The Role of Linguistic Labels in Categorization

Wei (Sophia) Deng (deng.69@osu.edu)
Center for Cognitive Science
The Ohio State University
209C Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Center for Cognitive Science
The Ohio State University
208C Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Abstract

Do young children treat labels as features or as category markers? The current study addressed this question by examining the effect of labels on young children's classification and induction. The first experiment replicated previous study on adults demonstrating that adults treat labels as category markers. The other two experiments applied the same paradigm to young children. Children were trained by classification in Experiment 2A and by induction in Experiment 2B, whereas both experiments used the classification and induction tasks that were identical to those in Experiment 1. The results from the three experiments indicated that adults treated labels as category markers, whereas no such evidence was found for young children.

Keywords: Cognitive Development, Classification, Induction, Label, Psychology, Human Experimentation.

Introduction

The ability to use linguistic labels to generalize from the known to the unknown is crucial for learning new information. Although a substantial body of experimental evidence has demonstrated that label has an impact on categorization and induction processes (Gelman & E. Markman, 1986; Sloutsky, Lo, & Fisher, 2001; Welder & Graham, 2001; Yamauchi and A. Markman, 1998, 2000), the mechanism underlying the role of labels is hotly debated. Are labels used as features (similar to other objects' properties) or as category markers representing category membership? This issue is particularly contentious with respect to the role of labels in early development and developmental changes in this role.

Some researchers have argued that labels are more than features. According to this view, labels are category markers used for representing a category. For example, E. Markman and Hutchinson (1984) found that children regarded words presented as count nouns changed the way young children grouped objects. Without labels children grouped objects thematically (e.g., a police car was grouped with a policeman), whereas when the same police car was referred to by a count noun, children grouped objects taxonomically (e.g., the police car with a passenger car). Gelman and Heyman (1999) demonstrated that young children were more willing to generalize properties from one person to another when both persons were referred to by a noun (i.e., "carrot-eaters") than when both were referred to by a descriptive sentence (e.g., "both like to eat carrots").

This evidence, however, does not lend unequivocal support to the idea that labels have to be category markers to make inductive inferences. For example, some researchers suggested that contribution of labels is driven by attentional rather than conceptual factors (Napolitano & Sloutsky, 2004; Sloutsky & Napolitano 2003). There is also evidence that labels contribute to the overall similarity of compared entities and thus to both categorization and induction (Sloutsky & Fisher, 2004). Sloutsky and Fisher (2004) also demonstrated that similarity computed over labels and appearances can accurately predict young children's responses with the Gelman and E. Markman (1986) task. These findings suggest that reliance on labels does not necessarily indicate that labels are more than features.

In a series of studies, Yamauchi and A. Markman (1998, 2000) designed a paradigm that could address this issue directly. Specifically, they compared participants' performance on classification tasks (e.g., is X a dax?) with that on induction tasks (e.g., given that X is a dax, does it have Y?). The tasks are structurally identical, except a critical difference. In the classification task participants predicted the category label of an item given all of its feature values. In contrast, in the induction task, participants predicted the value of a missing feature of an item given its category label and other feature values. These researchers argued that if the label is a feature then performance on classification and induction task should be symmetrical. However, if labels are more than features, then performance on induction tasks should be better than performance on classification tasks. Upon finding predicted asymmetries between the two conditions (i.e., participants were better at using the label to predict other features than at using other features to predict the label), these researchers concluded that participants are more likely to regard labels as category markers instead of object features.

However, this paradigm has not been applied to children. Does the asymmetry found in adults exist in children? Finding such an asymmetry would support the idea that labels are more than category features, whereas a symmetric performance in the classification and induction conditions would support the idea that labels are features. The primary goal of this study is to address these questions.

The current study consists of three experiments. Experiment 1 replicated Yamauchi and A. Markman's paradigm (2000) with adults. Based on their findings, it was hypothesized that adults would regard labels as category markers. Experiments 2A and 2B, using comparable learning and testing conditions, examined how labels would affect young children's performance.

Experiment 1

Method

Participants Sixteen adults participated in this experiment. Participants were undergraduate student from the Ohio State University participating for course credit. Three of them gave one type of response to over 95% of all trials. Their data were excluded from the analysis due to the response bias.

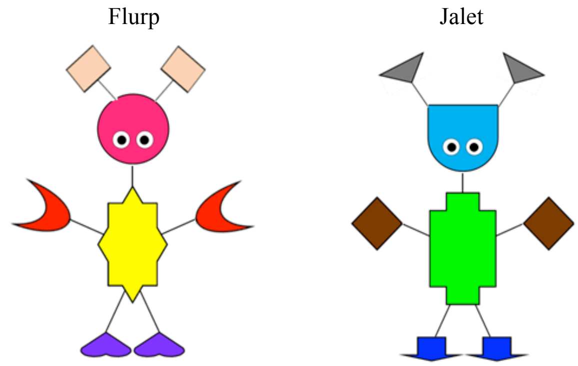

Figure 1. The prototypes of stimuli used in this study.

Stimuli The stimuli were artificial creatures accompanied by a category label ("Flurp" of "Jalet") and two categories of objects that were created using five features varying in color and shape (see Figure 1). As shown in Table 1 and 2, the two categories have a family-resemblance structure, which is derived from two prototypes (F0 and J0) by modifying the values of one of five features. For example, to produce the stimulus F1, the value of the antenna is changed from 1 to 0 so that it has four features consistent with the prototype F0 and one feature consistent with the prototype J0. The degree of similarity between test stimulus and the prototype is defined by the number of matching features of the test stimulus to the prototype of the corresponding category (Table 1 and 2).

Table 1. Category structure used in learning.

Flurp							Jalet						
Stimuli	Head	Body	Hands	Feet	Antenna	Label	Stimuli	Head	Body	Hands	Feet	Antenna	Label
F1	1	1	1	1	0	1	J1	0	0	0	0	1	0
F2	1	1	1	0	1	1	J2	0	0	0	1	0	0
F3	1	1	0	1	1	1	J3	0	0	1	0	0	0
F4	1	0	1	1	1	1	J4	0	1	0	0	0	0
F5	0	1	1	1	1	1	J5	1	0	0	0	0	0
F0	1	1	1	1	1	1	J0	0	0	0	0	0	0

Note. The value 1 = any of five dimensions identical to "Flurp" (see Figure 1). The value 0 = any of five dimensions identical to "Jalet" (see Figure 1). F = Flurp; J = Jalet. F0 and J0 are prototypes of each category.

Table 2. Stimulus structure used in testing.

Flurp								Jalet						
Stimuli	Head	Body	Hand	Feet	Antenna	Target Label	Match	Stimuli	Head	Body	Hand	Feet	Antenna	Target Label
F11	1	1	1	1	0	1		J11	0	0	0	0	1	0
F12	1	1	1	0	1	1		J12	0	0	0	1	0	0
F13	1	1	0	1	1	1	High	J13	0	0	1	0	0	0
F14	1	0	1	1	1	1		J14	0	1	0	0	0	0
F15	0	1	1	1	1	1		J15	1	0	0	0	0	0
F21	1	0	1	0	0	1		J21	0	1	0	1	1	0
F22	0	1	0	1	0	1		J22	1	0	1	0	1	0
F23	0	0	1	0	1	1	Low	J23	1	1	0	1	0	0
F24	1	0	0	1	0	1		J24	0	1	1	0	1	0
F25	0	1	0	0	1	1		J25	1	0	1	1	0	0

Note. High and low are two levels of feature match. F = Flurp; J = Jalet. Category-accordance responses were the ones consistent with the values indicated in the target features and target labels.

Similar to Yamauchi and A. Markman, there are two levels of similarity (or feature match) in current research: high and low. At the high level of feature match, each test stimulus has four features in common with the prototype of the corresponding category and one feature in common with the prototype of the contrasting category. Similarly, each test stimulus at the low level of feature match has two features in common with the prototype of the corresponding category and three features in common with the prototype of the other category.

Procedure The entire experiment consisted of two phases, learning and testing. During the learning phase, participants were instructed that they should try to remember and distinguish two groups of artificial creatures represented by the labels "Flurp" and "Jalet". And then participants were presented with 36 trials of creatures produced from stimulus structure shown in Table 1 and each stimulus had a correspondent label above it.

The testing phase was administered immediately after the learning phase. The Classification and Induction conditions differed in the type of features being predicted. In the Classification condition, participants predicted the category label of a stimulus given information about all five features with the label covered. In the Induction condition, participants predicted the value of one of five features given the other four features with the label uncovered. The classification question was phrased as "Which group do you think this creature is more likely to belong to, Flurp or Jalet?" The induction question was phrased as "Which antenna do you think this creature is more likely to have?" The order of the testing trials was randomized for each subject. Feedback was given in first 6 trials of each condition. No feedback in other 40 trials in both conditions. The proportion of responses in accordance with the category from which the exemplar was derived (called "category-accordance responses" by Yamauchi & A. Markman, 2000, see Table 2) was the dependent variable.

Results and Discussion

The main results of Experiment 1 are shown in Figure 2. The data were analyzed with 2 (testing type: Classification and Induction) × 2 (feature match: high and low) analysis of variance (ANOVA). There was a main effect of feature match, $F(1,12) = 165.39$, $MSE = 0.89$, $p < 0.01$, as well as an interaction between testing type and feature match, $F(1,12) = 38.61$, $MSE = 0.32$, $p < 0.01$. At the low level of feature match, category-accordance responses made in the Induction condition were more than in the Classification condition, $t(12) = 4.88$, $p < 01$. However, there was no significant difference in these two testing types at the high level of feature match, $t(12) = 2.12$, $p > 0.05$.

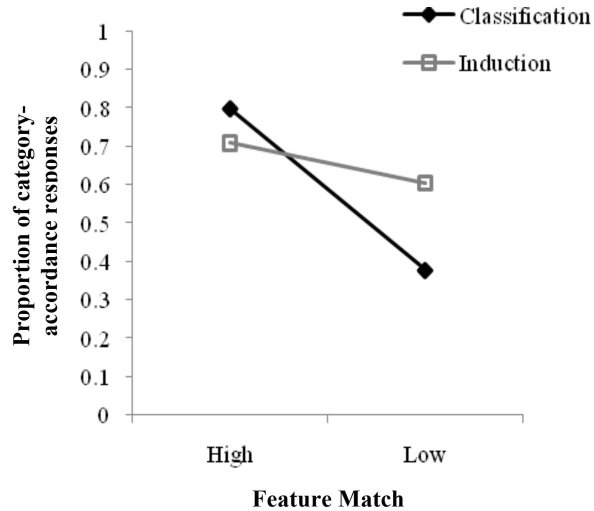

Figure2. Performance for classification and induction tasks in Experiment 1.

The results replicate Yamauchi and A. Markman (2000) pointing to the predicted asymmetry and suggesting that for adults labels are more than objects features. In Experiment 2, we expand this paradigm to young children.

Experiment 2A

Method

Participants There were thirteen preschool children (6 boys and 7 girls) with an average age of 55.8 months participating in this experiment. They were given Classification learning. In Classification learning, children were presented with all five features of a creature and told that it was a Flurp (or Jalet). A memory check was administered after main experiment to examine whether participants could remember the stimuli and correspondent labels. Children were given 5 trials in memory check by presenting a creature and asking them which group this creature came from. One of them answered less than 3 out 5 memory check questions correctly and these data were excluded from the analysis.

Stimuli and procedure The visual stimuli were identical to Experiment 1 (see Figure 1). The entire experiment consisted of two phases, classification learning phase and testing phase. During classification learning, in contrast to Experiment 1 with adults, children were instructed that there were two groups of creatures, Flurp and Jalet. And then they were trained by presenting creatures with category labels and told:" This is a Flurp (or Jalet)."

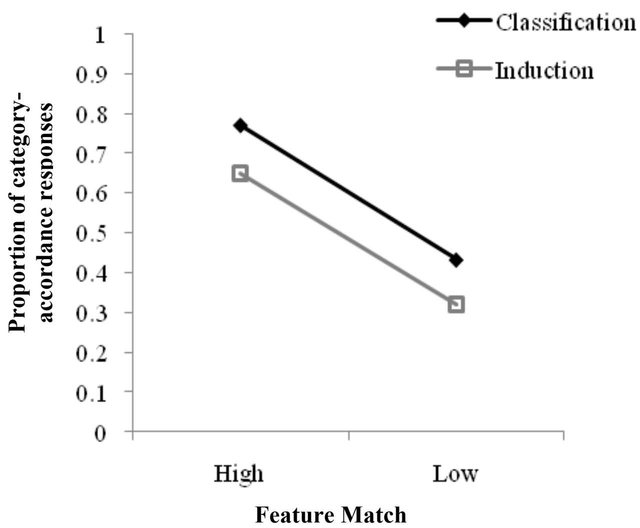

Figure 3. Performance for classification and induction tasks in Experiment 2A.

The testing phase was identical to Experiment 1 except how the questions were asked. Unlike the adults participants in Experiment 1 who read the questions presented on screen, children were asked both classification and induction questions by a female experimenter.

Results and Discussion

The main results of Experiment 2A are shown in Figure 3. The data were analyzed with 2 (testing type: Classification and Induction) × 2 (feature match: high and low) analysis of variance (ANOVA), with testing type and feature match as within-subjects factors. There was a main effect of feature match, $F(1,11) = 43.56$, $MSE = 1.33$, $p < 0.01$, as well as a main effect of testing type, $F(1,11) = 14.77$, $MSE = 0.16$, $p < 0.01$. However, unlike adults in Experiment 1, there was no significant interaction between testing type and feature match, $F(1,11) = 0.02$, $MSE = 0.00$, $p > 0.10$.

Children in Experiment 2A, unlike adults in Experiment 1, made more category-accordance responses on classification questions than on induction questions at both high and low level of feature match. These results indicate that, if anything, the Classification condition elicited better performance than the Induction condition. These results present little evidence that young children treated labels as category markers.

However, the learning type, in this experiment using classification learning, might have a facilitative effect on children's classification. In Experiment 2B, we explored the impact of learning type on children's performance in the classification and induction tasks by training them with induction instead of classification task.

Experiment 2B

Method

Participants There were fourteen preschool children (8 boys and 6 girls) with an average age of 54.00 months in this experiment. They were given Induction learning. In contrast to the children trained by classification in Experiment 2A, children in this experiment were trained by induction in which they were presented all five features of a creature and told that the creature had a Flurp (or Jalet) inside its body. A memory check, identical to Experiment 2B, was administered after main experiment to examine whether participants could remember the stimuli and correspondent labels. Two of them answered less than 3 out 5 memory check questions correctly and these data were excluded from the analysis.

Stimuli and procedure The visual stimuli were identical to previous experiments. The entire experiment consisted of two phases, induction learning phase and testing phase. During induction learning, in contrast to Experiment 2A, children were instructed that there were two groups of creatures and something special was inside each group of creatures. One group of creatures had Flurp while another group had Jalet. And then they were trained by presenting creatures and told:" This one has a Flurp (or Jalet)." The testing phase was identical to Experiment 2B.

Results and Discussion

The main results of Experiment 2B are shown in Figure 4. The data were analyzed with 2 (testing type: Classification and Induction) × 2 (feature match: high and low) ANOVA, with testing type and feature match as within-subjects factors. There was a main effect of feature match, $F(1,16) = 86.84$, $MSE = 2.50$, $p < 0.01$. In contrast to Experiment 2A, children in this experiment did not differ in the two testing types, $F(1,11) = 3.90$, $MSE = 0.03$, $p > 0.05$.

These results, compared to Experiment 2A, suggest that there was an effect of learning type and the induction learning facilitated children's performance on the induction questions. However, children's performance, similar to Experiment 2A, was symmetric in both testing conditions and there was no evidence that children treated differently induction and classification questions.

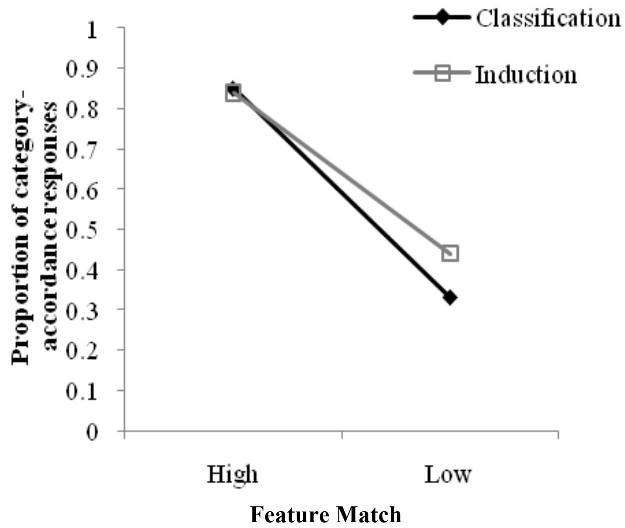

Figure 4. Performance for classification and induction tasks in Experiment 2B.

General Discussion

The results point to two main findings. First, there was an effect of learning. Children had better performance on classification questions when they were trained by classification (Experiment 2A). At the same time, when they were trained by induction (Experiment 2B), their performance on induction questions became equivalent to that on classification questions and there was no significant difference between these two testing types.

And more importantly, there were marked developmental differences in the role of linguistic labels. Adults exhibited better performance in inferring a feature by using a label than inferring a label by using features. These findings are consistent with previous research (Yamauchi & A. Markman, 2000) and suggest that adult may have used labels as category markers. However, labels had little facilitative effect on children's performance – their performance was equivalent whether they were asked to predict labels on the basis of other features (i.e., the classification condition) or to predict a feature on the basis of the label (i.e., the induction condition). Furthermore, regardless of the type of learning (i.e., Classification or Induction), children, in contrast to adults' asymmetry, consistently exhibited a symmetric pattern on classification and induction questions (see Figure 3 and 4). These results suggest that while labels may be different from category features for adults, this is not the case for young children.

These results have important implications for understanding of inter-relationships between language and cognition, and specifically the role of linguistic labels in categorization and category learning. Recall that according to some accounts, even early in development linguistic labels words affect categorization and inductive inference by marking the underlying category (e.g., Gelman & Heyman, 1999; Gelman & Markman, 1986). According to other accounts, early in development linguistic labels are features of entities. As a result, when two entities share a label, young children may perceive these entities as being more similar than when no labels are introduced (Sloutsky, et al, 2001; Sloutsky & Fisher, 2004). Yamauchi and A. Markman (1998, 2000) developed a procedure enabling the distinction between these accounts. This procedure, however, was never used with young children.

Our results with young children support the latter account, while generating little evidence that young children treat labels as category markers. In addition, although current research does not conclusively eliminate the possibility that for young children linguistic labels are category markers, it demonstrates that the role of linguistic labels changes in the course of development.

Acknowledgments

This research has been supported by grants from the grants from the NSF (BCS-0720135), from the Institute of Education Sciences, U.S. Department of Education (R305B070407), and from NIH (R01HD056105) to Vladimir M. Sloutsky.

References

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10,* 489 – 493.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition, 23,* 183-209.

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

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development, 75,* 1850 – 1870.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133,* 166 – 188.

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

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

Welder, A., & Graham, S. (2001). The influences of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development, 72,* 1653 – 1673.

Yamauchi, T., & Markman, A. B. (1998). Category-learning by inference and classification. *Journal of Memory and Language, 39,* 124-148.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26,* 776-795.