

# Linguistic Labels and the Development of Inductive Inference

**Vladimir M. Sloutsky (sloutsky.1@osu.edu)**

Center for Cognitive Science & School of Teaching & Learning  
Ohio State University  
21 Page Hall, 1810 College Road, Columbus, OH 43210, USA

**Ya-Fen Lo (lo.37@osu.edu)**

Center for Cognitive Science & School of Teaching & Learning  
Ohio State University  
21 Page Hall, 1810 College Road, Columbus, OH 43210, USA

## Abstract

The paper presents a model suggesting that inductive generalizations in young children could be a function of similarity among compared stimuli. Predictions derived from the model were tested in two experiments where young children and preadolescents were presented with triads of schematic faces (a Target and two Test stimuli) that varied in perceptual similarity, with one of the Test stimuli sharing a linguistic label with the Target. Participants were taught a biological property about the Target and asked to generalize the property to one of the Test stimuli. Results from both experiments support predictions, indicating that for young children, proportions of label-based generalizations varied with featural overlap among the compared stimuli. There were also developmental differences found in effects of labels: while for young children these effects varied with featural overlap, preadolescents relied solely on linguistic labels when performing inductive generalizations.

## Introduction

Inductive generalization is an important component of human thought. Furthermore, some believe that it the most important component because "inductive inference is the only process... by which new knowledge comes into the world" (Fisher, 1935). Therefore, understanding of the development of induction is an important step in understanding of human thought.

One theoretical proposal suggests that induction starts out as a category-based process (see Gelman & Coley, 1991; Gelman, Coley, & Gottfried, 1994, for reviews and discussions). In this case, independent of similarity, generalization within a theoretically-defined category is more likely than generalization across categories. For example, a person is more likely to generalize a property (e.g., the ability to drink) from a bird to another dissimilar looking bird than from a bird to a similar looking airplane (Mandler & McDonough, 1998; Gelman & Markman, 1986; Gelman & Coley, 1991).

The alternative, similarity-based approach, suggests that induction starts out as a special case of the "universal law of generalization" (Shepard, 1987). The law states that the probability of generalizing a response (e.g., fear) from one stimulus to another stimulus varies with featural similarity between the stimuli.

Although the similarity-based approach seems to be more

appealing on the basis of parsimony, it has been often criticized for the failure to constrain the notion of similarity (e.g., Goodman, 1992/1972). Indeed, with the increase of the complexity of predicate structure, it becomes unclear which of these predicates will be used in computing similarity.

Recently, we proposed a model suggesting that for young children, linguistic labels might be an important constraining factor (Sloutsky & Lo, 1999). In a series of experiments, we demonstrated that linguistic labels have larger weights in similarity judgment of young children than other perceptual attributes. We argued that similarity between stimuli patterns decreases as a function of exponential decay (cf. Estes, 1994; Medin, 1975). That is similarity between two labeled stimuli patterns could be calculated using Equation 1:

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

where  $N$  denotes the total number of visual attributes,  $k$  denotes the number of matches,  $S_{vis.attr}$  denotes values (weights) of a mismatch on a visual attribute,  $S_{Label}$  denotes values of label mismatches, and  $L$  denotes a label match. When there is a label match,  $L = 1$ , and  $S_{Label} = 1$ ; when there is a label mismatch,  $L = 0$ , and  $S_{Label} < 1$ . Because  $S$  varies between 0 and 1, similarity equals to one when there are no mismatches, otherwise it is smaller than 1.

We also suggested that when a 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 probability of choosing B could be predicted using Equation 2:

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

In this paper, we present evidence that the model can account not only for similarity judgement, but for inductive inference of young children as well.

Of course, it could be argued that reliance of young children on linguistic labels when performing induction is an indicator that they perform induction in a category-based manner, because they use linguistic labels as category markers. There is an important caveat, however. If they rely on linguistic labels as category markers, labels should affect

induction in a qualitative "all-or-none" manner (i.e., presence or absence of the shared label should be a critical factor in induction). We predict an alternative course: labels affect induction in a quantitative manner, in accordance with equations 1 and 2. In other words, proportions of label-based generalizations in young children should vary, with the number of visual attributes shared between the compared entities.

## Experiment 1

### Method

**Participants** A group of 87 children aged 4 to 12 years participated in the study. The participants represented three age groups: (1) 32 four-to-five year-olds ( $M = 4.5$  years,  $SD = 0.56$  years; 14 boys and 18 girls), 30 seven-to-eight year-olds ( $M = 8.1$  years,  $SD = 0.5$  years; 15 boys and 15 girls), and 25 eleven-to-twelve year-olds ( $M = 11.8$  years,  $SD = 0.5$  years; 15 boys and 10 girls). The participants were recruited from daycare centers, elementary and middle schools located in middle class suburbs of Columbus, Ohio.

**Materials** The materials included triads of 2" by 2" schematic faces, two of which were Test stimuli and one of which was a Target. Each schematic face had three distinct attributes (shape of head, shape of ears, and shape of nose), and each attribute had three values (e.g., "curve-lined" nose, "straight-lined" nose, and "angled" nose). These materials were identical to those used in Part 1 experiments (Sloutsky & Lo, 1999). Materials also included 36 artificial bisyllable labels (e.g., *Bala*, *Gula*, and so forth) and a set of unobservable biological properties of the Target. These properties were as follows:

1. Has pink bones
2. Has green brain
3. Has white heart
4. Has orange stomach
5. Has blue fat
6. Has yellow blood

Participants were asked which of the Test stimuli was more likely to share a biological property with the Target.

**Design and Procedure** The experiment had a mixed design with age and labeling condition (label vs. no-label) as between-subject factors and a stimulus pattern condition as a within-subject variable. For both levels of the labeling condition, participants were presented with the same triads of schematic faces, two of which were Test stimuli and one of which was a Target. The only difference was that in the label condition all stimuli were labeled, whereas in the no-label condition these stimuli were not labeled. The stimulus pattern condition included six levels, T-00, T-11, T-22, T-01, T-12, and T-02. Note that T refers to the Target, the first digit refers to the number of attributes shared by Test B with the Target, and the second digit refers to the number of attributes shared by Test A with the Target. In the label condition, the Target always shared labels with Test B and always had labels different from Test A. A female researcher interviewed children in a quiet room in their schools. Before the experimental task, children were introduced to some

warm-up questions and were given feedback. In the warm-up tasks, children were presented with Test and Target stimuli and were asked to choose the Test stimulus that shared a biological property with the Target.

**Warm up Trials.** In the first warm-up trial, participants were presented with a Target (a shark) and two Test stimuli (a bear and a tree branch). In the second warm-up trial, they were presented with a rabbit as a Target, and an apple and a dog as Test stimuli. In the third warm-up trial, children were presented with a fish as a Target, and a turtle and a spider as Test stimuli. In all these warm-up trials, children were first told that the Target stimuli either had bones, blood, or skeleton inside the body. Children then were asked to determine which of the two Test stimuli has the same thing inside the body as the Target. If a child failed to answer induction questions, the researcher explained how each of the Test stimuli could have the same thing as the Target.

**Experimental Trials.** If a child was capable of giving correct answers in two out of three warm-up trials, the researcher proceeded to the main experiment. No child was eliminated from the study since all participants provided satisfactory responses in at least two out of three warm-up trials. In the Label condition, children were first introduced to the labels for the Target and Test pictures and asked to repeat them. All labels used were the same artificial names (e.g. *Bala*, *Guga*) as in Part 1 experiments. After each stimulus was labeled, children were asked to repeat these labels. No labels were introduced in the no-label condition. Children were then introduced to an unobservable biological property that belonged to the Target stimuli and were asked which of the Test stimuli was likely to have this property. Positions of the two Test pictures were counterbalanced across the experimental trials. After children answered the questions, they were asked to provide their justification for their choices. In both conditions, participants of the two older groups had 24 experimental trials (6 within-subject stimulus patterns with 4 trials each), while participants of the youngest group had 18 experimental trials (6 within-subject stimulus patterns with 3 trials each). This reduction in the number of trials was important to avoid fatigue that could lead to random responding in young children. The order of presenting of stimulus patterns was randomized within participants.

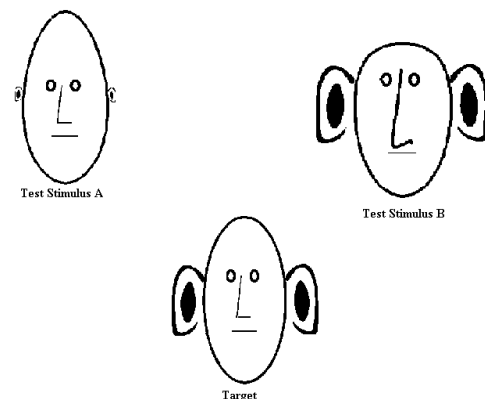


Figure 1: Example of stimuli presented in one trial in the T-

1-2 condition: The Target shares the overall shape and the nose with Test A and the size of the ears with Test B.

**Results and Discussion**

Results indicate that in the no-label condition, participants of all age groups based their inductive inferences on available perceptual information. Proportions of Test B choices broken down by stimulus pattern condition and age groups are presented in Figure 2. Recall that B-choices refer to the selection of the Test stimulus that in the label condition shares the label with the Target. As predicted (see Table 1), for all indeterminate stimulus pattern conditions, when the Target shared equal numbers of attributes with each Test stimulus (i.e., T-00, T-11, and T-22), the proportions of B-choices for all age groups were at chance (one-sample t-tests, all  $ps > .25$ ). Also as predicted, in all determinate stimulus pattern conditions where Test B shared fewer attributes with the Target than Test A (i.e., T-01, T-12, and T-02), the proportions of B-choices for all age groups were below chance (one-sample t-tests, all  $ps < .01$ ).

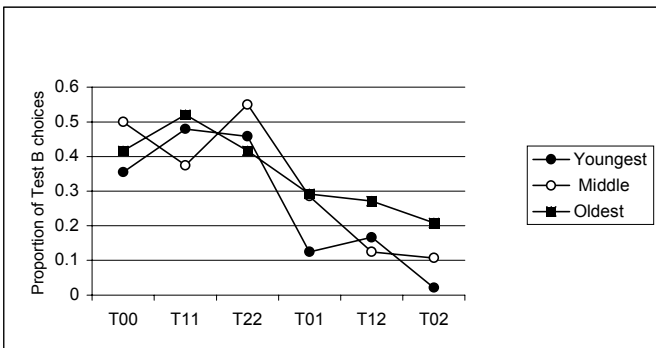


Figure 2: Induction in the No-Label condition broken down by stimulus pattern condition and age group.

To examine the direction of differences among the stimulus pattern conditions in the no-label condition, proportions of B-choices were subjected to a two-way ANOVA with age as a factor and stimulus pattern condition as a repeated measure. Because proportions of B-choices across the T-00, T-11, and T-22 conditions were very similar (all  $ts < 1$ ), proportions of B-choices were averaged across these conditions into a new aggregated variable T-Equal ("T" stands for the Target and "Equal" indicates that each of the Test stimuli shared equal number of features with the Target). While there were no significant differences in proportions of B-choices among the age groups,  $F(2, 35) = .8, p = .45$ , there was a significant main effect due to the stimulus pattern condition,  $F(3, 105) = 13.5, MSE = 0.1, p < .0001$ . Planned comparisons indicated that T-Equal exhibited the largest proportion of B-choices (46%), whereas T-02 exhibited the smallest proportion of B-choices (11%), all  $ts > 2.2, ps < .05$ . At the same time, T-12 and T-01 did not differ significantly,  $t < 1$ . These results indicate that when only perceptual information was available, participants based their inductive inference on this information: in all age groups inductive inference was a function of the number of attributes shared by the Target

with Test stimuli.

Introduction of labels, however, dramatically changed the proportions of B-choices that had been observed in the no-label condition. Recall that in the label condition, Test B always shared the label with the Target. Proportions of B-choices (i.e., label-based generalizations) in the label condition broken down by age group and stimulus pattern condition are presented in Figure 3. In the oldest group, all participants on all trials, with the exception of one participant on one trial, used labels as the only basis of their induction. At the same time, effects of labels in the two younger groups varied across stimulus pattern condition. Because participants in the older group exhibited no variability in their responses (311 out of 312 responses were label-based generalization), while participants in the two younger groups exhibited variability, the former were not included in the analysis of label-based generalizations across stimulus pattern condition.

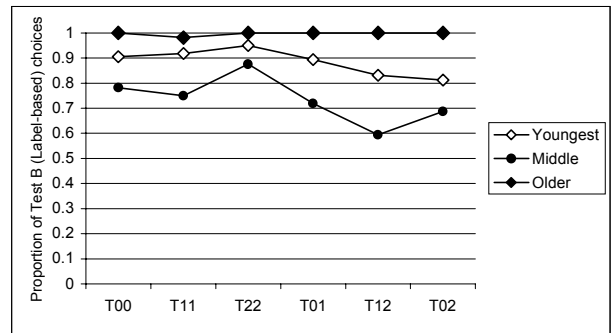


Figure 3: Induction in the Label condition broken down by stimulus pattern condition and age group.

Proportions of label-based choices across stimulus pattern condition in the two younger groups were subjected to a two-way (age by stimulus pattern condition) ANOVA with stimulus pattern as a repeated measure. Because proportions of label-based generalizations across the T-00, T-11, and T-22 conditions were statistically equivalent ( $t < 1$ ), these proportions were averaged across these conditions into a new aggregated variable T-Equal. The analysis indicates a significant main effect due to stimulus pattern condition,  $F(3, 102) = 2.8, MSE = 0.06, p < .05$ , whereas neither main effects of age group, nor the interaction of the two factors were significant ( $p = .15$  and  $p = .8$ , respectively). Planned comparisons pointed to significant differences between the T-Equal condition and the T-12 and T-02 conditions, all  $ts (35) > 2.1, ps < .05$ . In short, as predicted, in the oldest group the proportion of selecting Test B did not vary across the stimulus pattern conditions, whereas this proportion did vary as a function of stimulus pattern condition in the two younger groups. Although differences among the stimulus pattern conditions may appear relatively small (in particular, differences between T-02, on the one hand, and T-01 and T-12, on the other hand, fell short of statistical significance), the direction of these differences, except for the T-02 condition in the middle group, closely match predictions derived from Equations 1 and 2.

Overall fit between predicted probabilities and observed frequencies is presented in Figure 4. Each data point in Figure 4 represents a cell in Table 1, corresponding to performance of children in the two younger groups in the label and no label conditions. In addition to a high correlation between the predicted and observed probabilities ( $r = .96$ ), the proposed theoretical model accounts for approximately 92% of the observed variance ( $R^2 = .9196$ ). These findings support the notion that for younger children labels contribute to specific induction in a quantitative manner and that this contribution varies with the number of attributes shared by Test A and Test B with the Target. Note that the Figure does not include performance of the oldest group, because their induction was not derived from Equations 1 and 2. On the contrary, their induction was predicted to be category-based, as opposed to similarity-based induction of younger children.

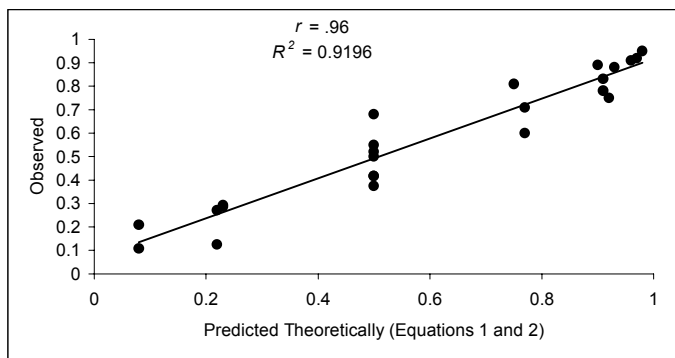


Figure 4: Theoretical probabilities (computed from the model using Equations 1 and 2) and observed probabilities of generalization of biological properties to Test B. Note: parameter  $S$  was estimated from our previous research.

Findings of this experiment support our predictions regarding quantitative contribution of labels to inductive inferences of young children. In the absence of labels, participants of all age groups based their inductive inference on perceptual information. However, when labels were introduced, the pattern of choices changed dramatically: preadolescents based their induction solely on labels, whereas younger children based their induction on a combination of labels and the number of overlapping attributes. Therefore, it seems reasonable to infer that preadolescents performed induction in a category-based manner, whereas younger children performed induction in a similarity-based manner.

However, it could be argued that these findings strongly support predictions for preadolescents' induction, while providing only tentative support of predictions for young children's induction. The support is tentative because some differences among the stimulus pattern conditions, while all in predicted directions, failed to reach significance. Because of this, we deemed it necessary to conduct a second experiment, replicating the current experiment for young children while simplifying the task, and increasing the sample sizes and the number of trials.

## Experiment 2

### Method

**Participants** A group of 30 four-year-old children ( $M = 4.3$  years,  $SD = 0.5$  years; 19 boys and 11 girls) participated in the experiment. The participants were recruited from daycare centers located in middle class suburbs of Columbus, Ohio.

**Materials** The materials included triads of schematic faces identical to those used in Experiment 1.

**Design and procedure** The design and procedure were identical to that in Experiment 1 with three exceptions. First, the current experiment included only a label condition. Second, the "why" questions that accompanied children's choices in Experiment 1 were dropped. Finally, the number of trials within each stimulus pattern condition was increased from three to four.

### Results and Discussion

Proportions of label-based generalizations across the stimulus pattern conditions are presented in Figure 5. Because proportions of label-based generalizations in T-00, T-11, T-22 conditions were statistically equivalent (87%, 85%, and 87% respectively,  $ts < 0.5$ ), participants' responses were averaged across these conditions into a new variable T-Equal. Proportions of label-based generalizations in T-Equal, T-01, T-12, and T-02 conditions were subjected to a one-way repeated measures ANOVA. The analysis points to significant differences among the stimulus pattern conditions,  $F(3, 87) = 16.744$ ,  $MSE = 1.15$ ,  $p < 0.0001$ . Planned comparisons pointed to the following order among the conditions in the proportion of label-based generalizations: T-Equal (86%) > T-12 (63%) = T-01 (56%) > T-02 (39%). All indicated differences were significant, all  $ts > 3.5$ , Bonferroni adjusted  $ps < .01$ , while the difference between T-12 and T-01 was not significant,  $t < 1$ . These results clearly indicate that the proportion of label-based generalizations varied as a function of the number of features shared by the Target with each of the test stimuli, thus further supporting the notion of similarity-based specific induction in young children.

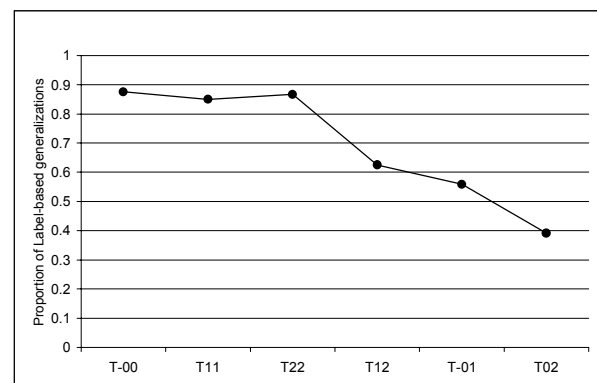


Figure 5: Proportions of label-based generalizations across the stimulus pattern conditions.

## General Discussion

Results of the two reported experiments are as follows. In Experiment 1, when labels were not provided, 4-5 year-olds, 7-8 year-olds, and 11-12 year-olds relied on perceptual similarity when making specific induction with novel entities. At the same time, when labels were introduced, preadolescents made inductive inferences based solely on the basis of the provided label, whereas specific induction of younger children varied with the number of attributes (which includes labels) shared by the Target and Test stimuli. Results of Experiment 2 further indicated that proportions of label-based generalizations in young children varied as a function of visual attributes shared by the Target with each of the Test stimuli. These results fit predictions, indicating that the model proposed by Sloutsky & Lo (1999) can account for specific induction of younger children and that specific induction of the younger children is similarity-based.

The results of the no-label condition indicate that for all three age groups, similarity-based specific induction is the default mechanism (cf. Keil, 1989). When no other information was available, participants of all age groups used perceptual similarity to generalize biological properties from the Target to Test stimuli. Therefore, if similarity-based induction is the default mechanism, it seems likely that it might developmentally precede category-based induction. This contention was supported by results of the label condition.

The results of the label condition supported the notion of different mechanisms underlying specific induction in young children and preadolescents, thus allowing the resolving of an apparent paradox of specific induction. The paradox is as follows. On the one hand, if specific induction is category-based, it should be dependent on general induction and the ability to perform induction-deduction coordination. On the other hand, even three year-olds are capable of performing specific induction (Gelman & Markman, 1987). The reported results suggest that specific induction does not have to be category-based -- it may start out as similarity-based and develop into category-based later. The reported experiments support this notion, suggesting that this shift may occur sometime between nine and eleven years of age. Indeed, while specific induction of 7-8 year-olds appeared to conform to the proposed model and to vary with a number of perceptual attributes shared by the Target and Test stimuli, specific induction of 11-12 year-olds appeared to be independent of shared attributes and to be a function of labels.

It is also important that for younger children, labels exert similar effects on similarity judgment and specific induction. At the same time, in preadolescents these effects are fundamentally different. While labels had no effect on similarity judgment of preadolescents (Sloutsky & Lo, 1999), in specific induction preadolescents relied solely on labels. These findings further support the possibility of a developmental shift from similarity-based to category-based induction occurring between 9 and 11 years of age.

This developmental shift may be a function of the development of a categorical structure: when two objects share a label they are more likely to be considered members

of the same category than to be considered members of different categories. When a categorical structure is in place, the probability that two remotely similar entities that have the same label would be considered members of different categories could be estimated by the base rate of homonyms (and homophones), and therefore is negligibly small. In fact, we drew a random sample of 200 most frequently used English nouns from Francis and Kucera (1982) and asked three native English speakers to mark those that have homonyms and homophones. While the overall rate of homonyms and homophones appeared to be relatively high (ranging from 20% to 30%), many of these homonyms and homophones were adjectives and verbs (e.g., horse/hoarse or board/bored). At the same time, the rate of noun-noun homonyms (e.g., case/case) was around 5%. Furthermore, the rate of a noun having a homonym within the same ontological class (e.g., living creature having a homonym that indicates a completely different living creature) was practically nonexistent. Hence, remotely similar entities that share the label should be interpreted as members of the same category and, therefore, to share unobservable properties as well. In short, the label-as-attribute model proposed by Sloutsky & Lo (1999) can account not only for similarity judgment of younger children, but also for their specific induction.

While the model provides a reasonable account of specific induction with artificial stimuli that are relatively similar on the overall scale, it remains unclear whether or not the model is capable of handling more naturalistic and diverse set of stimuli. Our most immediate concern is to test the model with these kinds of stimuli. Because our stimuli were quite similar overall (all pictures represent human-like faces) it is possible that results might have been different had the stimuli been more different. It is also possible that results might have been different if stimuli were not human-like: infants and young children have been shown to develop different types of representations of humans and animals (Quinn & Eimas, 1998). While the former could be represented as individual exemplars, the latter may have summary (i.e., category-based yet perceptual) representations. However, we believe that introduction of more different and more diverse stimuli would make differences between younger and older children even more apparent, because stimuli would increase perceptual-similarity-based variance in the younger groups without increasing this variance in the older group.

If induction in young children is based on overall similarity among compared entities then introduction of new attributes (both perceptual and non-perceptual) that contribute to overall similarity, should also contribute to inductive generalizations. It would be important to test this prediction and to estimate weights of different classes of attributes. It would be also important to trace changes in these weights with development and learning. Finally, it would be necessary to test the model on younger participants and have more dense developmental observations.

Because the proposed model is capable of formulating specific predictions, these predictions can be tested in future research. For example, we contend that specific induction in young children is similarity-based, whereas preadolescents it

is category-based. If this is true, then for younger children specific induction should be easier than general induction, while for older children it should be more difficult (because category-based specific induction requires more mental steps than general induction). However, if specific induction in younger children is also category-based, then in both younger and older children specific induction should be more difficult than general induction.

Recall that the label-as-attribute model also affords the computation of specific probabilities of inductive generalizations across stimuli that vary in overall similarity. In future research, we plan to test these predictions of the model with respect to naturalistic stimuli. Because it is impossible to individuate features and to precisely calculate featural overlap with complex naturalistic stimuli patterns, we will manipulate similarity by "morphing" naturalistic pictures into each other in a fixed number of steps. Morphing of naturalistic stimuli has been used in the past in study of categorization in humans (Beale & Keil, 1995) and non-human primates (Brown, 1999).

### Acknowledgements

This research has been supported by a grant from the James S. McDonnell Foundation to the first author.

### References

- Beale, J., & Keil, F. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217-239.
- Brown, D. (1999). Categorization of naturalistic stimuli by chimpanzees (*Pan Troglodytes*). Unpublished masters thesis. The Ohio State University.
- Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.
- Fisher, R. (1935). *The design of experiments*. Oliver & Boyd, London: UK.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman, S. & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (146-196). New York, NY: Cambridge University Press.
- Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*, 183-209.
- Gelman, S. A., Coley, J. D. & Gottfried, G. M. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234-254). New York, NY: Cambridge University Press.
- Goodman, N. (1992/1972). Seven strictures on similarity. In M. Douglas & D. Hull (Eds.), *How classification works* (pp. 13-23). Edinburgh, Scotland: Edinburgh University Press.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Mandler, J. M., & McDonough, L. (1998). Studies in inductive inference in infancy. *Cognitive Psychology*, *37*, 60-96.
- Medin, D. (1975). A theory of context in discrimination learning. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 263-314), Vol. 9. New York: Academic Press.
- Quinn, P., & Eimas, P. (1998). Evidence for global categorical representation of humans by young infants. *Journal of Experimental Child Psychology*, *69*, 151-174.
- Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.
- Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgement. *Developmental Psychology*, *6*, 1478-1492.