Psychological Science

Carrot Eaters or Moving Heads : Inductive Inference Is Better Supported by Salient Features Than by Category Labels

Wei (Sophia) Deng and Vladimir M. Sloutsky

Published by:

⑤SAGE

On behalf of:

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Association for Psychological Science

Additional services and information for *Psychological Science* can be found at:

Email Alerts: http://pss.sagepub.com/cgi/alerts

Subscriptions: http://pss.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Proof - Jan 6, 2012

What is This?

Carrot Eaters or Moving Heads: Inductive Inference Is Better Supported by Salient Features Than by Category Labels

Wei (Sophia) Deng and Vladimir M. Sloutsky
Department of Psychology and Center for Cognitive Science, The Ohio State University

Abstract

How do words affect generalization, and how do these effects change during development? One theory posits that even early in development, linguistic labels function as category markers and thus are different from the features of the stimuli they represent. Another theory holds that early in development, labels are akin to other features, but that they may become category markers in the course of development. We addressed this issue in two experiments with 4- to 5-year-olds and adults. In both experiments, participants performed a categorization task (in which they predicted a category label) and an induction task (in which they predicted a missing feature). In the latter task, the category label was pitted against a highly salient feature, such that reliance on the label and reliance on the salient feature would result in different patterns of responses. Results indicated that children relied on the salient feature when performing induction, whereas many adults relied on the category label. These results suggest that early in development, labels are no more than features, but that they may become category markers in the course of development.

Keywords

Inductive generalization is a critical aspect of cognition because it allows people to use knowledge creatively by extending it from known to novel situations. Two aspects of generalization are particularly important: categorization and projective induction. Imagine that X shares certain characteristics with Y. On learning that X is a member of category C, one may decide that Y is also a member of C (i.e., categorization), and on learning that X has property P, one may decide that Y also has P (i.e., projective induction).

There is much evidence showing that from early in development, both categorization and induction are affected by whether presented items are labeled and how they are labeled. For example, if multiple items are accompanied by the same label, young children are more likely to group the items together and generalize a property from one item to another than if no labels are provided (Gelman & Markman, 1986; Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001; Welder & Graham, 2001). However, the mechanism underlying the effect of labels on generalization, as well as possible ways in which this mechanism may change in the course of development, remains unclear.

One theory proposes that early in development, induction is based on category membership, which is communicated by a category label: "Children assume that every object belongs to a natural kind and that common nouns can convey natural kind status" (Gelman & Coley, 1991, p. 190), with names embodying children's intuitive theories. In one study demonstrating this point (Gelman & Markman, 1986), preschoolers were shown three items and given information about the insides of two of them (e.g., "this is a flower and it has tubes for water inside, and this is a sea anemone, it has muscles inside"; see Sloutsky & Fisher, 2004, for details of the stimuli). In this example, the third item looked like an anemone, but was referred to as a flower, and participants were asked about its insides. Researchers found that even 4-year-olds tended to make inferences on the basis of labeled category membership (but see Experiment 4 in Sloutsky & Fisher, 2004, and Fisher, 2010, for diverging evidence and counterarguments).

Corresponding Authors:
Vladimir M. Sloutsky, Department of Psychology and Center for Cognitive Science, The Ohio State University, 208C Ohio Stadium East, 1961 Tuttle Park Pl., Columbus, OH 43210
E-mail: sloutsky.1@osu.edu

Wei (Sophia) Deng, Department of Psychology and Center for Cognitive Science, The Ohio State University, 209C Ohio Stadium East, 1961 Tuttle Park Pl., Columbus, OH 43210
E-mail: deng.69@osu.edu

According to another theory, labels are features of items (similar to color or shape) rather than category markers (Anderson, 1990, 1991; Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999). Because labels may affect processing of visual input (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003), the use of matching labels in tasks in which visual items are presented simultaneously may contribute to the overall similarity of the compared entities and thus to induction (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999).

In an attempt to distinguish between labels being features and category markers, Yamauchi and Markman (1998, 2000) developed an innovative paradigm potentially capable of settling the issue: Imagine two categories—A and B—each having five binary dimensions (e.g., size: large vs. small, color: black vs. white). For the prototype of Category A, all values for these dimensions are denoted by "1" (i.e., A: 1, 1, 1, 1, 1), and for the prototype of Category B, all values for these dimensions are denoted by "0" (i.e., B: 0, 0, 0, 0, 0). There are two interrelated generalization tasks—classification and projective induction. 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, 1) and have to predict category label A or B.

In contrast, the goal of induction is to infer a feature on the basis of the item's category label and other presented features. For example, given an item (e.g., A: 1, ?, 1, 0, 1), participants have to predict the value of the missing feature. A critical manipulation that could illuminate the role of labels is low-match induction, in which participants are presented with an item that has a label from one category but most of the features from the prototype of the opposite category (e.g., A: ?, 0, 1, 0, 0); participants are then asked to infer the missing feature. For low-match classification, participants are also presented with an item similar to the prototype of the opposite category (e.g., ?: 1, 0, 1, 0, 0) and asked to infer the missing label.

In both of these examples, the items are more similar to Prototype B, but if labels are category markers, participants should be more likely to infer the missing feature as belonging to Category A in the induction 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 the attentional weights of labels compared with those of other features. Specifically, if there are features with a higher attentional weight than the label, then a classification task (in which a highly salient feature could be used to predict the label) should yield more "A" responses than should an induction task (in which the label is used to predict a missing feature).

There is much evidence supporting the idea that adults use labels as category markers (Hoffman & Rehder, 2010; Markman & Ross, 2003; Yamauchi, Kohn, & Yu, 2007; Yamauchi & Markman, 1998, 2000; Yamauchi & Yu, 2008). In particular, this evidence shows that in the example

discussed earlier, low-match induction is more likely than low-match classification to yield "A" responses (i.e., responses consistent with the prototype of Category A).

The goal of the study reported here was to use a variant of this paradigm to examine the role of labels early in development. To achieve this goal, we first added a highly salient feature that along with the label distinguished between two categories of stimuli and, second, we had participants perform both a classification and an 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 features, then a different pattern should emerge: If the added feature is more salient than the label, participants should rely on the salient feature rather than on the label. Two experiments with children and adults were conducted to test these competing hypotheses.

Experiment 1
Method

Participants. Thirteen preschool children (6 girls, 7 boys; mean age = 55.5 months, range = 48.6–59.5 months) were recruited from local childcare centers. They 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, so data from this participant were excluded from the analysis. In addition, 30 undergraduate students (16 women, 14 men) from The Ohio State University participated for course credit. One of these participants did not follow the instructions, so data from this participant were excluded from the analysis.

Materials. The materials were colorful drawings of artificial creatures accompanied by the novel labels "*flurp*" (Category A) and "*jalet*" (Category B). For these two categories, we created two prototypes (A0 and B0, respectively) that were distinct in the color and shape of five of their features: body, hands, feet, antennae, and head (see Fig. 1). As Table 1 shows, the two categories had a family-resemblance structure. Stimuli were derived from the two prototypes by modifying the values of one or more of four features—antennae, hands, body, or feet. For example, to produce Stimulus A1, the value of the head, hands, body, and feet were set to 1 (Category A), and the value of the antennae was set to 0 (Category B). As a result, four features were consistent with the features of Prototype A0, and one feature was consistent with the features of Prototype B0.

To set up a proper competition between the category label and a feature, we fixed the value of one feature (the head) within each category. In addition, to make the fixed feature highly salient, we animated the head using Macromedia Flash MX software. For *flurps*, the head was pink and moved up and down; for *jalets*, the head was blue and moved sideways (see
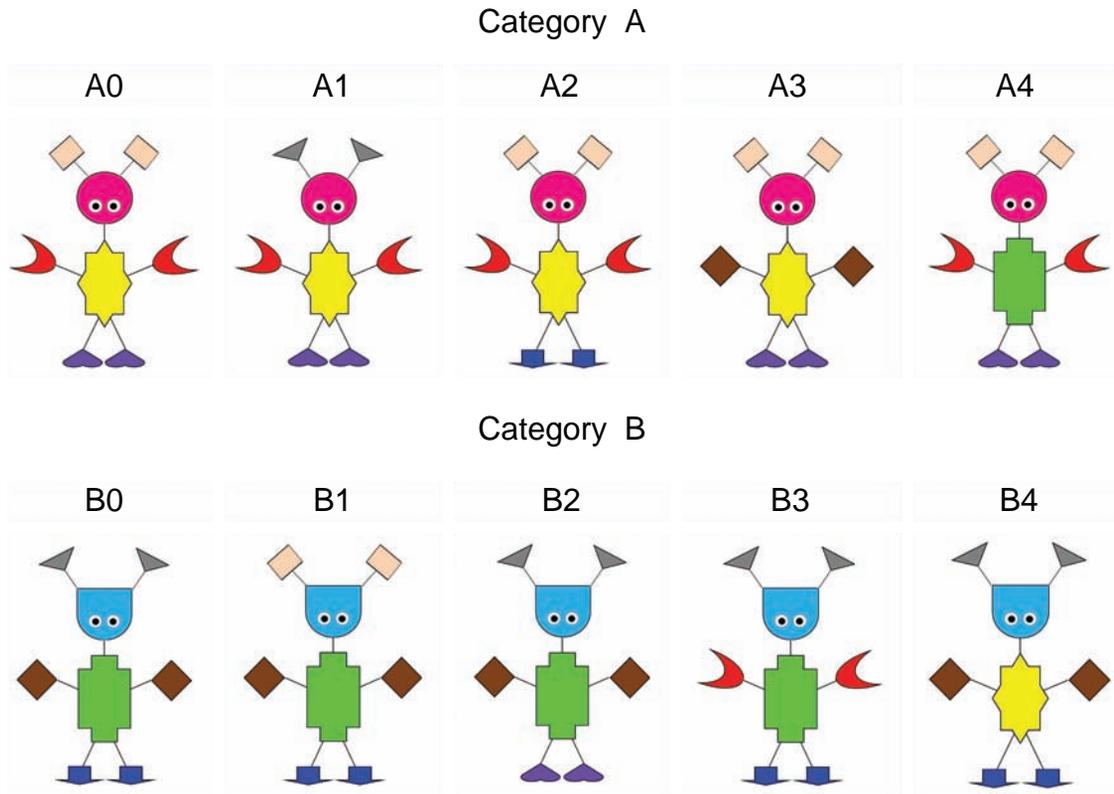
Category A

| A0 | A1 | A2 | A3 | A4 |

Category B

| B0 | B1 | B2 | B3 | B4 |

Fig. 1. Prototypes and high-match stimuli used in Experiments 1 and 2. For each of two categories (A and B), we created a prototype (A0 and B0, respectively) that was distinct in the color and shape of five features (antennae, feet, hands, body, and head). For each of the high-match stimuli (A1–A4, B1–B4), all but one of these features matched the corresponding prototype. The value of the nonmatching feature (never the head) was taken from the other category's prototype. Only the high-match stimuli were used in training.

http://cogdev.cog.ohio-state.edu/MovingHeadDemo.mov to view the stimuli). When asked after the experiment what they noticed about the items, all but one child and all adults mentioned the moving head. Three children and no adults also mentioned the category label. Therefore, it was concluded that the moving head was more salient than any other feature or the category label.

Two types of stimuli were created in each category: those whose features had a high match with the prototype of their category (see Fig. 1) and those whose features had a low match

with the prototype of their category. On low-match trials, stimuli had only one feature (i.e., the moving head) in common with the prototype of their category. On high-match trials, stimuli had four features in common with the prototype of their category.

Design and procedure. Participants were tested in two conditions. In the classification condition, participants were asked to predict a category label (i.e., to identify which group a creature in question was more likely to belong to, *flurp* or *jalet*). In the

Table 1. Structure of the Prototypes and Training Stimuli Used in Experiments 1 and 2

	Category A							Category B					
Stimulus	Body	Hands	Feet	Antennae	Head	Label	Stimulus	Body	Hands	Feet	Antennae	Head	Label
A0	I	I	I	I	I	I	B0	0	0	0	0	0	0
A1	I	I	I	0	I	I	B1	0	0	0	I	0	0
A2	I	I	0	I	I	I	B2	0	0	I	0	0	0
A3	I	0	I	I	I	I	B3	0	I	0	0	0	0
A4	0	I	I	I	I	I	B4	I	0	0	0	0	0

Note: A0 and B0 are the prototypes, from which features were drawn to create the training stimuli. A1 through A4 and B1 through B4 are training stimuli that matched their prototype in all but one feature, which was taken from the opposite category. Features from Category A are indicated with a 1; features from Category B are indicated with a 0.

induction condition, participants were asked to predict whether one of the four unfixed features (e.g., the antennae) would be from the prototypical *flurp* or the prototypical *jalet*. (See Tables 2 and 3 for the structure of stimuli used in the classification and induction conditions, respectively.)

The experiment had a 2 (test condition: classification vs. induction) × 2 (feature match: high vs. low) within-subjects design. The experiment was administered on a computer and controlled by E-Prime software (Version 2.0; Schneider, Eschman, & Zuccolotto, 2002). There were two consecutive phases: training and testing. During training, participants were instructed to remember and distinguish two groups of creatures labeled "*flurp*" and "*jalet*," respectively. The experimenter read these instructions aloud to children, and adults silently read the instructions to themselves. Then, participants were given 24 training trials, each lasting for 5,000 ms and presenting one of the high-match stimuli shown in Table 1. On each training trial, participants saw a stimulus with a corresponding label printed above it, and the label was spoken by the computer (e.g., "This is a *flurp*"). The labeling phrase started at the onset of the trial and lasted for approximately 1,800 ms.

Training was followed immediately by testing (see Fig. 2 for examples of test trials), in which participants completed both high-match and low-match test trials in both the classification and induction conditions. Adults responded to test questions by pressing a key on the keyboard, and children made verbal responses, which were recorded by the experimenter.

The order of the classification and induction conditions was counterbalanced across participants, and within those conditions, the order of the high- and low-match testing trials was randomized for each participant. The first six testing trials in each condition were used as a warm-up and were high-match trials, during which yes/no feedback was provided. The remaining 16 testing trials in each condition were not accompanied by feedback and were used for data analysis.

The critical trial type was low-match induction, in which the only feature that was in common between each stimulus and its prototype was the moving head (whereas the category label and the other three given features of each stimulus were in common with the prototype of the contrasting category). Therefore, if participants rely on the category label to predict the value of a missing feature, they should choose the feature from the contrasting category, thus exhibiting a high level of label-based responding. In contrast, if they rely on the moving head, they should choose the feature from the same category, thus exhibiting a low level of label-based responding. In all other trial types (i.e., high-match induction and low- and high-match classification), there was no conflict between the label and the moving head, and thus reliance on the moving head would result in a high level of label-based responding.

The proportion of label-consistent responses was the dependent variable in our analysis. In the classification condition, responses were identified as label consistent if participants correctly predicted the category label of the presented stimulus. In

Table 2. Structure of the Testing Stimuli Used in the Classification Condition

Stimulus	Specific feature					
	Body	Hands	Feet	Antennae	Head	Label
High match with prototype						
A11	1	1	1	0	1	?
A12	1	1	0	1	1	?
A13	1	0	1	1	1	?
A14	0	1	1	1	1	?
B11	0	0	0	1	0	?
B12	0	0	1	0	0	?
B13	0	1	0	0	0	?
B14	1	0	0	0	0	?
Low match with prototype						
A21	0	1	0	0	1	?
A22	1	0	0	0	1	?
A23	0	0	0	1	1	?
A24	0	0	1	0	1	?
B21	1	0	1	1	0	?
B22	0	1	1	1	0	?
B23	1	1	1	0	0	?
B24	1	1	0	1	0	?

Note: High-match stimuli shared most of their features with their prototype, whereas low-match stimuli shared most of their features with the prototype of the contrasting category. Features drawn from the prototype of Category A are denoted by a 1; features drawn from the prototype of Category B are denoted by a 0. In the classification condition, participants were instructed to predict each stimulus's category label.

Table 3. Structure of the Testing Stimuli Used in the Induction Condition

Stimulus	Specific feature					Label
	Body	Hands	Feet	Antennae	Head	
High match with prototype						
A11	?	1	1	0	1	1
A12	1	?	0	1	1	1
A13	1	0	?	1	1	1
A14	0	1	1	?	1	1
B11	?	0	0	1	0	0
B12	0	?	1	0	0	0
B13	0	1	?	0	0	0
B14	1	0	0	?	0	0
Low match with prototype						
A21	0	?	0	0	1	0
A22	?	0	0	0	1	0
A23	0	0	0	?	1	0
A24	0	0	?	0	1	0
B21	1	?	1	1	0	1
B22	?	1	1	1	0	1
B23	1	1	1	?	0	1
B24	1	1	?	1	0	1

Note: High-match stimuli shared most of their features with their prototype, whereas low-match stimuli shared most of their features with the prototype of the contrasting category. The label for Category A and each associated feature are denoted by a 1; the label for Category B and each associated feature are denoted by a 0. In the induction condition, the category label was provided, but each stimulus was presented with one feature covered (the value denoted here by a question mark). Participants were asked to predict what the feature would be.

the induction condition, responses were identified as label consistent if participants correctly predicted the feature associated with the presented category label. Recall that if the label is a category marker, then participants should rely on the label even when it is pitted against a highly salient feature (i.e., in low-match induction). However, if the label is akin to other features, participants may fail to rely on the label when it is pitted against a highly salient feature. Thus, if they relied on the moving head in low-match induction trials, they should exhibit a low level of label-based responding.

A memory check was administered after the experiment to determine whether participants remembered the 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 category 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.

Results and discussion

The main results are presented in Figure 3. In the classification condition, regardless of whether the level of feature match was high or low, children generated a high level of label-consistent responses (Fig. 3a). Perhaps not surprisingly, children accurately predicted labels for both high- and low-match stimuli by relying on the moving head. Most important, when the moving head pointed to one response, and the label pointed to another response (i.e., in low-match induction), children relied primarily on the moving head.

Children's data were analyzed with a 2 (test condition: classification vs. induction) × 2 (feature match: high vs. low) within-subjects analysis of variance (ANOVA). There was a significant Test Condition × Feature Match interaction, $F(1, 11) = 82.92$, $MSE = 1.02$, $p < .01$, $\eta_p^2 = .883$: Children made comparably high proportions of label-consistent responses in high- and low-match classification, $p > .10$, whereas in the induction condition, children made more label-consistent responses on high-match trials than on low-match trials, paired-samples $t(11) = 12.85$, $p < .01$, $d = 5.27$. Furthermore, when the label was pitted against the salient feature (i.e., in low-match induction), children performed significantly below chance in relying on the label to infer missing features; they relied instead on the moving head, one-sample $t(11) = 10.56$, $p < .01$, $d = 3.05$.

For adults, there was also a Test Condition × Feature Match interaction, $F(1, 28) = 5.90$, $MSE = 0.20$, $p < .05$, $\eta_p^2 = .176$: Adults were likely to make label-consistent responses in the classification condition, regardless of whether the level of
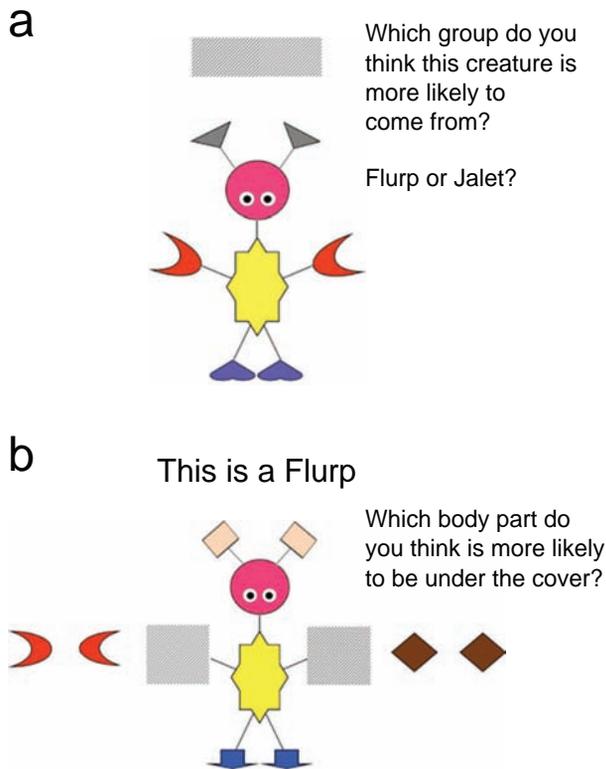
Fig. 2. Examples of (a) a classification test trial and (b) an induction test trial in Experiments 1 and 2. On classification trials, participants were shown a creature and asked to indicate the category to which it was more likely to belong. On induction trials, participants were shown a creature with one feature covered (the hands in this example) and were asked to predict the appearance of the covered feature. The two response options, shown to the left and the right of the creature, were taken from the two category prototypes.

feature match was high or low, $p > .10$, whereas in the induction condition, they made more label-consistent responses on high-match trials than on low-match trials, paired-samples $t(28) = 2.94$, $p < .01$, $d = 0.82$ (Fig. 3b). However, in contrast with children's performance, adults' performance was not significantly different from chance when the label was pitted against the salient feature (i.e., in low-match induction), $p > .10$.

Because adults performed near chance on low-match trials in the induction condition, we deemed it necessary to analyze individual patterns of responses in this condition. Data from adults who made at least 75% (six out of eight testing trials) label-consistent responses on high-match induction trials (19 out of 30 participants) were selected for the analysis of the response pattern on the low-match induction trials. Adults 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% (6 participants) were feature-based responders and 37% (7 participants) were label-based responders, with the remaining 31.5% being mixed responders.

In addition, 11 out of 12 children passed the criterion to be included in the analysis of response patterns, and 91% of them (10 participants) were feature-based responders. This pattern was different from that of adults, $\chi^2(2, N = 31) = 11.023$, $p < .01$. That is, children uniformly relied on a highly salient feature (i.e., the moving head) rather than on the category label to make inductive inferences, even when the salient feature was the single cue that was pitted against the combination of label and other features.
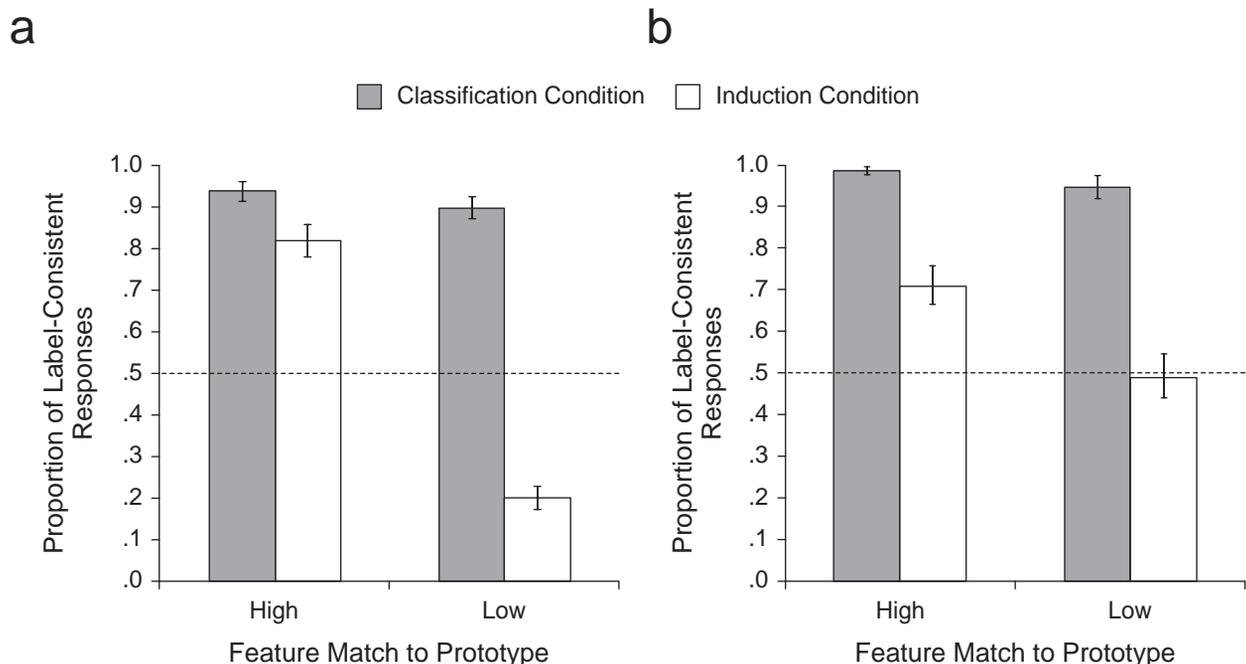

Fig. 3. Results from Experiment 1: proportion of label-consistent responses as a function of condition and the degree to which stimuli matched the prototype of their category. Responses were considered label consistent in the classification condition if participants correctly predicted the category label of the presented stimulus. Responses were considered label consistent in the induction condition if participants correctly predicted the feature associated with the presented category label. Results are shown for (a) children and (b) adults. The dashed lines indicate chance-level responding. Error bars represent standard errors of the mean.
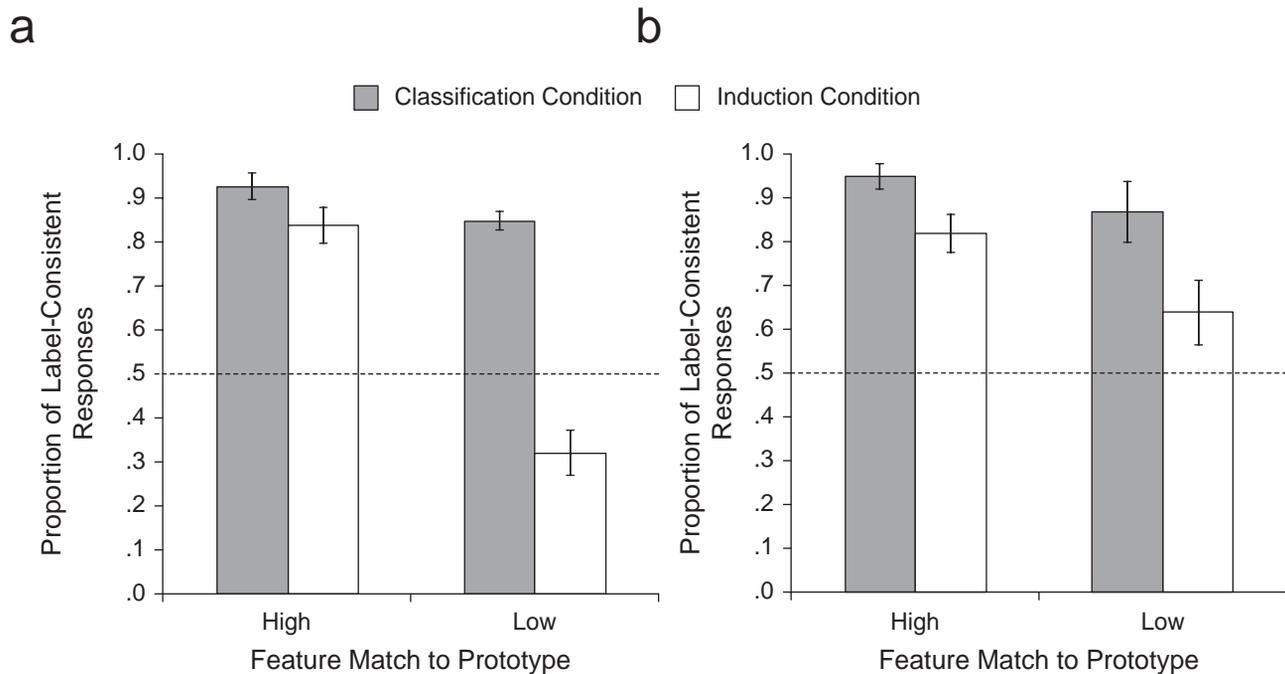
a

b

Fig. 4. Results from Experiment 2: proportion of label-consistent responses as a function of condition and the degree to which stimuli matched the prototype of their category. Responses were considered label consistent in the classification condition if participants correctly predicted the category label of the presented stimulus. Responses were considered label consistent in the induction condition if participants correctly predicted the feature associated with the presented category label. Results are shown for (a) children and (b) adults. The dashed lines indicate chance-level responding. Error bars represent standard errors of the mean.

Overall, children relied on the salient feature (i.e., the moving head) rather than on the category 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 category label, as evidenced by the trimodal distribution on low-match induction trials. 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 one-third of the adults consistently relied on category labels, thus suggesting that these participants treated the label as a category marker. However, it could also be argued that children and many adults failed to rely on the label because the labels were novel (e.g., Davidson & Gelman, 1990). Experiment 2 was designed to test this possibility by using familiar labels, some of which were used in a previous study (Gelman & Heyman, 1999).

Experiment 2

Method

Participants. Seventeen preschool children (10 girls, 7 boys; mean age = 54.9 months, range = 49.7–58.5 months) and 15 undergraduate students (4 women, 11 men) participated in Experiment 2. As in Experiment 1, children were recruited from local childcare centers, and students participated for course credit.

Materials, design, and procedure. The stimuli and procedure in Experiment 2 were similar to those used in Experiment 1, except that the familiar labels "carrot eater" and "meat eater" were used instead of "*flurp*" and "*jalet*." When the computer spoke each of these labels, the labeling phrase lasted for approximately 2,700 ms. Similar to Experiment 1, a memory check was administered after Experiment 2; Participants accurately recalled labels of training items (84% for children and 98% for adults). No participant answered less than three out of five memory-check questions correctly.

Results and discussion

The main results are presented in Figure 4. Data from children and adults were submitted to 2 (test condition: classification vs. induction) × 2 (feature match: high vs. low) within-subjects ANOVAs. As Figure 4a shows, children's performance was similar to that in Experiment 1: There was a significant Test Condition × Feature Match interaction, $F(1, 16) = 44.44$, $MSE = 0.830$, $p < .01$, $\eta_p^2 = .735$. There was a difference in label-consistent responding across the feature-match levels in the classification condition (93% vs. 85% for high and low matches, respectively), paired-samples $t(16) = 2.28$, $p < .05$,

$d = 0.74$; however, there was a substantially greater difference across the feature-match levels in the induction condition (84% vs. 32% for high and low matches, respectively), paired-samples $t(16) = 7.41$, $p < .01$, $d = 2.75$. In addition, when the category label was pitted against the moving head (i.e., in low-match induction), children performed below chance in relying on the label; instead, they relied on the moving head, one-sample $t(16) = 3.57$, $p < .01$, $d = 0.87$.

The results for adults revealed significant main effects of test condition and feature match on label-consistent responding, with no interaction between these two factors (Fig. 4b). Adults made more label-consistent responses in the classification condition than in the induction condition, $F(1, 14) = 7.38$, $MSE = 0.482$, $p < .05$, $\eta_p^2 = .345$, and more label-consistent responses on high-match trials than on low-match trials, $F(1, 14) = 14.72$, $MSE = 0.250$, $p < .01$, $\eta_p^2 = .513$. When the label was pitted against the moving head (i.e., on low-match induction trials), reliance on the label was marginally above chance, one-sample $t(14) = 1.91$, $p = .076$, $d = 0.49$.

As in Experiment 1, we analyzed individual patterns of responses. The analysis revealed that of those 11 adults who passed the 75% criterion, 18% (2 participants) were feature-based responders and 64% (7 participants) were label-based responders, with the remaining 18% being mixed responders. In contrast, of the 15 children who passed the 75% criterion, 67% (10 participants) were feature-based responders and 7% (1 participant) were label-based responders, with the remaining 26% being mixed responders, $\chi^2(2, N = 26) = 10.124$, $p < .01$. Therefore, the use of familiar labels in Experiment 2 resulted in both adults and children exhibiting somewhat greater reliance on labels than when novel labels were used in Experiment 1; this result may have stemmed from the increased salience of familiar labels. However, similar to children in Experiment 1, children in Experiment 2 remained predominantly feature-based responders.

These findings, together with the results of Experiment 1, show that children generate similar patterns of responses for both familiar and novel labels—in the induction condition, children relied on the highly salient perceptual feature rather than on the label. In contrast, about one-third of the adults in Experiment 1 and more than two-thirds in Experiment 2 exhibited consistent label-based performance. These results point to an important developmental difference in the role of labels: Although many adults treat familiar labels as category markers, this is not the case for young children.

General Discussion

In the research reported here, we examined the role of labels in early generalization by extending the paradigm pioneered by Yamauchi and Markman (1998) to young children. Recall that this paradigm was based on the following reasoning. If labels are category markers, 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).

Our research showed that young children exhibit over-whelming reliance on a highly salient feature and not on a category label, whether the label was novel (Experiment 1) or familiar (Experiment 2). The results are more complicated in adults: Some adults exhibited consistent reliance on the salient feature and some relied on the label. Taken together, these results indicate that for young children (and for some adults), category labels may function as features, as little reliance on the category label was observed when it was pitted against the highly salient feature. At the same time, for some adults, labels may be category markers.

The idea that for young children, linguistic labels function as features raises interesting questions regarding the role of labels in infants' inductive generalizations. For example, some researchers (e.g., Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Waxman & Markow, 1995) have demonstrated that labels may facilitate categorization in infants. At the same time, other researchers (Graham, Kilbreath, & Welder, 2004; Welder & Graham, 2001) have demonstrated that labels may facilitate infants' ability to make inductive inferences. These researchers concluded that even for young infants, labels are category markers.

How can labels be category markers for infants but not for young children and some adults? We believe that labels are in fact not category markers for either infants or young children. First, many of the studies examining the effects of labels on infants' category learning compared the effects of labels with the effects of unfamiliar sounds, but not with learning in a no-auditory-input (i.e., silent) baseline condition. When a silent baseline was introduced (e.g., Robinson & Sloutsky, 2007), labels did not facilitate category learning above that baseline (see also Robinson & Sloutsky, 2008, for similar findings on individuation tasks). Second, none of the studies examining the effects of labels on categorization and induction in infancy demonstrated that these effects are greater than those of highly salient features. This latter issue has to be addressed in future research.

Note that in all previous research using Yamauchi and Markman's (1998) paradigm, the relation between classification and induction was fixed, with performance on low-match induction trials exceeding that in low-match categorization trials. This fixedness (as well as differences in goals between classification and induction) suggests that classification and induction may result in different category representation, and there is much research supporting this possibility in adults (Hoffman & Rehder, 2010; see also Markman & Ross, 2003, for a review). The findings reported here suggest that the relation between classification and induction is context-specific rather than fixed: Relative performance on classification and induction tasks may depend on the attentional weights of labels compared with those of other features. These findings may have important implications for the understanding of how classification and induction affect category representation and how these representations may change in the course of development.

The question regarding the role of language in generalization has generated considerable debate, with some researchers arguing that linguistic labels have the special status of category markers and others arguing that labels are akin to features. The research reported here indicates that when labels are pitted against salient perceptual features, young children (and some adults) rely on the salient features, which should not have happened if labels are category markers. These results cast doubt on the view that labels start out as category markers, suggesting instead that labels are features early in development, but may become category markers in the course of development.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*, 3–26.

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

Ferry, A., Hespos, S. J., & Waxman, S. (2010). Language facilitates category formation in 3-month-old infants. *Child Development*, *81*, 472–479.

Fisher, A. V. (2010). What's in the name? Or how rocks and stones are different from dogs and puppies. *Journal of Experimental Child Psychology*, *105*, 198–212.

Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). New York, NY: Cambridge University Press.

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

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

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*, 319–340.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613.

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

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*, 1387–1401.

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*, 233–253.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*, 869–881.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

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

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology*, *35*, 1478–1492.

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

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

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

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

Yamauchi, T., Kohn, N., & Yu, N. Y. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, *35*, 852–863.

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

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

Yamauchi, T., & Yu, N. (2008). Category labels versus feature labels: Category labels polarize inferential predictions. *Memory & Cognition*, *36*, 544–553.