Journal of Experimental Psychology: General
2004, Vol. 133, No. 2, 166–188

0096-3445/04/$12.00 DOI: 10.1037/0096-3445.133.2.166

Induction and Categorization in Young Children: A Similarity-Based Model

Vladimir M. Sloutsky and Anna V. Fisher
The Ohio State University

The authors present a similarity-based model of induction and categorization in young children (SINC). The model suggests that (a) linguistic labels contribute to the perceived similarity of compared entities and (b) categorization and induction are a function of similarity computed over perceptual information and linguistic labels. The model also predicts young children's similarity judgment, induction, and categorization performance under different stimuli and task conditions. Predictions of the model were tested and confirmed in 6 experiments, in which 4- to 5-year-olds performed similarity judgment, induction, and categorization tasks using artificial and real labels (Experiments 1–4) and recognition memory tasks (Experiments 5A and 5B). Results corroborate the similarity-based account of young children's induction and categorization, and they support both qualitative and quantitative predictions of the model.

Inductive inference, or extending knowledge from known to novel instances, is ubiquitous in human cognition. For example, if one learned that a particular lion has a certain neurotransmitter in its brain, one would expect another lion also to have this neurotransmitter, even if one did not have factual knowledge of the brain physiology in lions. Some believe that induction is the most important component of thinking because "inductive inference is the only process . . . by which new knowledge comes into the world" (Fisher, 1935/1951, p. 7). Examples of inductive inference are (a) X_1 has property Y, therefore X_2 has property Y, or (b) X_1 is Z, therefore X_2 is Z. We refer to the former as *projective induction* and to the latter as *categorization*. The term *induction* is used to refer to both projective induction and categorization.

Induction requires at least two stimuli or stimuli sets, the base and the target, and it is clear that not all kinds of stimuli support induction. For example, if X_1 and X_2 are lions, they have much more in common than if X_1 and X_2 are white things, and therefore much more can be induced from one lion to another than from one white thing to another white thing (see Markman, 1989, for related discussion). The former entities constitute a natural kind: a grouping of entities existing in nature and having many commonalities. Natural kinds form taxonomies or hierarchically organized categories (e.g., dogs are mammals, and mammals are animals). A part of a natural kind taxonomy is presented in Figure 1. The taxonomy captures class-inclusion relationships, such that information that is known to be true of higher level nodes can be deduced (or inferred with certainty) downward to lower level nodes, whereas information that is true of lower level nodes can be induced (or inferred with some likelihood) to the same-level nodes or upward to nodes of higher levels. Deductive inferences are represented in Figure 1 with black arrows, whereas inductive inferences are represented with gray arrows.

There is little disagreement that natural kind categories, as well as some nominal kinds (e.g., even → natural → rational → real numbers in the number system), are organized taxonomically and that these taxonomies are often present in people's knowledge systems (e.g., Anderson & Bower, 1973; Collins & Quillian, 1969; McClelland & Rogers, 2003). The contested part pertains to acquisition of taxonomic knowledge, its representation, and its use in generalization and reasoning tasks. In what follows, we review several approaches to induction, which vary in their assumptions about reliance on taxonomic knowledge. We then propose a new similarity-based model that assumes no use of taxonomic knowledge by young children.

Similarity-Based Versus Knowledge-Based Approaches to Induction

One approach that assumes very little reliance on taxonomic knowledge is the feature-based model (Sloman, 1993, 1998). According to this model, induction is based on the featural overlap between entities in premises and in the conclusion, regardless of the place of these entities in the taxonomy. The model accounts for some of the taxonomic effects, such as the differential strength of inference to the immediate versus a more remote category. For example, the argument "Sparrows have X. Eagles have X. Therefore birds have X" is judged stronger than "Sparrows have X. Eagles have X. Therefore animals have X." The model also makes a nontrivial prediction that poses challenges to its competitors. The model predicts that feature overlap mediates not only upward

Vladimir M. Sloutsky, Center for Cognitive Science, The Ohio State University; Anna V. Fisher, Center for Cognitive Science and Department of Psychology, The Ohio State University.

This research was supported by Grant BCS 0078945 from the National Science Foundation to Vladimir M. Sloutsky. We thank Margie Spino and Aaron Yarlas for their helpful comments on earlier versions of this article. We are especially grateful to Susan A. Gelman for promptly providing stimuli and the list of properties used in Gelman and Markman (1986, Studies 1 and 3).

Correspondence concerning this article should be addressed to Vladimir M. Sloutsky, Center for Cognitive Science, The Ohio State University, 208C Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210. E-mail: sloutsky.1@osu.edu

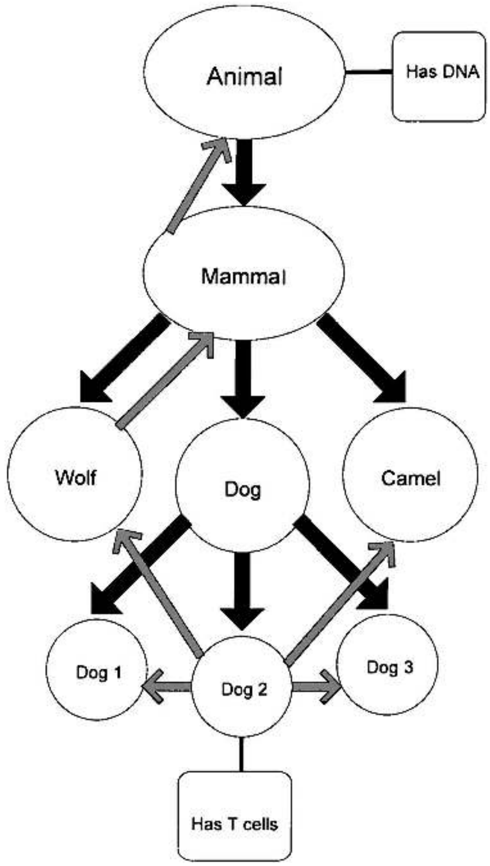


Figure 1. A component of a natural kind taxonomy. Black arrows represent deductive inferences; gray arrows represent inductive inferences.

inference but also downward inference, leading to the neglect of class-inclusion relationships. In particular, participants in Sloman's (1993) studies considered the argument "All animals have X. Therefore all mammals have X" to be stronger than the argument "All animals have X. Therefore all reptiles have X," whereas according to the taxonomic model, both arguments should have equal (and maximal) strength.

Another approach, the similarity-coverage model (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), assumes that induction relies both on the similarity of entities in premises and in the conclusion and on the place of these entities in a taxonomy. For example, when presented with the premise "Dogs have T cells" and asked "Do camels have T cells?" participants compute similarity (a) between dogs and camels and (b) between dogs and mammals, which is the nearest superordinate category that includes both dogs and camels. It is clear that the latter computation requires some knowledge of and reliance on the taxonomical organization.

However, it has been argued that similarity-based models cannot explain several empirical phenomena (see Goldstone, 1994a; Rips, 2001, for related discussions; see also McClelland & Rogers, 2003, for a similarity-based model that can account for these empirical phenomena). First, people may consider some features to be more central (or essential) for category membership than others, and hypothetical elimination of less central features had smaller effects on categorization than did elimination of more central features. For example, children and adults are more likely to categorize a tiger without stripes as a tiger than they are a tiger whose genetic makeup or insides have been changed (Gelman & Wellman, 1991; Keil, 1989; Keil, Smith, Simons, & Levin, 1998; Rips, 1989). Similarly, test animals differing from target animals in that they missed features that were deemed causal, such as a protein that causes bones to grow, were less likely to be considered to belong to the same kind as target animals than were test animals that lacked noncausal features, such as large bones (Ahn, Gelman, Amsterlaw, Hohenstein, & Kalish, 2000).

Second, the importance (or centrality) of features may differ across contexts. In particular, the same feature, such as curvedness, may be more central for categorizing an object as a boomerang than it is for categorizing an object as a banana: Although a straight banana is still a banana, a straight boomerang is not a boomerang (Medin & Shoben, 1988). Similarly, when asked to induce anatomical features (e.g., type of liver), people are more likely to induce from whales to bears than from whales to tuna, whereas when asked to induce behavioral features (e.g., type of locomotion), people are more willing to induce from whales to tuna than from whales to bears (Heit & Rubinstein, 1994).

Hence several alternatives to similarity-based models have been proposed. These alternative approaches argue that in both categorization and induction, people rely more on taxonomic information than they do on similarity (see Rips, 2001, for a review). For example, the hypothesis-testing model (e.g., McDonald, Samuels, & Rispoli, 1996) presupposes knowledge of the biological taxonomy, and it does not posit reliance on similarity. The model suggest that when presented with the premise "Dogs and camels have T cells" and asked "Do mammals have T cells?" participants base their induction on the amount of evidence (i.e., information in the premises), the scope of the hypothesis (i.e., information in the conclusion), and the number of plausible alternative hypotheses (i.e., that T cells are a result of the immunization of domestic animals).

Another proposal, which we refer to as *causal essentialism*, presupposes that categorization and projective induction are affected by beliefs about (a) feature centrality, with some features being core or essential and others being peripheral (Keil et al., 1998; Medin & Shoben, 1988; Rips, 1989), and (b) causal status of features in question (Keil et al., 1998; Ahn et al., 2000). The causal essentialism proposal also has been extended to induction in children, suggesting that even young children base their induction on intuitions about the taxonomic organization of natural kinds (e.g., Gelman, 2000). The proposal assumes that, similar to adults, children believe that there is a hierarchy of features, some of which are more central for the category membership than others (Ahn et al., 2001; Gelman & Wellman, 1991; Keil et al., 1998). Because the same property could be central in one domain of knowledge and peripheral in another domain (e.g., as argued by Medin & Shoben, 1988), induction, according to this view, is a domain-specific process.

In addition, young children expect individuals to be members of some categories, and their category membership is conveyed by category labels presented as nouns. "Children assume that every object belongs to a natural kind and that common nouns convey natural kind status (as well their accompanying properties) . . . Names are embodiment of our theories" (Gelman & Coley, 1991,

p. 190). This position is presented schematically in Figure 2. In the absence of category labels (or other theoretically central information), young children may perform induction on the basis of similarity (i.e., induction from Individual 2 to Individual 3 in Figure 2). They first include Individuals 2 and 3, who are similar, into assumed Template Category Y (gray arrows in Figure 2 represent similarity-based categorization) and then infer common properties for members of the same category (black arrows represent inference from a category to an individual). However, when entities are accompanied by category labels or when labels are known (i.e., induction from Individual 1 to Individual 2 in Figure 2), young children first include entities into assumed Template Category X on the basis of common labels (clear arrows in Figure 2 represent label-based categorization) and then infer common properties for members of the same category (black arrows in Figure 2).

However, the taxonomy-based models of induction in general and the causal essentialism position in particular also have their weaknesses (see Goldstone, 1994a; Sloutsky, 2003, for related discussions). First, as mentioned before, even adult participants often neglect class-inclusion relations (Sloman, 1993, 1998), and the adults' failure to rely on category membership weakens the assumption that young children typically rely on category membership when performing induction (Sloman, 1993, 1998).

Second, there is evidence that domain-general attentional and perceptual mechanisms play a greater role in conceptual development than that assumed by the causal essentialism position (Sloutsky, Lo, & Fisher, 2001; Smith & Heise, 1992; Smith, Jones, & Landau, 1996; see also Sloutsky, 2003, for a review). Furthermore, some of the experiments demonstrating that appearance is less central than "essential" features are have relied on verbal descriptions of appearance or on greatly impoverished perceptual stimuli (see Jones & Smith, 1993, for a discussion). In addition, some authors have criticized the essentialist approach for assuming

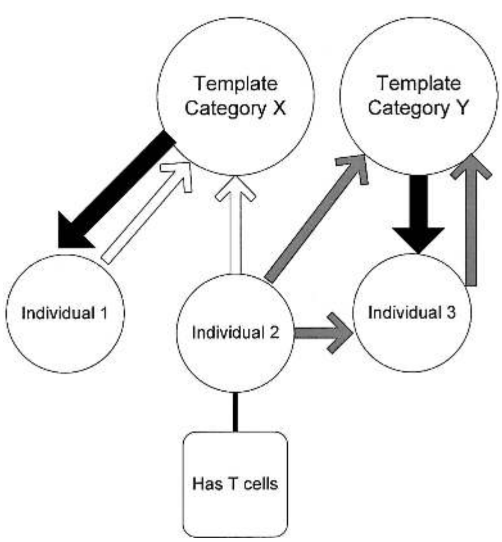


Figure 2. Category templates that, according to the causal essentialist position, underlie young children's induction. Black arrows represent inferences from a category to an individual; gray arrows represent similarity-based categorization; clear arrows represent label-based categorization.

too much, when existing data could be explained with fewer assumptions (e.g., Strevens, 2000).

Third, under some conditions, even for adults, essences fail to predict category membership. For example, some solutions known to participants to have lower proportions of H_2O (H_2O is an assumed essence of water) were more likely to be judged as water than were solutions known to have higher proportions of H_2O (Malt, 1994).

In addition, computer simulations of categorization and induction (McClelland & Rogers, 2003) also weaken somewhat the causal essentialism proposal. Simple neural networks that have little prior knowledge and powerful learning mechanisms that make them sensitive to feature co-occurrences successfully approximate human performance on categorization and induction tasks.

Therefore, we suggest dropping the taxonomic assumption altogether and instead recommend considering a model that explains young children's induction, categorization, and similarity computation as essentially variants of the same process. The model is abbreviated as SINC for similarity, induction, and categorization. In what follows, we present SINC and six experiments designed to test it. SINC assumes that young children consider linguistic labels to be attributes of objects that contribute to similarity among compared entities. This assumption has been supported empirically (Sloutsky & Lo, 1999; Sloutsky et al., 2001) and is further tested in the current research. The model also generates a wide range of testable predictions about mechanisms underlying induction and categorization in young children.

SINC: A Model of Similarity-Based Induction in Young Children

Qualitatively, SINC suggests that linguistic labels contribute to the similarity of compared entities and that similarity drives induction and categorization in young children. The model is based on the product-rule model of similarity (Estes, 1994; Medin, 1975) that specifies similarity among nonlabeled feature patterns. In the product-rule model, similarity is computed using Equation 1:

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

where i and j represent any two stimulus patterns, 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 one is presented with two visual patterns (e.g., schematic faces A and B). Further suppose that these patterns consist of four distinct features (i.e., the shape of the face, eyes, and nose and the size of ears) and that the patterns share two of these features (i.e., the shape of the face and eyes) and differ on the other two. Assuming that $S = .5$ (the value frequently derived empirically; Estes, 1994), similarity between A and B would be equal to .25 (i.e., $.5^2$). Note that similarity between entities decreases very 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., $.5^3$). However, if the faces shared all four features, they would be identical, and their similarity would be equal to 1 (i.e., $.5^0$).

According to SINC, similarity of labeled feature patterns could be calculated using Equation 2:

$$Sim(i, j) = W_{\text{label}}^{1-L} S_{\text{vis.attr}}^{N-k} \quad \left\{ \begin{array}{ll} L = 1 & \text{if } L_i = L_j \\ L = 0 & \text{otherwise} \end{array} \right\}. \quad (2)$$

Again, 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, W_{label} denotes values of label mismatches, and L denotes a label match. When there is a label match, $L = 1$ and $W_{\text{label}} = 1$; when there is a label mismatch, $L = 0$ and $W_{\text{label}} < 1$. Note that S and W ($0 \leq S \leq 1$) denote attentional weights of mismatches and that the contribution of S and W is large when these parameters are close to 0 and is small when they are close to 1. This is because the closer the value of these parameters to 1, the smaller the contribution of a mismatch to the detection of difference, whereas the closer the value of the parameters to 0, the greater the contribution to the detection of 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 the model, when neither entity is labeled (i.e., $W_{\text{label}} = 1$), similarity between the 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 weight than are most visual attributes (Sloutsky & Lo, 1999). In the case that the weight of a label does not differ from that of other attributes, the label will become one of the attributes in the computation of similarity, and Equation 2 turns into Equation 1.

Why would labels contribute to similarity? And what might be a mechanism underlying the greater weight of labels for children demonstrated in previous research (e.g., Sloutsky & Lo, 1999)? Although this research does not directly address these issues, several possibilities are worth mentioning here. We briefly return again to these possibilities in the General Discussion section. One possibility is that labels have larger weights because they are presented auditorily, and auditory modality dominates visual modality in infancy and early childhood, but this dominance decreases with age (Lewkowicz, 1988a, 1988b; Sloutsky & Napolitano, 2003). Alternatively, it is possible that the larger weights of labels are grounded in a special status of sounds of human speech (Balaban & Waxman, 1997; Waxman & Markow, 1995).

Finally, SINC 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 choice for one of the test patterns (e.g., Test B) could be predicted using a variant of the Luce's choice rule presented in Equation 3:

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

We argue that if induction and categorization in young children are indeed similarity based, then this model that predicts similarity judgment in young children (e.g., Sloutsky & Lo, 1999) should be able to predict their induction.

However, for the majority of naturalistic visual stimuli patterns, it is impossible to individuate features and calculate feature overlap (e.g., think about photographs of two animals and the multiplicity of perceptual features that they have). At the same time, perceptually rich naturalistic stimuli constitute the most interesting and informative test of the proposed model. Because neither N nor k presented in Equation 1 are determinable a priori for perceptually rich naturalistic stimuli, we made several additional steps to apply the model to naturalistic stimuli. Denoting similarity of Test Stimuli A and B to the target as S^x and S^y, respectively, and performing simple derivations from Equation 3 allow us to get equations predicting categorization and induction performance. First, consider the case when entities are not labeled. By substituting $Sim(T, B)$ and $Sim(T, A)$ with S^x and S^y, we get Equation 4:

$$P(B) = \frac{S^x}{S^x + S^y} = \frac{S^x}{S^x(1 + S^{y-x})} = \frac{1}{1 + \frac{S^y}{S^x}}. \quad (4)$$

For the labeled entities, derivations remain essentially the same, except for the W_{Label} parameter. The parameter equals 1 if there is a label match, otherwise it varies from 0 to 1, and the smaller the value of W, the greater the contribution of label mismatch. Therefore, in the case of labeled entities, the probability of selecting the item that shares the same label (say Test Stimulus B if it shares the label with the target) could be derived as follows:

$$P(B) = \frac{S^x}{S^x + WS^y} = \frac{S^x}{S^x(1 + WS^{y-x})} = \frac{1}{1 + \frac{WS^y}{S^x}}. \quad (5)$$

In short, Equations 4 and 5 should predict participants' induction responses in label and no-label conditions, respectively. In other words, their willingness to induce from Test Stimulus B to the target should be a function of the ratio of S^y/S^x (i.e., of the similarity of Test Stimuli A and B to the target) when no labels are provided, and it should be a joint function of S^y/S^x and W (i.e., the attentional weight of label) when labels are provided. Note that Equation 5 reflects a situation where the target and Test Stimulus B have the same labels and Test Stimulus A has a different label. For the purpose of expository convenience, in the remainder of this article, we refer to the test stimulus sharing the label with the target as *Test B*. However, the equation could be easily modified to reflect a situation in which Test A rather than Test B shares the label with the target. In this case, S^x but not S^y should be multiplied by W. Because we can estimate W from our prior research and S^y/S^x can be measured directly for specific stimuli triads, we can use Equations 4 and 5 for predicting specific probabilities of induction and categorization.

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 induction as well. Predicted probabilities of inducing from Test B to the target under different similarity ratios and under different weights of label (W) are presented in Figure 3. The figure showcases the probability of choosing Test B over Test A as a function of similarity ratio [$S^y = Sim(\text{A, T})/S^x = Sim(\text{B, T})$] and the weight of the label. If the label is ignored (or no label is introduced) and decisions are made solely on the basis of perceptual similarity, the label's weight equals 1 (see $W = 1$ condition). At the same time, if the label is attended to and decisions are made solely on the basis of linguistic label, its weight is very small and its contribution is

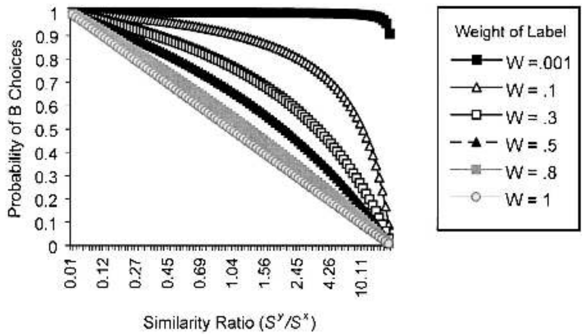


Figure 3. Hypothetical functions: probability of B choice as a function of similarity ratio and the weight of label.

very large (see $W = .001$ condition). The rest of the curves fall between these two extremes. Functions in the graph demonstrate that if the similarity ratio (S^y/S^x) is very small (i.e., the target is much more similar to Test B), the majority of induction responses should be Test B rather than Test A, regardless of the weight of the labels. At the same time, if the ratio is very large (i.e., the target is much more perceptually similar to Test A than to Test B), the majority of induction responses, except for very small weights of the label (e.g., $W = .001$ condition), should be Test A rather than Test B, again regardless of the weight of the labels. Figure 3 also suggests that these functions could be most clearly differentiated when similarity ratios fall between 1 and 9. These considerations determined the experimental approach for Experiments 1–3. In Experiments 4–5, we tested predictions of the model under different stimuli and task conditions.

Overview of Reported Experiments

In what follows, we present six experiments. In Experiment 1, we examined similarity judgments of young children, whereas in Experiment 2, we examined young children's induction and categorization. In these experiments, artificial labels were used. To examine the generalizability of findings of Experiments 1 and 2 to real labels, we introduced labels of animals familiar to participants (e.g., *cat*, *dog*, *rabbit*) in Experiment 3.

In these experiments, participants were presented with triads of stimuli, with each triad consisting of two test stimuli (Test A and Test B) and a target. Similarity ratios S^y/S^x were estimated in a preliminary experiment. These triads were selected from morphed sequences of images, in which one animal was morphed into another in a fixed number of steps, in the manner described in Sloutsky et al. (2001). An example of a morphed sequence is presented in Figure 4. The triads selected from these sequences were subjected to a calibration study, in which 4- to 5-year-olds were asked to estimate the similarity of each of the test stimuli to the target. Those triads that yielded the ratios of $S^y/S^x = 1$, $S^y/S^x = 1.22$, $S^y/S^x = 1.86$, and $S^y/S^x = 9$ were selected for the presented experiments. Examples of each type of triad are depicted in Figure 5.

Predictions of SINC based on these ratios are presented in Figure 6. Solid lines in Figure 6 represent the weights of labels derived from previous research: $W = .35$ when the task is similarity judgment (Sloutsky & Lo, 1999), whereas $W = .1$ when the task is inductive projection (Sloutsky et al., 2001) and $W = 1$ when labels are absent or ignored. Dashed lines exemplify hypothetical scenarios when induction is performed solely on the basis of labels while the similarity ratio is ignored (i.e., $W = .001$) and when labels have weights different from those derived from our previous research (i.e., $W = .5$ and $W = .8$). In short, on the basis of weights of labels estimated from our previous research, the following predictions can be made: The $W = 1$ function predicts similarity judgment, induction, and categorization when no labels are introduced, the $W = .35$ function predicts participants' similarity judgment when entities are labeled, and the $W = .1$ function predicts induction and categorization when entities are labeled.

In Experiment 4, we tested the generality of SINC by testing its predictions on a different set of stimuli, which were used previously by other researchers (Gelman & Markman, 1986, Study 1). Similar to Experiments 1–3, similarity ratios were derived empirically in a separate calibration experiment, and these ratios were used to predict induction performance. Finally, in Experiment 5, we tested qualitative predictions of the similarity-based model and of category-based approaches to induction, examining effects of induction on the recognition memory of children and adults.

Experiment 1: Similarity Judgment

Our goal in this experiment was to test the model's assumption that for young children, linguistic labels contribute to the similarity of compared entities. To achieve this goal, we presented young children and undergraduate students with triads of pictures of animals with similarity ratios varying systematically across triads and asked them to make similarity judgments. In one condition, pictures were accompanied by linguistic labels presented as count

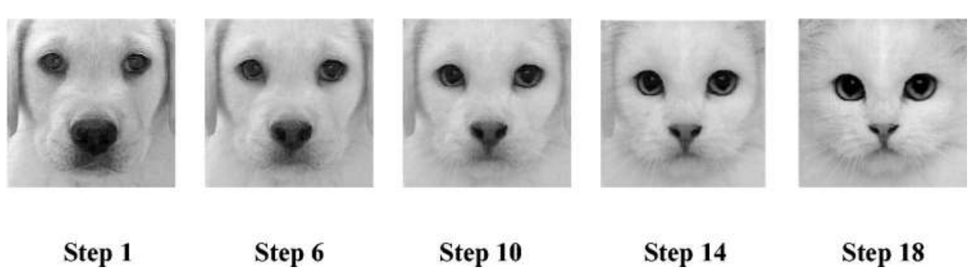


Figure 4. Examples of 5 steps in a 20-step morphing sequence. (These stimuli were presented to participants in color, and a color version of these stimuli is available in the online version of this article, which is part of the PsycARTICLES database.)

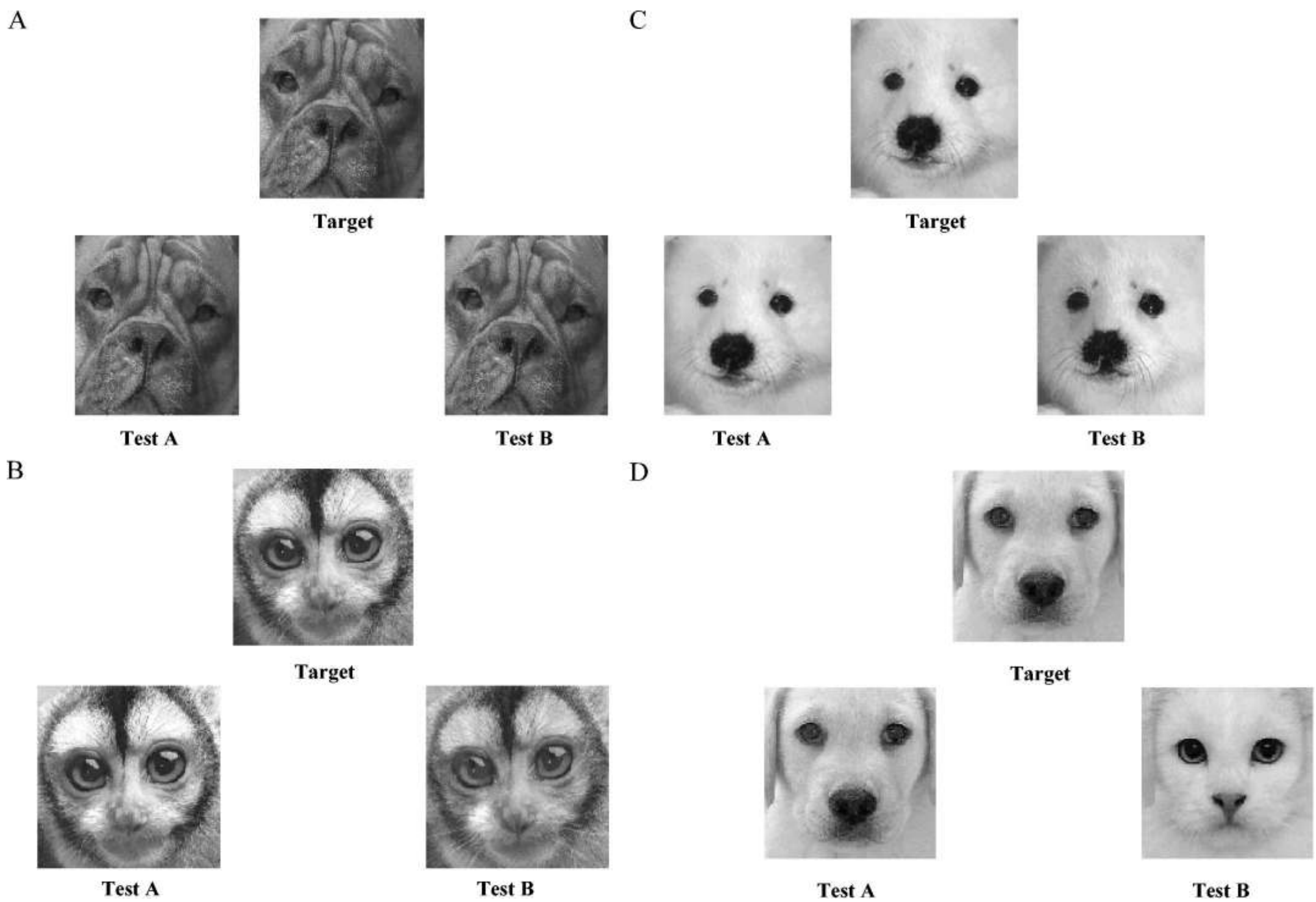


Figure 5. Examples of picture triads used in Experiments 1–3. A: Similarity ratio = 1. B: Similarity ratio = 1.22. C: Similarity ratio = 1.86. D: Similarity ratio = 9. (These stimuli were presented to participants in color, and a color version of these stimuli is available in the online version of this article, which is part of the PsycARTICLES database.)

nouns, such that the more perceptually similar test stimulus (Test A) always had a different label from the target, whereas the less perceptually similar test stimulus (Test B) always had the same label as the target. In another condition, no labels were introduced.

Method

Participants. Participants were 35 preschool children (14 girls and 21 boys; M = 4.58 years, SD = 0.46 years) recruited from several day-care centers located in middle-class suburbs of Columbus, Ohio, and 39 college undergraduates (19 women and 20 men; M = 19.15 years, SD = 2.76 years) from a large midwestern university participating in the experiment in partial fulfillment of a course requirement. In the group of young children, there were 18 participants in the label condition (M = 4.69 years, SD = 0.39 years; range = 4.07–5.37 years) and 17 participants in the no-label condition (M = 4.54 years, SD = 0.48 years; range = 4.07–5.34 years), whereas in the adult group, there were 18 participants in the label condition and 21 participants in the no-label condition.

Design and materials. The experiment had a mixed design, with age and labeling condition (label vs. no label) as between-subject factors and similarity ratio as a within-subject variable. Materials consisted of triads of 10 cm × 10 cm pictures of animal faces. Each triad included the target,

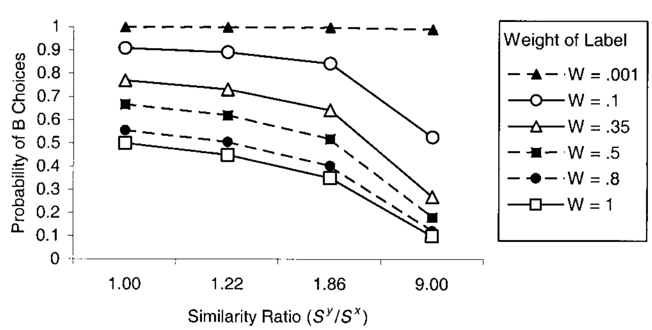


Figure 6. Predicted probabilities of B choices under different task conditions. This figure presents probabilities of B choices when either labels are absent or Test B shares the label with the target. Solid lines represent predictions of the model based on values derived from previous research using similar tasks (W = .1 in induction and categorization tasks, W = .35 for the similarity judgment task, and W = 1 when entities are not labeled). Dashed lines represent possible values of W. If W is very small (e.g., W = .001), decisions are based exclusively on the shared label.

located at the center, and two test stimuli, located equidistantly below the target. In the label condition, pictures were accompanied by artificial two-syllable linguistic labels that were presented as count nouns (e.g., a lolo, a tippy). In each pair, one test stimulus (Test B) had the same label as the target, whereas another (Test A) had a different label.

The picture stimuli were created as follows. First, five pairs of animal pictures were used to create five 20-step sequences, in which one member of a pair was morphed into another member (see Figure 4 for an example of part of morphed sequence). These morphing sequences were created using MorphMan 1.1 (1995) software. Then 32 triads were selected from the set of morphed images, such that one member of the triad was designated as the target and two other members were designated as test stimuli. These triads were subjected to a calibration study to quantify the similarity of each of the test stimuli to the target. A group of twenty 4- to 5-year-old children (10 girls and 10 boys) were presented with the 32 triads of animal faces, 1 triad at a time, and asked whether the target stimulus looked more like Test A or more like Test B. For each triad, the proportion of choices of Test A over Test B averaged across participants was the measure of similarity between Test A and the target (i.e., S^y), whereas its compliment, the proportion of choices of Test B over Test A, was the measure of similarity between Test B and the target (i.e., S^x). On the basis of this calibration study, 16 triads were selected, representing four similarity ratios: (a) $S^y/S^x = 1$, (b) $S^y/S^x = 1.22$, (c) $S^y/S^x = 1.86$, (d) $S^y/S^x = 9$. Each of the four ratios included 4 different triads.

Procedure. A female researcher interviewed each child individually in a quiet room in their day-care center. Undergraduate students were interviewed in a laboratory room on campus. In this experiment and the other experiments, triads of pictures were presented to each participant on a computer screen using specially written software. The computer was also used to record participants' responses: The experimenter recorded these responses by pointing the cursor to a selected item.

In the label condition, the target and test stimuli were labeled auditorily, and participants were asked to repeat these labels. No labels were introduced in the no-label condition. Then participants were asked whether the target looked more like Test A or Test B. Positions of the two test stimuli were randomized across trials, and the order of trials was randomized for each participant. No feedback was given to the participants on their performance.

The important part of the instructions read as follows:

> Now we are going to play a game about animals from faraway places. I am going to show you pictures of those animals, tell you their names, and then ask you to repeat them to me. Then I will ask you a question about those animals. Are you ready to start? Please, look at them. This is a lolo [points to the target]. This is a tippy [points to Test A]. This is a lolo [points to Test B]. Could you please repeat their names? Does this lolo [points to the target] look more like this tippy [points to Test A] or like this lolo [points to Test B]?

Instruction for adults did not make reference to a game. Note that special care was taken to emphasize that the labels were count nouns and not proper names. Each animal was referred to as "a lolo" and as "this lolo." Also note that in the no-label conditions, all stimuli were referred to as "this one."

Before the experimental task proper, participants were presented with two practice trials. These trials (in terms of both stimuli and the task) were identical to the experimental trials, and the sole purpose of these trials was to present participants with practice opportunities. Similar to the experimental trials, no feedback was given to the participants on their performance in the practice trials, and no participant was eliminated from the study on the basis of his or her performance in practice trials. At the end of the experiment, each child received a small toy as a reward for participation (children were not informed in advance that they were going to receive a toy).

Results and Discussion

For young children, proportions of B choices (i.e., the choice of Test B as more similar to the target) by similarity ratio and labeling condition are presented in Figure 7A. Solid lines represent observed values, whereas the dashed line represents predicted values. Data in the figure indicate that both linguistic labels and featural overlap contributed to young children's similarity judgment. These data also point to good fit between proportions of B choices predicted by the model and observed values (a quantitative estimate of the fit is presented in Experiment 3).

Proportions of B choices for the children's group were subjected to a two-way mixed analysis of variance (ANOVA) with labeling condition as a factor and similarity ratio as a repeated measure. The analysis reveals a significant main effect of labeling condition, such that the proportion of B choices in the label condition was significantly higher than that in the no-label condition, $F(1, 32) = 16.63$, $MSE = 0.09$, $p < .0001$. There was also a significant main effect of similarity ratio, $F(3, 96) = 21.14$, $MSE = 0.06$, $p < .0001$. More specifically, the analysis pointed to the following ordinal effects in the proportion of B choices across similarity ratios: (ratio = 1) > (ratio = 1.86) > (ratio = 9) and (ratio = 1.22) > (ratio = 9), all one-tail paired-sample $ts(33) > 2.12$, $ps < .05$. At the same time, the Label Condition × Similarity Ratio interaction did not approach significance, $F(3, 96) < 1$, $p > .75$.

To estimate effect sizes of contributions of linguistic labels and of perceptual similarity to children's similarity judgments, we

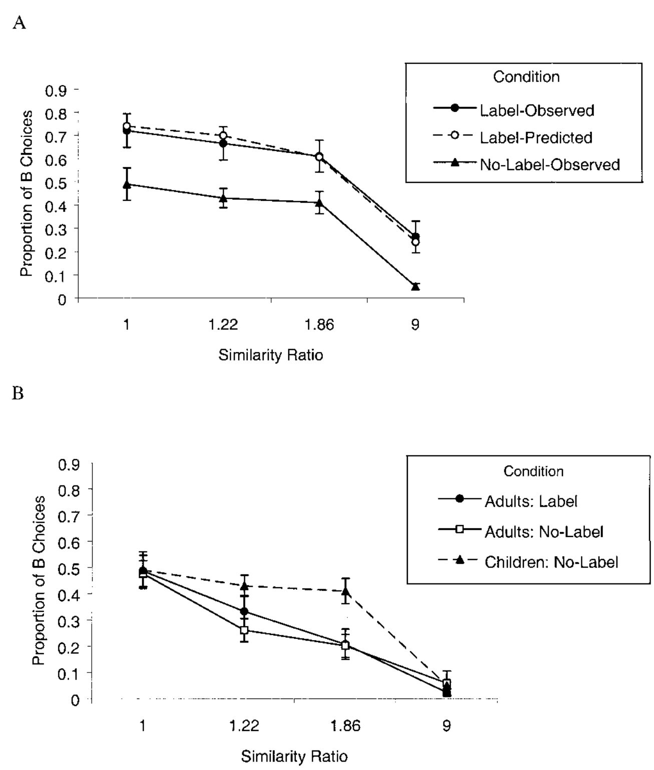


Figure 7. Predicted and observed probabilities of B choices as a function of similarity ratio and labeling in the similarity judgment task, Experiment 1. A: Young children's performance. B: Adults' and young children's performance. Error bars represent standard errors of the mean.

computed Cohen's *d*s of the label condition and of the similarity ratio condition. To calculate *d*s, differences between means for each condition (e.g., $M_{\text{label}} - M_{\text{no label}}$ and $M_{\text{ratio} = 1} - M_{\text{ratio} = 9}$) were divided by the pooled standard deviations (Cohen, 1988). Effects sizes due to both conditions were large, $d_{\text{label}} = 1.17$, $d_{\text{similarity ratio}} = 1.24$, thus suggesting that linguistic labels and perceptual similarity were both important for young children's similarity judgment. Although the contribution of appearance to similarity is hardly surprising, the contribution of linguistic labels to similarity is a nontrivial finding corroborating the assumption of the model. Finally, and most important, there was a close correspondence between similarity judgments predicted by the model and observed similarity responses.

For adults, proportions of B choices by similarity ratio and labeling condition are presented in Figure 7B. First, adults exhibit somewhat better discrimination than children (see the differences between children's and adults' data). These differences are worth mentioning as a note of caution against calibrating on adults stimuli to be used with children. At the same time, the differences are not surprising and should not be a concern here. More important, data in the figure point to nearly identical performance by adults in the label and no-label conditions ($F < 1$, $p > .8$), thus indicating that linguistic labels made no contribution to similarity judgments of adults.

Why do labels contribute to similarity judgments of young children but not of adults? One possibility is that as children grow older, the ability of labels to predict similarity gets degraded. First, in the course of development, children acquire homonyms and homophones (same label accompanying different-looking entities) and synonyms (different labels accompanying similar-looking entities). Second, they get exposed to concepts that exhibit high variability and low similarity in their appearance (e.g., "fans of the Chicago Bulls" or "furniture"), in which the same label accompanies different-looking entities. As a result of these developments, attentional weights of labels for similarity decrease. Therefore, because labels do not contribute to similarity for adults, SINC in its current form is limited to young children. Further discussion of developmental changes underlying these differences will be presented in the General Discussion section.

It is worth noting that the present findings should be distinguished from the line of research on acquired equivalence and distinctiveness (e.g., Katz, 1963; Norcross & Spiker, 1957; Spiker & Norcross, 1962; see also Hall, 1991, for a review). These studies demonstrated that discrimination learning can be facilitated when to-be-discriminated visual stimuli are accompanied by different linguistic labels. Unlike these studies, our experiments did not include learning. Furthermore, incidental learning seems unlikely because (a) participants were presented with a large number of different stimuli, (b) the order of trials was randomized for each participant, and (c) effects of similarity ratios did not decrease in the label condition as compared with the no-label condition. Therefore, the perception of greater similarity among entities sharing the label did not seem to be acquired in the course of learning; rather, it stemmed from an online computation of similarity.

Of course, one may counterargue that young children's responses could have stemmed from their inability to ignore task-irrelevant information. If one assumes that for both young children and adults, labels are irrelevant for the task of similarity judgment, then children should have greater difficulty ignoring labels. Therefore, there is little surprise that children cannot ignore labels when making similarity judgments but adults can. This possibility was addressed and eliminated by Sloutsky and Lo (1999). In particular, these researchers presented children with a variety of task-irrelevant cues (e.g., differently colored dots and different gestures) that accompanied pictures. Similar to labels, the cue of one of the test items was the same as the cue of the target, whereas another test item had a different cue. It was found that only labels and not other cues contributed to similarity judgments of young children.

It could be also argued that the results of Experiment 1 could have stemmed (a) from children's confusion of "looking like" and "being like" or (b) from young children having difficulty inhibiting their knowledge of what things are when asked similarity questions.[1] Although the wording of the question (i.e., whether the target looked more like Test A or more like Test B) was clear and explicit, we deemed it necessary to address this issue directly.

We therefore conducted Experiment 1A, in which young children were presented with conflict and no-conflict triads. In the conflict (or experimental) triads, perceptual similarity was pitted against the ontological kind. For example, in one triad, the target was a lemon, Test A was a tennis ball that looks similar to the lemon, and Test B was a bunch of grapes that are of the same ontological kind as the lemon (see Appendix for triad pictures). In the no-conflict (or baseline) triads, there was no conflict between appearance and ontological kind (e.g., the target was a bird, with one test stimulus being a similar-looking bird and another being a table). For both types of triads, the task was to select the test item that looked more like the target. To ascertain that young children knew what the items were and could label the items, we conducted a naming experiment with a separate group of young children. It was argued that if in Experiment 1, young children failed to understand the similarity question or failed to inhibit their knowledge, then in Experiment 1A, young children should exhibit confusion when presented with conflict triads. At the same time, if young children clearly understand the similarity question, they should reliably select the similar-looking test item in both no-conflict and conflict triads.

Experiment 1A: Control for Understanding the Similarity Question

Method

Participants. Participants were 40 preschool children (23 girls and 17 boys), with 22 participants in the naming condition ($M = 4.84$ years, $SD = 0.41$ years; age range = 4.01–5.49 years) and 18 participants in the similarity judgment condition ($M = 4.85$ years, $SD = 0.55$ years; age range = 4.02–5.52 years). The participants were recruited from several day-care centers located in middle-class suburbs of Columbus, Ohio.

Materials and procedure. Materials were 36 color photographs of natural kind and artificial objects. All pictures were organized into six conflict (or experimental) and six no-conflict (or baseline) triads. In the experimental triads, the target was perceptually similar to Test A and of the same ontological kind as Test B. In the baseline triads, the target was both perceptually similar to Test A and of the same ontological kind as Test A. The triads were selected in a preliminary calibration experiment: Only

[1] We thank an anonymous reviewer for suggesting this possibility.

those items that were correctly labeled consistently by young children were selected. Triads used in Experiment 1A are presented in the Appendix.

In the naming condition, children were presented with 12 triads, one at a time, and asked to label each item. After naming all the items in a triad, children were asked an additional question probing their knowledge about whether depicted objects belonged to different ontological kinds. For Triads 1, 5, 6, and 11, children were asked to identify the ones "that a person can eat"; for Triads 2, 4, 7, 8, 9, 10, and 12, they were asked to identify the ones "that can grow"; and for Triad 3, they were asked to identify the ones "that can tell time" (see each triad in the Appendix). For each triad, the correct answer was to select all items that had the property in question. In the similarity judgment condition, children were presented with the same triads as in the naming condition and were asked whether the target looked more like Test A or more like Test B.

All children were tested by female hypothesis-blind experimenters in a quiet room in their day-care centers. In both conditions, stimuli were presented to participants in random order on the screen of a laptop computer.

Results and Discussion

In the naming condition, children exhibited high accuracy in both naming the objects (93% correct for the experimental triads and 96% correct for the baseline triads) and answering questions about them correctly (95% and 94%, respectively). Therefore, all items used in Experiment 1A were familiar to children, and children knew what these items were.

In the similarity judgment condition, 3 participants produced fewer than 5 correct responses out of 6 baseline trials, and their data were excluded from the analysis. The rest of the participants were very accurate on the baseline trials, averaging 92% correct on their responses. Participants were also very accurate on the experimental trials, making on average 89% of correct choices (selecting perceptually similar Test A over the ontologically similar Test B). Not only were participants performing better than chance on both baseline and experimental trials, both ts > 14.4, ps $< .0001$, their performance on the experimental trials was not statistically different from performance on the baseline trials, which could be considered a functional ceiling, $F(1, 17) = 2.8$, $p > .1$.

Experiment 1A clearly demonstrated that children are capable of distinguishing the question about objects' identities from the question about objects' appearances: The majority of the children correctly labeled every object presented to them and were highly accurate in answering questions about the ontological membership of objects, yet this knowledge of what things were did not prevent children from performing well on the similarity judgment task. These results strongly indicate that effects of labeling on similarity reported in Experiment 1 did not stem from children's misinterpretation of the similarity question or their inability to inhibit their prior knowledge.

Having supported the SINC model's assumption regarding labels and similarity in Experiments 1 and 1A, we conducted Experiment 2, in which we tested predictions of the model for induction and categorization tasks.

Experiment 2: Categorization and Induction

Our goal in Experiment 2 was to test quantitative predictions of the model: Overall similarity, computed over linguistic labels and perceptual attributes, should accurately predict young children's categorization and induction. The task was similar to the task used in Experiment 1, except that children were asked categorization and induction questions. Similar to Experiment 1, in one condition, pictures were accompanied with count nouns, whereas in another condition, no labels were introduced. It was predicted that in the label condition, categorization and induction judgments would be a function of perceptual similarity and linguistic label.

Method

Participants. Participants were preschool children recruited from several day-care centers located in middle-class suburbs of Columbus, Ohio. A group of 71 children (32 girls and 39 boys) participated in this experiment. There were 41 participants in the induction condition ($M = 5.2$ years, $SD = 0.4$ years; age range = 4.1–5.6 years) and 30 participants in the categorization condition ($M = 5.1$ years, $SD = 0.5$ years; age range = 4.1–5.6 years).

Design and materials. The experiment had a mixed design, with task (categorization vs. induction) and labeling condition (label vs. no label) as between-subject factors and similarity ratio as a within-subject variable. Materials included the same 16 picture triads and 32 artificial labels used in Experiment 1. In addition, 32 biological properties were introduced. The list of names and properties used can be found in Table 1.

Procedure. The overall procedure was similar to that in Experiment 1 except that Experiment 2 included induction and categorization tasks. For the induction task, children were presented with picture triads (see Figure 5 for an example), which in the label condition were accompanied by auditorily presented linguistic labels, such that the less perceptually similar test stimuli shared labels with the target. Children were told that each of the test stimuli had a certain biological property and were asked which of these properties are likely to be shared by the target. These properties were introduced after participants were familiarized with the pictures and labels. The important part of the instructions in this condition read as follows:

> This is a lolo [points to the target]. This is a tippy [points to Test A]. This is a lolo [points to Test B]. Could you please repeat their names? This tippy [points to Test A] has blue blood inside his body. What does he have inside his body? This lolo [points to Test B] has green blood inside his body. What does he have inside his body? What does this lolo [points to the target] have inside his body? Does he have blue blood like this tippy [points to Test A] or green blood like this lolo [points to Test B]?

Table 1
Labels and Biological Properties Used in Experiments

Artificial labels used in Experiment 2	Biological properties used in Experiments 2 and 3
A lolo, a tippy	Blue blood, green blood
A vila, a saba	Square heart, round heart
A gaga, a mosso	Warm blood, cold blood
A poffy, a reepa	Big brain, small brain
A bolo, a mekky	Red bubbles, blue bubbles
A jaja, a sita	Small heart, big heart
A lopa, a zizi	Small bubbles, big bubbles
A tibo, a reeny	Yellow bones, red bones
A lono, a veeda	Square bubbles, round bubbles
A zoony, a luko	Short bones, long bones
A tiggy, a cula	Brown brain, green brain
A kebo, a yama	White heart, green heart
A meeta, a gapo	Big muscles, small muscles
A boto, a daza	Red muscles, blue muscles
A toky, a googa	Blue brain, red brain
A meega, a reefo	Green muscles, yellow muscles

Another task was categorization, in which children were asked whether the target was the same kind of animal as Test A or as Test B. The important part of the instructions in the categorization condition read as follows:

> This is a lolo [points to the target]. This is a tippy [points to Test A]. This is a lolo [points to Test B]. Could you please repeat their names? Is this lolo [points to the target] the same kind of animal as this tippy [points to Test A] or this lolo [points to Test B]?

Note that in the no-label condition of both the induction and the categorization tasks, the tests and target were referred to as "this one." Similar to Experiment 1, experimental trials were preceded by two practice trials. At the end of the experiment, each child received a small toy as a reward for participation (children were not informed in advance that they were going to receive a toy).

Results and Discussion

Overall results of this experiment along with values predicted by the model are presented in Figure 8. Dashed lines represent predicted values, whereas solid lines represent observed values. The figure points to several important regularities.

First, for both categorization and induction, the proportion of B choices is affected by linguistic labels and by similarity ratios. Proportions of B choices in induction and categorization tasks were subjected to two separate two-way mixed ANOVAs, with labeling condition as a factor and similarity ratio as a repeated measure. For induction, the analyses point to a significant main effect of labeling condition, with the proportion of B choices in the label condition being significantly higher than that in the no-label condition, $F(1, 39) = 25.65$, $MSE = 0.15$, $p < .0001$. There was also a significant main effect of similarity ratio, $F(3, 117) = 9.20$, $MSE = 0.06$, $p < .0001$. Planned comparisons pointed to the following ordinal effects in the proportion of B choices across similarity ratios: (ratio = 1) > (ratio = 1.2) = (ratio = 1.86) > (ratio = 9) for all differences, one-tail paired-sample $ts(40) > 1.8$, $ps < .05$. At the same time, the Label Condition × Similarity Ratio interaction did not approach significance, $F(3, 117) < 1$, $p > .75$.

For categorization, the analysis points to a significant main effect of labeling condition, with the proportion of B choices in the label condition being significantly higher than that in the no-label condition, $F(1, 28) = 29.81$, $MSE = 0.13$, $p < .0001$. There was also a significant main effect of similarity ratio, $F(3, 84) = 18.12$, $MSE = 0.06$, $p < .0001$. Planned comparisons pointed to the following ordinal effects in the proportion of B choices across similarity ratios: (ratio = 1) = (ratio = 1.2) > (ratio = 1.86) > (ratio = 9) for all differences, one-tail paired-sample $ts(29) > 2.4$, $ps < .01$. At the same time, the Label Condition × Similarity Ratio interaction was not significant, $F(3, 84) = 1.43$, $p > .2$.

In the induction task, effect sizes due to both conditions were large, $d_{\text{label}} = 1.24$, $d_{\text{similarity ratio}} = 0.9$. In the categorization task, effect sizes due to both conditions were also large, $d_{\text{label}} = 1.15$, $d_{\text{similarity ratio}} = .98$. These analyses indicate that both linguistic labels and perceptual similarity were important for young children's similarity judgments, with labels having somewhat larger effects on induction and categorization performance than perceptual similarity.

The analyses also indicate that induction and categorization tasks elicit very similar performance, $r_{\text{induction–categorization}}(8) = .97$, $p < .01$. In addition, for both label and no-label conditions, the model rather accurately predicts induction and categorization performance (a quantitative estimate of the fit is presented in Experiment 3). And finally, both induction and categorization strongly correlate with similarity judgments presented in Experiment 1, $r_{\text{similarity induction}}(8) = .84$ and $r_{\text{similarity categorization}}(8) = .89$, both $ps < .05$. Therefore, it seems reasonable to conclude that for young children, induction and categorization performance is driven by the overall similarity computed over linguistic labels and visual similarity.

However, it could be argued that these findings are limited to artificial labels and cannot be generalized to induction with real categories denoted by familiar labels. For example, Davidson and Gelman (1990) argued that category labels may affect children's induction only when the labels map onto coherent, familiar categories. These authors argued that familiar labels, such as *cows*, are more likely to support induction than novel or artificial labels, such as *zavs*. To examine the applicability of the SINC model to familiar labels, we conducted Experiment 3.

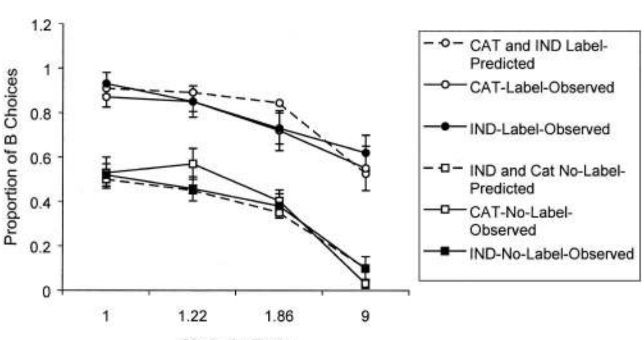


Figure 8. Predicted and observed probabilities of B choices as a function of similarity ratio and labeling in induction and categorization tasks, Experiment 2. IND = induction; CAT = categorization. Error bars represent standard errors of the mean.

Experiment 3: Induction and Categorization With Realistic Labels

This experiment was similar to Experiment 2, except that instead of artificial labels, real labels that were familiar to the children were used.

Method

Participants. Participants were preschool children recruited from several day-care centers located in middle-class suburbs of Columbus, Ohio. A group of 31 children participated in this experiment. There were 15 participants in the induction condition ($M = 5.21$ years, $SD = 0.52$ years; age range = 4.26–5.67 years; 9 girls and 6 boys), and 16 participants in the categorization condition ($M = 5.50$ years, $SD = 0.36$ years; age range = 4.88–5.84 years; 8 girls and 8 boys).

Materials and procedure. Stimuli were identical to those used in Experiments 1 and 2, with one important difference: Real animal names were used instead of artificial labels. The tasks were identical to the induction and categorization tasks in Experiment 2, except that the task included only the label condition. The stimuli set for this experiment (as well as for Experiments 1–2) included pictures of two different cats, three different dogs, a bear, a seal, a lion, a chipmunk, and an angwantibo. We assumed that children would be unfamiliar with some of these animals, and therefore we deemed it necessary to substitute names of unfamiliar animals with other real names that children were familiar with. In so doing, our goal was to avoid a conflict between pictures and labels, such that the majority of children would identify a picture as X, whereas label Y would be used in the experiment.

To achieve this goal, we conducted a calibration experiment in which a group of eighteen 4.5- to 5.5-year-old children (none of whom participated in Experiments 1, 2, or 3) were asked to name pictures of the animals used in experimental triads. Results of the calibration experiment are presented in Table 2. The table presents the percentage of correct naming responses (out of 18), other labels used by children to name depicted animals, and labels selected for Experiment 3.

As shown in Table 2, when the majority of children correctly identified an animal (e.g., by identifying a dog as "a dog"), we used this label when referring to this animal. However, when the majority of children did not correctly identify an animal (e.g., a chipmunk was identified as "a squirrel," "a cat," or "a rat"), we used the real label that the majority of children used to name that animal (i.e., "a squirrel"). If there was no clear tendency (e.g., a lion was referred to as "a tiger," "a cat," or "a cheetah"), the experimenter selected a word that should be well-known by young children (e.g., "a lion"), with familiarity of the words checked against several databases (e.g., Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; MRC Psycholinguistic Database, 1997).

Results and Discussion

Proportions of B choices by similarity ratios and label types (artificial labels from Experiment 2 and realistic labels from the current experiment) are presented in Figure 9. The overall pattern of results is quite similar to that in the label condition of Experiment 2. In the categorization condition (Figure 9A), the proportion of B choices when the children were given realistic labels varied with the similarity ratio, $F(3, 45) = 6.1$, $MSE = 0.05$, $p < .001$. Paired comparison pointed to the following differences in the proportion of B choices: (ratio = 1) > (ratio = 1.86) > (ratio = 9) and (ratio = 1.22) > (ratio = 9) for all differences, one-tailed $ts(15) > 1.9$, all $ps < .05$. At the same time, when performance observed in Experiment 3 was compared with that observed in Experiment 2, there were no differences in proportions of B choices between artificial and real labels, $F(1, 29) < 1$.

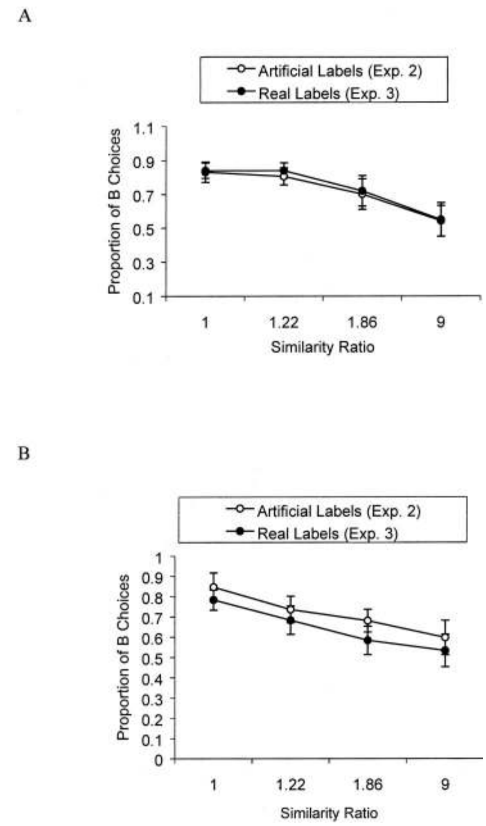


Figure 9. Probability of B choices as a function of similarity ratio with artificial labels (Experiment 2) and real labels (Experiment 3). A: Categorization condition. B: Induction condition. Exp. = Experiment. Error bars represent standard errors of the mean.

In the induction condition (Figure 9B), the proportion of B choices also varied with the similarity ratio, $F(3, 42) = 6.23$, $MSE = 0.03$, $p < .001$. Paired comparison pointed to the following differences in the proportion of B choices: (ratio = 1) > (ratio = 1.86) > (ratio = 9) and (ratio = 1.22) > (ratio = 9) for all differences, one-tailed $ts(14) > 3$, all $ps < .01$. At the same time, similar to the categorization condition, when performance observed in Experiment 3 was compared with that observed in Experiment 2, there were no differences in proportions of B choices between artificial and real labels, $F(1, 32) = 1.13$, $p > .29$.

A scatter plot of values predicted by the SINC model for similarity judgment, induction, and categorization (for both artificial and realistic labels) is presented in Figure 10. Results point to a good fit between predicted and observed values: Across experiments, predicted values correlated highly with observed values, $r(28) = .96$, with approximately 91% of observed variance explained by SINC. Note that if instead of weights of labels derived from our previous experiments, we plug $W = .001$ into the model (recall that $W = .001$ if decisions are made overwhelmingly on the basis of linguistic labels), the quantitative fit of the model decreases markedly, with R^2 dropping from .91 to .64. Therefore, it could be concluded that the SINC model proposing that labels

Table 2
Label Selection for Experiment 3: Picture Naming by Children and Selected Labels

Depicted animal	% of correct responses	Other labels used by children	Labels selected for Experiment 3
Dog 1	100.00		Dog
Cat 1	100.00		Cat
Dog 2	27.77	Raccoon, bear, squirrel, fox, skunk, baboon	Dog
Angwantibo	0.00	Snake, frog, owl, mouse, cat, bat, lizard, monkey	Raccoon
Bear	22.22	Dog, kitten, seal	Bear
Seal	0.00	Bear, fox, cat, puppy	Rabbit
Dog 3	94.44	Hippo	Dog
Lion	61.11	Tiger, cat, cheetah	Lion
Cat 2	94.44	Tiger	Cat
Chipmunk	0.00	Squirrel, monkey, gorilla, cat, rat, opossum	Squirrel

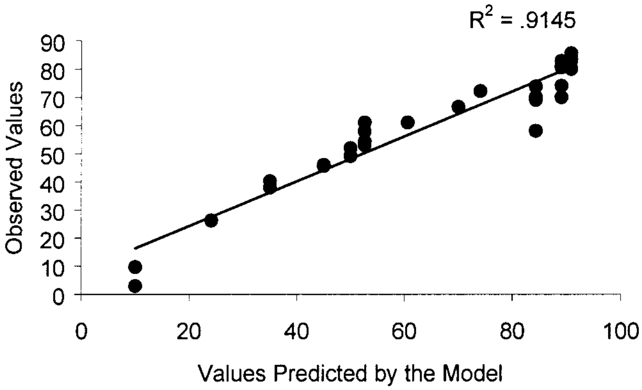


Figure 10. Overall fit of the SINC (similarity, induction, and categorization) model: proportions of B choices predicted by the model and observed in Experiments 1–3.

make a partial (and quantifiable) contribution to induction and categorization is more accurate than an alternative that assumes that labels are the sole contributor. Furthermore, the fact that appearances and labels make comparable contributions to categorization and induction casts doubt on the causal essentialist claim that linguistic labels are more central than appearances. We return to this issue in the General Discussion section.

Taken together, results of Experiments 1–3 indicate that (a) for young children, matching labels contribute to the similarity of compared items; (b) categorization and induction in young children are driven by the overall similarity computed over perceptual and labeling information, and SINC accurately predicts young children's similarity judgment, induction, and categorization; and (c) induction and categorization performance of young children does not differ for artificial and realistic labels.

However, a critical test of the generality of the SINC model would be its ability to predict performance with stimuli that differ from ours in two respects. First, would the model predict performance with a broad range of natural kind pictures, or is it limited to human and animal faces? And second, would the model be able to predict performance with stimuli where both depicted animals and their labels are somewhat familiar to children? For example, there is a set of stimuli used in a Gelman and Markman (1986) study in which children were presented with stimuli that differed from our stimuli in these two respects. Therefore, we conducted Experiment 4, which tested SINC's predictions using Gelman and Markman's (1986) stimuli.

Experiment 4: The Ability of the SINC Model to Predict Performance With Earlier Used Stimuli

The reported experiment tested predictions of the SINC model on Gelman and Markman's (1986) stimuli depicting a subset of living things. First, we conducted a calibration similarity judgment experiment affording the estimation of similarity ratios. We then conducted an induction experiment yielding observed probabilities that could be compared with probabilities predicted by the model. The present induction study was identical to that of Gelman and Markman (1986, Study 1), except that in the reported experiment, stimuli presentation was controlled by computer. In our experiment, the calibration similarity judgment and induction tasks were administered by different hypothesis-blind researchers.

Method

Participants. Participants were preschool children recruited from several day-care centers located in middle-class suburbs of Columbus, Ohio. A group of fifty-nine 4- to 5-year-olds (27 girls and 32 boys; $M = 4.51$ years, $SD = 0.31$ years; age range = 4.0–5.0 years) participated in a calibration similarity judgment task. Another group of 31 children (18 girls and 13 boys; $M = 4.43$ years, $SD = 0.30$ years; range = 4.0–5.0 years) participated in the induction task.

Materials. In the experiment, we used Gelman and Markman's (1986, Study 1) stimuli. These stimuli included a set of 10 triads of pictures, with each picture measuring approximately 7 cm × 4 cm. Triads and respective labels are presented in Figure 11 (note that the labels were presented auditorily by the experimenter and did not appear on screen). All picture triads were presented on the computer screen of a portable computer (Dell Inspiron or Dell Latitude) in a random order. Presentation of stimuli was controlled by Super Lab Pro 2.0 (1999) software. In the induction task, pictures were accompanied by auditorily presented labels, and biological properties were introduced for each of the test stimuli. The linguistic labels and biological properties were those used by Gelman and Markman (1986, Study 1). No labels were introduced in the similarity judgment (i.e., calibration) task.

Procedure. A female hypothesis-blind researcher interviewed children individually in a quiet room in their day-care center. At the end of the experiment, each child was praised and received a small prize for participating in the experiment (children were not informed in advance that they were going to receive a toy).

In the similarity judgment task, each triad appeared on screen, and the child was asked which of the two items on the top (i.e., the test items) looked more like the item on the bottom (i.e., the target). On each trial, items were referred to as "this one," and the order of introduction of test items was randomized across trials.

In the induction task, the procedure was similar to that described by Gelman and Markman (1986, Study 1), except that in our reported experiment, picture triads were presented on a computer screen, and the researcher recorded participants' responses by pressing a button corresponding to a participant's response. Triads were presented one at a time and were followed by the labeling of each item in the presented triad. Then the participant was asked to repeat these labels. Finally, participants were told that each of the test stimuli had a certain biological property and were asked to repeat these properties (e.g., "This bird gives its baby mashed-up food. What does it give its baby?"). When properties were repeated successfully, the participant was asked which of these properties was likely to be shared by the target. After the child answered the induction question, the experimenter said "Okay" and proceeded to the next trial. The experiment included one practice trial that was introduced prior to experimental trials. The practice trial was identical to experimental trials, except that data were not recorded.

Results and Discussion

Recall that the goal of this experiment was to estimate the similarity of each of the test stimuli to the target for stimuli used by Gelman and Markman (1986), calculate similarity ratios, and derive estimates of young children's induction performance by plugging these ratios into the model. The induction experiment was conducted to test predictions of the model for induction responses. The similarity of each of the test stimuli to the target derived from children's responses and SINC's predictions for induction performance are presented in Table 3.

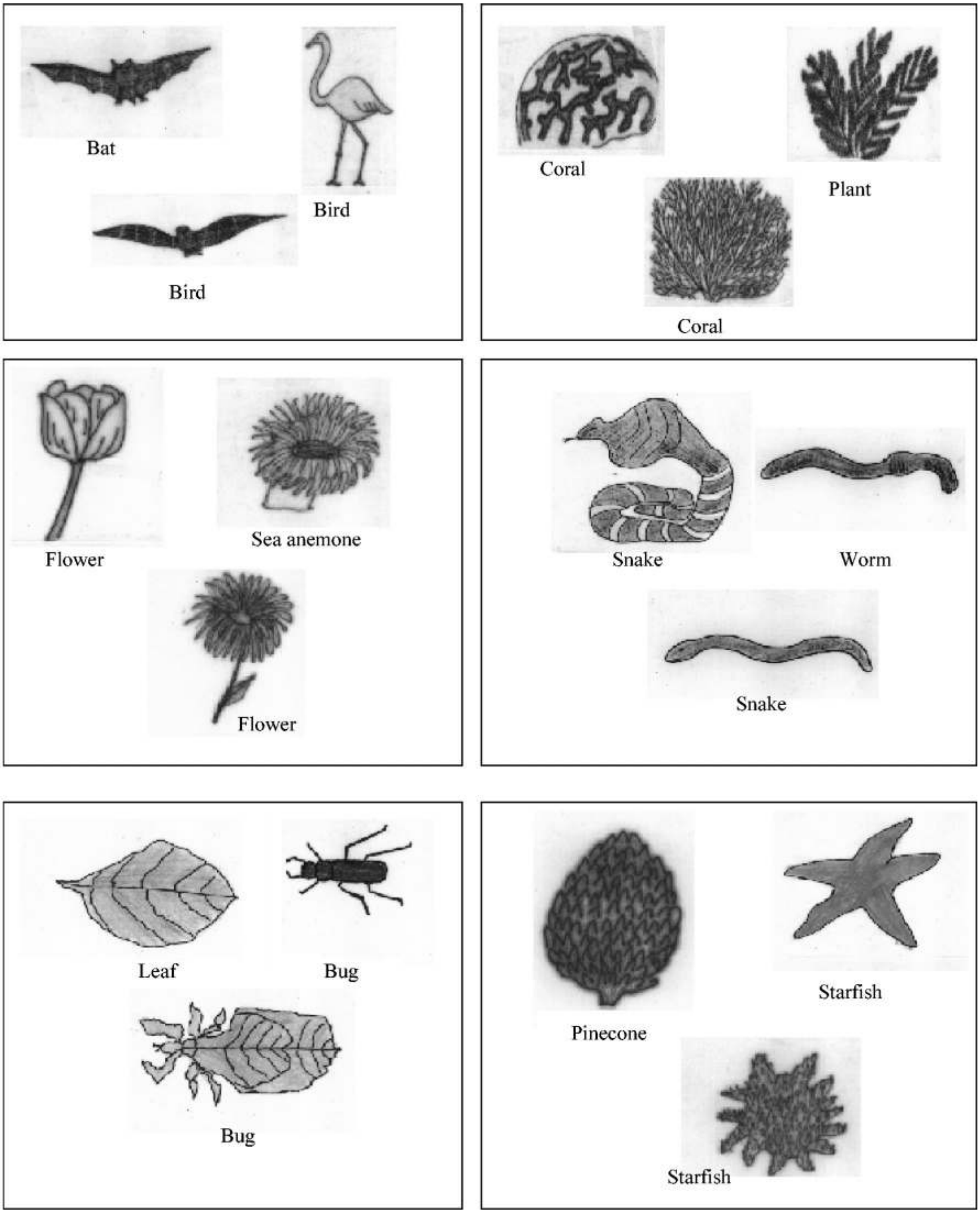


Figure 11. Triads of pictures and respective labels used in Experiment 4. These stimuli were kindly provided by Susan A. Gelman. (These stimuli were presented to participants in color, and a color version of these stimuli is available in the online version of this article, which is part of the PsycARTICLES database.)

Note that for the entire set presented in Table 3, SINC predicts a bimodal distribution of induction responses: When Test A is markedly more similar to the target than Test B is and the similarity ratio is large (i.e., in the 11–19 range), SINC predicts that only 40% to 50% of induction responses will be label based. At the same time, when Test A is just somewhat more similar to the target than Test B is or even less similar and the similarity ratio is small (i.e., in the 0.7–1.5 range), SINC predicts that over 70% of induction responses will be label based. Note that these predictions differ markedly from Gelman

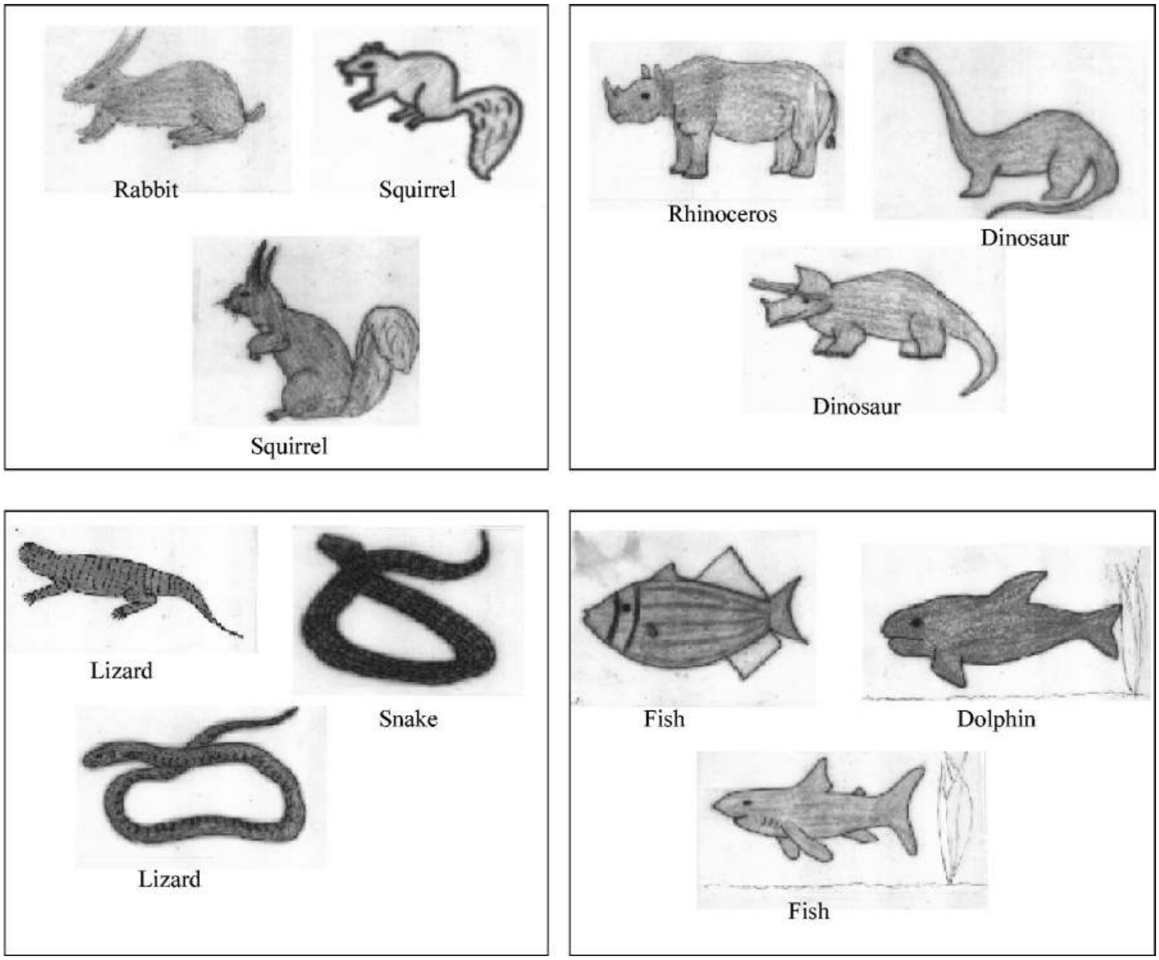


Figure 11 (continued)

and Markman's (1986) findings that young children induce properties from Test B reliably above chance and their conclusion that category membership, rather than appearance, drives young children's induction responses.

Predicted and observed proportions of B choices across stimuli triads are presented in Figure 12. First, as predicted by the model for this stimuli set, there was a bimodal distribution of induction responses. Incidentally, Gelman and Markman (1986) reported

Table 3
Stimuli Triads Depicted in Figure 11, Similarity of Each of the Test Stimuli to the Target, Similarity Ratios, and the SINC Model's Predictions for Induction Across Identically Labeled Items (B Choices), Experiment 4

Triad (target, Test A, Test B)	S^y = Sim(A, target)	S^x = Sim(B, target)	Similarity ratio (S^y/S^x)	Predicted probability of B choice
Bird, bat, bird	.92	.08	11.5	.47
Coral, plant, coral	.54	.46	1.2	.89
Flower, sea anemone, flower	.92	.08	11.5	.47
Snake, worm, snake	.88	.12	7.3	.58
Bug, leaf, bug	.56	.44	1.3	.88
Starfish, pinecone, starfish	.44	.56	0.79	.92
Squirrel, rabbit, squirrel	.76	.24	3.2	.75
Dinosaur, rhinoceros, dinosaur	.92	.08	11.5	.47
Lizard, snake, lizard	.95	.05	19	.34
Fish, dolphin, fish	.88	.12	7.3	.58
Average	.78	.22	3.5	.63

Note. SINC = similarity, induction, and categorization.

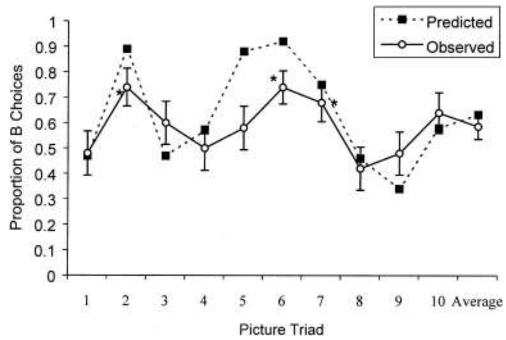


Figure 12. Predicted and observed proportions of B choices by stimuli triads in the induction task, Experiment 4. Predictions were made using the SINC (similarity, induction, and categorization) model. Triad 1: bird–bat–bird; Triad 2: coral–plant–coral; Triad 3: flower–sea anemone–flower; Triad 4: snake–worm–snake; Triad 5: bug–leaf–bug; Triad 6: starfish–pinecone–starfish; Triad 7: squirrel–rabbit–squirrel; Triad 8: dinosaur–rhinoceros–dinosaur; Triad 9: lizard–snake–lizard; Triad 10: fish–dolphin–fish. Error bars represent standard errors of the mean. *—above chance, $p < .05$.

data averaged across the 10 triads, thus treating children's responses as if they were uniform (which, according to our data, they are not). As shown in the figure, SINC accurately predicts proportions for 9 out of 10 triads (the model failed to predict responses for the bug–leaf–bug triad). Even with this triad being inaccurately predicted, the overall fit of the model is good, likelihood ratio $\chi^2(9, N = 382) = 2.174$, $p = .99$, with the model accounting for approximately 77% of observed variance in induction responses ($R^2 = .77$). When the bug–leaf–bug triad is removed, the model yields even tighter fit, likelihood ratio $\chi^2(8, N = 336) = 1.1$, $p = .99$, with the model accounting for approximately 88% of observed variance in induction responses ($R^2 = .88$). Note that the alternative model that assumes that induction is performed overwhelmingly on the basis of labels (i.e., $W = .001$) predicts only 3% of the observed variance ($R^2 = .03$).

Several important findings stem from the reported experiment. First, the proposed model makes very accurate qualitative predictions of young children's induction performance with stimuli used in previous experiments (e.g., Gelman & Markman, 1986), accurately predicting performance on 9 out of 10 triads, and it makes reasonably accurate quantitative predictions. Second, the SINC model that suggests reliance on both perceptual information and labels makes markedly more accurate predictions than a competitor assuming reliance solely on linguistic labels ($R^2 = .88$ vs. $R^2 = .03$).

Third, both predicted and observed values differ from those reported by Gelman and Markman (1986, Study 1). Recall that Gelman and Markman (1986, Study 1) reported that on average, Test B (i.e., the item that shared the label with the target) was selected in 68% of responses, which is above chance, whereas in our experiment, the average was 59%, which did not differ significantly from chance, although, because of the bimodality of induction responses, triad-by-triad analysis seems more informative than the average. In particular, although some of the triads (e.g., the coral–plant–coral triad) generate above-chance label-based inductions, many others (including the often-cited bird–bat–bird triad) generated only chance responding (e.g., predicted and observed values for the bird–bat–bird triad were .47 and .48, respectively). Note that only triads with small similarity ratios (i.e., those where the same-label choice was at least as similar to the target as the different-label choice was) generated above-chance label-based induction. Therefore, on the basis of these item-by-item analyses, it would be inaccurate to conclude that children's inductions are driven by category membership rather than by similar appearance.

In short, results of Experiment 4 corroborated predictions of the model and findings of Experiments 1–3 that linguistic labels and perceptual similarity make comparable contributions to young children's induction. These results do not support the causal essentialist claim that linguistic labels are more central than appearances. Finally, the results of Experiment 4 indicated that predictions of the proposed model are not limited to stimuli used in Experiments 1–3.

Experiments 1–4 provided a quantitative test for the SINC model and its ability to account for data across various tasks. However, it could be argued that a variety of alternative models considering a combination of appearance and label information would make similar quantitative predictions. Although some of these models would have difficulty accounting for the effects of labels on perceived similarity, we deemed it necessary to derive additional qualitative predictions from the similarity-based model, predictions that cannot be accounted for by an alternative class of models: those assuming young children's reliance on taxonomic knowledge in the course of induction.

Experiment 5: Pitting Similarity-Based Against Category-Based Induction

Recall that similarity-based and category-based induction assume different kinds of processing in the course of induction, with the former positing induction as a function of similarity among compared entities and the latter positing induction as a function of identification of a common category among compared entities. Are there any observable outcomes with respect to which of these approaches generates qualitatively different predictions? We believe that one such outcome is memory: Similarity-based induction and category-based induction may result in different memories.

If induction is category-based, participants would first categorize items, thus forming a category-based or "gist" representation. Because perceptual details of each individual item are irrelevant for the task, they would fail to encode these details or encode them poorly. As a result, when presented with a surprise recognition task, they would rely on the gist representation, erroneously recognizing critical lures, or new items from studied categories (Brainerd, Reyna, & Forrest, 2002; Koutstaal & Schacter, 1997). If, however, induction is similarity-based, participants would compute similarity between the target and each of the test items, thus processing the items perceptually and forming item-specific representations. As a result, they would correctly discriminate between studied items and critical lures, accepting the former and rejecting the latter. Therefore, similarity-based induction is likely to result in accurate memories, whereas category-based induction is likely to lead to memory distortions, such as poor discrimination of presented and nonpresented members of presented categories.

We specifically predict that unlike adults exhibiting memory distortions as a result of induction tasks, young children should exhibit accurate memory for individual items. If a difference in representation underlies differential memory accuracy in children and adults, then explicitly teaching young children to perform category-based induction may result in an increase in false recognition of critical lures and a decrease of their memory accuracy to the level of adults. In fact, these predictions have received initial support (Sloutsky & Fisher, in press), and the goal of the reported study was to replicate these highly surprising findings and to further examine effects of induction on recognition memory in young children.

The former prediction was tested in Experiment 5A, in which we examined effects of induction on recognition memory of children and adults. The latter prediction was tested in Experiment 5B: Young children were trained to perform category-based induction, and the effects of this training on their recognition were examined.

Experiment 5A

Method

Participants. Participants were 61 young children (M = 5.3 years, SD = 0.3 years; age range = 4.8–5.8 years) and 57 introductory psychology students (M = 20.8 years, SD = 1.6 years) from a large midwestern university participating in the experiment in partial fulfillment of a course requirement. Young children were recruited from middle-class and upper-middle-class suburban neighborhoods of Columbus, Ohio, on the bases of returned parental consent forms.

Materials, design, and procedure. Materials were 44 color photographs of animals presented against a white background. During the study phase, participants were presented with a set of 30 pictures of animals, 1 picture at a time, from three different categories (10 cats, 10 bears, and 10 birds). During the recognition phase, they were presented with 28 pictures of animals, 1 picture at a time, and were asked whether they had seen that exact picture during the study phase. Half of these pictures were presented during the study phase, and the other half were new pictures. The new pictures included 7 control items representing novel categories that were not presented during the study phase and 7 critical lures, new items from the studied categories. These recognition animals represented three categories: cats (7 of which were studied items and 7 were critical lures), bears (all 7 of which were old), and squirrels (that were control items, all 7 of which were new). To ascertain that all of these animals were well-known by children and that young children could name the depicted animals, we presented these pictures to young children in an earlier naming study. Only those pictures that were consistently named by a basic level name (i.e., "cat," "bear," "bird," or "squirrel") by more than 85% of the children were selected for this study.

The experiment included three between-subject conditions: baseline, induction, and blocked categorization, with each condition consisting of a study phase and a recognition phase. The recognition phase was identical in all three conditions, whereas the study phase differed across the conditions. Participants were randomly assigned to one of the three conditions.

In the study phase of the baseline condition, participants were presented with 30 pictures of animals, 1 picture at a time, and their task was to remember these pictures as accurately as possible for a subsequent recognition test. Participants were explicitly informed about the recognition test.

In the study phase of the induction condition, participants were first presented with a picture of a cat and informed that the animal had "beta cells inside its body." Participants were then presented with 30 pictures of animals (identical to those presented in the baseline condition), 1 picture at a time. Participants were asked whether each of the presented animals also had beta cells inside its body. After responding, the participant was provided with "yes/no" feedback indicating that only cats, not bears or birds, had beta cells inside their bodies. The recognition test was not mentioned in the study phase of this condition. After the induction phase, participants were presented with a surprise recognition task.

In the study phase of the blocked categorization condition, participants were first presented with a picture of a cat and informed that the animal was young. After that, participants were presented with 30 pictures of animals (identical to those presented in the baseline and induction conditions), 1 picture at a time. Participants were asked whether each of the presented animals was young or mature. Participants were provided with random "yes/no" feedback. The purpose of this random feedback was to block inferences based on the animal kind information and to force both children and adults to focus on perceptual features of individual items. The recognition test was not mentioned in the study phase of the latter two conditions. Similar to the induction condition, after the induction phase, participants were presented with a surprise recognition task.

The recognition task was presented immediately after the study phase. Pictures were presented for recognition one at a time, in a self-paced manner. Participants were asked to determine whether each picture presented during the recognition phase was old (i.e., exactly the same as the one presented during the study phase) or new. No feedback was provided in the recognition phase of the experiment.

Young children were tested individually in their day-care centers by hypotheses-blind female experimenters. All stimuli were presented to children on the screen of a laptop computer, and all instructions were presented verbally by the experimenters. Undergraduate students were tested individually in a laboratory on campus and had all stimuli and instructions presented to them on a computer screen. For all participants, the experiment was controlled by Super Lab Pro 2.0 (1999) software.

Results and Discussion

Several participants (2 children and 6 adults in the induction condition, 1 adult in the blocked categorization condition, and 5 children and 3 adults in the baseline condition) did not reliably reject control items, and their data were excluded from the analysis. In addition, two children were not different from chance in performing induction, and their data were not included in the analyses.

Recall that young children were expected to perform induction by comparing each animal with the target animal and thus to remember study phase animals well, accurately accepting old animals and rejecting new ones. At the same time, it was expected that adults would spontaneously categorize animals when performing induction and form gist or category-level memory traces, and, as a result, they would poorly discriminate between old members of the target category and critical lures.

After several trials, the majority of young children and adults realized that the property of having beta cells should be induced to cats but not to bears or birds, and they accurately performed this induction: The average rate of correct induction was over 80% for both children and adults. Also, across conditions, participants exhibited high recognition accuracy in accepting old targets (71% and 82% for children and adults, respectively) and rejecting distractors from an unstudied category (92% and 94%, respectively). These data indicate that participants took the task seriously and paid attention to stimuli during the study phase.

Recall that three categories of items were presented during the study phase, and, in this section, we focus on participants' discrimination of old items and new items belonging to the studied categories. Recognition memory accuracy (i.e., hits − false alarms [FA]) by age group and condition are presented in Figure 13. Data in the figure indicate that although children exhibited equivalently high accuracy (i.e., hits − FA) across the conditions (i.e., .28 in the induction condition, .21 in the blocked categorization condition,

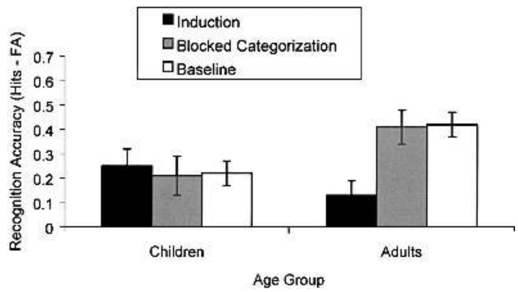


Figure 13. Recognition memory accuracy in children and adults in the induction, blocked categorization, and baseline conditions, Experiment 5A. FA = false alarms. Error bars represent standard errors of the mean.

and .22 in the baseline condition), $F(2, 49) < .3$, $p > .8$, the accuracy of adults in the induction condition (hits − FA = .13) was dramatically lower than their accuracy in the other two conditions (.41 and .42, respectively), $F(2, 47) = 7.4$, $MSE = 0.058$, $p < .003$, both ps $< .01$ on post hoc Tukey tests.

It is interesting to note that in the induction condition, young children exhibited higher accuracy than adults (.28 vs. .13), although this difference does not reach significance. However, if the accuracy of adults in the baseline condition (i.e., accuracy > .4) represents a functional ceiling for this task, then there are more young children than adults exhibiting this near-ceiling accuracy in the induction condition (36% vs. 6%), $\chi^2(1, N = 31) = 4.4$, $p < .05$.

In short, the induction task markedly attenuated adults' recognition accuracy, whereas young children remained accurate. Furthermore, the induction task resulted in a greater proportion of highly accurate young children than highly accurate adults. These results suggest that although adults performed category-based induction, young children performed similarity-based induction.

However, it could be argued that children were accurate across conditions because of extraneous factors. For example, children could have been more interested in the pictures than were the adults, or children could have forgotten gist information faster, while adults could have forgotten item-specific information faster. The goal of Experiment 5B was to eliminate these explanations by training children to perform category-based induction. If our hypothesis is correct, this training should differentially affect young children's memory in the induction and baseline conditions: Although their accuracy should drop in the induction condition (analogous to the drop for adults in Experiment 5A), it should not be negatively affected by training in the baseline condition.

Experiment 5B

Method

Participants. Participants in Experiment 5B were 36 young children ($M = 5.19$ years, $SD = 0.22$ years; range = 4.9–5.7 years). These participants were selected in the same way as were participants in Experiment 5A.

Materials, design, and procedure. The design of Experiment 5B included two between-subject conditions: induction and baseline. Participants were randomly assigned to one of the two conditions. Materials in both conditions were identical to those in Experiment 5A. The procedure of this experiment differed from that of Experiment 5A in that prior to the study and recognition phases, 5-year-olds were presented with training in which they were taught to perform category-based induction. They were first taught that animals that have the same name belong to the same category (i.e., that "they are the same kind of animal"). Children were then presented with pictures of lions, rabbits, and dogs (none of these categories were presented during the main experiment). All presented pictures were subjected to a naming study prior to this experiment that revealed that each of the depicted animals could be reliably named by 5-year-olds. Children were given three boxes, with each box identified by a black outline of a lion, a rabbit, or a dog, and were told that animals that have the same name are the same kind of animal, and they could be placed in the same box. They were then presented with six categorization trials. On each trial, they were given a picture of a lion, a rabbit, or a dog and asked to put the picture in a box that had animals of the same kind. Pictures were placed in boxes face down. On each trial, participant's response was followed by "yes/no" feedback. Both types of feedback were accompanied by an explanation that animals that have the same name are the same kind of animal and should be placed in the same box.

The categorization training task was followed by an induction training task, in which children were told that a particular animal had a particular biological property referring to the insides of an animal (e.g., "this dog has thick blood inside its body"). After being reminded that animals that have the same name are the same kind of animal and further told that animals of the same kind have "the same stuff inside," participants completed six induction trials. On each trial, they had to put the picture in a box where other animals may also have the same property. Each trial was followed by "yes/no" feedback. Feedback was accompanied by an explanation that animals of the same kind have the same name and same stuff inside. All children completed training successfully, giving either 5 out of 6 correct answers or 4 correct answers in a row in the induction training task. At the conclusion of the training session, each child was praised and told that now he or she knew that "animals that have the same name are the same kind of animal, and these animals have the same stuff inside." They were then presented with the main experiment that consisted of the study and recognition phases. During the study phase, participants were presented with either the induction or the baseline condition. These conditions were identical to those in Experiment 5A.

Results and Discussion

Across conditions, participants exhibited high recognition accuracy in accepting old targets and rejecting distractors from an unstudied category (84% and 98%, respectively). In the induction condition, participants were highly accurate in inducing the property in question: The average rate of correct induction was over 90%. These data indicate that participants took the task seriously and paid attention to stimuli during the study phase.

Recognition accuracy (hits − FA) is presented in Figure 14.

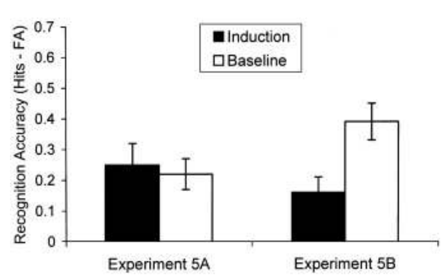


Figure 14. Pre- and posttraining recognition memory accuracy in children in the induction and baseline conditions, Experiments 5A and 5B. FA = false alarms. Error bars represent standard errors of the mean.

Data in the figure point to marked differences between young children's high accuracy in the baseline condition (hits − FA = .39) and their lower accuracy in the induction condition (hits − FA = .16), $F(1, 34) = 8.57$, $MSE = 0.057$, $p < .01$. Although accuracy in the baseline condition increased compared with that in Experiment 5A, accuracy in the induction condition dropped to the level of adults in the induction condition of Experiment 5A. In addition, in the induction condition, the proportion of highly accurate children dropped compared with this proportion in Experiment 5A (11% versus 36%), $\chi^2(1, N = 33) = 3.63$, $p = .058$.

These findings support the hypothesis indicating that in Experiment 5A, young children, unlike adults, did not perform category-based induction, and as a result they exhibited greater discrimination between old items and critical lures than adults did. Taken together, results of Experiments 5A and 5B support the notion that in the absence of training, young children perform similarity-based induction, while challenging the notion that spontaneous induction in young children is category based.

General Discussion

The reported experiments revealed several important findings. First, as predicted by the SINC model, linguistic labels contribute to the overall similarity among compared stimuli. Second, the model of similarity accurately predicts categorization and induction performance of young children. Third, there are high correlations among similarity judgment, induction, and categorization across experiments. Fourth, young children exhibited very similar performance when entities were accompanied by artificial labels or by realistic labels. And finally, the analysis of young children's memory indicates that they perform induction in a similarity-based rather than in a category-based manner.

At the broadest level, the reported experiments support quantitative and qualitative predictions of the SINC model, which explains induction and categorization performance in young children without assuming their reliance on taxonomic knowledge. The model fits the data well, accounting for approximately 91% of the variance across Experiments 1–3 and for more than 77% of the variance in Experiment 4, which used stimuli different from those used in Experiments 1–3. Furthermore, findings of Experiment 5 indicate that SINC is capable of generating qualitative predictions that pose challenges to the category-based approach. These results have several theoretical implications, and we discuss some of these implications in the next sections.

The Role of Linguistic Labels in Categorization and Induction

It has been well established that labels contribute to induction and categorization; however, the manner in which they contribute remains contested. Current research brings new evidence to this debate: (a) The contribution of labels is quantitative and limited and (b) matching labels contribute to similarity among compared entities. Both findings seem to dispute the causal essentialist claims of the centrality of labels.

Although attentional weights of linguistic labels (i.e., the overall contribution of labeling information to similarity, induction, and categorization decisions) vary across tasks and these weights could be larger than weights of any single visual attribute, the overall contribution of appearance is not much smaller than the contribution of labels (recall that both effects sizes were comparably large). Therefore, it is difficult to agree with the causal essentialist claims that "names convey . . . all the properties that go along with category membership" (Gelman & Coley, 1991, p. 184) and that for young children, labels represent causally central information.

In addition, there is mounting evidence that the weights of labels are not fixed; rather, they can vary depending on the visual information they compete against. First, reliance on linguistic labels may be affected by the quality of visual information (see Jones & Smith, 1993, for a related discussion). In particular, Deak and Bauer (1996) demonstrated that relative contributions of appearances and labels to young children's induction and categorization differ when entities are perceptually rich, three-dimensional objects and when entities are perceptually impoverished line-drawing pictures. Hence, if stimuli are perceptually impoverished (or visual stimuli are substituted by their verbal descriptions), then effects of visual information may decrease while effects of labels may increase, thus leaving the impression that visual information is of secondary importance. It seems that such nonfixedness of the importance of labels runs counter to the idea of their fixed, causally central role.

Second, reliance on linguistic labels may be affected by the type of perceptual information. For example, Mak and Vera (1999) presented 3- and 4-year-olds with induction tasks where patterns of motion were pitted against the same linguistic label. Participants were asked to induce a property (e.g., the ability to see in the dark) from a target animal to a test animal that shared a pattern of motion with the target or shared a label with the target. Researchers found that young children were more likely to rely on similarity of motion than on the matching linguistic label. Again, the importance of labels exhibits variability that seems to undermine the idea of the causal centrality.

Third, it has been recently demonstrated that after participating in an experiment where matching labels were poor predictors of biological properties while similar appearances were good predictors, young children ignored labels in favor of appearances in a different induction task presented 3 months later by a different experimenter, which was not the case for adolescents (Sloutsky & Spino, in press). The fact that young children retained outside the training situation the increased importance of appearances (that are supposed to be peripheral) and the decreased importance of labels (that are supposed to be to be theoretically central) casts further doubt on the idea that children assume the centrality of labels.

In short, the fact that the quantity, quality, and type of perceptual information affect the reliance on labels seriously undermines the essentialist idea of causal centrality because properties of peripheral information should not affect the centrality of causally central features. A more low-level attentional mechanism underlying induction and categorization is a parsimonious alternative to the causal-essentialist position, and this alternative deserves serious consideration.

In particular, recent findings that young children are more likely to process auditory than visual information (Sloutsky & Napolitano, 2003) may indicate that the somewhat larger attentional weights of labels may stem from their auditory presentation. It is also possible that sounds of human speech may further amplify these attentional effects (cf. Balaban & Waxman, 1997).

Although the contribution of linguistic labels to similarity has been corroborated, the precise mechanisms underlying this contribution remain unknown. On the one hand, it is possible that labels are considered attributes of objects and that they contribute to similarity directly, in a manner similar to other attributes, such as size, color, or shape. Alternatively, it is possible that labels contribute to similarity indirectly by facilitating the comparison of identically labeled entities (Namy & Gentner, 1999, 2002). Under this view, the comparison results in the alignment of feature structures, and it leads to greater attention to perceptual and conceptual attributes that correlate with labels. In short, in the case of direct contribution, labels attract much attention to themselves, whereas in the case of indirect contribution, they attract attention to corresponding attributes rather than to themselves.

The Role of Similarity in Categorization and Induction

Recall that we considered two distinct approaches to induction and categorization in young children: one that assumes the existence of taxonomic knowledge and another that does not make this assumption. The fact that the model of similarity accounts for induction and categorization supports the similarity-based approach to categorization and induction in young children. As predicted by the SINC model, both linguistic labels and perceptual similarity systematically contribute to young children's induction. Reported results support the contention of SINC that categorization and induction in young children are similarity-based processes, and two aspects of similarity seem to be critically important: (a) relationships between the overall similarity and induction and (b) the contribution of labels to the overall similarity.

First, not only did the proposed similarity-based model accurately predict performance on categorization and induction tasks (in fact, it was markedly more accurate than a competitor model assuming an overwhelming reliance on label), but also both induction and categorization strongly correlated with similarity judgments studied in Experiment 1. Therefore, it seems reasonable to conclude that for young children, induction and categorization performance is driven by the overall similarity computed over linguistic labels and visual similarity, and that the same process of computation of similarity may underlie similarity judgments, categorization, and induction in young children (see Hampton, 1997, 2001, for the discussion of the role of similarity in adults' categorization).

In addition to these predictions, SINC can also predict performance with different stimuli sets: The model accurately predicted children's performance on the Gelman and Markman (1986) task. In particular, when we derived similarity ratios for the 10 living-thing triads used by Gelman and Markman (1986) in an experiment with young children, fit the ratios into the model, and conducted induction experiment with another group of young children, the model quite accurately predicted induction performance averaged across the 10 triads as well as performance for 9 out of 10 individual triads. Not only did the SINC model account for more than 80% of the variance in children's responses, it also fared markedly better than a competitor model that assumes an overwhelming reliance on labels, with the competitor model explaining 3% of the variance.

Induction and Categorization: What Develops?

Recall that SINC predicts young children's but not adults' performance on induction and categorization tasks. What changes in the course of cognitive development? Although current research does not address these issues, we deem it necessary to provide some tentative considerations.

We believe that the similarity-based induction and categorization depicted by the SINC model hinges on similarity computed over visual and linguistic cues. Recall that although linguistic labels contributed to perceived similarity in young children, they did not contribute to perceived similarity in adults. Similar findings were obtained in earlier experiments using different stimuli (Sloutsky & Lo, 1999). Therefore, it seems that in its present form, the model predicts performance of only those participants whose similarity is affected by linguistic labels; that is, it predicts performance of young children and not older participants. In fact, there is evidence (e.g., Sloutsky et al. 2001) that older participants tend to ignore perceptual similarity in favor of labels when performing induction.

It seems plausible that the SINC model accurately describes young children's similarity, categorization, and induction because young children process complex information in a holistic manner without attending to specific dimensions of stimuli (Shepp, 1978; Smith, 1989; although see Ward et al., 1989, for diverging evidence). Young children's perceived similarity is based on an aggregate of multiple sources: Young children are more likely to notice overall similarity across several dimensions, such as similar shape and similar color, than they are to notice the identity on a single dimension, such as the same color (Smith, 1989). At the same time, adults exhibit dimensional processing, such that they can attend to a single cue, even if this cue is no more salient than competing cues, and they often have knowledge of which cues are more predictive than others. In addition, older participants' induction is affected by their knowledge of feature centrality, feature stability, and feature variability (Medin & Shoben, 1988; Rips, 1989); causal status of available features (Ahn et al., 2000); and the importance of different features in different contexts (Heit & Rubinstein, 1994).

In short, in the course of development, children acquire requisite capacity to selectively attend to a separate dimension, feature, or cue (Shepp, 1978; Shepp & Swartz, 1976). In addition, they acquire information that degrades the predictive power of perceptual similarity: They encounter objects that are different in spite of a similar appearance (e.g., bat and bird or dolphin and shark). They also acquire knowledge of homonyms and synonyms, degrading the contribution of labels to similarity. Finally, in the course of school-based learning, they acquire domain-specific knowledge of predictive properties in various knowledge domains, as well as get training in performing categorization based on "deep" theoretical properties. Such training may lower the predictive value of appearances and increase the predictive value of more deep properties. Results of Experiment 5B suggest that it is possible to train children to perform induction in a category-based manner; however, young children may not perform category-based induction spontaneously.

Therefore, even if similarity plays an important role in older participants' induction (e.g., Hampton, 1997, 2001; Osherson et al., 1990; Sloman, 1993, 1998), there are at least two important developmental changes to be considered: (a) selective attention to a small number of important cues and (b) knowledge of which cues are more central in particular contexts and knowledge domains.

Limitations of the SINC Model and Further Directions

The proposed model makes accurate predictions, which are limited, however, for certain classes of stimuli. Recall that the model's predictions are based on the similarity ratio S^x/S^y, and the same ratio can be derived from an infinitely large number of absolute values of S^x and S^y. The model is limited to situations in which each of the test stimuli is somewhat similar to the target, considering possible perceptual variability in the universal set. Because of this limitation, the model cannot be applied to situations in which both test stimuli are very dissimilar from the target. For example, a triad consisting of a yellow jellyfish (target), yellow tractor (Test A), and red carpet (Test B) may yield a similarity ratio based on the same color, but it is doubtful that this similarity ratio would be used for induction or categorization decisions. It has been demonstrated that such out-of-place feature matches (the yellow color of jellyfish and a tractor) have small effects on absolute similarity (Goldstone, 1994b). Similarly, a triad having a similarity ratio equal to 1—for example, a white cloud (target), white bird (Test A), and white box (Test B)—with the box sharing the label with the cloud, would hardly elicit label-based categorization and induction responses. It is more likely that children would interpret the labels as homonyms and balk at the question. Therefore, we must restrict the model to (a) entities representing the same ontological kind (animals, plants, or artifacts) and (b) entities having at least some degree of absolute similarity, as measured, for example, by a same–different immediate recognition task. Entities that have nearly no similarity are likely to be placed into different "folders" by the perceptual system, and, unless attention is explicitly attracted to a particular dimension (e.g., "look, this one is yellow and this one is yellow"), these entities are not likely to be compared spontaneously.

A number of issues require future research. In particular, it remains unclear how children weigh even high perceptual similarity for items belonging to different ontological kinds. In particular, the fact that the SINC model failed to predict children's responses for the bug–leaf–bug triad may suggest that they were using knowledge of animate–inanimate distinction. However, when presented with this triad, young children were not discounting perceptual similarity either. Unfortunately, Gelman and Markman's (1986) task that uses one trial per triad does not allow us to establish whether some children were relying on ontological kind and some on perceptual similarity or whether all children were relying on both sources of information.

Another issue that requires future research is whether young children may differentially weigh different features for different types of to-be-inferred properties. There is evidence that such differential weighting is present in adults (e.g., Heit & Rubinstein, 1994), and it would be interesting to examine this issue in young children. However, given that effect sizes due to different to-be-inferred properties were small even in adults, this possibility seems rather remote.

There are several interesting extensions of the SINC model that should be considered in the future. First, it is important to establish whether a label is a feature. If so, is it a categorical "on–off" feature, or is it a quantitative one? If the former is the case, then regardless of similarity among labels, any deviation from a perceived identity of labels should constitute a mismatch. However, if the latter is the case, then the similarity of nonmatching labels should contribute to the overall similarity of compared entities. Although the latter possibility appears less likely than the former, indirect evidence supports the latter possibility: Phonological similarity of discriminable words (e.g., *dog* and *wog*) was found to affect young children's mappings of words onto referents (Merriman & Marazita, 1995).

Second, as mentioned above, there could be a trade-off between the weights of labels and those of perceptual attributes, and if such a trade-off is firmly established, it should be included in the model. Third, it would be informative to test the model by using a wider range of categories and properties and by having greater variability in linguistic input (e.g., using not only nounlike but also adjectivelike and verblike labels).

Finally, it would be interesting to extend the SINC model to situations in which participants learn predictive values of various cues. One possibility would be to provide participants with feedback about predictive values of different cues and to examine predicted and observed changes in the weights of these cues. Several models are applicable to learning situations and could be potentially informative, including Rescorla-Wagner (see Hall, 1991, for an extensive review), ALCOVE (attention learning covering map; Kruschke, 1992), or network models (e.g., Gluck & Bower, 1988; McClelland & Rogers, 2003). Another possibility would be to include analogy-based learning (Namy & Gentner, 1999, 2002), in which the process of comparison rather than feedback drives learning about predictive cues.

Conclusion

The proposed model, SINC, offers a similarity-based account of categorization and induction in young children. Six experiments using a wide range of tasks and stimuli corroborated predictions of the model, indicating that (a) for young children, linguistic labels contribute to similarity among compared entities, and (b) categorization and induction in young children are similarity-based processes.

References

Ahn, W., Gelman, S. A., Amsterlaw, J. A., Hohenstein, J., & Kalish, C. W. (2000). Causal status effect in children's categorization. *Cognition*, *76*, B35–B43.

Ahn, W., Kalish, C. W., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., et al. (2001). Why essences are essential in the psychology of concepts. *Cognition*, *82*, 59–69.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.

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.

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development*, *73*, 1363–1377.

Cohen, J. (1988). *Statistical power analysis*. Hillsdale, NJ: Erlbaum.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *8*, 240–247.

Cycowicz, Y., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237.

Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development*, *5*, 151–176.

Deak, G. O., & Bauer, P. J. (1996). The dynamics of preschoolers' categorization choices. *Child Development*, *67*, 740–767.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

Fisher, R. (1951). *The design of experiments*. New York: Hafner. (Original work published 1935)

Gelman, S. A. (2000). The role of essentialism in children's concepts. *Advances in Child Development and Behavior*, *27*, 55–98.

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

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the obvious. *Cognition*, *38*, 213–244.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247.

Goldstone, R. L. (1994a). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125–157.

Goldstone, R. L. (1994b). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 3–28.

Hall, G. (1991). *Perceptual and associative learning*. New York: Oxford University Press.

Hampton, J. A. (1997). Associative and similarity-based processes in categorization decisions. *Memory & Cognition*, *25*, 625–640.

Hampton, J. A. (2001). The role of similarity in natural categorization. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 13–28). New York: Oxford University Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 411–422.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development*, *8*, 113–139.

Katz, P. A. (1963). Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology*, *66*, 423–428.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 103–135.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555–583.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*, 155–171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: II. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology*, *24*, 172–182.

Mak, B. S. K., & Vera, A. H. (1999). The role of motion in children's categorization of objects. *Cognition*, *71*, B11–B21.

Malt, B. (1994). Water is not H_2O. *Cognitive Psychology*, *27*, 41–70.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*, 310–322.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. *Cognition*, *59*, 199–217.

Medin, D. L. (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 Press.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*, 158–190.

Merriman, W. E., & Marazita, J. M. (1995). The effect of hearing similar-sounding words on young 2-year-olds' disambiguation of novel reference. *Developmental Psychology*, *31*, 973–984.

MorphMan 1.1. (1995). [Computer software]. Moscow: Stoik Corporation.

MRC Psycholinguistic Database. (1997). [Database]. Retrieved April 2001 from http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm

Namy, L. L., & Gentner, D. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487–513.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*, 5–15.

Norcross, K. J., & Spiker, C. C. (1957). The effects of type of stimulus pretraining on discrimination performance in preschool children. *Child Development*, *28*, 79–84.

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.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Orthony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Rips, L. J. (2001). Necessity and natural categories. *Psychological Bulletin*, *127*, 827–852.

Shepp, B. E. (1978). From perceived similarity to dimensional structure: A new hypothesis about perceptual development. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 135–167). Hillsdale, NJ: Erlbaum.

Shepp, B. E., & Swartz, K. B. (1976). Selective attention and the processing of integral and nonintegral dimensions: A developmental study. *Journal of Experimental Child Psychology, 22*, 73–85.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, *35*, 1–33.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*, 246–251.

Sloutsky, V. M., & Fisher, A. V. (in press). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*.

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 judgment. *Developmental Psychology*, *6*, 1478–1492.

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

Sloutsky, V. M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*, 822–833.

Sloutsky, V. M., & Spino, M. A. (in press). Naïve theory and transfer of learning: When less is more and more is less. *Psychonomic Bulletin & Review*.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*, 125–144.

Smith, L. B., & Heise, D. (1992). Perceptual similarity and conceptual structure. In B. Burns (Ed.), *Percepts, concepts and categories: The representation and processing of information* (pp. 233–272). Amsterdam, the Netherlands: North-Holland.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*, 143–171.

Spiker, C. C., & Norcross, K. J. (1962). Effects of previously acquired stimulus names on discrimination performance. *Child Development*, *33*, 859–864.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition*, *74*, 149–175.

Super Lab Pro 2.0. (1999). [Computer software]. San Pedro, CA: Cedrus Corporation.

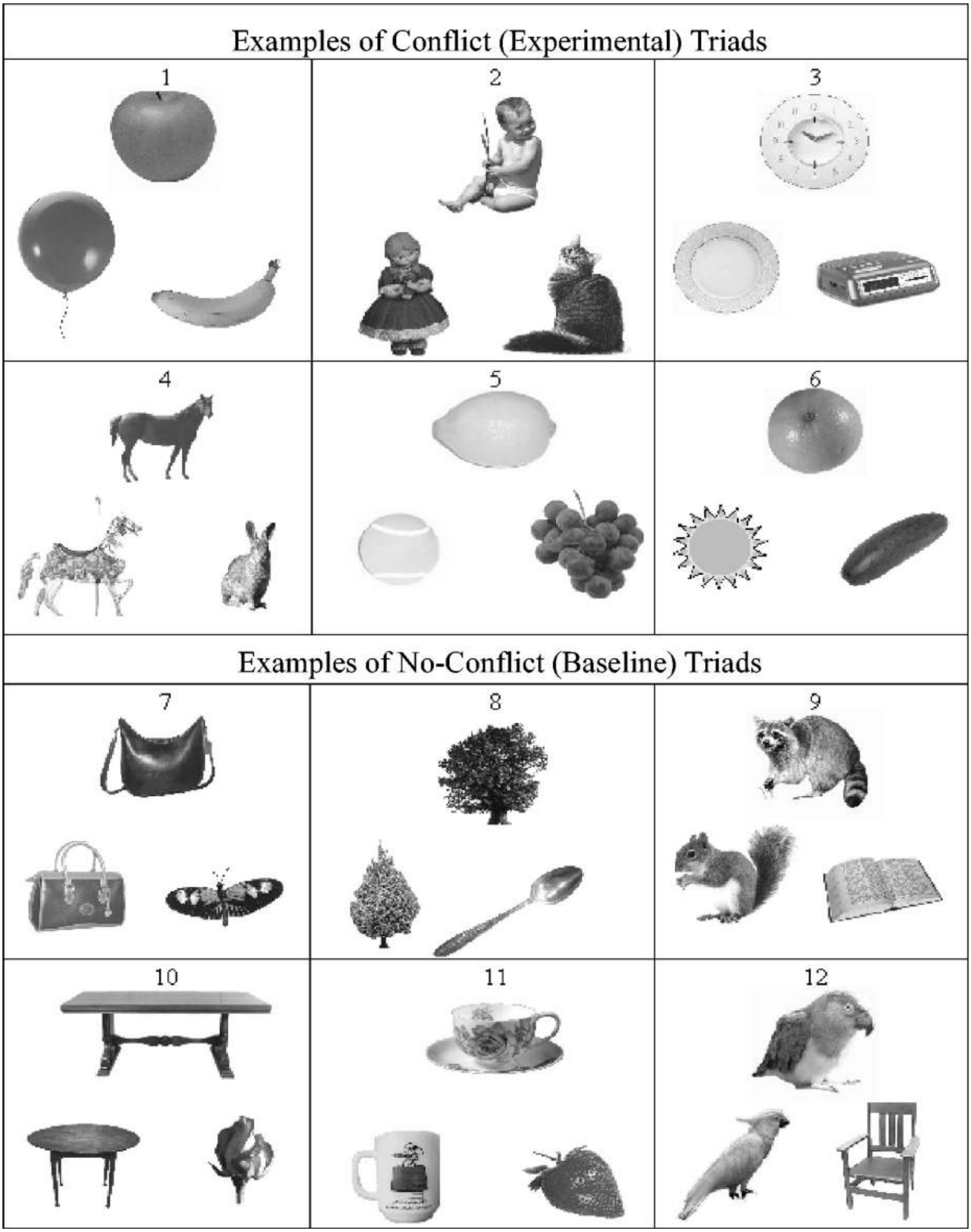
Ward, T. B., Vela, E., Peery, M. L., Lewis, S. N., Bauer, N. K., & Klint, K. A. (1989). What makes a vibble a vibble? A developmental study of category generalization. *Child Development*, *60*, 214–224.

Waxman, S. R., & Markow, D. B. (1995). Words as invitation to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

(*Appendix follows*)

Appendix

Picture Triads Used in Experiment 1A



(These stimuli were presented to participants in color, and a color version of this appendix is available in the online version of this article, which is part of the PsycARTICLES database.)

Received January 30, 2003
Revision received November 17, 2003
Accepted December 10, 2003 ■