Running Head: Salient Features vs. Category Labels

Carrot-Eaters and Moving Heads: Salient Features Provide Greater Support for Inductive Inference than Category Labels

Wei (Sophia) Deng and Vladimir M. Sloutsky

Department of Psychology and Center for Cognitive Science,

The Ohio State University

July 20, 2011

In press: *Psychological Science*

Word count: 3,938

Please address correspondence to

Vladimir M. Sloutsky
Center for Cognitive Science
208C Ohio Stadium East
1961 Tuttle Park Place
The Ohio State University
Columbus, OH 43210
Phone: (614) 688-5855
Fax: (614) 292-0321
Email: Sloutsky.1@osu.edu

ABSTRACT

How do words affect generalization and how do these effects change with development? According to one theoretical position, even early in development linguistic labels are category markers and, as such, they differ from other features. According to another position, labels start out as features, but they may become category markers in the course of development. This issue was addressed in two experiments with 4-5-year-olds and adults. In both experiments, participants were presented with either a categorization or induction task and the category label was either consistent with a highly salient feature or was pitted against it. Results indicated that children 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, but they may become category markers in the course of development.

Carrot-eaters and moving heads: Salient features provide greater support for inductive inference than category labels

Inductive generalization is a critical aspect of cognition as 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. Upon learning that X is a member of category C, one may decide that Y is also a member of C (i.e., categorization) and upon learning that X has property P, one may decide that Y also has P (i.e., projective induction).

There is much evidence 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 items are accompanied by the same label, young children are more likely to group the items and generalize a property from one item to another than if no labels are provided (Gelman & E. Markman, 1986; Sloutsky & Fisher, 2004; Sloutsky, Lo, & Fisher, 2001; Welder & Graham, 2001). However, the mechanism underlying the effect of labels as well as possible changes in this mechanism in the course of development remain unclear.

According to one theoretical position, early 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 convey natural kind status (as well their accompanying properties) . . .Names are embodiment of our theories” (Gelman & Coley, 1991, p. 190). In one study demonstrating this point (Gelman & Markman, 1986), preschoolers were shown three items and were informed about 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”). The third item looked like the 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 Sloutsky & Fisher, 2004, Experiment 4; Fisher, 2010, for diverging evidence and counterarguments).

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

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 five binary dimensions (e.g., Size: large vs. small, Color: black vs. white, etc.). The prototype of Category A has all values denoted by "1" (i.e., "A", 1, 1, 1, 1, 1) and the prototype of Category B has all values denoted by "0" (i.e., "B", 0, 0, 0, 0, 0). There are two inter-related 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 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 the "low-match" condition. For low-match induction, 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.

In both cases, items are more similar to prototype B and, if labels are category markers, participants should be more likely to infer the missing feature as belonging to A (i.e. the induction task) than to infer label "A" (i.e., 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 with a higher attentional weight than the label, then a classification task (when a highly salient feature could be used to predict the label) should yield more A-responses than an induction task (when the label is used to predict the highly salient feature).

There is much evidence supporting the idea that for adults labels are category markers (Hoffman & Rehder, 2010; Markman & Ross, 2003; Yamauchi & Markman, 1998, 2000; Yamauchi & Yu, 2008; Yamauchi, Kohn, & Yu, 2007). In particular, low-match induction was more likely to yield A-responses (i.e., responses consistent with the prototype of Category A) than low-match classification.

The goal of the current study was to apply a variant of this paradigm to examine the role of labels early 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 features, then a different pattern should emerge – if the added feature is more salient than the label, participants should rely 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 (*M* = 55.5 months, range 48.6-59.5 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 did not follow instruction and these data were excluded from the analysis.

Materials

The materials were colorful drawings of artificial creatures and novel labels "flurp" and "jalet". Two categories were created using five features varying in color and shape (see Figure 1). As shown in Table 1, the two categories had a family-resemblance structure, which was derived from two prototypes (A0 and B0) by modifying the values of one of four features – antenna, hands, body, or 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 label (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. For “flurps”, the head was pink and moved up and down; whereas, for “jalets”, the head was blue and moved sideways. 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 label. Therefore, it was concluded that the moving head was more salient than any other feature or the label.

Two levels of feature match between the test item and the prototype of the corresponding category were used, high and low (see Table 2). 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 label were common with the prototype of the contrasting category. Therefore, if participants rely on the label, 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 exhibit a low level of label-based responding. In all other conditions, 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 (see Table 2).

Design and Procedure

The experiment had a two (Test Condition: Classification vs. Induction) by two (Feature Match: High vs. Low) within-subjects design, was administered on a computer and controlled by E-prime 2.0 software. There were 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 training trials, each lasting for 5000 msec and presenting a creature produced from the structure shown in Table 1. On each training trial, participants saw a stimulus with a corresponding label printed above it and heard the label (e.g., “This is a flurp”) presented by the computer. The labeling phrase started at the onset of the trial and lasted for approximately 1800 msec.

Training was followed immediately by testing (see Figure 2 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 an antenna of a prototypical flurp and that 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 testing trials in each condition were used as a warm-up and were high-match trials in 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 proportion of *label-consistent responses* was the dependent variable. 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 presented label 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 3. These results indicated that 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. And most importantly, 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) by 2 (Feature Match: High vs. Low) within-subjects ANOVA. There was a significant test condition by feature match interaction, $F(1, 11) = 82.92$, $MSE = 1.02$, $p < 0.01$, $\eta^2 = 0.883$: 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, paired-samples $t(11) = 12.85$, $p < 0.01$, $d = 5.27$. Furthermore, when the label was pitted against the salient feature (i.e., in low-match induction), children were significantly below chance in relying on the label to infer missing features, relying instead on the moving head, one-sample $t(11) = 10.56$, $p < 0.01$, $d = 3.05$.

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, paired-samples $t(28) = 2.94$, $p < 0.01$, $d = 0.82$. However, in contrast to children, when the label was pitted against the salient feature (i.e., in low-match induction), adults were not different from chance, $p > 0.10$.

Because adults in the low-match induction were 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. This pattern was different from that of adults', $\chi^2(2, 31) = 11.023$, $p < 0.01$. That is, children uniformly relied on a highly salient feature (i.e., the moving head) rather than on the 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.

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 one-third of the adults 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). Experiment 2 was designed to test this possibility by using familiar labels, some of which were used previously (Gelman & Heyman, 1999).

EXPERIMENT 2

Method

Participants

Seventeen preschool children (*M* = 54.9 months, range 49.7-58.5 months; 10 girls) and 15 undergraduate students (4 women) participated. Recruitment procedures were identical to Experiment 1.

Materials, Design, and Procedure

The stimuli and procedure were similar to those used in Experiment 1 except that familiar labels "Carrot-eater" and "Meat-eater" were used instead of "Flurp" and "Jalet". Each labeling phrase lasted for approximately 2700 msec. Similar to Experiment 1, a memory check was administered after the main experiment, with participants accurately recalling labels of training items (84% for children and 98% for adults). No participant answered correctly less than three out of five memory check questions.

Results and Discussion

The results are presented in Figure 4. Data from children and adults were submitted to 2 (Test Condition: Classification vs. Induction) by 2 (Feature Match: High vs. Low) within-subjects ANOVAs. As shown in Figure 4A, children's performance was similar to that in Experiment 1: there was a significant test condition by feature match interaction, $F(1, 16) = 44.44$, $MSE = 0.830$, $p < 0.01$, $\eta^2 = 0.735$. 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, paired-samples $t(16) = 2.28$, $p < 0.05$, $d = 0.74$), there was a substantially greater difference across the feature match levels of the Induction condition (84% vs. 32%, for high- and low-match, respectively, paired-samples $t(16) = 7.41$, $p < 0.01$, $d = 2.75$). In addition, when the label was pitted against the moving head (i.e., in low-match induction), children were below chance in relying on the label, relying instead on the moving head, one sample $t(16) = 3.57$, $p < 0.01$, $d = 0.87$.

Adults' performance is shown in Figure 4B. 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, $F(1,14) = 14.72$, $MSE = 0.250$, $p < 0.01$, $\eta^2 = 0.513$. When the label was pitted against the moving head (i.e., in low-match induction), reliance on the label was marginally above chance, one sample $t(14) = 1.91$, $p = 0.076$, $d = 0.49$.

As in Experiment 1, we analyzed individual patterns of responses. The analysis revealed that of those 11 adults who passed the 75% criterion, 18% (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, which was different from adults, $\chi^2(2, 26) = 10.124$, $p < 0.01$. Therefore, familiar labels used in Experiment 2 resulted in both adults and children exhibiting somewhat greater reliance on labels than with novel labels used in Experiment 1, which may have stemmed from increased salience of familiar labels. However, similar to Experiment 1, children remained predominantly feature-based responders.

These findings, together with the results of Experiment 1, indicated that for children, both familiar and novel labels generated similar pattern of responses – when performing induction, 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: while many adults may treat familiar labels as category markers, this is not the case for young children.

General Discussion

The reported research examined the role of labels in early generalization by extending the paradigm pioneered by Yamauchi & 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).

Current research indicates that young children exhibited overwhelming 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 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 generalization. For example, some researchers (e.g., Balaban & Waxman, 1997; Ferry, Hespos, & Waxman, 2010; Waxman & Markow, 1995) have demonstrated that labels may facilitate category learning in infants. At the same time, other researchers (Graham, Kilbreath, Welder, 2004; Welder & Graham, 2001) have demonstrated that labels may facilitate infants' inductive inference. 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 effects of labels on infants' category learning compared effects of labels with effects of unfamiliar sounds, but not with learning in a no-auditory input (i.e., silent) baseline. When a silent baseline was introduced (e.g., Robinson & Sloutsky, 2007), it turned out that labels did not facilitate category learning above the silent baseline (see also Robinson & Sloutsky, 2008, for similar findings on individuation tasks). And second, none of the studies examining effects of labels on categorization and induction in infancy demonstrated that these effects are greater than those of highly salient features. This latter question has to be addressed in future research.

Note that in all previous research using Yamauchi and Markman's paradigm, the relation between classification and induction was fixed, with performance in low-match induction exceeding that in low-match categorization. This fixedness (as well as differences in goals between classification and induction) suggested 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). Current findings suggest that the relation between classification and induction is context-specific rather than fixed: relative performance on classification and induction tasks may depend on attentional weights of labels compared to those of other features. These findings may have important implications for our 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 arguing that linguistic labels have the special status of category markers and others arguing that labels are akin to other features. 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 early in development labels are features, but they 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: 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). 13-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.

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 12-13-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.

Author Note

This research is supported by the NSF grant BCS-0720135 and by NIH grant R01HD056105 to VMS. We thank Catherine Best, Anna Fisher, Chris Robinson, and two anonymous reviewers for helpful comments. Address correspondence to Vladimir M. Sloutsky (sloutsky.1@osu.edu) or Sophia Deng (deng.69@osu.edu).

Table 1. Structure of training stimuli used in Experiments 1 and 2.

Category A							Category B						
Stimuli	Body	Hands	Feet	Antenna	Label	Head	Stimuli	Body	Hands	Feet	Antenna	Label	Head
A1	1	1	1	0	1	1	B1	0	0	0	1	0	0
A2	1	1	0	1	1	1	B2	0	0	1	0	0	0
A3	1	0	1	1	1	1	B3	0	1	0	0	0	0
A4	0	1	1	1	1	1	B4	1	0	0	0	0	0
A0	1	1	1	1	1	1	B0	0	0	0	0	0	0

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. A0 and B0 are prototypes of each category and A1/B1 –A4/B4 are individual exemplars.

Table 2. Structure of testing stimuli in Classification used in Experiments 1-2.

Category A								Category B					
Stimuli	Body	Hand	Feet	Antenna	Label	Head	Match	Stimuli	Body	Hand	Feet	Antenna	Label	Head
A11	1	1	1	0	?	1	High	B11	0	0	0	1	?	0
A12	1	1	0	1	?	1		B12	0	0	1	0	?	0
A13	1	0	1	1	?	1		B13	0	1	0	0	?	0
A14	0	1	1	1	?	1		B14	1	0	0	0	?	0
A21	0	1	0	0	?	1	Low	B21	1	0	1	1	?	0
A22	1	0	0	0	?	1		B22	0	1	1	1	?	0
A23	0	0	0	1	?	1		B23	1	1	1	0	?	0
A24	0	0	1	0	?	1		B24	1	1	0	1	?	0

Note. High and low are two levels of feature match. A = Category A; B = Category B.

Table 3. Structure of testing stimuli in Induction used in Experiments 1-2.

Category A								Category B					
Stimuli	Body	Hand	Feet	Antenna	Label	Head	Match	Stimuli	Body	Hand	Feet	Antenna	Label	Head
A11	?	1	1	0	1	1	High	B11	?	0	0	1	0	0
A12	1	?	0	1	1	1		B12	0	?	1	0	0	0
A13	1	0	?	1	1	1		B13	0	1	?	0	0	0
A14	0	1	1	?	1	1		B14	1	0	0	?	0	0
A21	0	?	0	0	0	1	Low	B21	1	?	1	1	1	0
A22	?	0	0	0	0	1		B22	?	1	1	1	1	0
A23	0	0	0	?	0	1		B23	1	1	1	?	1	0
A24	0	0	?	0	0	1		B24	1	1	?	1	1	0

Note. High and low are two levels of feature match. A = Category A; B = Category B.

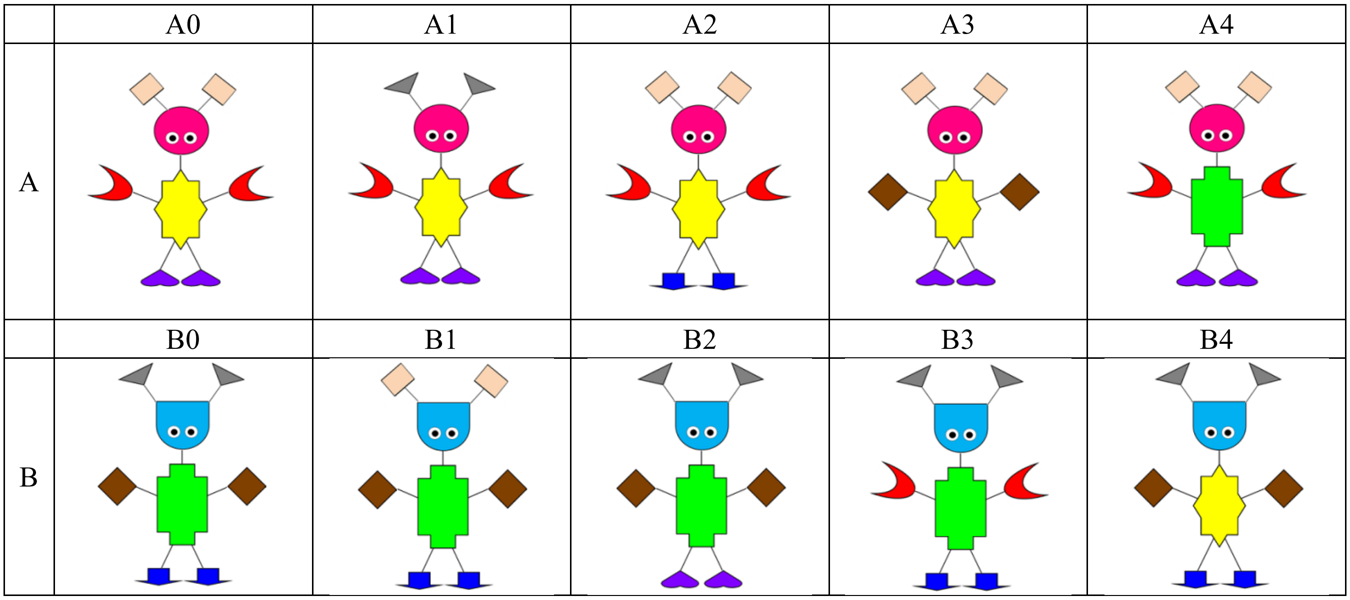
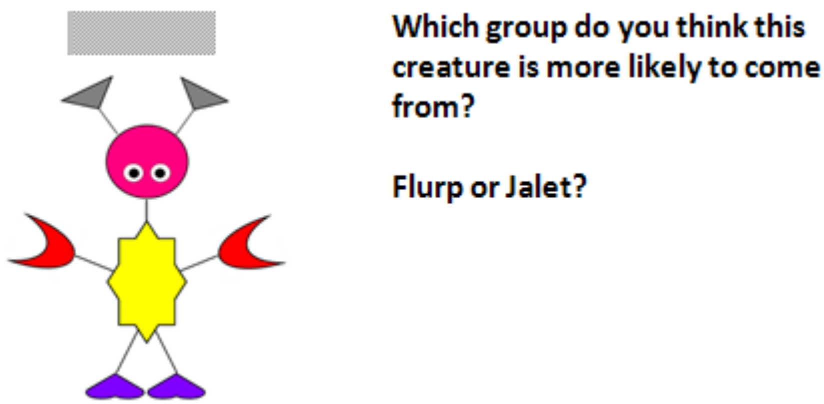


Figure 1. Examples of stimuli of Category A and Category B used in Experiments 1-2.

A. Classification



B. Induction

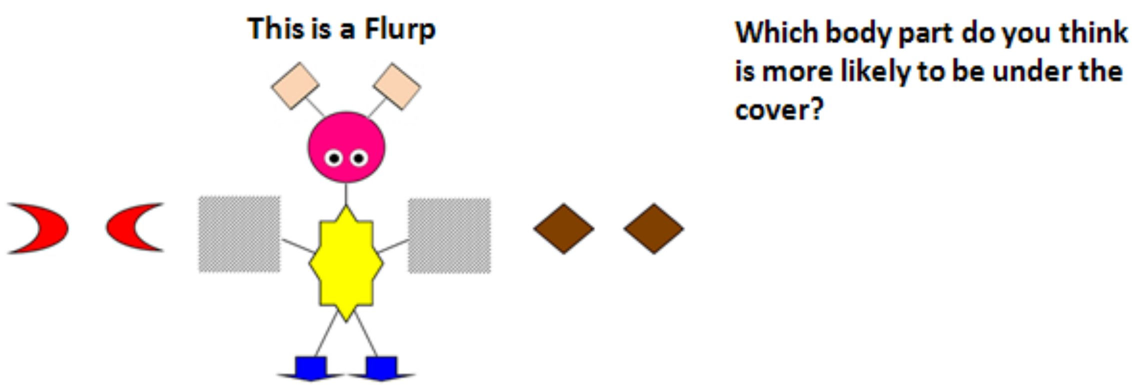
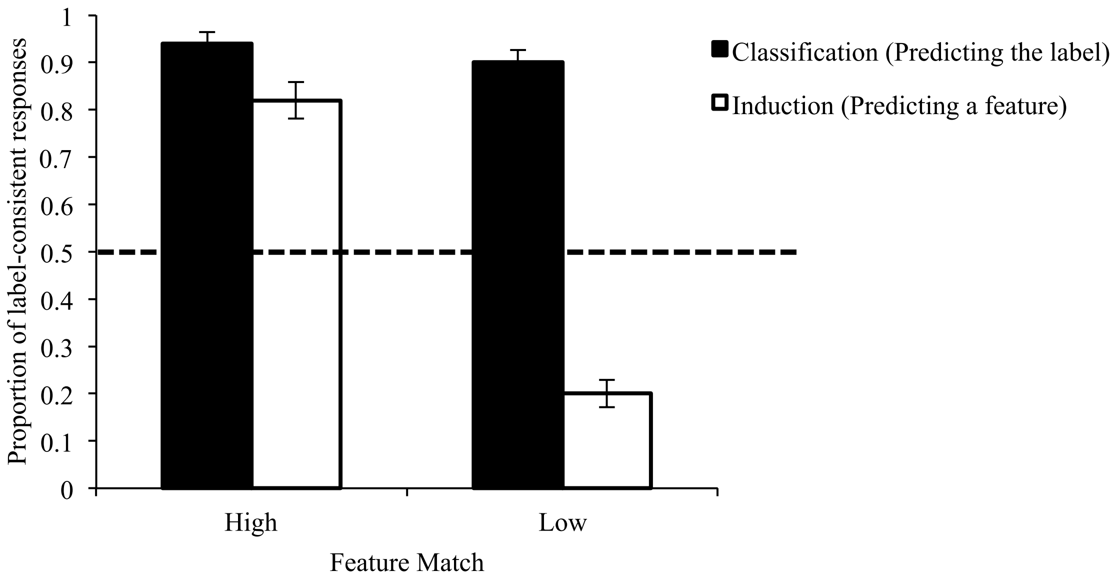


Figure 2. Examples of Classification and Induction test trials in Experiments 1-2

A. Children



B. Adults

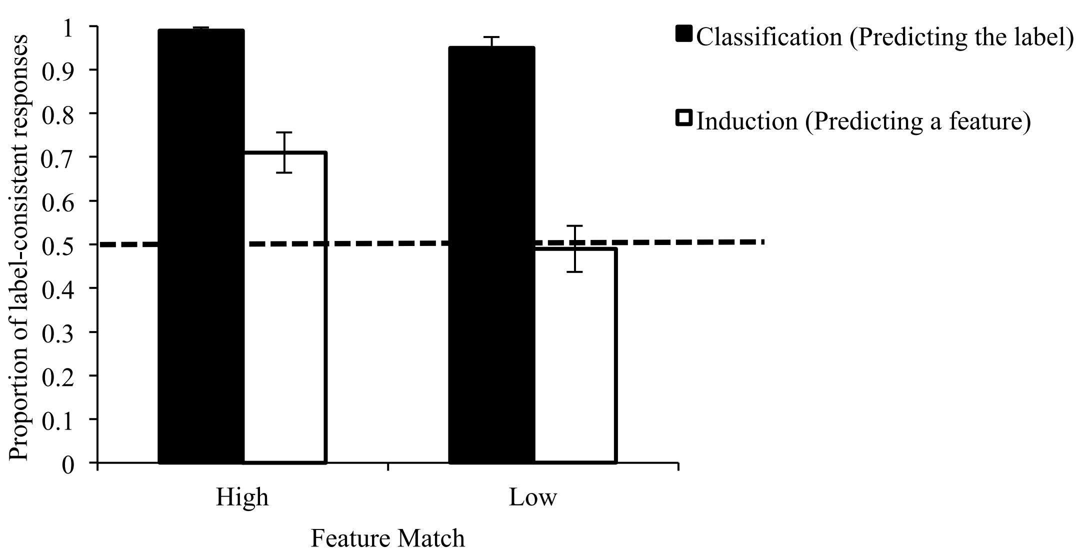
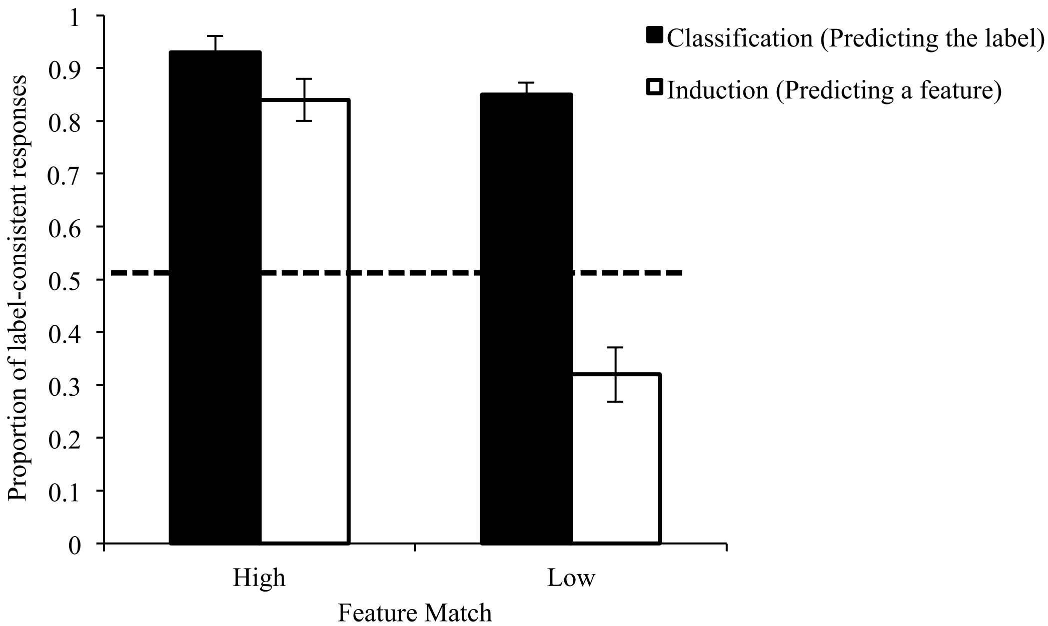


Figure 3. Proportion of label-consistent responses by feature match and testing condition in Experiment 1.

A. Children



B. Adults

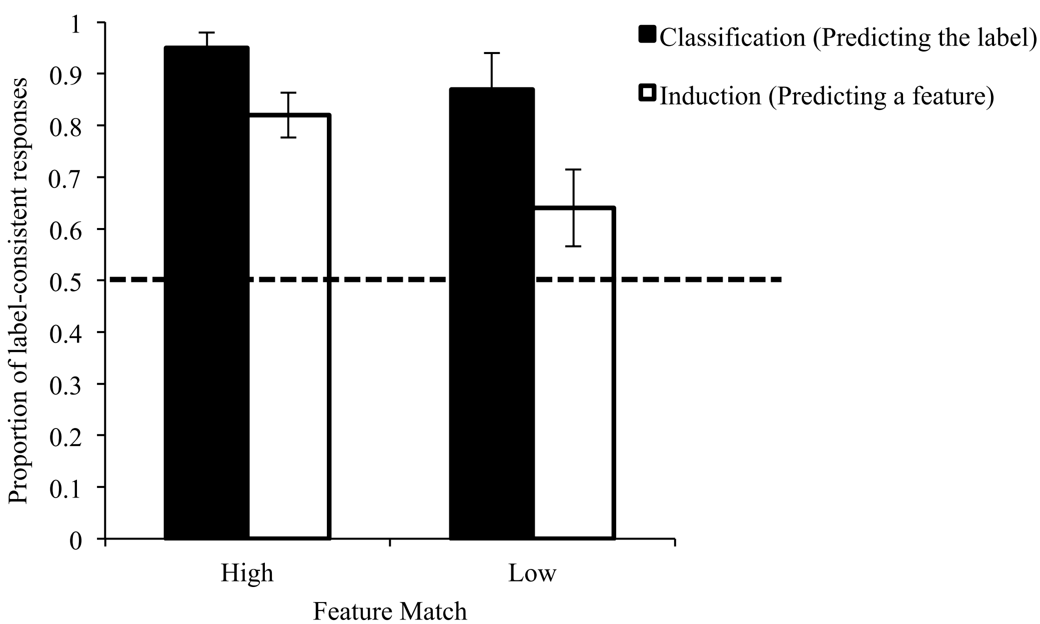


Figure 4. Proportion of label-consistent responses by feature match and testing condition in Experiment 2.