The Role of Linguistic Labels in Categorization

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Wei Deng

Graduate Program in Psychology

The Ohio State University

2011

Master's Examination Committee:

Vladimir M. Sloutsky, Advisor; Susan C. Johnson; John E. Opfer
Abstract

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? According to one approach labels denote categories, and as such, they differ from object features, whereas according to the other approach, labels start out as features, but they may become category markers in the course of development. This issue was addressed in seven experiments with 4- to 5-year-olds and adults. In Experiments 1-3, we attempted to replicate Yamauchi & Markman’s findings (1998, 2000) with adults and to extend the paradigm to young children, while manipulating the role of labels. In Experiment 4, we manipulated the way learning was introduced. In Experiment 5-7, we compared effects of labels to those of highly salient visual features. Results indicated that adult made category-accordance responses to induction questions more often than to classification questions at the low level of feature match and an asymmetric performance was found between these two tasks; whereas children’s performance in Classification and Induction tasks were equivalent and they relied on similarity among features rather than labels, which are supposed to represent category membership, to make inductive inferences. In addition, children uniformly relied on the salient feature; whereas, many adults relied on the category label. These results suggest that early in development, labels are no more than features and are processed the same way as other features, but they may become category markers in the course of development.
Acknowledgments

I am heartily thankful to my advisor, Dr. Vladimir Sloutsky, whose inspiration, encouragement, guidance and support from the initial to the final level enabled me to complete this project and to develop an understanding of this project, as well as of the joy and challenge of doing science.

Deepest gratitude is also to the Committee members, Dr. Susan Johnson and Dr. John Opfer, for their insightful comments and critics on this project.

Besides, I would like to express my gratitude and thanks to Cynthia Walter and Marie Valentine-Elam, the research assistants in the Cognitive Development Lab, for their assistance in scheduling the participants and collecting data in daycare centers. I would also like to convey thanks to the Department of Psychology and the Center for Cognitive Science for providing the financial means and laboratory facilities.

Special thanks also to my graduate friends, especially my labmates, Hyungwook Yim and Xin Yao, for sharing the literatures, ideas and invaluable assistance. And I’d like to thank Catherine Best and Christ Robinson for their discussions and comments on this project. Also, I’m grateful to my friend Yinghao Sun for the technical discussions on the statistical analyses.

Lastly, I wish to express my love and gratitude to my beloved families; for their understanding and endless love, through the duration of my studies.
Vita

2005 ......................................................... No. 11 High School, Wuhan, China

2009 .......................................................... B.S. Psychology, Zhejiang University,

                                   Hangzhou, China

July-September 2009 ................................. Graduate Research Associate, Center for

                                   Cognitive Science, The Ohio State University

2009 to present .......................... Graduate Teaching Associate, Department

                                   of Psychology, The Ohio State University

Publications

Deng. W., & Sloutsky, V. M. (2010). The role of linguistic labels in categorization. In S.

                                   Ohlsson & R. Catrambone (Eds.), Proceedings of the XXXII Annual Conference of the


Fields of Study

Major Field: Psychology
# Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgments ...................................................................................................................... iii

Vita .............................................................................................................................................. iv

Table of Contents ....................................................................................................................... v

List of Tables ............................................................................................................................. vi

List of Figures ............................................................................................................................ vii

Introduction ................................................................................................................................ 1

Experiments ............................................................................................................................... 10

General Discussion ..................................................................................................................... 60

References ................................................................................................................................. 70
List of Tables

Table 1. Category structure used in learning in Experiment 1-4. .......................... 12
Table 2. Stimulus structure used in testing in Experiment 1-4 ............................... 12
Table 3. Category structure used in learning in Experiments 5 ............................... 44
Table 4. Structure of testing stimuli in Classification used in Experiments 5-7 ......... 46
Table 5. Structure of testing stimuli in Induction used in Experiments 5-7 ............... 46
List of Figures

Figure 1. The prototypes of stimuli used in this study.................................................. 11

Figure 2. Examples of stimuli of Category A and B used in Experiments 1-4............ 11

Figure 3. Examples of Classification and Induction test trials in Experiments 1-7........ 14

Figure 4. Proportion of category-consistent responses by feature match and testing condition in Experiment 1.................................................................................................. 16

Figure 5. Proportion of category-consistent responses in Experiment 2A ................. 20

Figure 6. Proportion of category-consistent responses in Experiment 2B............... 25

Figure 7. Proportion of category-consistent responses in Experiment 3A ............... 30

Figure 8. Proportion of category-consistent responses in Experiment 3B............... 35

Figure 9. Proportion of category-consistent responses in Experiment 4 ................. 40

Figure 10. Examples of stimuli of Category A and B used in Experiments 5-7......... 45

Figure 11. Proportion of category-consistent responses in Experiment 5 ............... 48

Figure 12. Proportion of label-consistent responses in Experiment 6 ..................... 54

Figure 13. Proportion of label-consistent responses in Experiment 7 ..................... 57
INTRODUCTION

Have we any reason, assuming that they (laws like the law of gravitation) have always held in the past, to suppose that these laws will hold in the future? If we are to know of the existence of matter, of other people, of the past before our individual memory begins, or of the future, we must know general principles of some kind by means of which such inferences can be drawn. On our answer to this question must depend the validity of the whole of our expectations as to the future, the whole of the results obtained by induction, and in fact practically all the beliefs upon which our daily life is based. (Russell, 1959, chapter 6)

Inductive inference is a critical component of our lives: upon learning that a rose produces attractive scent, a child might start smelling other flowers. As such, inductive inference, or generalizing knowledge from the known to the unknown, is crucial for applying learned knowledge to new situations. Examples of inductive inference are (a) inferring a property of a novel item given that a known item has this property or (b) inferring a category of a novel item given category membership of a known item. The former is referred to as projective induction and the latter as categorization. The term induction is used to refer to both projective induction and categorization (Sloutsky & Fisher, 2004a).
It has been well established that induction appears early in development (Gelman & E. Markman, 1986; Mandler & McDonough, 1996; Sloutsky & Fisher, 2004a), and 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). However, 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 linguistic labels in early development and developmental changes in this role.

Some researchers have argued that even infants expect labels to mark categories (Waxman, 1995, 2003). According to this “knowledge-based” approach, labels affect induction because they function as category markers used for representing a category and induction is driven by conceptual knowledge—the category information. People including young children, when they perform an inductive generalization, first identify the category of an entity and then generalize properties of that entity to other members of the category. This position is mainly supported by the experimental evidence from the innovative research by Gelman and E. Markman (1986), which showed that young children predicted that objects grouped by a common label share a property even when they differ in appearance. In a series of experiments, young children were given a triad task in which stimuli consisted of one target and two test items. One test item shared the label with the target item but looked dissimilar from it, whereas the other test item looked similar to the target but had a different label. Child participants were informed that one test item had a
special hidden property (e.g., “hollow bones”) and the other test item had a different hidden property (e.g., “solid bones”); and then they were asked to generalize a hidden property to the target item. The results indicated that children were more likely to generalize the property of the test item which shared the same label as the target than the property of the test item which shared the similar appearance of the target (but see Sloutsky & Fisher, 2004a, Experiment 4, for diverging evidence and counterarguments). This finding was interpreted as evidence that children’s induction is based on category membership which is denoted by a particular label.

According to this position, a common label plays a critical role in identifying the common category. In particular, E. Markman and Hutchinson (1984) presented evidence that without labels 2- and 3-year-olds may group 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 the same objects taxonomically (e.g., the police car with a passenger car). There is also evidence (e.g., Gelman & Heyman, 1999) that young children are 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"). These results suggest that words (in the form of a noun) might be important category markers guiding both categorization and induction in young children.

This evidence, however, does not lend unequivocal support to the idea that words are category markers. For example, some researchers suggested that the contribution of labels is driven by attentional rather than conceptual factors (Napolitano & Sloutsky,
There is also evidence that labels contribute to the overall similarity of compared entities and thus to both categorization and induction. In one experiment using items that had been previously used by Gelman and E. Markman (1986), Sloutsky and Fisher (2004a) found little evidence that in their induction responses, 4- and 5-year-olds relied exclusively on linguistic labels. Sloutsky and Fisher (2004a) 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.

If this is the case, then early in development induction could be driven by low-level information (i.e., cross-item similarity) rather than by top-down conceptual knowledge. According to this view, conceptual knowledge (e.g., knowledge that members of the same category share important properties) is a product rather than a precondition of learning. Furthermore, generalization processes, including induction of properties, are grounded in perceptual and attentional learning mechanisms, such as statistical and attentional learning (French, Mareschal, Mermillod, & Quinn, 2004; Mareschal, Quinn, & French, 2002; McClelland & Rogers, 2003; Sloutsky & Fisher, 2004a; Sloutsky & Fisher, 2008; Smith, 1989; Smith, Jones, & Landau, 1996). Therefore, early in development, labels, like other features (e.g., shape, color, size, etc.), function as one of the characteristics or properties of an object rather than a category marker that represents category membership. The role of labels, however, may change in the course of development as the conceptual knowledge on categorization has been learned.
Some of these ideas have been implemented in a model of early generalization (SINC for “Similarity-Induction-Naming-Categorization”) proposed by Sloutsky and colleagues (Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004a; Sloutsky, Lo, & Fisher, 2001). The model accurately predicted similarity and categorization judgment in young children when labels were and were not introduced. According to SINC, early in development, both induction and categorization are based on the overall similarity of compared entities, with labels being features of objects that contribute to their overall similarity, rather than symbols denoting category membership.

In a set of follow-up experiments examining mechanisms underlying early induction, Sloutsky, Kloos, and Fisher (2007) tested 4-and 5-year-old children with two types of tasks: categorization and induction. Both tasks were performed with artificial animal-like categories in which appearance was pitted against category membership. The results indicated that children’s induction was driven by appearance similarity (and not by category membership) although they had readily acquired category-membership information during training. These results support the idea of similarity-based induction early in development and present evidence that labels, for children, are no more than features.

In short, according to one approach labels denote categories, and as such, they differ from object features, whereas according to the other approach, early in development labels do not qualitatively different from other features of entities. In an attempt to distinguish between labels being features and category markers, Yamauchi and A. Markman (1998; 2000) developed an innovative paradigm potentially capable of settling
the issue. The paradigm is based on the following idea. Imagine two categories A (labeled “A”) and B (labeled “B”), each having four binary dimensions (e.g., Size: large vs. small, Color: black vs. white, Shape: square vs. circle, and Texture: smooth vs. rough). The prototype of Category A has all values denoted by “1” (i.e., “A”, 1, 1, 1, 1) and the prototype of Category B has all values denoted by “0” (i.e., “B”, 0, 0, 0, 0). There are two inter-related generalization tasks – classification and projective induction, or inference. The goal of classification is to infer category membership (and hence the label) on the basis of presented features. For example, participants are presented with all the values for an item (e.g., ?, 0, 1, 1, 1) and have to predict category label “A” or “B”. In contrast, in the inference task participants have to infer a feature on the basis of category label and other presented features. For example, given an item (e.g., “A”, 1, ?, 1, 0), participants have to predict the value of the missing feature. A critical manipulation that could illuminate the role of labels is the “low-match” condition. For low-match inference, participants were presented with an item “A”, ?, 0, 1, 0, 0 (which was more similar to the prototype of Category B) and asked to infer the missing feature. For low-match classification, participants were presented with an item ?, 1, 0, 1, 0, 0 (which again was more similar to the prototype of Category B) and asked to infer the missing label.

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 and are treated as category markers, then inferring a label when features are given (i.e., a classification task) should elicit different performance from a task of inferring a feature when the label is given (i.e., a feature induction task). Specifically,
category-based responding should be more likely in induction tasks (where participants could rely on the category label) than in categorization tasks (where participants had to infer the category label). Upon finding predicted asymmetries between the two conditions, these researchers concluded that category labels differed from other features in that adult participants are more likely to treat labels as category markers instead of object features. These findings have been replicated in a series of follow-up studies (Yamauchi, Kohn, & Yu, 2007; Yamauchi & Yu, 2008; A. Markman & Ross, 2003). For example, Yamauchi, Kohn, and Yu (2007) examined patterns of mouse-tracking (a procedure that is similar to eye tracking) to examine attention allocated to labels when labels were introduced as category markers (e.g., “This is a dax”) or as denoting category features (e.g., “This one has a dax”). Results indicated that participants viewed these visually presented labels more often in the former condition than in the latter condition. There is a body of evidence indicating that adults tend to treat the category label as a category marker rather than a category feature.

Does the asymmetry between categorization and induction tasks found in adults also 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. We argue that labels function differently across development: whereas labels may denote category membership and act as category markers in adults, they are not likely to do so in children. Instead, labels are treated as perceptual features that contribute to the perceived similarity of compared entities in children and early induction is driven by similarity.
In addition, in the category A and B example above, both items presented in low-match classification and low-match induction are more similar to prototype B and if labels are more than features, participants should be more likely to infer the missing feature as belonging to category A in the inference task than to infer label “A” in the classification task. In contrast, if the label is just another feature, then a different pattern should emerge: relative performance on classification and induction tasks should depend on attentional weights of labels compared to those of other features. Specifically, if there are features that have a higher attentional weight than the label, then classification task (when a high saliency feature could be used to predict the label) should yield more A-responses than induction (when the label is used to predict the high salience feature). The present studies aim to address these questions and to provide evidence for this supposition by applying the paradigm in adults (Yamauchi and A. Markman, 2000) to young children.

The reported experiments were designed to address these issues. In Experiments 1-3, we attempted to replicate Yamauchi & Markman’ findings with adults and to extend the paradigm to young children, while manipulating the role of labels. In Experiment 4, we manipulated the way learning was introduced. In Experiment 5-7, we applied a variant of this paradigm to examining the role of labels in development. To achieve this goal, we (a) added a highly salient feature that along with the label distinguished between two categories and (b) gave participants either a classification or induction task. If labels are category markers, then introducing highly salient features should generate the same pattern of responses as reported by Yamauchi and Markman -- participants should rely on
labels and not on salient features. However, if labels are no more than features, then a different pattern should emerge—if a feature is more salient than the label, participants should rely on a salient feature rather than on the label. Overall, results indicate that while adult performance is equivocal, there is little evidence that young children treat labels as anything but object features.
EXPERIMENTS

EXPERIMENT 1

The goal of Experiment 1 was to replicate Yamauchi and A. Markman’s paradigm (2000) with adults by contrasting the impact of category labels and category features at the low level of feature match. Adult participants learned two categories of creatures and then responded to 40 classification questions and 40 induction questions, which were divided into high and low levels of feature match. Based on Yamauchi and A. Markman’s findings (2000), it was hypothesized that adults would regard labels as category markers. Therefore, we expected that participants would make category-consistent responses to induction questions more often than to classification questions at the low level of feature match.

Method

Participants

Seventeen adults (10 women) participated in this experiment. Participants were undergraduate students from the Ohio State University participating for course credit. One of them gave one was an outlier and these data were excluded from the analysis.

Materials

In all experiments reported here, the materials were colorful drawings of artificial creatures accompanied by a category label ("Flurp" or "Jalet"). Two categories of objects
Figure 1. The prototypes of stimuli used in this study.

Figure 2. Examples of stimuli of Category A and Category B used in Experiments 1-4.

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 (A0 and B0) by modifying the values of one of five features (see Figure 2). For example, to produce the stimulus A1, the value of the antenna is changed from 1 to 0 so that it has four features consistent with the prototype A0 and one feature
consistent with the prototype B0. 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 in Experiment 1-4.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>H</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
<th>Stimuli</th>
<th>H</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>B1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>B2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>B3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>B4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>B5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>B0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The value 1 = any of six dimensions identical to Category A (see Figure 1). The value 0 = any of six dimensions identical to Category B (see Figure 1). A = Category A; B = Category B. H = Head; B = Body; Hn = Hand; F = Feet; A = Antenna; L = Label. A0 and B0 are prototypes of each category and A1/B1–A5/B5 are individual exemplars.

Table 2. Stimulus structure used in testing in Experiment 1-4.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>H</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
<th>Match</th>
<th>Stimuli</th>
<th>H</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>B11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>B12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A13</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>High</td>
<td>B13</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>B14</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A15</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>B15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A21</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td>B21</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A22</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>B22</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A23</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Low</td>
<td>B23</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A24</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>B24</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>B25</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. High and low are two levels of feature match. A = Category A; B = Category B. H = Head; B = Body; Hn = Hand; F = Feet; A = Antenna; L = Label. Category-consistent responses were the ones consistent with the values indicated in the target features and target labels.
Similar to Yamauchi and A. Markman, there were two levels of similarity (or feature match) in current research: high and low. In the high-match condition, 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. In the low-match condition, each test stimulus had two features in common with the prototype of the corresponding category and three features in common with the prototype of the other category.

**Design and Procedure**

This and all other experiment reported here had a two (Test Condition: Classification vs. Induction) by two (Feature Match: High vs. Low) within-subjects design. The experiment was administered on a computer and controlled by E-prime 2.0 software. The procedure consisted of two phases, training and testing. During the training phase, participants were instructed that they should try to remember and distinguish two groups of artificial creatures, "Flurps" and "Jalets". 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 training phase (see Figure 3 for examples of testing trials). 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
A. Classification

Which group do you think this creature is more likely to come from?

Flurp or Jalet?

B. Induction

This is a Flurp

Which body part do you think is more likely to be under the cover?

Figure 3. Examples of Classification and Induction test trials in Experiments 1-7.

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 participant. Feedback was given in first six trials of each condition in order to help
participants get familiar with the specific task. No feedback in the other 40 trials in both conditions. The proportion of responses consistent with the category (category-consistent responses) from which the exemplar was derived (called "category-accordance responses" by Yamauchi & A. Markman, 2000) was the dependent variable (see Table 2).

Results and Discussion

The main results of Experiment 1 are shown in Figure 4. As can be seen in the figure, in the high-match condition there were no differences between Classification and Induction, whereas there was a marked difference in the high-match condition. Specifically, in the high-match condition, participants were more likely to produce category-consistent responding in the Induction condition than in the Classification condition.

The data were subjected to a 2 (testing type: Classification and Induction) × 2 (feature match: high and low) within-subjects ANOVA. The analysis revealed a significant testing type by feature match interaction, $F(1, 15) = 36.92, MSE = 0.37, p < 0.01$. In the high-match condition, there was no difference between the Classification and Induction condition, $t(15) = 1.71, p = 0.108$. In contrast, in the low-match condition, participants were significantly more likely to make category-consistent responses in the Induction condition than in the Classification condition, $t(15) = 4.95, p < 0.01$.

The results replicate Yamauchi and A. Markman (2000) pointing to the predicted asymmetry and suggesting that for adults labels might be more than objects features. In the following experiments, this paradigm was expanded to young children aiming to examine whether the asymmetry between categorization and induction tasks found in
Figure 4. Proportion of category-consistent responses by feature match and testing condition in Experiment 1.

adults also exist in children. Finding such an asymmetry would support the idea that labels are more than category features for children as well, whereas a symmetric performance in the classification and induction conditions would support the idea that labels are features.

EXPERIMENT 2

The goal of Experiment 2 was to examine the role of labels in early generalization by expanding the Yamauchi and A. Markman’s (2000) paradigm to young children. The general procedure in Experiment 2 was identical to that Experiment 1 except (1) the way in which the labels were introduced and (2) the way training was administered. In
contrast to Experiment 1 in which the label was presented as a single written word, labels in Experiment 2 were presented in a carrier phrase (e.g. “This is a Flurp.”).

Different kinds of training regimes were given to participants in Experiment 2A and Experiment 2B. In Experiment 2A, both adult and child participants were trained by Classification, with category labels (“Flurp” and “Jalet”) denoting groups of creatures. In Experiment 2B participants were trained by Induction, and the labels referred to something inside the creatures’ bodies (e.g., this one has a flurp inside its body).

As discussed earlier, we expected that adult participants would make category-consistent responses to induction questions more often than to classification questions at the low level of feature match and an asymmetric performance would be found between these two tasks in both training conditions. For young 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 no more than features.

EXPERIMENT 2A

Method

Participants

There were twelve adults (3 women) and thirteen preschool children (\(M = 55.8\) months; 7 girls) participating in this experiment. In this and all other experiments reported here, adult participants were undergraduate students from the Ohio State University participating for course credit, whereas child participants were recruited from
childcare centers, located in middle-class suburbs of a Midwestern city. Also in all experiments reported here, child participants were tested in a quiet room in their preschool by a female experimenter.

**Materials, Design, and Procedure**

The visual stimuli were identical to Experiment 1 (see Figure 2). The experiment consisted of two consecutive phases, training and testing. During training, both adult and child participants were instructed that there were two groups of creatures labeled as Flurp and Jalet. They were then presenting with creatures (one at a time) each accompanied by a category label presented in a carrier phrase “This is a Flurp (or Jalet).” The testing phase was identical to Experiment 1. In the Classification task, participants predicted the category label of a stimulus given information about all five features with the label covered. In the Induction task, participants predicted the value of one of five features given the other four features with the label uncovered. Similar to Experiment 1, the current experiment had a two (Test Condition: Classification vs. Induction) by two (Feature Match: High vs. Low) within-subjects design.

The procedures were identical for both adult and child participants except the way the instructions were presented and the questions were asked. Adult participants read instructions and questions on screen, whereas for children instructions as well as classification and induction questions were presented verbally by a female experimenter.

In addition, a memory check was administered after the main experiment to examine whether participants remembered two categories after completing all the tasks. Participants were presented with five trials of stimuli randomly generated from the
training structure (see Table 1) and were asked to recall the corresponding label of each stimulus. Children and adults exhibited memory accuracy of 91.7% and 78% respectively. One of the adults answered less than 3 out 5 memory check questions correctly and these data were excluded from the analysis.

Results and Discussion

The main results are presented in Figure 5. Patterns of adult responding were very similar to those in Experiment 1. Whereas in the high-match condition there were no differences between Classification and Induction, there was a marked difference in the high-match condition. In particular adults were more likely to produce category-consistent responding in the Induction condition than in the Classification condition. In contrast, young children were likely to produce high levels of category-consistent responding in the high-match condition; this was not the case in the low-match condition.

Children’s and adults’ data were analyzed with two separate 2 (testing type: Classification and Induction) × 2 (feature match: high and low) within-subjects ANOVAs. For adults, there was a significant testing type by feature match interaction, $F(1, 10) = 24.63, MSE = 0.32, p < 0.01$. A paired-samples t test indicated that in the high-match condition there were no differences between the Induction and Classification conditions $t(10) = 0.30, p = 0.772$, whereas in the low-match condition participants were more likely to make category-consistent responses the Induction condition than in the Classification condition, $t(10) = 3.89, p < 0.01$. 

19
A. Adults

B. Children

Figure 5. Proportion of category-consistent responses by feature match and testing condition in Experiment 2A.
For children, there was a main effect of feature match (high-match > low-match), $F(1, 11) = 43.56$, $MSE = 1.33$, $p < 0.01$, as well as a main effect of testing type (Classification > Induction), $F(1, 11) = 14.77$, $MSE = 0.16$, $p < 0.01$. However, unlike adults in this experiment, there was no significant interaction between testing type and feature match, $F(1, 11) = 0.02$, $MSE = 0.00$, $p = 0.90$. In particular, in contrast to adults, low-match Induction in children did not exceed low-match Classification. In fact, the opposite was true – participants were somewhat more likely to generate category-accordant responses in low-match Classification than in low-match Induction.

Given that children exhibited a somewhat higher level of category-consistent responses in low-match Induction than low-match Classification, we performed analyses of individual patterns of response. Participants who made at least 13 out of 20 testing trials category-consistent responses (above chance binomial $p = .07$) in the high–match induction were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of category-based responses (15 out of 20 testing trials) were classified as category-based responders, whereas those providing at least 75% of responses based on the overall similarity were classified as feature-based responders. Of those 8 adults who were included in the analysis, 75% (six participants) were category-based responders with the remaining being mixed responders. No participant made their inductive inference based on overall similarity. In contrast, there were also 6 out of 12 children passing the criterion to be included in the analysis of response patterns, 50% (three participants) were feature-based responders with the remaining of them being mixed responders. Therefore, in contrast to adults who were performing induction on the
basis of category information, young children were more likely to perform inductive inferences on the basis of overall similarity.

Therefore, adult results replicate those of Yamauchi & Markman (2000) suggesting that adults are likely to treat labels as category markers. In contrast, children data provided little evidence that for these participants labels were more than features.

In sum, asymmetry between low-match Induction and low-match classification in adults coupled with symmetry in young children suggests that (a) for adults labels are processed differently from other features and (b) for young children, labels are no more than features. However, the fact that the experiment used only Classification training might have put Induction during testing at a disadvantage compared with Classification. In Experiment 2B, attempted to make Induction testing easier for children by making it consistent with the training condition. To achieve this goal, we administered Induction training to participants.

EXPERIMENT 2B

Method

Participants

There were eleven adults (6 women) and twelve preschool children (M = 54.0 months; 4 girls) participating in this experiment. In contrast to the participants trained by Classification in Experiment 2A, participants in this experiment were trained by Induction in which they were presented all five features of a creature and told that the creature had something special (Flurp or Jalet) inside its body.
Materials, Design, and Procedure

The visual stimuli were identical to those in reported above experiments. Similar to other experiments reported here, the current experiment consisted of two consecutive phases, training and testing. During the Induction training, in contrast to Experiment 2A, both adult and child participants were instructed that there were two groups of creatures, with member of each group having something special inside its body. One group of creatures had a Flurp inside its body, whereas the other had a Jalet. During training, participants were presented with one creature at a time and told:" This one has a Flurp (or Jalet) inside its body." The testing phase was identical to Experiment 2A, with participants presented with both the Classification task and the Induction task. In the Classification task, they were asked to predict the feature label of a stimulus given information about all five features with the feature label covered (e.g., Is it from the group with Flurp or Jalet inside the body?). In the Induction task, they were asked to predict the value of one of five features given the other four features and the feature label. Similar to the way of presenting instructions and questions in Experiment 2A, the adult participants read the instructions and the questions presented on screen, whereas for children both instruction and questions were presented verbally by a female experimenter.

Similar to Experiment 2A, a memory check was administered after the main experiment with all child and adult participants exhibiting memory accuracy of 83.3% and 72.7 % respectively. Two of the adults answered less than 3 out 5 memory check questions correctly and these data were excluded from the analysis.
Results and Discussion

The main results of Experiment 2B are shown in Figure 6. Patterns of responding were very similar to those in Experiment 2A. In adults, there were no differences between Classification and Induction in the high-match condition, whereas there was a marked difference in the low-match condition. In particular adults were more likely to produce category-consistent responding in the Induction condition than in the Classification condition. In contrast, young children were likely to produce high levels of category-consistent responding only in the high-match condition, but not in the low-match condition.

The data were subjected to two separate 2 (testing type: Classification and Induction) by 2 (feature match: high and low) within-subjects ANOVAs. For adults, there was a significant testing type by feature match interaction, $F(1, 8) = 9.92, MSE = 0.26, p < 0.05$. Specifically, in the low-match condition, participants were more likely to provide category-consistent responses in the Induction condition than in the Classification condition, $t(8) = 2.79, p < 0.05$. However, there was no significant difference in these two testing types in the high-match condition, $t(8) = 0.53, p = 0.613$.

For children, there was a main effect of feature match, $F(1, 11) = 86.84, MSE = 2.50, p < 0.01$, with participants being more likely to provide category-consistent responses in the high-match than in the low match condition. At the same time, neither the main effect of testing nor the interaction reached significance, both $p$s > 0.06.

An analysis of individual responses in both adults and children was performed to compare individual patterns in the Induction condition. For adults, participants who made
A. Adults

Figure 6. Proportion of category-consistent responses by feature match and testing condition in Experiment 2B.

B. Children

Figure 6. Proportion of category-consistent responses by feature match and testing condition in Experiment 2B.
at least 13 out of 20 testing trials category-consistent responses (above chance binomial $p = .07$) in the high-match induction (6 out 9 participants) were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of category-based responses (15 out of 20 testing trials) were classified as category-based responders, whereas those providing at least 75% of responses based on the overall similarity were classified as feature-based responders. Of the 13 adults who were included in the analysis, 83.3% (five participants) were category-based responders with the remaining being mixed responders. No participant made their inductive inference based on overall similarity. In contrast, there were also 11 out of 12 children passing the criterion to be included in the analysis of response patterns, 18.2% (two participants) were feature-based responders with the remaining of them being mixed responders. Therefore, in contrast to adults who were performing induction on the basis of category information, young children were mixed-responders.

These results showed similar pattern to that of Experiment 2A. Adults in Experiment 2B showed an asymmetric performance in Classification and Induction tasks, thus replicating findings of Yamauchi & Markman (2000). In contrast, 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.

A close examination of the difference between children’s performance in Classification and Induction tasks in both Experiment 2A and Experiment 2B suggests that child participants may be affected by different training procedures which may
facilitate children’s performance in corresponding testing tasks. In Experiment 2A, children were trained by classification in which they learned two categories denoted by different labels. In contrast, children in Experiment 2B were given induction training in which they were instructed to infer a special element inside the creature from each category. A better performance in Classification task than in Induction task was found in Experiment 2A may be elicited by the classification training. This issue is further addressed in Experiment 4.

In contrast to adults, children’s performance in Classification and Induction tasks were equivalent when they were trained either by classification or induction. These results suggest whereas adults may treat labels as category markers, for young children, labels are not different from other features. As a result, children relied on the overall similarity rather than on labels in both Classification and Induction tasks.

However, if the label is just one of the features, then category information might not be accessible for children in Experiment 2. Would children use category information if category membership is represented by a more general category marker than a novel label? We address this issue in Experiment 3.

**EXPERIMENT 3**

In Experiment 3, category information was communicated by familiar descriptions rather than by novel labels. The same stimuli structure was used as previous experiments but the labels (Flurp and Jalet in previous experiments) were not introduced. Participants learned two categories either by classification (in Experiment 3A) or induction (in
Experiment 3B). In contrast to the classification training in Experiment 2A, participants in Experiment 3A were instructed that creatures were divided into two groups -- friendly pets and wild creatures. In Experiment 3B, participants were trained by induction and were told that there were two groups of creatures living in different places. Classification and Induction tasks were administered after training and the category-consistent responses were recorded. Participants’ response patterns as a function of two levels of feature match (i.e., high and low) were examined and compared with the performance patterns when labels were provided in Experiment 2.

EXPERIMENT 3A

Method

Participants

There were thirteen adults (2 women) and nine preschool children ($M = 56.6$ months; 5 girls) participating in this experiment. All participants were given Classification training with two groups of stimuli categorized as “friendly pet” and “wild creature”.

Materials, Design, and Procedure

The visual stimuli were identical to previous experiments. The experiment consisted of two consecutive phases, the Training and Testing. During training, both adults and children participants were instructed that there were two groups of creatures, friendly pets and wild creatures. And then they were trained by presenting trials of stimuli and told: "This is a friendly pet (or a wild creature)." The testing phase was identical to Experiment 2A. The participants were given both the Classification task and the Induction task. In the
Classification task, participants were asked to predict the category membership of an item given information about all five features (e.g., Is this a friendly pet or a wild creature?). In the Induction task, they were asked to predict the value of one of five features given the other four features and the category membership. Similar to the way of presenting instructions and questions in Experiment 2A, the adult participants read the instructions and the questions presented on screen, whereas children were instructed and asked both classification and induction questions by a female experimenter.

A memory check, identical to Experiment 2A, was administered after main experiment to examine whether participants could remember the stimuli and correspondent category memberships. Both children and adults exhibited high memory accuracy (83.3% and 86.2% respectively), with no participant answering less than three out of five memory check questions correctly.

Results and Discussion

The main results of Experiment 3A are shown in Figure 7. The overall pattern is strikingly similar to that in Experiment 2. In adults, there were no differences between Classification and Induction in the high-match condition, whereas there was a marked difference in the high-match condition. In particular adults were more likely to produce category-consistent responding in the Induction condition than in the Classification condition. In contrast, young children were likely to produce high levels of category-consistent responding only in the high-match condition, but not in the low-match
A. Adults

B. Children

Figure 7. Proportion of category-consistent responses by feature match and testing condition in Experiment 3A.
condition. These data were analyzed with two separate 2 (testing type: Classification vs. Induction) by 2 (feature match: High vs. Low) within-subjects ANOVAs.

For adults, there was a significant interaction between testing type and feature match, $F(1, 12) = 36.23, MSE = 0.48, p < 0.01$. Similar to previous experiments and to Yamauchi and Markman (2000), in the low-match condition, adults were more likely to provide category-consistent responses in the Induction condition than in the Classification condition, $t(12) = 4.78, p < 0.01$. At the same time, there was no significant difference in these two testing types at the high level of feature match, $t(12) = 0.76, p = 0.461$.

For children, there was a significant main effect of feature match, $F(1, 8) = 35.66, MSE = 2.15, p < 0.01$. Specifically, children were more likely to provide category-consistent responses in the high-match than in the low-match condition. At the same, neither the main effect of testing, nor the interaction approached significance, both $p s > 0.28$.

An analysis of individual responses in both adults and children was performed to compare individual patterns in the Induction condition. For adults, participants who made at least 13 out of 20 testing trials category-consistent responses (above chance binomial $p = .07$) in the high-match induction (all 13 adults passed the criterion) were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of category-based responses (15 out of 20 testing trials) were classified as category-based responders, whereas those providing at least 75% of responses based on the overall similarity were classified as feature-based responders. Of the 13 adults who were included in the analysis, 69.2% (nine participants) were category-based responders with
the remaining being mixed responders. No participant made their inductive inference based on overall similarity. In contrast, there were also 8 out of 9 children passing the criterion to be included in the analysis of response patterns, 62.5% (five participants) were feature-based responders with the remaining of them being mixed responders. Therefore, in contrast to adults who were performing induction on the basis of category information, young children were more likely to perform inductive inferences on the basis of overall similarity.

The relative contribution of category membership and category features in inductive inference was compared in this experiment. These results revealed the same pattern as in Experiment 2. While adults reliably used category membership when appearance information was pitted against category information, children ignored the category information that was denoted by general category markers and relied on similarities among features when performing induction. In Experiment 3B, we extend this paradigm to Induction training.

**EXPERIMENT 3B**

**Method**

*Participants*

There were twelve adults (6 women) and nine preschool children (M = 53.6 months; 3 girls) participating in this experiment. All participants were given the Induction training. In contrast to the Induction training in Experiment 2B in which the labels “Flurp” and “Jalet” were introduced, participants in this experiment were trained by Induction in
which they were presented all five features of a creature and told that the creature lived in different placed (i.e., in the forest or in the sea).

**Materials, Design, and Procedure**

The visual stimuli were identical to previous experiments and the experiment consisted of two consecutive phases, Training and Testing. During Training, in contrast to Experiment 2B, both adults and children participants were instructed that there were two groups of creatures and they lived in different places. One group of creatures lived in the forest while the other lived in the sea. On each trial, they were presented with an item and told: "This one lives in the forest (or in the sea)." Their task was to predict a missing feature. Upon responding, they were given yes/no feedback. The testing phase was identical to that in Experiment 2B: participants were presented with both Classification and Induction testing. In the Classification task, participants were asked to predict where each item lives, given information about all five features. In the Induction task, they were asked to predict the value of one of five features given the other four features and information where it lives. Similar to the way of presenting instructions and questions in Experiment 2B, the adults participants read the instructions and the questions presented on screen, whereas children were instructed and asked both classification and induction questions by a female experimenter.

A memory check was administered after main experiment to examine whether participants could remember the stimuli and correspondent living places. Adults exhibited memory accuracy of 76.7% with one of the adults answered less than 3 out 5 memory check questions correctly and these data were excluded from the analysis.
Children exhibited high memory accuracy (84.4%), with no participant answering less than three out of five memory check questions correctly.

Results and Discussion

The main results are shown in Figure 8 and again the overall pattern is similar to that in the above reported experiments: adults exhibited asymmetry between low-match Induction and Classification, whereas children exhibited symmetric responding. The data were analyzed with two separate 2 (testing type: Classification and Induction) by 2 (feature match: High vs. Low) between-subjects ANOVAs.

For adults, there was a significant testing type by feature match interaction, $F(1, 10) = 37.56, MSE = 0.38, p < 0.01$. In the low-match condition, category-consistent responses were more likely to be given in the Induction condition than in the Classification condition, $t(10) = 4.08, p < 0.01$, which was not the case for the high-match condition.

For children, in contrast to previous experiments, there was a significant testing type by feature match interaction, $F(1, 8) = 15.75, MSE = 0.07, p < 0.01$. Similar to adults, in the low-match condition, they were more likely to provide category-consistent response in the Induction than in the Classification condition, $t(8) = 2.51, p < 0.05$, which was not the case in the low-match condition, $t(8) = 0.92, p = 0.38$. Furthermore, unlike adults who were above chance in relying on category information in low-match induction, $p < 0.05$, young children were not different from chance, $p = 0.56$.

An analysis of individual responses in both adults and children was performed to compare individual patterns in the Induction condition. For adults, participants who made
Figure 8. Proportion of category-consistent responses by feature match and testing condition in Experiment 3B.
at least 13 out of 20 testing trials category-consistent responses (above chance binomial p = .07) in the high-match induction (7 out of 11 participants) were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of category-based responses (15 out of 20 testing trials) were classified as category-based responders, whereas those providing at least 75% of responses based on the overall similarity were classified as feature-based responders. Of those 7 adults who were included in the analysis, 71.4% (five participants) were category-based responders with the remaining being mixed responders. No participant made their inductive inference based on overall similarity. In contrast, there were also 6 out of 9 children passing the criterion to be included in the analysis of response patterns and all of them were mixed responders. Therefore, in contrast to adults who were performing induction on the basis of category information, young children were mixed responders.

Overall, results of Experiment 3B replicated and further extended results of Experiments 1, 2, and 3A. Most critically, children’s performance improved in the low-match Induction. This effect might be indicative of the fact that when category information is communicated through familiar description rather than through novel labels, young children are more likely to rely on this information when performing induction. However, it is also possible the improvement stemmed from the fact that participants were trained by induction. It is conceivable that training in induction facilitated their performance in induction testing. This possibility was examined in Experiment 4.
EXPERIMENT 4

The goal of Experiment 4 was to eliminate the possibility of condition-specific training effects by intermixing induction and classification training. There were two training conditions, with participants randomly assigned either to classification-label and induction-descriptor condition or to classification-descriptor and induction-label condition. As a result, all participants received both classification and induction training.

Method

Participants

There were twenty-six adults (8 women) and twenty preschool children ($M = 55.6$ months; 13 girls) participating in this experiment. One of the children was interrupted during the experiment. These data were excluded from the analysis.

Materials, Design, and Procedure

The experiment had two between-subjects training conditions: (1) classification-label (CL) and induction-descriptor (ID), (2) classification-descriptor (CD) and induction-label (IL). Across the conditions, participants were presented with the same visual stimuli used in previous experiments. Participants in CL and ID condition were trained with 20 classification trials and 20 induction trials. The correspondent testing trials (40 trials for each testing condition) were administered after all training trials. The CL trials were identical to those of Experiment 2A and the ID trials were identical to those of Experiment 3B. For example, participants in the CL-ID condition would be trained by classification with labels (e.g., “This is a Flurp.”) and induction with descriptions (e.g.,
“This one lives in the forest.”) first. They were then presented with test trials of both Classification with labels (to predict the label given all other features) and Induction with descriptions (to predict a missing feature given other four features and the living place).

The order of CL and ID was counterbalanced between subjects. The procedure of CD and IL condition was similar, but the CD trials were identical to those of Experiment 2B and the IL trials were identical to those of Experiment 3A. Participants in the CD-IL condition would be trained by classification with descriptors (e.g., “This is a friendly pet.”) and induction with labels (e.g., “This one has a Jalet.”) first and then were given Classification (with descriptors) and Induction (with labels) tasks.

Similar to previous experiments, a memory check was administered after the main experiment. Participants were presented with ten trials of stimuli randomly generated from the training structure (see Table 1) and were asked to recall either the corresponding category label or descriptor of each stimulus (i.e., “Flurp/Jalet” or “friendly pet/wild creature” for classification conditions) or the corresponding category information either in terms of label or description of each stimulus (i.e., “has Flurp/Jalet inside” or “living in the forest/sea” for induction conditions). Child and adult participants exhibited memory accuracy of 88.4% and 72.8% respectively, with five of the adults and one of the children answering less than 6 out of 10 memory check questions correctly. These data were excluded from following analysis.
Results and Discussion

Analyses across conditions were performed to investigate the patterns of category-consistent responses to Classification and Induction questions in high and low levels of feature match both in adults and young children. Specifically, participants’ performance in different types of tasks was compared when there were labels and when there were no labels but descriptors respectively. The main results are shown in Figure 9 and again the overall pattern is similar to that in the above reported experiments: adults exhibited asymmetry between Induction and Classification, whereas children exhibited symmetric responding. For each age group, adults and children, the data were analyzed with two separate 2 (Testing Type: Classification vs. Induction) × 2 (Feature Match: High vs Low) mixed ANOVAs, with testing type as a between-subjects factor and feature match as the within-subjects factor.

**Adults** When labels were introduced, the analyses revealed an interaction between testing type and feature match for adults, $F(1, 19) = 16.31$, $MSE = 0.26$, $p < 0.01$. At the low level of feature match, for adults, category-consistent responses made in the Induction questions were more than in the Classification questions, $t(19) = 3.15$, $p < 0.01$. The results revealed similar patterns for Classification and Induction when there were no labels but descriptors, with an interaction between testing type and feature match, $F(1, 19) = 13.53$, $MSE = 0.25$, $p < 0.01$. At the low level of feature match, for adults, category-consistent responses made in the Induction questions were more than in the Classification questions, $t(19) = 2.32$, $p < 0.05$; whereas category-consistent responses
A. Adults

Figure 9. Proportion of category-consistent responses by feature match and testing condition in Experiment 4.
Figure 9 continued

B. Children

![Graph showing proportion of category-consistent responses for children in feature match conditions.](image)

- Classification (with label)
- Induction (with label)

![Graph showing proportion of category-consistent responses for children in feature match conditions with descriptor.](image)

- Classification (with descriptor)
- Induction (with descriptor)
made in Classification were more than in the Induction at the high level of feature match, \(t(19) = 2.27, p < 0.05\).

**Children** The results revealed similar patterns for Classification and Induction either when labels were introduced or when there were no labels but descriptors. In contrast to adults, children did not show an interaction between testing type and feature match when labels were introduced, \(F(1, 16) = 0.11, MSE = 0.001, p = 0.742\) (with labels); \(F(1, 16) = 1.58, MSE = 0.007, p = 0.277\) (with descriptors).

By manipulating the way of learning and comparing the performance across conditions in Experiment 4, we trained subjects by classification and induction and compared their performance across conditions. The patterns of category-consistent responses for adults and children were consistent with those in Experiment 2 and Experiment 3, except an interaction of Classification and Induction, which might have reflected effects of induction training, in children in Experiment 3B. While adults relied on labels to make inductive inferences when label was pitted against similarities among other features, children’s induction was similarity-based. These results suggest that labels, for adults are category markers; whereas for young children, they are no more than category features.

**EXPERIMENT 5**

In Experiments 1-4, adult and child participants consistently showed different patterns of responding. The asymmetric performance found in adults suggests that labels
are regarded as category markers denoting category-membership information for adults. However, this is not the case for young children. The symmetry between Classification and Induction provides little evidence that labels are treated as category markers for young children. Instead, they are processed the same way as other features.

It seems that children’s similarity-based induction reflects that they need to collect information from various features in order to make inductive judgments. Do children have the ability to rely on a single source of information (e.g., a salient feature) to make inductive inferences, just like adults’ relying on label alone? The goal of Experiment 5 was to address this question. The stimuli were created as animations in which the head of the creature was moving in order to increase the salience of one feature of the stimuli. The procedure consisted of two consecutive phases: training and testing. Children were told that there were two groups of creatures one living in the forest and one in the sea. The primary difference with previous experiments was that one of the features was animated, thus becoming highly salient. After training participants were presented with Classification and Induction test questions. Similar to previous experiments, we compared the proportion of category-consistent responses in Classification and Induction at different levels of feature-match.

Method

Participants
Twelve preschool children (\(M = 55.9\) months; 5 girls) participated in this experiment.
Table 3. Category structure used in learning in Experiments 5.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>B</th>
<th>Hn</th>
<th>F</th>
<th>A</th>
<th>L</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The value 1 = any of six dimensions identical to Category A (see Figure 1). The value 0 = any of six dimensions identical to Category B (see Figure 1). A = Category A; B = Category B. H = Head; B = Body; Hn = Hand; F = Feet; A = Antenna; L = Label. A0 and B0 are prototypes of each category and A1/B1 – A4/B4 are individual exemplars.

**Materials, Design, and Procedure**

The visual stimuli were identical to previous experiments except the animated heads. Two categories were created using five features varying in color and shape (see Figure 10). As shown in Table 3, the two categories have a family-resemblance structure, which is derived from two prototypes (A0 and B0) by modifying the values of one of four features – antenna, hands, body and feet. For example, to produce stimulus A1, the value of the head, the hands, body and feet were set to 1; whereas, the value of the antenna was set to 0. As a result, the former four features were consistent with prototype A0, whereas the latter feature was consistent with prototype B0.

To set up a proper competition between the category information (which did not vary across the exemplars), and a feature, the value of one feature (the head) was also fixed within each category. In addition, to make the fixed feature highly salient, the head was animated using Macromedia Flash MX software. The head of one category was pink and moved up and down; whereas, for the other one, the head was blue and moved sideways.
The entire experiment consisted of two phases, identical to Experiment 3B, the learning phase and the testing phase. Two levels of feature match between the test item and the prototype of the corresponding category were used, high and low (see Table 4 and 5). As shown in the table, in the low-match condition there was only one feature (i.e., the moving head) in common with the respective prototype; whereas, in the high-match condition there were four such features. The critical condition was low-match induction where only the moving head was in common with the prototype of the corresponding category; whereas, three features and the category information (living place) were common with the prototype of the contrasting category. Therefore, if participants rely on the category information, they should choose the feature from the contrasting category, thus exhibiting a high level of category-based responding. In contrast, if they rely on the moving head, they should exhibit a low level of category-based responding. In all other conditions, there was no conflict between the category information and the moving head,
Table 4. Structure of testing stimuli in Classification used in Experiments 5-7.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Category A</th>
<th>Category B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Hn</td>
</tr>
<tr>
<td>A11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A13</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A22</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. High and low are two levels of feature match. A = Category A; B = Category B. H = Head; B = Body; Hn = Hand; F = Feet; A = Antenna; L = Label.

Table 5. Structure of testing stimuli in Induction used in Experiments 5-7.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Category A</th>
<th>Category B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Hn</td>
</tr>
<tr>
<td>A11</td>
<td>?</td>
<td>1</td>
</tr>
<tr>
<td>A12</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>A13</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>A14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A21</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>A22</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>A23</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. High and low are two levels of feature match. A = Category A; B = Category B. H = Head; B = Body; Hn = Hand; F = Feet; A = Antenna; L = Label.

and thus reliance on the moving head would result in a high level of category-based responding (see Table 4 and 5).

Similar to previous experiments, a memory check was administered after the main experiment with all participants exhibiting high memory accuracy (93%), with no participant answering correctly less than three out of five memory check questions correctly.
Results and Discussion

The main results of Experiment 5 are shown in Figure 11. The data were analyzed with a 2 (testing type: Classification vs. Induction) × 2 (feature match: High vs. Low) within-subjects ANOVA. Most importantly, there was a significant testing type by feature match interaction, $F(1, 11) = 129.76$, $MSE = 1.51$, $p < 0.01$. In the high-match condition, there was no difference between Classification and Induction, $t(11) = 0.84$, $p = 0.417$, whereas in the low-match condition, participants were more likely to category-consistent responses in Classification than in Induction condition, $t(11) = 19.90$, $p < 0.01$. Most importantly, when category information and the moving head pointed to the same response (i.e., high-match Induction), children were above chance in relying on these two features, one sample $t (11) = 11.86$, $p < 0.01$, whereas when the category membership was pitted against the salient feature (i.e., in low-match induction), children were significantly above chance in relying on the moving head to infer missing features, one-sample $t (11) = 13.68$, $p < 0.01$. Therefore, while in all previous experiments, there was no evidence that children rely on label or category information in the low-match induction, they had no difficulty relying on a singly highly salient feature.

An analysis of individual responses was performed to analyze the individual patterns in the Induction condition. Participants who made at least 75% (6 out of 8 testing trials) category-consistent responses in the high-match induction (11 out of 12 participants) were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of category-based responses were classified as category-based responders, whereas those providing at least 75% of responses based on the moving head.
were classified as feature-based responders. Of those 11 children who were included in the analysis, all of them were feature-based. Therefore, young children were able to reliably perform inductive inference on the basis of a highly salient feature.

These results indicate that children can rely on a single salient feature and that they are more likely to rely on this feature than on category information. Overall, across all the reported 5 experiments children either failed to rely on label or category information (Experiments 2-4) or relied on a salient feature (Experiment 5), thus providing little evidence that they treated labels as category markers. At the same time, tended to rely on label (or category information) and not on the overall similarity. Taken together these
results suggest that early in development labels are treated as features, but they may become category markers in the course of development.

Furthermore, if labels are category markers, then introducing highly salient features should generate the same pattern of responses as reported by Yamauchi and Markman -- participants should rely on labels and not on salient features. However, if labels are no more than features, then a different pattern should emerge— if a feature is more salient than the label, participants should rely a salient feature rather than on the label. Specifically, if there are features that have a higher attentional weight than the label, then classification task (when a high saliency feature could be used to predict the label) should yield more category-consistent responses than induction (when the label is used to predict the high salience feature). Experiment 6 and 7 will address this issue by pitting label against a highly salient feature.

**EXPERIMENT 6**

**Method**

**Participants**

Thirteen preschool children ($M = 55.5$ months; 6 girls) recruited from local childcare centers were tested in a quiet room in their preschool by a female experimenter. One of these participants was unable to finish because of school activities, and these data were excluded from the analysis. In addition, 30 undergraduate students (16 women) from the Ohio State University participated for course credit. One of these participants answered a phone call during the experiment and these data were excluded from the analysis.
Materials

The stimuli and structures were identical to those in Experiment 5 (see Figure 10 and Table 3-4) except novel labels ("flurp" and "jalet") were introduced in order set up a proper competition between the label (which did not vary across the exemplars), and a salient feature, the value of one feature (the head) was also fixed within each category.

Design and Procedure

The experiment, identical to Experiment 5, had a two (Test Condition: Classification vs. Induction) by two (Feature Match: High vs. Low) within-subjects design. The experiment was administered on a computer and controlled by E-prime 2.0 software and it consisted of two consecutive phases, training and testing. During training, participants were instructed to try to remember and distinguish two groups of creatures, labeled "flurp" and “jalet". The experimenter read aloud the instructions to children; whereas, adults silently read instructions to themselves. Then participants were given 24 trials, each presenting a creature produced from the structure shown in Table 3. On each training trial, both child and adult participants saw a stimulus with a corresponding label printed above it and heard the label (e.g., “This is a flurp”).

Training was followed immediately by testing (see Figure 3 for examples of testing trials). In the Classification condition, participants were asked to predict a label (i.e., which group a creature in question was more likely to belong to, flurp or jalet). In the Induction condition, participants were asked to predict the value of one of four unfixed features (e.g., the antenna) by choosing between the presented in isolation antenna of a
prototypical flurp and of a prototypical jalet. Adults responded by pressing the keyboard; whereas, children’s verbal responses were recorded by the experimenter.

A memory check was administered after the main experiment to examine whether participants remembered two categories after completing all the tasks. Participants were presented with five trials of stimuli randomly generated from the training structure (see Table 1) and were asked to recall the corresponding label of each stimulus. Both children and adults exhibited high memory accuracy (94% and 100% respectively), with no participant answering less than three out of five memory check questions correctly.

The order of Classification and Induction conditions was counterbalanced across participants and the order of the high- and low-match testing trials within conditions was randomized for each participant. The first six trials in each condition were used as a warm-up and were high-match trials in which feedback identifying the correct response was provided. The remaining 16 testing trials in each condition were not accompanied by feedback and were used for data analysis. In contrast to the dependent variable we used in previous experiments, that is, the proportion of the category-consistent responses, the proportion of label-consistent responses was the dependent variable in this and following experiments. Specifically, in the Classification task (where participants predicted labels) the responses consistent with the respective prototype were identified as label-consistent. In the Induction task, the responses that were in accordance with the label presented were identified as label-consistent. If the label is a category marker, then participants should rely on the label, even when the label is pitted against a highly salient
feature (i.e., in low-match induction). However, if the label is a feature, participants may fail to rely on the label, when the label is pitted against a highly salient feature.

Results and Discussion

The main results are presented in Figure 12. As can be seen in the figure, in the Classification condition, regardless of the level of feature match, children generated a high level of label-consistent responses. Perhaps not surprisingly, across the levels of feature match participants accurately predicted labels by relying on the moving head. In contrast, when the moving head pointed to one response and the label pointed to another response (in low-match induction), there was a marked drop in reliance on the label with children relying primarily on the moving head.

Children’s data were analyzed with a two (Test Condition: Classification vs. Induction) by two (Feature Match: High vs. Low) within-subjects ANOVA. There was a significant test condition by feature match interaction, $F(1, 11) = 82.92$, $MSE = 1.02$, $p < 0.01$, $η^2 = 0.883$: participants made comparable proportions of label-consistent responses in high- and low-match classification, $p > 0.10$; whereas, in the Induction condition they made more label-consistent responses in the high-match than in the low-match condition, $t(11) = 12.85$, $p < 0.01$, $d = 5.27$. Furthermore, when the label was pitted against the salient feature (i.e., in low-match induction), children were significantly above chance in relying on the moving head to infer missing features, one-sample $t(11) = 10.56$, $p < 0.01$, $d = 3.05$. 
As can be seen in Figure 12A, for adults there was also a test condition by feature match interaction, $F(1, 28) = 5.90$, MSE =0.20, $p < 0.05$, $\eta^2 = 0.176$: participants were likely to make label-consistent responses in the Classification condition, regardless of the feature match, $p > 0.10$; whereas, in the Induction condition they made more label-consistent responses in the high-match than in the low-match condition, $t(28) = 2.94$, $p < 0.01$, $d = 0.82$. However, in contrast to children, when the label was pitted against the salient feature (i.e., in low-match induction), adults’ responding was not different from chance, $p > 0.10$.

Because adults’ performance in the low-match induction was at chance, we deemed it necessary to analyze individual patterns of responses in the Induction condition. Participants who made at least 75% (6 out of 8 testing trials) label-consistent responses in the high–match induction (19 out of 30 participants) were selected for the analysis of the response pattern in the low-match induction. Those providing at least 75% of label-based responses were classified as label-based responders, whereas those providing at least 75% of responses based on the moving head were classified as feature-based responders. Of those 19 adults who were included in the analysis, 31.5% (six participants) were feature-based responders and 37% (seven participants) were label-based responders, with the remaining 31.5% being mixed responders.

There were also 11 out of 12 children passing the criterion to be included in the analysis of response patterns and 91% of them (ten participants) were feature-based responders. That is, unlike adults who exhibited a trimodal response pattern, children uniformly relied on a highly salient feature (i.e., the moving head) rather than on the label.
A. Adults

![Graph showing proportion of label-consistent responses by feature match and testing condition in Experiment 6 for adults.]

B. Children

![Graph showing proportion of label-consistent responses by feature match and testing condition in Experiment 6 for children.]

Figure 12. Proportion of label-consistent responses by feature match and testing condition in Experiment 6.
to make inductive inferences, even when the salient feature was the single cue that was pitted against the combination of label and other features.

Overall, children relied on the salient feature (i.e., the moving head) rather than on the label, regardless of the condition and the level of feature match, thus providing little evidence that they treated labels as category markers. Adults’ performance was sensitive to the competition between the salient feature and the label, as evidenced by the trimodal distribution in low-match induction. This trimodal distribution raises a question concerning the role of labels in adults’ induction. When there was a salient feature competing with the label, only a third of adult participants consistently relied on category label, thus suggesting that these participants treated the label as a category marker.

However, it could be also argued that children and many adults failed to rely on the label because the labels were novel (e.g., Davidson & Gelman, 1990). These researchers argued that familiar labels have a lower memory load. Experiment 7 was designed to test this possibility by using familiar labels, some of which were used previously (i.e., “carrot-eater” in Gelman & Hayman, 1999).

**EXPERIMENT 7**

**Method**

**Participants**

Seventeen preschool children ($M = 54.4$ months; 10 girls) and 15 undergraduate students (4 women) participated.
Materials, Design, and Procedure

The stimuli and procedure were similar to those used in Experiment 6 except that familiar labels “Carrot-eater” and “Meat-eater” were used instead of “Flurp” and “Jalet”. Similar to Experiment 6, a memory check was administered after the main experiment with all participants exhibiting high memory accuracy (84% for children and 98% for adults), with no participant answering correctly less than three out of five memory check questions correctly.

Results and Discussion

The results are presented in Figure 13. Data in the figure were submitted to a 2 (Test Condition: Classification vs. Induction) by 2 (Feature Match: High vs. Low) between-subjects ANOVA. As can be seen in the Figure 13B, children’s performance was similar to that in Experiment 6: there was a significant test condition by feature match interaction, $F(1, 16) = 44.44$, $MSE = 0.83$, $p < 0.01$, $\eta^2 = 0.735$. While there was a difference in label-consistent responding across the feature match levels of Classification condition (93% vs. 85% for high- and low-match, respectively, $t(16) = 2.28$, $p < 0.05$, $d = 0.74$), there was a substantially bigger difference across the feature match levels of the Induction condition (84% vs. 32%, for high- and low-match, respectively $t(16) = 7.41$, $p < 0.01$, $d = 2.75$). In addition, when the label was pitted against the moving head (i.e., in low-match induction), participants were above chance in relying on the salient feature, $t(16) = 3.57$, $p < 0.01$, $d = 0.87$. 
A. Adults

B. Children

Figure 13. Proportion of label-consistent responses by feature match and testing condition in Experiment 7.
Adults’ performance is shown in Figure 13A. The results revealed significant main effects of test condition and feature match, with no interaction. Adult participants made more label-consistent responses in Classification than in Induction, $F(1,14) = 7.38$, $MSE = 0.482$, $p < 0.05$, $\eta^2 = 0.345$, and more label-based responding in the high-match than in the low-match condition, $F(1,14) = 14.72$, $MSE = 0.25$, $p < 0.01$, $\eta^2 = 0.513$. When the label was pitted against the moving head (i.e., in low-match induction), reliance on the label was marginally above chance, $p = 0.076$, $d = 0.49$.

As in Experiment 6, we analyzed individual patterns of responses. The analysis revealed that of those 11 adults who passed the 75% criterion, 18% (two participants) were feature-based responders and 64% (seven participants) were label-based responders, with the remaining 18% being mixed responders. In contrast, of those 15 children who passed the 75% criterion, 67% (ten participants) were feature-based responders and 7% (one participant) were label-based responders, with the remaining 26% being mixed responders. Therefore, with the familiar labels used in Experiment 2, adults were more likely to produce label-based responding than with novel labels used in Experiment 6, whereas children remained predominantly feature-based responders, although with familiar labels the proportion of feature-based responders was somewhat lower than with novel labels.

These findings, together with the results of Experiment 6, indicate for children, both familiar and novel labels generate similar pattern of responses – when performing induction, children relied on the highly salient perceptual feature rather than on the label. In contrast, about a third of adults in Experiment 6 and more than two-thirds in
Experiment 7 exhibited consistent label-based performance. These results point to an important developmental difference in the role of labels, suggesting that while many adults may treat familiar labels as category markers, this is not the case for young children.
GENERAL DISCUSSION

The reported research presented eight experiments designed to examine the role of labels in early generalization and changes in this role in the course of development. To achieve this goal, we used a paradigm pioneered by Yamauchi & Markman (1998, 2000). Recall that their paradigm was based on the following reasoning. If the label is a category marker, then participants should exhibit greater reliance on the label (when it is a sole predictor) than on a feature (when it is a sole predictor). Overall, the reported results present little evidence that early in development labels are category markers, whereas the results suggest that labels may become category markers in the course of development.

Summary of Findings

Several major findings stem from the reported experiments. First, across Experiment 1-4 adults exhibited the ability to rely on category label even when the label was pitted against appearance similarity, which was not the case for young children. In contrast, under no conditions did young children exhibit sole reliance on category label in their induction. At the same time, when a salient visual feature was introduced (Experiment 5), young children did perform induction by relying on this feature. Third, young children exhibited overwhelming reliance on a highly salient feature and not on a category label, whether the label was novel (Experiment 6) or familiar (Experiment 7). The results are more complicated in adults. When labels were novel, some adults exhibited consistent relying on the salient feature and some relied on the label. And fourth, familiar labels were more likely than novel ones to be used as category markers (see Davidson &
Gelman, 1990, for a discussion of potential reasons), but only by adults. Therefore, at least for some adults labels may become category markers. Taken together these results indicate that for young children and for some adults a category label is simply another feature as little reliance on category label was observed when it was pitted against a highly salient feature. In addition, under no condition induction performance exceeded classification performance, thus suggesting that the relation between classification and induction is not fixed. However, if labels are category markers, this relation should be fixed: in this case, it should be easier to predict a single feature knowing the category (marked by the label) than it is to predict the label knowing a single feature. In contrast, if the label is a just another feature, then the relation between classification and induction should be context-specific: if the label is pitted against less salient features, then induction is an easier task than classification, whereas if the label is pitted against a more salient feature, then classification becomes an easier task. Recall that there is much research with adults demonstrating that classification and induction tasks result in different performance and different category representation (see Markman & Ross, 2003, for a review; Hoffman & Rehder, 2010). Although the current research did not examine category representation, it did examine performance on classification and inference tasks. The current research presents clear evidence that when a category includes a highly salient feature, performance on the classification task does not drop below that of induction tasks (in fact, under all conditions, classification performance exceeded that of induction performance).
Current results (Experiment 6 and 7) demonstrating higher label-consistent performance in the classification condition, coupled with previous research demonstrating higher performance in induction condition (Yamauchi & Markman, 1998, 2000) demonstrate that the relative performance in classification and induction tasks depends on the relative salience of the label and other features. Although additional research is needed to examine category representation when category members include a highly salient feature, the current findings provide evidence that for young children and for some adults labels are no more than features.

The question regarding the role of language in generalization has generated considerable debate, with some arguing that words denoting categories have the special status of category markers and others arguing that words are akin to other features. Research reported here indicates that for young children (and for some adults) reliance on labels drops dramatically when labels are pitted against salient perceptual features, which should not have happened if labels were category markers. These results cast doubt on the view that early in development labels are category markers, suggesting instead that early in development labels are features, but they may become category markers in the course of development.

*Labels and the Mechanism of Generalization*

Although there is little disagreement that from early in development people are capable of performing inductive generalization, the mechanism underlying induction is a matter of debate. Some researchers have argued that inductive inference is category-based in that when making induction, people access the category of items in question.
Other researchers presented an alternative argument that at least early in development, inductive inference is driven by similarity of compared entities rather than by a common category membership. There is also little disagreement that under typical circumstances, it is difficult to distinguish between these possibilities. There have been at least three proposals as to how such a distinction could be made.

First, there is an argument that category-based and similarity based induction may result in different memory traces for studied items, with similarity-based induction resulting in more detailed verbatim memories and category-based induction resulting in less detailed gist-type memory. This argument has resulted in a set of studies demonstrating that (a) young children (who presumably perform similarity-based induction) retain more accurate memories of the studied items than adults (who presumably perform category-based induction) and (b) training young children to perform category-based induction attenuates their memory accuracy to the level of adults (Sloutsky & Fisher, 2004b; Fisher & Sloutsky, 2005).

The second idea is to experimentally dissociate category membership and similarity. If induction is category-based, it should follow category information, whereas if it is similarity-based, it should follow similarity information. In one such study (Sloutsky, Kloos, & Fisher, 2007), 4-5-year-olds learned two rule-based categories, with similarity not being predictive of category membership. Upon learning the categories, participants were presented with a set of induction trials in which they could rely on either category information or similarity information. Despite the fact that children successfully acquired
the categories and retained this knowledge throughout the experiment, their induction was similarity based.

The third proposal focuses on the role of labels in induction and is based on the idea that understanding of this role is critical for distinguishing among the proposed the mechanisms. The overall argument is that in order for induction to be category-based, people have to have access to category information that is independent of similarity. This is because if people access category information through similarity, then similarity-based induction and category-based induction are the same. One way of accessing category information is by category label. Therefore, in order for participants to perform category-based induction labels have to be category markers. Conversely, if labels are features contributing to similarity, participants are likely to be engaged in similarity-based induction.

The problem has been well formulated by Yamauchi & Markman (2000, p. 777). These researchers argue that if participants make inductive judgments based solely on the similarity (i.e., feature overlap) between a test item and the category representation, then performance on the inference task should be the same as performance on the classification task. In contrast, if category membership (communicated by the label) is not treated as just another feature, then performance on an inference task should differ from that on a classification task. Yamauchi and Markman (1998, 2000) presented extensive evidence that for adults labels are likely to be category markers. The current work replicates these findings with adults, while demonstrating that for young children labels are no more than features. If labels are indeed features contributing to similarity,
then the ability of young children to perform category-based induction is highly questionable.

Therefore, current research, in conjunction with earlier reported research (Sloutsky & Fisher, 2004a; Sloutsky & Fisher, 2004b; Fisher & Sloutsky, 2005; Sloutsky, et al, 2007) presents further evidence that evidence that early induction is similarity-based but it may become category-based in the course of development.

*Language and Cognition: Are Labels Features of Category Markers*

There has been much debate about the role of labels in generalization. Some believe that from very early labels are features of entities, whereas others believe that labels are category markers. While this question is critically important for understanding the mechanism of generalization, it has broader implications for understanding of the role of language in cognition and cognitive development. At the computational level of analysis (Marr, 1982), this approach assumes that words are not merely a part of stimulus input, but rather fulfill the role of supervisory signals directing and guiding learning. Thus, if two discriminable items share the same count noun (e.g., both are called “a dax”), the name serves as a top-down signal that the items are equivalent in some way (cf. Gliga, Volein, & Csibra, 2010). Another possibility is that early in development, words, just like any other perceptual feature, are first and foremost part of the input (Colunga & Smith, 2005; Plunkett, et al, 2008; Sloutsky & Fisher, 2004). The third possibility is that words start out as part of the stimulus input, but eventually become supervisory signals as well (Casasola & Bhagwat, 2007; Casasola, 2008; Arias-Trejo & Plunkett, 2010; Sloutsky, 2010).
Each of these possibilities presumes a different mechanism and dedicated neural architecture, and a different developmental trajectory. Distinguishing among them and understanding the mechanisms underlying the effect of words on category learning is of critical importance for understanding cognitive development. If from early in development words serve as supervisory signals, then top-down effects have to play a significant role in early cognitive development. Perhaps the most important implication is that at both the cognitive and the neural levels, the lower-level processes (such as discrimination and generalization) are subject to top-down control. Alternatively, if words become supervisory signals in the course of development, then top-down control does not have to exhibit early onset and could be itself the product of development. Therefore, understanding the role of labels in generalization has implication for most fundamental aspects of cognitive development as well as for understanding of the interaction between language and cognition.

Although additional research is needed, present research indicates that even at 4-5 years of age, labels function more like features than like category markers. Assuming that labels eventually become category markers for adults (e.g., Yamauchi & Markman, 1998, 2000, as well as experiments presented here), a theoretical and empirical challenge is to establish the developmental mechanism of this process.

*From Features to Markers? The Changing Role of Category Label in Generalization*

Recall that there is significant body of evidence that for adults labels are category markers rather than simple features. In addition to work by Yamauchi and Markman reviewed above, there is more recent evidence supporting this idea. For example,
Yamauchi, Kohn, and Yu (2007) examined patterns of mouse-tracking (a procedure that is similar to eye tracking) to examine attention allocated to labels when labels referred to the entire category (e.g., “This is a dax”) or as denoting category features (e.g., “This one has a dax”). Results indicated that participants viewed these visually presented labels more often in the former condition than in the latter condition. If labels are features as argued here and elsewhere (Sloutsky & Fisher, 2004a; Sloutsky, Kloos, & Fisher, 2007), how do labels become category markers?

One possible idea that was discussed elsewhere (e.g., Sloutsky, 2010) is that the contribution of labels to categorization and category learning hinges on (a) the ability to process cross-modal information and (b) the ability to attend selectively. Although neither of these abilities might be sufficient, both seem to be necessary.

First, there is a growing body of evidence that auditory input may affect attention allocated to corresponding visual input (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008). In particular, linguistic labels may strongly interfere with visual processing in prelinguistic children, but these interference effects may weaken when children start acquiring language (Sloutsky & Robinson, 2008, see also Robinson & Sloutsky, 2007a; 2007b). Given that category learning depends critically on visual processing, labels may hinder learning of new categories in both infants and young children. Therefore, the ability to efficiently process and integrate auditory and visual input appears to be a critical (yet by no means sufficient) step in labels becoming category markers.
And second, in order for a label to be used as a category marker, participants should be able to selectively attend to the label and to selectively search for commonalities of identically labeled items. However, there is little evidence that infants or young children can to use the label in guiding their search: instead there is evidence that labels interfere in infants’ and young children’s search for commonalities. At the same time, there is much evidence that labels may contribute to similarity of items (e.g., Sloutsky & Fisher, 2004a; Sloutsky & Lo, 1999). For example, in one set of experiments young children were presented with triads of items (a Target and two Test items) and were asked which of the Test items looked more similar to the Target. One of the test items (e.g., Test A) was very similar to the Target, whereas similarity of the other test item (say Test B) varied across trials from very similar to very different. In the Baseline condition, labels were not provided, whereas in the Label condition, one of the Test items shared the label with the Target, whereas the other Test item did not. The labels were artificial bi-syllabic count nouns. Results indicate that the presence of labels increased similarity for all levels of similarity. However, when the same task was given to, labels had no effect on similarity judgment.

There is also evidence that effects of labels on categorization are graded rather than rule-like, with labels affecting, but not overriding perceptual similarity (e.g., Sloutsky & Fisher, 2004a). In several experiments conducted by Sloutsky and Fisher, 4- and 5-year-olds performed a match-to-sample categorization task. On each trial, they were presented with a triad of pictures, a target and two test items. All items were labeled and only one of the test items shared the label with the target. Participants were asked to decide which
of the test items belongs to the same kind as the Target. Strikingly similar patterns were observed for categorization and feature induction tasks in young children: again, participants’ categorization and induction responses were affected by the similarity ratio, with labels contributing to these effects of similarity rather than overriding them. In yet another experiment, Sloutsky and Fisher (2004a) used items that had been previously used by Gelman and Markman (1986), which turned out to vary widely in terms of appearance similarity. Again, there was little evidence that in their induction responses, 4- and 5-year-olds relied exclusively on linguistic labels.

In sum, it seems that labels function differently across development: whereas labels are likely to function as features early in development (e.g., Sloutsky & Fisher, 2004a; Sloutsky & Lo, 1999), they may become category marker later in development (cf. Yamauchi & A. Markman, 2000). Although the ability to integrate cross-modal information and to attend selectively could be necessary steps in the changing role of labels, precise mechanisms underlying this transition remain unknown. Therefore, much research is needed to understand why, how, and when labels become category markers.

Conclusion

Current research presents extensive evidence that labels are features rather than category markers early in development, but they may become category markers in the course of development. The remaining challenge is to understand why and how this transition takes place.
References


