eggs), whereas some properties cannot be generalized to other members (e.g., is dirty). In fact, 4- to 5-year-old children have been shown to understand that, at least for some stable categories (e.g., natural kinds), properties such as “fell on the floor” are not generalizable, whereas properties such as “has hard bones inside the body” are generalizable (Gelman, 1988, Experiment 1; Gutheil, Vera, & Keil, 1998).

Therefore, although both positions are logically possible, the question of which of these possibilities actually takes place remains an open empirical question. In this article, we present arguments that the similarity-based position is capable of explaining inductive generalization in young children. If the argument is correct, and inductive generalization in young children is similarity based, then at least some of the existing models of similarity should be able to predict performance of young children on induction tasks. One such model has been proposed recently to account for young children’s similarity judgment (Sloutsky & Lo, 1999). In what follows, the model and its predictions are specified, and the experiments designed to test predictions of the model are presented.

The label-as-attribute model proposes a low-level mechanism underlying similarity judgment. The model is based on the product-rule model of similarity (Estes, 1994; Medin, 1975) that specifies similarity among non-labeled feature patterns. In the product-rule model, similarity is computed using equation 1:

$$\text{Sim}(i, j) = S^{N-k}, \quad (1)$$

where  $N$  denotes the total number of relevant attributes,  $k$  denotes the number of matches, and  $S$  ( $0 \leq S \leq 1$ ) denotes values (weights) of a mismatch. For example, suppose that a child is presented with two schematic faces, A and B (see Figure 1). Further, suppose that these faces consist of four distinct features

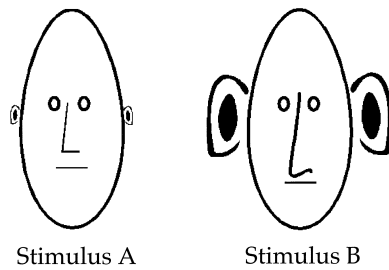


Figure 1 A pair of stimulus patterns with distinct facial features.

(i.e., the shape of the face, eyes, nose, and the size of ears), that they share two of these features (i.e., the shape of the face and eyes), and differ on the other two. Finally, suppose that  $S = .5$ , the value frequently derived empirically in past research (Estes, 1994). In this case, similarity between A and B would be equal to .25 (i.e.,  $.5^2$ ). Note that similarity between entities decreases rapidly with a decrease in the number of mismatches, approximating the exponential decay function discussed elsewhere (Nosofsky, 1984). For example, if the faces shared only one of the four features, their similarity would be equal to .125 (i.e.,  $.5^3$ ). On the other hand, if the faces shared all four features, they would be identical, and their similarity would be equal to 1 (i.e.,  $.5^0$ ).

The label-as-attribute model suggests that (1) linguistic labels are considered by young children as attributes of entities, and (2) linguistic labels have greater contribution to similarity judgment for younger children than do other attributes. Both suggestions have been supported empirically (see Sloutsky & Lo, 1999). Why would labels weigh more for younger children and what might be a mechanism underlying the greater weight of labels at an earlier age? One possible explanation is that labels have larger weights because they represent the auditory modality, and the auditory system matures earlier than the visual system. In particular, the auditory system starts functioning during the last trimester of gestation (Birnholtz & Benaceraff, 1983; for a review, see also Jusczyk, 1998), whereas the visual system does not start functioning until after birth. As a result, even though the neural bases of visual perception are fully developed at quite a young age (e.g., Aslin & Smith, 1988), auditory stimuli may still have a privileged processing status for younger children, thus resulting in larger weights of auditory stimuli. In this case, auditorily presented labels are not only attended to, but they may also attract attention to visual attributes that correlate with the labels. In fact, it has been demonstrated that 15-month-olds grouped objects together when the objects shared a label or a

nonlinguistic auditory input (instrumental music input) when either input perfectly correlated with an infant's fixation of an object (Roberts, 1995; Roberts & Jacob, 1991; but see Balaban & Waxman, 1997). Additionally, it was found that in a learning task, auditory stimuli overshadowed visual stimuli for 4-year-olds, whereas visual stimuli overshadowed auditory stimuli for adults (Napolitano, Sloutsky, & Boysen, 2001).

According to the label-as-attribute model, similarity of labeled feature patterns could be calculated using equation 2:

$$\text{Sim}(i, j) = S_{\text{Label}}^{1-L} S_{\text{Vis.attr.}}^{N-k} \quad (2)$$

where  $N$  denotes the total number of visual attributes,  $k$  denotes the number of matches,  $S_{\text{Vis.attr.}}$  denotes values (attentional weights) of a mismatch on a visual attribute,  $S_{\text{Label}}$  denotes values of label mismatches, and  $L$  denotes a label match. When there is a label match,  $L = 1$ , and  $S_{\text{Label}} = 1$ ; when there is a label mismatch,  $L = 0$ , and  $S_{\text{Label}} < 1$ . Note that  $S$  ( $0 \leq S \leq 1$ ) denotes attentional weights of mismatches, and the contribution of  $S$  is large if  $S$  is close to 0 and small if  $S$  is close to 1. In other words, the closer the value of  $S$  to 1, the smaller the contribution of a mismatch to overall difference of compared entities; whereas the closer the value of  $S$  to 0, the greater its contribution to overall difference. When two entities are identical on all dimensions (i.e., there are no mismatches), their similarity should be equal to 1; otherwise, it is smaller than 1. Note that according to this model, when neither entity is labeled (i.e.,  $S_{\text{Label}} = 1$ ), similarity between entities is determined by the number of overlapping visual attributes, thus conforming to equation 1. Labels are presented as a separate term in the equation because they are expected to have larger attentional weights than most visual attributes. In the case in which the weight of a label does not differ from that of other attributes, the label simply becomes one of the attributes in the computation of similarity, and equation 2 turns into equation 1. Of course, there could be another case of nonlabeling, when two unfamiliar entities share a label, and one unfamiliar entity is not labeled at all. Should the model treat an absent label as a mismatch? On the one hand, in terms of distal stimuli, absence of a label should constitute a mismatch; on the other hand, in terms of mental representation, it may not. We, therefore, consider this an open question and defer answering it until further empirical investigation.

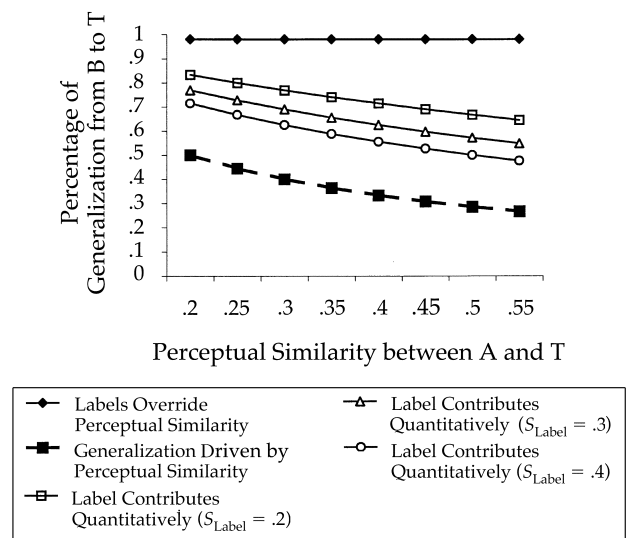
Finally, the model suggests that if the child is presented with a Target feature pattern (T) and Test feature patterns (A and B) and asked which of the Test patterns is more similar to the Target, the child's choices could be predicted using equation 3:

$$P(B) = \frac{\text{Sim}(T,B)}{\text{Sim}(T,B) + \text{Sim}(T,A)}. \quad (3)$$

In short, we argue that if induction in young children is indeed similarity based, then the same model that predicts similarity judgment in young children (e.g., Sloutsky & Lo, 1999) should be able to predict their inductive inference. The proposal is not unreasonable, as there is a large body of evidence indicating that (at least for artifacts) overall similarity explains young children's specific induction (Gentner, 1978; Landau, Smith, & Jones, 1998; Smith, Jones, & Landau, 1996).

One important (and testable) consequence of this proposal is that because linguistic labels contribute to similarity in a quantitative manner rather than in a qualitative "all-or-nothing" manner, they should also make a quantitative contribution to young children's induction. To make this difference between the qualitative and quantitative contributions clear, suppose that a young child is presented with a set of unique triads of stimuli. Each triad comprises Test stimuli A and B and Target T. Furthermore, let us, for the purposes of this illustration, hold perceptual similarity of B to T constant across trials, while varying similarity of A to T. On each trial, the child is informed that the Target has some nonobvious property (e.g., has pink bones) and is asked which of the Test stimuli was likely to have the property in question. If stimuli A and B are not labeled, proportions of inductive generalization from A to T should change as a function of perceptual similarity between A and T (for a related discussion, see Keil, 1989; for corroborating evidence, see Gelman & Markman, 1987; Springer, 1992, Experiment 1). This type of induction is represented in Figure 2 (see the dashed line). However, if B and T share a linguistic label, patterns of induction should differ. First, labels could contribute qualitatively to inductive inference, such that induction is performed strictly along the shared labels (see the solid line with black diamonds in Figure 2). Alternatively, labels could contribute quantitatively, in which case the slope and intercept should depend on the weight of the label (see lines with clear diamonds, circles, and triangles in Figure 2, with each line representing a different weight of the label).

Note that the label-as-attribute model predicts a quantitative contribution of linguistic labels to induction, and the critical prediction of the model is variability in label-based inductive generalization as a function of similarity among A, B, and T. In the case of qualitative effects of labels on induction, the proportion of label-based generalizations should be invariantly near ceiling, independent of perceptual similarity. Such qualitative effects should be observed when

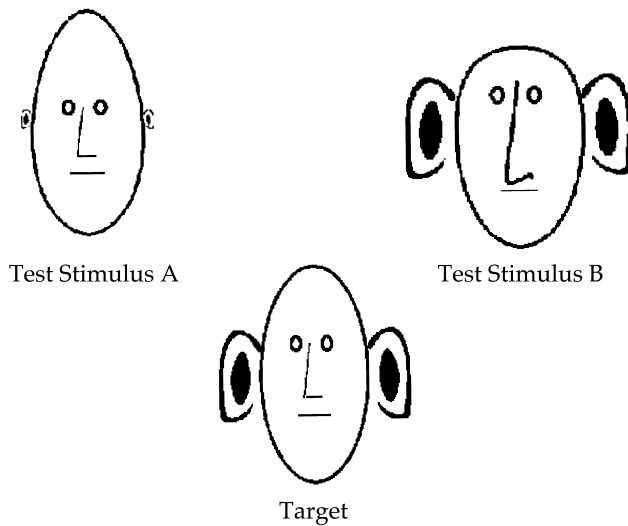


**Figure 2** Possible patterns of induction in young children: induction driven by labels alone, induction driven by perceptual similarity alone, and induction jointly driven by labels and perceptual similarity with different weights of labels. A = Test stimulus A; B = Test stimulus B; T = Target.

a person knows that the category label is the best predictor of certain properties of members of the category. For example, if a person familiar with the biological taxonomy is told that two animals are mammals, that person would be able to generalize many (although, of course, not all) biological properties from one animal to another, independent of these animals' appearance. Therefore, labels may have qualitative effects when induction is based on domain-specific knowledge.

In addition to the qualitative prediction (i.e., variability in label-based generalization as a function of similarity among A, B, and T), the label-as-attribute model is capable of predicting ordinal differences across different levels of perceptual similarity. Suppose that patterns A, B, and T have distinct features (see Figure 3). Each of the Test stimuli might share zero visual attributes, one visual attribute, or two visual attributes with the Target, so that Test A always shares either the same number or more attributes with the Target than Test B. Combinations of attributes shared by each of the Test stimuli and the Target define several distinct stimulus pattern conditions. For example, a T-00 pattern means that the Target shares zero attributes with both Test stimuli, whereas a T-20 pattern means that the Target shares two attributes with Test A and zero attributes with Test B. If we know  $S_{\text{vis.attr.}}$  and  $S_{\text{Label}}$ , we could compute similarity of Test A and the Target, and Test B and the Target for each of these stimulus pattern conditions, using equation 2. Plugging these similarities into equation 3, we could pre-





**Figure 3** Example of stimuli presented in one trial in the T-21 condition: The Target shares the shape of the face and the nose with Test stimulus A, and the size of the ears with Test stimulus B.

dict probabilities of inductive generalization between the Target and the Test stimulus that shares the label with the Target (i.e., Test B) for each of the stimulus pattern conditions. Values of  $S_{\text{Vis.attr.}} = .5$  and  $S_{\text{Label}} = .4$  were derived from Sloutsky & Lo's (1999) data on 6- to 7-year-olds' similarity judgment using these exact stimuli patterns. These estimates were plugged into equation 2, thus yielding similarity estimates that were plugged, in turn, into equation 3, thus yielding the presented probabilities. These predicted probabilities for each stimulus pattern condition are presented in Table 1.

The model presented in equation 2 was based on the assumption that different visual attributes have approximately equal attentional weights, something that represents a convenient simplification. For exam-

ple, overall shape of the face may be a more important feature than the size of ears. Thus, to avoid systematic biases stemming from this simplification, we (1) randomly assigned features to each of the stimulus pattern conditions, and (2) calibrated the stimuli. To calibrate the stimuli, we conducted a calibration study with all T-00, T-11, and T-22 stimuli patterns (each based on a unique combination of features). In the study, young children and college undergraduates decided which of the Test stimuli was more similar to the Target. Only those stimulus patterns in which young children and undergraduate participants deemed each of the Test stimuli to be equally similar to the Target were selected for the reported experiment. Note that T-00, T-11, and T-22 conditions were similar in that in each of these conditions, Test stimuli A and B were equally similar to the Target. Because attentional weights do not have to remain the same across different tasks and experimental conditions, predictions in Table 1 were ordinal rather than precise quantitative ones.

Based on the model of label-as-attribute we made the following critical predictions:

1. Independent of labeling, inductive generalization in young children should vary as a function of featural overlap among the compared entities.
2. Specific ordinal relations among the stimulus pattern conditions should be as follows (see Table 1):  $P(\text{T-00}) = P(\text{T-11}) = P(\text{T-22}) > P(\text{T-10}) = P(\text{T-21}) > P(\text{T-20})$ .
3. Labels should contribute quantitatively to inductive inference of young children (see Figure 2 for possible patterns of quantitative contribution).

In what follows, two experiments that tested predictions of the model with artificial stimuli (Experiment 1) and with naturalistic stimuli (Experiment 2) are presented. The goals of these experiments were: (1) to test the formulated critical predictions outlined above, (2) to examine the overall fit of the label-as-attribute model, and (3) to provide descriptive accounts of the development of induction.

## EXPERIMENT 1

### Method

#### Participants

A group of 97 children age 4 to 12 years participated in the experiment. The participants represented three age groups: (1) 42 four- to five-year-olds ( $M = 4.34$  years,  $SD = .49$  years; 25 males and 17 females), 30 seven- to eight-year-olds ( $M = 8.1$  years,  $SD = .5$  years; 15 males and 15 females), and 25 eleven- to twelve-year-olds ( $M = 11.8$  years,  $SD = .5$  years; 15

**Table 1** Predicted Probability of Generalizing an Unobservable Property from the Target to Test Stimulus B by Labeling Condition and Stimulus Pattern Condition

Stimulus Pattern Condition	Labeling Condition	
	No Label	Label Shared with B
T-00	.5	.71
T-11	.5	.71
T-22	.5	.71
T-10	.33	.56
T-21	.33	.56
T-20	.2	.38

males and 10 females). The participants were recruited from day-care centers and elementary and middle schools located in middle-class suburbs of Columbus, Ohio.

### Materials

The materials included triads of 5 cm × 5 cm schematic faces, two of which were Test stimuli and one of which was a Target, with the Target on the bottom and the two Test stimuli located equidistantly above (see Figure 3). Each schematic face had three distinct attributes (shape of head, shape of ears, and shape of nose), and each attribute had three values (e.g., "curved-lined" nose, "straight-lined" nose, and "angled" nose). Materials also included 36 artificial bisyllable labels (e.g., *Bala*, *Gula*, and so forth) and a set of unobservable biological properties of the Target. These unobservable properties included bones, brain, heart, stomach, fat, and blood, varying in size, shape, or color. Participants were asked which of the Test stimuli was more likely to share an unobservable property with the Target.

### Design and Procedure

The experiment had a mixed design with labeling condition (Label versus No Label) and age as between-subject factors and stimulus pattern as a within-subject variable. For both levels of the labeling condition, participants were presented with the same triads of schematic faces, two of which were Test stimuli and one of which was a Target. The only difference was that in the Label condition all stimuli were labeled, whereas in the No-Label condition these stimuli were not labeled. In the Label condition, the Target always shared labels with Test B and always had a different label than Test A. The stimulus pattern condition included six levels, T-00, T-11, T-22, T-10, T-21, and T-20. As mentioned previously, T refers to the Target, the first digit refers to the number of attributes shared by Test A with the Target, and the second digit refers to the number of attributes shared by Test B with the Target.

Children were seen by a female experimenter in a quiet room in their day-care centers. Before the experimental task, children were introduced to some warm-up questions and were given feedback.

*Warm-up trials.* In the warm-up trials, children were presented with Test and Target stimuli and were asked to choose the Test stimulus that shared a biological property with the Target. In the first warm-up trial, participants were presented with a Target (a shark) and two Test stimuli (a bear and a tree branch). In the

second warm-up trial, they were presented with a rabbit as a Target, and an apple and a dog as Test stimuli. In the third warm-up trial, children were presented with a fish as a Target, and a turtle and a spider as Test stimuli.

In all warm-up trials, children were first told that the Target stimuli either had bones, blood, or a skeleton inside the body. Children then were asked to determine which of the two Test stimuli had the same thing inside the body as the Target. If a child failed to make an inductive generalization, the researcher explained how each of the Test stimuli could have the same thing as the Target. The goal of the warm-up trials was to eliminate those participants who did not understand the task and produced random responses. If a child was capable of making inductive generalization in two out of three warm-up trials, the researcher proceeded to the experimental trials. No child was eliminated from the study because all participants provided satisfactory responses in at least two out of three warm-up trials.

*Experimental trials.* In the Label condition, children were first auditorily introduced to the labels for the Target and Test pictures. After each stimulus was labeled, children were asked to repeat these labels. No labels were introduced in the No-Label condition. Children were then introduced to an unobservable biological property of the Target stimulus and were asked which of the Test stimuli was more likely to have this property. Positions of the two Test pictures (i.e., to the left or right of the Target) were counterbalanced across the experimental trials. In both conditions, participants had 24 experimental trials (six within-subject stimulus patterns with four trials each). The order of stimulus patterns was randomized within participants. The important part of the instructions was as follows:

I am going to show you some pictures of Aliens, so you'll learn more about them. Are you ready to start? Let's start! Here we have three Alien pictures [pictures were introduced at this point]. They come from different planets . . . for example, Guga and Bala. Could you please repeat these names? Look at this one. This is a Guga [points to the target]. This is a Bala [points to Test stimulus A], and this is a Guga [points to Test stimulus B]. This Guga has yellow blood inside the body. Which one of them [points to two Test stimuli] has yellow blood inside the body like this Guga? Does this Guga [points to Test stimulus B] have yellow blood inside the body like this Guga [points to the target] or does this Bala [points to Test stimulus A] have yellow blood inside the body like this Guga [points to the target]?

Note that in the No-Label condition all stimuli were

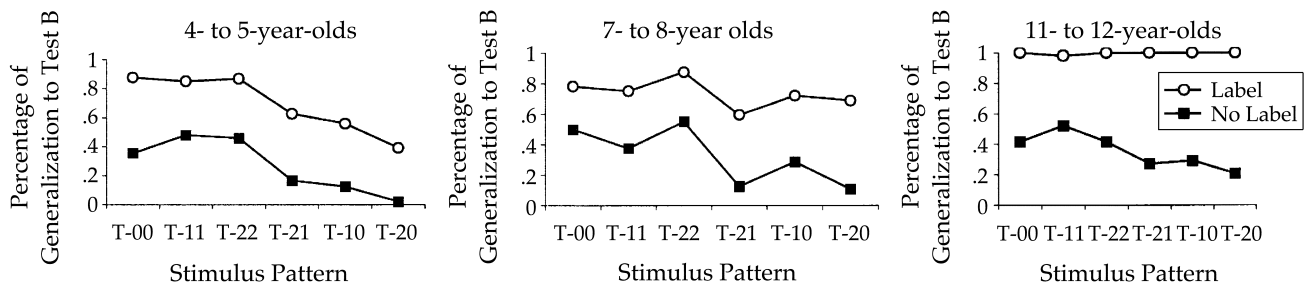


Figure 4 Percentage of Test B generalizations broken down by age, stimulus pattern, and labeling conditions (Experiment 1).

referred to as “this one.” The order of introduction of the Test stimuli and the order of induction questions were randomized.

## Results and Discussion

Proportions of Test B choices broken down by age, stimulus pattern, and labeling conditions are presented in Figure 4. Recall that B choices refer to the selection of the Test stimulus that in the Label condition always shared the label with the Target.

To examine overall age effects, proportions of B choices were subjected to a three-way mixed ANOVA with labeling condition and age as factors and stimulus pattern condition as the repeated measure. The analysis revealed a significant main effect for the stimulus pattern condition,  $F(5, 455) = 14.6$ ,  $MSE = .08$ ,  $p < .001$ ; a significant main effect for labeling condition,  $F(1, 91) = 114.4$ ,  $MSE = .03$ ,  $p < .001$ ; and a significant main effect for age,  $F(2, 91) = 6.1$ ,  $MSE = .02$ ,  $p < .005$ . There was also a marginally significant Age  $\times$  Labeling Condition interaction,  $F(2, 91) = 2.6$ ,  $MSE = .03$ ,  $p = .08$ , with larger effects for labeling condition in the group of 11- to 12-year-olds ( $M_{\text{Label}} - M_{\text{NoLabel}} = 64\%$ ) than in the group of 4- to 5-year-olds ( $M_{\text{Label}} - M_{\text{NoLabel}} = 42\%$ ) or in the group of 7- to 8-year-olds ( $M_{\text{Label}} - M_{\text{NoLabel}} = 40\%$ ). Finally, there was a significant Age  $\times$  Stimulus Pattern Condition interaction,  $F(10, 455) = 2.8$ ,  $MSE = .08$ ,  $p < .005$ , with smaller differences among the stimulus pattern conditions in the group of 11- to 12-year-olds (Max.Diff = 14%) than in the group of 4- to 5-year-olds (Max.Diff = 46%) or the group of 7- to 8-year-olds (Max.Diff = 46%). No other interactions were significant. Having established overall age effects, we also deemed it necessary to conduct more detailed analyses within each age group.

Proportions of Test B generalizations in the group of 4- to 5-year-olds are presented in the leftmost panel of Figure 4. These data were subjected to a mixed ANOVA with labeling condition as a factor and stimulus pattern condition as a repeated measure. The

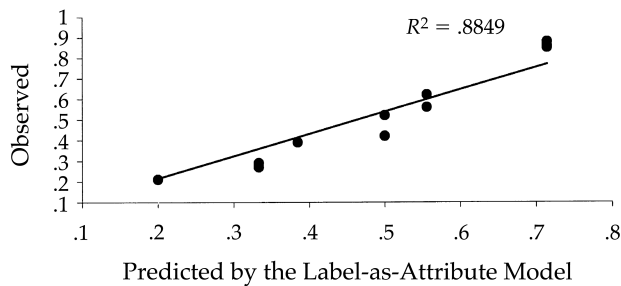
analysis revealed a significant main effect for labeling condition,  $F(1, 40) = 174.22$ ,  $MSE = .05$ ,  $p < .002$ , and a significant main effect for the stimulus pattern condition,  $F(5, 200) = 14.6$ ,  $MSE = .09$ ,  $p < .001$ ; the interaction was not significant,  $F < 1$ ,  $p = .9$ .

To test the hypothesized predictions, a set of planned comparisons was performed contrasting proportions of Test B generalizations across labeling and stimulus pattern conditions. As predicted (see Table 1), the proportions of inductive generalizations to Test B in T-00, T-11, and T-22 conditions were statistically equivalent, all  $ts(42) < 1$ . At the same time, as predicted, proportions of Test B generalizations differed significantly across the T-10, T-21, and T-20 conditions. Planned comparisons pointed to the following order among the conditions in the proportion of Test B generalizations: T-00 = T-11 = T-22 > T-21 = T-10 > T-20, for all indicated differences  $ts(42) > 3.7$ , Bonferroni-adjusted  $ps < .01$ . As predicted, differences between T-21 and T-10 were not significant,  $t(42) < 1$ . These results clearly support both critical predictions: the qualitative prediction that the proportion of label-based generalizations varies as a function of the number of features shared by the Target with each of the test stimuli, and the ordinal prediction. Participants' responses closely followed predictions presented in Table 1, exhibiting the predicted order in the proportion of Test B generalizations.

Overall fit between predicted probabilities (Table 1) and observed frequencies is presented in Figure 5. Predictions of the label-as-attribute model fit the observed data well: the model of label-as-attribute accounted for approximately 88% of the observed variance,  $R^2 = .884$ .

Proportions of Test B generalizations in the group of 7- to 8-year-olds are presented in the middle panel of Figure 4. These data were subjected to a mixed ANOVA with labeling condition as a factor and stimulus pattern condition as the repeated measure. The analysis revealed a significant main effect for labeling condition,  $F(1, 28) = 15.11$ ,  $MSE = .5$ ,  $p < .002$ , and a





**Figure 5** Theoretical probabilities (computed from the model using Equation 3) and observed probabilities of generalization of biological properties to Test B (Experiment 1).

significant main effect for the stimulus pattern condition,  $F(5, 140) = 5.7$ ,  $MSE = .1$ ,  $p < .001$ , whereas the interaction was not significant,  $F < 1$ ,  $p = .48$ . Planned comparisons indicated that there were no differences in Test B generalizations among the indeterminate conditions (T-00, T-11, and T-22),  $ps > .25$ , and among the determinate conditions (T-10, T-21, and T-20),  $ps > .49$ , whereas there were significant differences between the aggregated scores on determinate and indeterminate conditions,  $ts > 2.1$ ,  $ps < .05$ . Thus, the primary difference between the responses of 4- to 5-year-olds and 7- to 8-year-olds was that older children lacked ordinal differences in proportions of Test B generalizations across T-10, T-21, and T-20 conditions: in the group of 4- to 5-year-olds the direction was  $P(T-10) = P(T-21) > P(T-20)$ , whereas in the group of 7- to 8-year-olds there was no direction,  $P(T-10) = P(T-21) = P(T-20)$ .

Proportions of Test B generalizations in the group of 11- to 12-year-olds are presented in the rightmost panel of Figure 4. These data were subjected to a mixed ANOVA with labeling condition as a factor and stimulus pattern condition as the repeated measure. The analysis revealed a significant main effect for labeling condition,  $F(1, 23) = 534$ ,  $MSE = .03$ ,  $p < .001$ , a significant main effect for the stimulus pattern condition,  $F(5, 115) = 2.5$ ,  $MSE = .03$ ,  $p < .05$ , and a significant interaction,  $F(5, 115) = 3$ ,  $MSE = .03$ ,  $p < .05$ . The interaction was driven by the fact that in the No-Label condition there were significant differences between aggregated scores for the indeterminate stimulus pattern conditions (T-00, T-11, and T-22) and those for the determinate conditions (T-10, T-21, and T-20),  $t(11) = 3.6$ ,  $p < .01$ , whereas there were no such differences in the Label condition,  $t < 1$ . In fact, with the exception of one participant on one trial (i.e., 311 out of 312 responses), all participants on all trials in the Label condition used labels as the only basis for their induction.

Note that 11- to 12-year-olds exhibited important differences compared with 4- to 5- and 7- to 8-year-olds. In the No-Label condition, responses of 11- to 12-year-olds were similar to those in the two younger groups. In particular, there were no differences in Test B generalizations among the indeterminate conditions (T-00, T-11, and T-22),  $ps > .9$ , and among the determinate conditions (T-10, T-21, and T-20),  $ps > .7$ ; whereas, as noted above, there was a significant difference between the aggregated scores on determinate and indeterminate conditions. In the Label condition, however, proportions of Test B generalizations were statistically equivalent across the stimuli patterns with participants using labels as the only basis for their induction.

Overall, results presented in Figure 4 indicate that when only perceptual information was available, participants across all age groups based their inductive inference on this information: in all age groups inductive inference was a function of the number of attributes shared by the Target with the Test stimuli. Introduction of labels (recall that the Test B stimulus always shared the label with the Target), however, dramatically changed the proportions of B choices in the group of 11- to 12-year-olds, but not in 4- to 5-year-olds or in 7- to 8-year-olds. Whereas labels contributed qualitatively to induction performed by 11- to 12-year-olds and these participants relied exclusively on labels when performing induction, labels contributed quantitatively to induction performed by 4- to 5- and 7- to 8-year-olds, who relied on overall similarity computed over visual attributes and labels.

Recall that one of the goals of this experiment was to examine developmental differences in inductive inference. Results of the present experiment indicated that although younger children performed induction across labeled entities in a similarity-based manner, 11- to 12-year-olds performed such induction in a category-based manner. At the same time, 7- to 8-year-olds appeared to be a transitional group: their patterns of responses in the Label condition did not exhibit the ordinal relations observed in 4- to 5-year-olds, and yet they were not invariably at ceiling as were the responses of the 11- to 12-year-olds. To further examine the developmental transition, we compared proportions of participants in each of the three age groups who consistently performed label-based generalizations. Performance was considered consistent if a participant made the same choice on 18 out of 24 trials. This criterion was adopted because 18 out of 24 trials, with chance level equal to .5, constitutes a reliable, above-chance performance, binomial test,  $p < .01$ . Thirty nine percent of 4- to 5-year-olds, 56% of 7- to 8-year-olds, and 100% of 11- to 12-year-olds exhibited consistent label-based

generalizations. To examine the significance of these age-related differences, proportions of consistent responders were subjected to a  $\chi^2$  analysis. The analysis pointed to significant differences between the groups,  $\chi^2(2, N = 57) = 13.5, p < .001$ , with 11- to 12-year-olds demonstrating the largest proportion of consistent label-based responders,  $z_s > 1.99, p_s < .05$ . Thus, it seems reasonable to infer a developmental shift from similarity-based to category-based induction, occurring between 8 and 11, i.e., around 9 to 10 years of age.

In short, findings of this experiment supported predicted ordinal differences among the stimulus pattern conditions and indicated a good overall fit between predicted and observed data. In addition, findings pointed to important developmental differences in inductive generalization of 4- to 5-year-olds, 7- to 8-year-olds, and 11- to 12-year-olds.

However, it could be argued that the label-as-attribute account of young children's induction could be stimuli specific: the results of the reported experiment were based on artificial, perceptually impoverished artificial stimuli. Although these stimuli have important advantages in that they afford individuation of features and thus make these features tractable, these stimuli also have important limitations. In particular, it is unclear whether these stimuli are representative of naturalistic stimuli, and, therefore, it is unclear whether the results of Experiment 1 are generalizable to naturalistic, perceptually rich stimuli.

It also could be argued that the results might have been different if stimuli were not humanlike: infants and young children have been shown to develop different types of representation for human and nonhuman animals (Quinn & Eimas, 1998). Although the former could be represented as individual exemplars, the latter may have summary (i.e., category-based yet perceptual) representations.

Therefore, having supported the label-as-attribute model, we deemed it necessary to conduct another experiment to examine whether the regularities observed in Experiment 1 held for naturalistic, nonhumanlike stimuli. Because it is impossible to individuate features and to precisely calculate featural overlap with complex naturalistic stimuli patterns, we manipulated similarity by "morphing" naturalistic pictures into each other in a fixed number of steps.

## EXPERIMENT 2

The goal of this experiment was to replicate the findings of Experiment 1 with naturalistic stimuli. Findings would corroborate those of Experiment 1 if naturalistic stimuli exhibited patterns of variance similar to those observed in the leftmost panel in Figure 4. In other

words, we predicted that with naturalistic stimuli, the proportion of Test B generalizations in 4- to 5-year-olds should exhibit a main effect for stimulus pattern condition.

## Method

### Participants

A group of 26 children age 4 to 5 years ( $M = 4.8$  years,  $SD = .7$  years; 17 males and 10 females) and 40 undergraduate students ( $M = 21.9$  years,  $SD = 2.9$  years; 17 males and 10 females) participated in the experiment. Child participants were recruited from day-care centers located in middle-class suburbs of Columbus, Ohio and Houston, Texas, whereas undergraduate participants were students at the Ohio State University participating in the experiment in partial fulfillment of an introductory psychology course requirement.

### Materials

The materials included triads of  $5.6 \text{ cm} \times 5.6 \text{ cm}$  colored pictures of animals, two of which were Test stimuli and one of which was a Target, with the Target on the top and the two Test stimuli located equidistantly below. These materials were created in the following manner. First, 20 unique animal photographs were selected from a variety of Internet sites and software collections (e.g., Corel Gallery 1,300,000). These photographs were grouped into 10 pairs. Then each member of the pair was morphed into another member of the pair using MorphMan 1.1 (Stoik Corporation, Moscow, Russia, 1999). All morphing sequences consisted of eight intermediate steps. Each step in the sequence resulted in an approximately equal proportion of changed features as compared with a preceding step. An example of a morphed sequence with a cat morphed into a dog is presented in Figure 6. When morphed sequences were created, each member of the sequence was paired with each end point of the sequence (i.e., cat and dog are end points). These pairs were presented to 60 undergraduate students in the following preliminary experiment. Each undergraduate participant was presented with one pair of stimuli (an end point of a sequence and a step in the sequence) at a time and was asked to make a "same-different" judgment. Each pair was presented six times. These stimuli were presented on a PC computer monitor. The order of trials was randomized, and the experiment was controlled by SuperLab 2.0 (Cedrus Corporation, San Pedro, CA, 1999). The proportion of "same" responses within a pair aggregated across trials and participants was used as a measure of similarity between the mem-

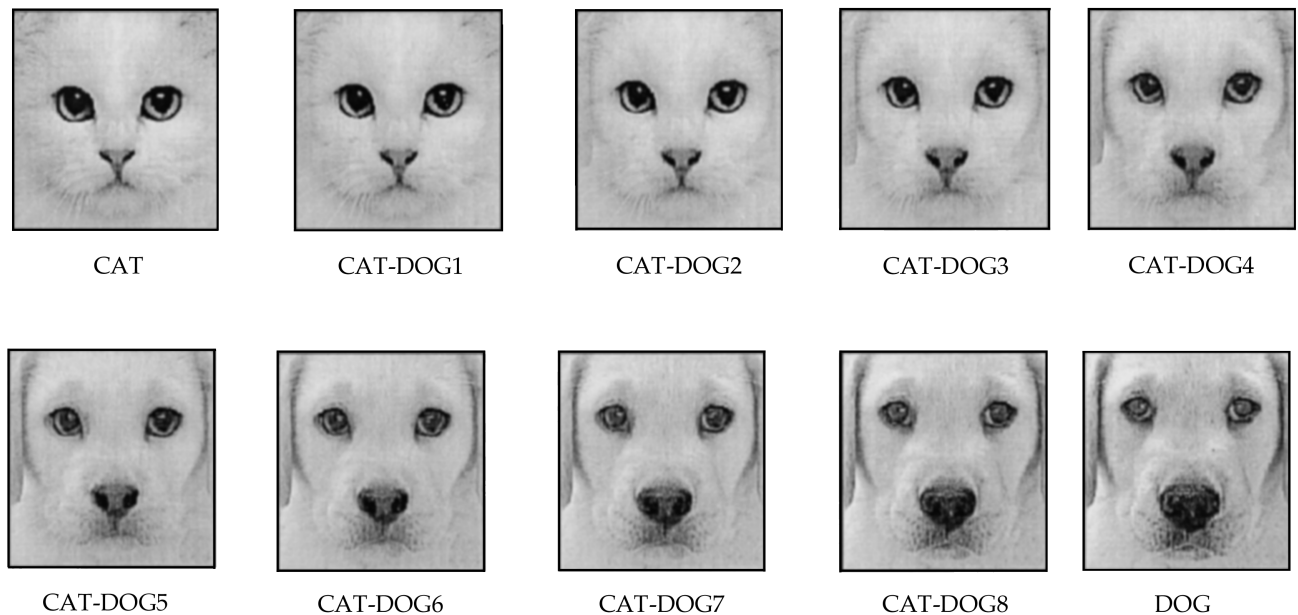


Figure 6 An example of an eight-step morphing sequence with a cat morphed into a dog (Experiment 2).

bers of the pair. This preliminary experiment was necessary because stimuli used in Experiment 2 did not afford individuation and calculation of overlapping features used in Experiment 1. Next, experimental triads (see Figure 7) were created with end points of sequences serving as Targets and morphed images serving as Test stimuli. The creation of triads was based on similarity between each of the Test stimuli and the Target. Three types of unique triads were created: (1) those in which both Test stimuli were judged to be equally similar to the Target [ $\text{Sim}(\text{Test A, Target}) - \text{Sim}(\text{Test B, Target}) = 0$ ], (2) those in which Test A was judged to be somewhat more similar to the Target than Test B [ $\text{Sim}(\text{Test A, Target}) - \text{Sim}(\text{Test B, Target}) = .5 \text{ to } .6$ ], and (3) those in which Test A was judged to be much more similar to the Target than Test B [ $\text{Sim}(\text{Test A, Target}) - \text{Sim}(\text{Test B, Target}) = .8 \text{ to } .9$ ]. These triads were referred to as T-0, T-.5, and T-.9 stimuli pattern conditions, respectively. The T-0, T-.5, and T-.9 stimuli pattern conditions consisted of four triads each (with each of the four triads made up of different animal pairs), thus resulting in 12 triads. Materials also included the set of artificial bisyllable labels and the set of unobservable biological properties used in Experiment 1.

#### Design and Procedure

The design and procedure used in Experiment 2 were similar to those used in Experiment 1, except that

in Experiment 2, stimuli were presented on a computer screen, presentation of the stimuli was controlled by a specially written computer program, and there were no warm-up trials. In both conditions, participants had 12 experimental trials (three within-subject stimulus patterns with four trials each). Children were seen in a quiet room in their day-care centers, whereas undergraduate students were seen in a research lab on campus.

#### Results and Discussion

Proportions of Test B choices broken down by age, stimulus pattern, and labeling conditions are presented in Figure 8. Recall that B choices refer to the selection of the Test stimulus that in the Label condition shared the label with the Target.

To examine overall age effects, proportions of B choices were subjected to a three-way mixed ANOVA with labeling condition and age as factors and stimulus pattern condition as the repeated measure. The analysis revealed a significant main effect for the stimulus pattern condition,  $F(2, 126) = 23.1$ ,  $MSE = .04$ ,  $p < .001$ ; a significant main effect for labeling condition,  $F(1, 63) = 225$ ,  $MSE = .06$ ,  $p < .001$ ; and a significant main effect for age,  $F(1, 63) = 4.99$ ,  $MSE = .06$ ,  $p < .05$ . There was also a significant Age  $\times$  Labeling Condition interaction,  $F(1, 63) = 47.75$ ,  $MSE = .06$ ,  $p < .001$ , with larger effects for labeling condition in

**Triad 1:**  $\text{Sim}(A, T) - \text{Sim}(B, T) \approx 0$



**TARGET**



**TEST A**

$\text{Sim}(A, \text{Target}) = .93$



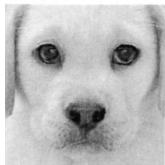
**TEST B**

$\text{Sim}(B, \text{Target}) = .99$

**Triad 2:**  $\text{Sim}(A, T) - \text{Sim}(B, T) \approx .5$

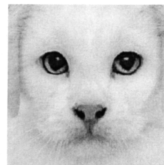


**TARGET**



**TEST A**

$\text{Sim}(A, \text{Target}) = .82$



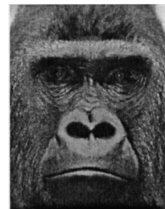
**TEST B**

$\text{Sim}(B, \text{Target}) = .26$

**Triad 3:**  $\text{Sim}(A, T) - \text{Sim}(B, T) \approx .8$

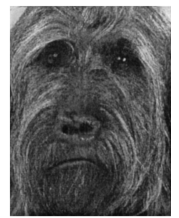


**TARGET**



**TEST A**

$\text{Sim}(A, \text{Target}) = .93$



**TEST B**

$\text{Sim}(B, \text{Target}) = .15$

**Figure 7** Examples of stimuli presented in the three stimulus pattern conditions (Experiment 2).

the group of undergraduates ( $M_{\text{Label}} - M_{\text{NoLabel}} = 76\%$ ) than in the group of 4- to 5-year-olds ( $M_{\text{Label}} - M_{\text{NoLabel}} = 28\%$ ). Finally, there was a significant Label  $\times$  Stimulus Pattern Condition interaction,  $F(2, 126) = 5.9$ ,  $MSE = .04$ ,  $p < .005$ , with smaller differences among the stimulus pattern conditions in the Label condition (Max.Diff = 12%) than in the No-Label condition (Max.Diff = 39%), and a significant three-way (Age  $\times$  Labeling Condition  $\times$  Stimulus Pattern Condition) interaction,  $F(2, 126) = 9.6$ ,  $MSE = .04$ ,  $p < .001$ . No other interactions were significant. Having established overall age effects, we deemed it necessary to conduct more detailed analyses within each age group.

Data in Figure 8 were subjected to two mixed ANOVAs with labeling condition as a factor and

stimulus pattern condition as the repeated measure. For the 4- to 5-year-olds, the analysis revealed a significant main effect for the labeling condition,  $F(1, 25) = 15.5$ ,  $MSE = .1$ ,  $p < .005$ , and a significant main effect for the stimulus pattern condition,  $F(2, 50) = 4$ ,  $MSE = .06$ ,  $p < .05$ , whereas the interaction was not significant,  $F < 1$ ,  $p = .9$ . Planned comparisons indicated the following direction of the proportion in label-based generalization:  $T-0 > T-.5 = T-.9$ , for all indicated differences  $t(15) > 2.4$ , Bonferroni-adjusted  $ps < .08$ , for  $T-.5$  versus  $T-.9$ ,  $t(15) < 1$ , Bonferroni adjusted  $p = 1$ .

For the undergraduate group, the analysis revealed a significant main effect for the labeling condition,  $F(1, 38) = 59.6$ ,  $MSE = .03$ ,  $p < .001$ , and a significant main effect for the stimulus pattern condition,  $F(2, 76) = 28.4$ ,  $MSE = .03$ ,  $p < .001$ . The interaction was

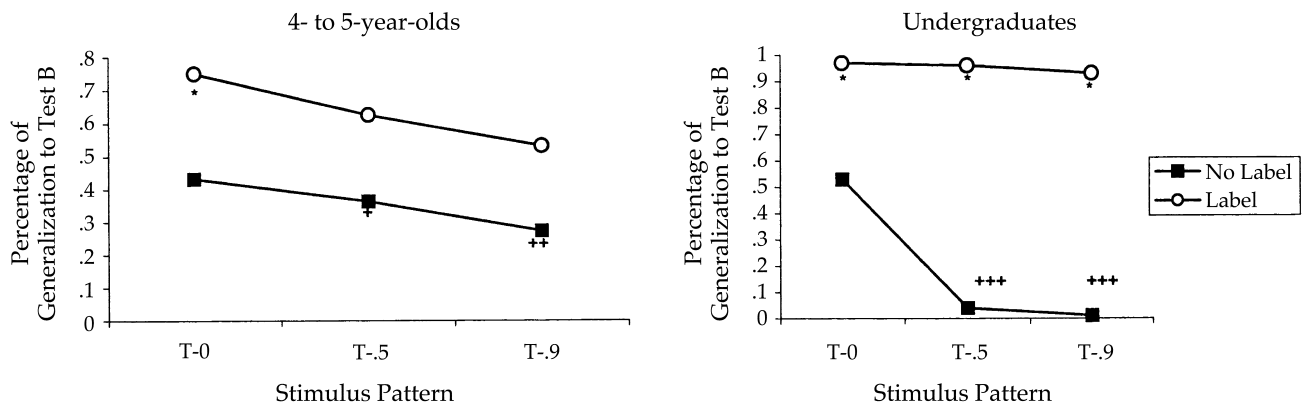


Figure 8 Percentage of Test B generalizations broken down by age, stimulus pattern, and labeling conditions (Experiment 2). \*Above chance,  $p < .001$ ; +below chance,  $p < .08$ ; ++below chance,  $p < .01$ ; +++below chance,  $p < .001$ .

also significant,  $F(2, 76) = 34.4$ ,  $MSE = .03$ ,  $p < .001$ , which indicates that there were differences across the stimulus pattern conditions in the No-Label condition ( $M_{T-0} = .53$ ,  $M_{T-5} = .04$ , and  $M_{T-9} = .01$ ), whereas there were no such differences in the Label condition ( $M_{T-0} = .97$ ,  $M_{T-5} = .96$ , and  $M_{T-9} = .93$ ).

To test differences among the stimulus pattern conditions, a set of planned comparisons was performed contrasting proportions of Test B generalizations across labeling and stimulus pattern conditions. Planned comparisons pointed to the following order among the conditions in the proportion of Test B generalizations in the No-Label condition:  $T-0 > T-5 = T-9$ , for all indicated differences  $ts(19) > 6.5$ , Bonferroni-adjusted  $ps < .001$ , for T-5 versus T-9,  $t(19) = 1$ , Bonferroni-adjusted  $p = .99$ . At the same time, in the Label condition there were no differences across the stimulus pattern conditions,  $ts(19) < 1$ , with the majority of participants performing label-based generalization independent of the stimulus pattern condition.

Interestingly enough, results point to differences in the proportion of Test B generalizations in the No-Label conditions between adult and children participants: In 4- to 5-year-olds, these proportions decreased gradually with an increase of similarity between Test A and the Target; whereas in adults, these proportions decreased steeply (see Figure 8). These different rates of decrease may have stemmed from the increased ability to discriminate stimuli by adults than by children; in which case, under T-5, and T-9 conditions, stimuli A and B appear to be very different to adults, whereas they appear to be somewhat similar to children. Although such differences in the discrimination ability are not unexpected (e.g., for an extensive theoretical treatment, see Gibson, 1969), additional research is required.

In short, results of Experiment 2 were similar to those of Experiment 1: when labels were available, adult participants used labels as the only basis of their induction, whereas young children's induction varied as a function of overall similarity among the Test stimuli and the Target.

## GENERAL DISCUSSION

Critical results of the two reported experiments support predictions of the label-as-attribute model and point to marked differences in induction of young children, on the one hand, and preadolescent and adults, on the other. In Experiment 1, young children performed induction in accordance with predictions of the label-as-attribute model. In both Label and No-Label conditions, their inductive generalization varied as a function of featural overlap among compared entities. Furthermore, observed probabilities of label-based generalizations in young children exhibited predicted ordinal relations among the stimulus pattern conditions:  $P(T-00) = P(T-11) = P(T-22) > P(T-10) = P(T-21) > P(T-20)$ . Results of Experiment 1 also pointed to developmental differences in the contribution of labels to induction: whereas the majority of 4- to 5-year-olds performed induction by relying on the overall similarity as determined by a combination of label and perceptual similarity, the majority of preadolescents relied exclusively on labels. At the same time, 7- to 8-year-olds appeared to be a transitional group, in that there were more consistent label responders among 7- to 8-year-olds than among 4- to 5-year-olds, but less than among 11- to 12-year-olds.

Experiment 2 corroborated results of Experiment 1, indicating that the regularities observed in Experiment 1 held for naturalistic, perceptually rich stimuli.

As in Experiment 1, for both labeled and nonlabeled stimuli, inductive generalization performed by 4- to 5-year-olds varied as a function of similarity among the compared stimuli. At the same time, adults, similar to preadolescents in Experiment 1, relied almost exclusively on labels when performing induction.

Recall that these experiments were designed to (1) test critical predictions derived from the label-as-attribute model, (2) test the overall fit of the model by comparing predicted and observed values, and (3) provide descriptive accounts of the development of induction from early childhood to preadolescence. The reported results suggest that the experiments met these goals.

First, these results support critical predictions of the label-as-attribute model. As predicted by the model, young children exhibited variability in generalization from Target to Test B as a function of similarity of Test A and Test B, on the one hand, and the Target, on the other hand. Furthermore, the observed responses exhibited predicted ordinal differences in proportions of Test B generalizations across the stimulus pattern conditions for both Label and No-Label conditions. These differences were as follows:  $P(T-00) = P(T-11) = P(T-22) > P(T-10) = P(T-21) > P(T-20)$ . In addition, the model of label-as-attribute accounted for approximately 88% of the observed variance in young children's inductive generalizations.

Sloutsky and Lo (1999, Experiment 1) proposed the label-as-attribute model as a model of similarity. It accounted for approximately 75% of the observed variance in young children's similarity judgments. Furthermore, the similarity judgment responses of young children that were observed by Sloutsky and Lo (1999) correlate highly with the induction responses of young children observed in Experiment 1 of the current study,  $r(12) = .96$ . In their study, Sloutsky and Lo (1999) used the same stimuli, the same stimulus pattern conditions, and the same labeling conditions as those used in Experiment 1 of the current study. In addition, their task and procedure were very similar to the current study, except that participants were asked to determine which of the test stimuli was more similar to the Target. Thus, a single model predicted both similarity judgment and inductive inference in young children, and young children's similarity judgment correlated highly with their induction. These findings support the notion that there might be a single mechanism underlying similarity judgment and inductive inference in young children, offering a parsimonious explanation for young children's induction and similarity judgment.

Furthermore, young children's reliance on overall similarity when performing induction, rather than on a

single most predictive attribute, such as category label, is similar to their reliance on holistic similarity when making similarity judgments (Smith, 1989a, 1989b). Holistic similarity, or overall proximity of stimuli on several dimensions, has been contrasted with dimensional identity, or attribute match on a single dimension. For example, in terms of holistic or overall similarity, a black circle is more similar to a gray ellipse than it is to a white circle (identity of shape) or a black rectangle (identity of color). In Smith's studies, young children were found to rely on holistic similarity, whereas in older children's similarity judgment, dimensional identity (e.g., a black circle and a white circle) was found to play a more prominent role. As with Smith's (1989a, 1989b) results with similarity judgment, the results of this study indicate that although younger children tended to perform induction based on the overlap of several features, older children focused on a single attribute (such as category labels) that they deemed to be most predictive of unobservable properties. Taken together, these results further support the contention that induction and similarity in young children are interrelated.

The reported results also point to an important developmental trend in the direction of reliance on a single most predictive attribute, or knowledge-based induction. Although there was a sizable minority of 4- to 5-year-old children who performed induction relying exclusively on labels, all preadolescents and adults performed induction in this manner.

The fact that in young children the proportion of label-based generalization varied as a function of the overall similarity, whereas preadolescents and adults invariantly relied on the common linguistic labels, points to developmental differences between inductive generalization in young children and preadolescents. The reported results suggest that specific induction may start out as a knowledge-lean similarity-based process and develop into a knowledge-based process, in which category label is used as a single most predictive attribute; of course, the same regularity has to be established across several knowledge domains to provide conclusive evidence to support this suggestion. This developmental shift may occur sometime between 8 and 11 years of age. Indeed, whereas specific induction of 4-year-olds is similarity based, 7- to 8-year-olds appear to be a transitional group, and 11- to 12-year-olds perform induction on the basis of common labels, independent of the overall similarity.

It seems plausible that this developmental shift may be a function of the development of (1) a domain-general record of probabilities, and (2) a domain-specific biological knowledge. The record of probabilities suggests that two remotely similar objects sharing a label



are more likely to be considered members of the same category than to be considered members of different categories. This probability could be estimated by its complement, or the probability that two remotely similar entities that have the same label would be considered members of different categories. This probability of the complement could be estimated by the base rate of homonyms (and homophones), and, therefore, would be negligibly small: it is difficult to come up with a noun having a homonym within the same ontological class (e.g., living creature having a homonym that indicates a completely different living creature). Hence, as children grow, they acquire more evidence confirming that remotely similar entities sharing the label are likely to be members of the same category. In addition, biological knowledge acquired at school may further corroborate these probabilistic intuitions, suggesting that a category label (e.g., *mammal*) is the single best predictor of most anatomical and physiological properties of mammalian animals.

Therefore, in the absence of labels, both children and preadolescents (and adults) rely on the overall perceptual similarity among compared entities. However, when labels are available, young children continue to rely on the overall similarity, whereas preadolescents and adults switch to labels as the most predictive attributes. Interestingly enough, the predictive value of labels for induction is independent of similarity: Sloutsky & Lo (1999) demonstrated that shared labels made virtually no contribution to similarity judgments of preadolescents (but made a sizable, yet quantitative, contribution to similarity judgments of young children). In other words, labels affect both similarity and induction in young children, but affect only induction in preadolescents and adults. Therefore, it seems reasonable to conclude that in young children (1) similarity and induction are related, and (2) labels contribute quantitatively to both similarity and induction. At the same time, in older children and adults (1) similarity and induction are independent, and (2) labels make no contribution to similarity, but contribute qualitatively to their induction.

There are several issues to be examined in future research. If induction performed by young children is based on overall similarity among compared entities, then introduction of new classes of attributes (both perceptual and nonperceptual) that contribute to overall similarity should also contribute to inductive generalizations. For example, we have evidence that some unobservable informational attributes, such as kinship information, contribute to induction of young children. Of course, it is reasonable to expect that different classes of attributes would have different attentional weights, and it would be important to estimate these

weights. It would be important also to trace changes in these weights with development and learning. In addition, it seems important to examine whether a lack of information about an attribute constitutes a mismatch. For example, how would occluded parts of objects be added to the computation of similarity? Furthermore, if a Target shares the label with Test B, but Test A is not labeled, would this constitute a label mismatch? Finally, it would be necessary to test the model of label-as-attributes on younger participants and have more densely spaced developmental observations.

In short, although the ability of the label-as-attribute model to account for younger children's induction under different stimuli conditions requires further testing, the reported findings support predictions that with both contrived and naturalistic stimuli, young children perform specific induction in a similarity-based manner. At the same time, results indicate that preadolescents and adults rely almost exclusively on labels as a single most predictive attribute when performing specific induction. Thus, reported results point to a developmental shift from treating linguistic labels as an attribute that contributes to similarity to treating them as markers of a common category, a shift that occurs between 8 and 11 years of age.

## ACKNOWLEDGMENTS

This research was supported by grants from the James S. McDonnell Foundation and the National Science Foundation (BCS #0078945) to the first author. The authors would like to thank James L. Dannemiller, Brad Morris, Aaron Rader, Margie Spino, Aaron Yarlas, Paul Harris, and two anonymous reviewers for their feedback. They are also thankful to Jenny Volvovski for creating the stimuli for Experiment 1, and to Dave Brown for creating and letting them use the cat-dog sequence in Experiment 2.

## ADDRESSES AND AFFILIATIONS

Corresponding author: Vladimir M. Sloutsky, Center for Cognitive Science, Ohio State University, 21 Page Hall, 1810 College Road, Columbus, OH 43210-1336; e-mail: sloutsky.1@osu.edu. Ya-Fen Lo is at Rice University, Houston, TX; Anna V. Fisher is also at Ohio State University.

## REFERENCES

- Aslin, R., & Smith, L. (1988). Perceptual development. *Annual Review of Psychology*, 39, 435-474.
- Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, 64, 3-26.

- Birnholz, J. C., & Benaceraff, B. B. (1983). The development of human fetal hearing. *Science*, 222, 516–518.
- Eimas, P. (1994). Categorization in early infancy and the continuity of development. *Cognition*, 50(1–3), 83–93.
- Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.
- Fisher, R. (1951). *The design of experiments*. New York: Hafner Publishing Company. (Original work published 1935)
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20, 65–95.
- Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). New York: Cambridge University Press.
- Gelman, S. A., Coley, J. D., & Gottfried, G. M. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.
- Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, 23, 183–209.
- Gelman, S. A., & Markman, E. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, 58, 1532–1541.
- Gentner, D. (1978). A study of early word meaning: What looks like a jiggy but acts like a zimbo? *Papers and Reports on Child Language Development*, 15, 1–6.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Goodman, N. (1992). Seven strictures on similarity. In M. Douglas & D. Hull (Eds.), *How classification works* (pp. 13–23). Edinburgh, Scotland: Edinburgh University Press. (Original work published 1972)
- Gutheil, G., Vera, A., & Keil, F. C. (1998). Do houseflies think? Patterns of induction and biological beliefs in development. *Cognition*, 66, 33–49.
- Jusczyk, P. W. (1998). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, 36, 1–27.
- Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition*, 59, 307–335.
- Mandler, J. M., & McDonough, L. (1998). Studies in inductive inference in infancy. *Cognitive Psychology*, 37, 60–96.
- Medin, D. (1975). A theory of context in discrimination learning. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 263–314). New York: Academic.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Napolitano, A., Sloutsky, V. M., & Boysen, S. T. (2001). Modality preference and its change in the course of development. In J. Moore & K. Stenning (Eds.), *Proceedings of the XXIII Annual Conference of the Cognitive Science Society* (pp. 681–685). Mahwah, NJ: Erlbaum.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339–363.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104–114.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185–200.
- Quinn, P., & Eimas, P. (1998). Evidence for global categorical representation of humans by young infants. *Journal of Experimental Child Psychology*, 69, 151–174.
- Roberts, K. (1995). Categorical responding in 15-month-olds: Influence of the noun-category bias and the covariation between visual fixation and auditory input. *Cognitive Development*, 10, 21–41.
- Roberts, K., & Jacob, M. (1991). Linguistic versus attentional influences on nonlinguistic categorization in 15-month-old infants. *Cognitive Development*, 6, 355–375.
- Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323.
- Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1. Linguistic labels and the development of similarity judgement. *Developmental Psychology*, 6, 1478–1492.
- Smith, L. (1989a). A model of perceptual classification in children and adults. *Psychological Review*, 96, 125–144.
- Smith, L. (1989b). From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosnoudou & A. Orthony (Eds.), *Similarity and analogical reasoning* (pp. 146–178). New York: Cambridge University Press.
- Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism. *Cognition*, 60, 143–171.
- Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development*, 63, 950–959.
- Thompson, L. A., & Massaro, D. (1989). Before you see it, you see its parts: Evidence for feature encoding and integration in children and adults. *Cognitive Psychology*, 21, 334–362.
- Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.