Developmental Psychology
2015, Vol. 51, No. 3, 392–405

0012-1649/15/$12.00 http://dx.doi.org/10.1037/a0038749

The Development of Categorization: Effects of Classification and Inference Training on Category Representation

Wei (Sophia) Deng and Vladimir M. Sloutsky
The Ohio State University

Does category representation change in the course of development? And if so, how and why? The current study attempted to answer these questions by examining category learning and category representation. In Experiment 1, 4-year-olds, 6-year-olds, and adults were trained with either a classification task or an inference task and their categorization performance and memory for items were tested. Adults and 6-year-olds exhibited an important asymmetry: they relied on a single deterministic feature during classification training, but not during inference training. In contrast, regardless of the training condition, 4-year-olds relied on multiple probabilistic features. In Experiment 2, 4-year-olds were presented with classification training and their attention was explicitly directed to the deterministic feature. Under this condition, their categorization performance was similar to that of older participants in Experiment 1, yet their memory performance pointed to a similarity-based representation, which was similar to that of 4-year-olds in Experiment 1. These results are discussed in relation to theories of categorization and the role of selective attention in the development of category learning.

Keywords: categorization, attention, learning, cognitive development

The ability to form categories, or equivalence classes, of discriminable entities is a central component of human cognition: Categorization enables abstract thought and promotes expansion of knowledge to novel situations. For example, having learned that a person's heart has four chambers, one may expect other humans (and perhaps great apes) to have similar hearts. It has been well established that at least a rudimentary ability to form categories appears in early infancy (Eimas & Quinn, 1994; Oakes, Madole, & Cohen, 1991) and is manifested in a variety of species (Lazareva, Freiburger, & Wasserman, 2004; Smith et al., 2012). There is also evidence of remarkable development in the ability to form categories (e.g., Kloos & Sloutsky, 2008; Smith, 1989; see also Sloutsky, 2010, for a review). It is hardly controversial that adults can acquire exceedingly abstract categories, whereas there is little evidence that infants or even young children (i.e., children younger than 6 years of age) can acquire categories of similar levels of abstraction. Although many agree that categorization does develop, there is less agreement as to what changes and why.

Possible answers to the *what* question range from (a) profound qualitative (often stage-like) changes in category representations, such as theory change (e.g., Carey, 1999; Inagaki & Hatano, 2002) or *characteristic-to-defining shift* (Keil, 1992; Keil & Batterman, 1984) to (b) relatively continuous representational change (e.g., Eimas, 1994). According to the shift view, *immature* representations are replaced by more *mature* representations, whereas according to the continuity view, the development consists of enrichment rather than replacement of immature representations.

Possible answers to the *why* question range from the acquisition of domain-specific (or even concept-specific) knowledge (e.g., Carey, 1999; Inagaki & Hatano, 2002; Keil & Batterman, 1984; Keil, 1992) to more domain-general explanations, such the development of selective attention enabling people to focus on relevant information (e.g., Sloutsky, 2010; Smith, 1989). In the former case, development is a function of knowledge acquisition: novices start with more *characteristic* representations, but shift to *defining* representations as more knowledge is acquired. In the latter case, development involves changes in basic cognitive processes.

The goal of present research is to better understand what changes in the course of development and why. In an attempt to answer these questions, we start with evidence that in adults representation of the same category structure may depend on the way the category is learned (Hoffman & Rehder, 2010; Love, Medin, & Gureckis, 2004; Sakamoto & Love, 2010; Yamauchi, Love, & Markman, 2002; see also Markman & Ross, 2003, for a review). Specifically, if they learn the category by classification (i.e., by predicting a label of each item) they tend to represent the structure in a more rule-based (or *defining feature* based) manner. At the same time, if they learn the category by inference (i.e., by predicting a missing feature of each item) they tend to represent the category in a more similarity-based (or *characteristic features* based) manner. Therefore, developmental change may not occur in a shift-like manner, with immature representations being replaced by more *mature* representations. Instead, the development may consist of acquiring the ability to form rule-based representations, and forming different category representations under different task conditions.

This article was published Online First January 19, 2015.

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

This research is supported by the NSF Grant BCS-1323963 and NICHD R01HD078545 to VMS. We thank Keith Apfelbaum, Anna Fisher, and John Opfer for helpful comments.

Correspondence concerning this article should be addressed to Vladimir M. Sloutsky, Department of Psychology, 239 Psychology Building, 1835 Neil Avenue, The Ohio State University, Columbus, OH 43210. E-mail: sloutsky.1@osu.edu

In addition, there is evidence that effects of the learning task or *learning regime* (i.e., learning by classification vs. learning by inference) on category representation stem from a domain general process—the way attention is allocated in the course of category learning (e.g., Hoffman & Rehder, 2010). Therefore, examining how these effects of learning regime on category representation emerge in the course of development may provide some answers pertaining to the mechanism of developmental change.

Effects of Learning Regime on Category Representation

The learning regime most frequently used in the lab studies is *classification* learning. In classification learning, participants learn a category by predicting the label of a given item on the basis of presented features: on each trial, a participant is presented with an item and has to predict how the item is labeled. In the case of learning two categories, A and B, the participant predicts whether the item is labeled A or B.

However, the ways people learn categories are not limited to classification learning. For example, in *inference* learning participants have to infer a missing feature on the basis of category label and other presented features. On each trial an item is presented and labeled, but one of the features is not revealed to the participant. The participant has to predict whether the nonrevealed feature comes from features of Category A or Category B.

There are two lines of evidence that for adults classification and inference learning are not equivalent and these learning regimes may result in different representations of the same category. First, in order for classification and inference learning to be equivalent and result in equivalent representations, labels have to be equivalent to other features (see Markman & Ross, 2003; Yamauchi & Markman, 1998, for extensive arguments). This is because classification requires one to predict the category label when features are given, whereas inference requires one to predict a feature, when the label and the rest of the features are given. However, there is much evidence stemming from classification/inference judgment tasks (these tasks do not involve learning) that, at least for adults, labels are not equivalent to features (Yamauchi & Markman, 2000; see also Markman & Ross, 2003, for a review).

Another source of evidence of representational differences between classification and inference learning pertains to differences in allocating attention under these two learning regimes (Hoffman & Rehder, 2010). Note that most theories of categorization agree that adult category learning results in increased attention to the dimension(s) that separate the studied categories. For example, learning of two categories, such as squirrels versus chipmunks, may result in attention shifting to stripes (which is a diagnostic feature) and away from the tail (which is not diagnostic).

This attentional selectivity has consequences: While learning to attend to the diagnostic dimension (i.e., presence or absence of stripes), participants also learn to ignore nondiagnostic dimensions—the phenomenon known as learned inattention (see Hoffman & Rehder, 2010, for a review). If after learning the two categories, a learner embarks on a new categorization task—differentiating between squirrels and hamsters—the tail that was nondiagnostic for previous learning becomes diagnostic for current learning. In other words, as a result of allocating attention selectively in the first task, participants may have difficulty shifting attention to a previously ignored dimension.

Using a combination of behavioral and eye tracking methodology, Hoffman and Rehder (2010) found profound differences between classification and inferences learning. Whereas classification learners *optimized* attention (i.e., shifted attention to the category-relevant or diagnostic dimension) in Phase 1 and exhibited learned inattention in Phase 2, neither optimization nor learned inattention was the case for inference learners. It was concluded that, in contrast to classification learners who attend selectively, trying to extract the diagnostic dimension, inference learners attend diffusely, trying to learn multiple dimensions and the ways these dimensions interrelate. These findings suggest that classification and inference learning lead to differences in allocation of attention and subsequently to differences in category representation. In classification learning, participants are likely to extract the most diagnostic (or rule) feature, whereas in inference learning they are more likely to extract within-category similarity.

Classification Versus Inference Learning: Do Representational Differences Emerge in the Course of Development?

Although adults exhibit these representational differences, there are two sources of evidence suggesting that these difference are a product of development rather than a developmental starting point: (a) developmental differences in how labels are treated, and (b) developmental differences in selective attention. First, there is a growing body of developmental work suggesting that, in contrast to the findings with adults, early in development labels may function as features. For example, it has been demonstrated that early in development, labels contribute to similarity of compared entities and the contribution is quantitative, feature-like (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004, 2007; Sloutsky & Fisher, 2004, 2012; Sloutsky & Lo, 1999; but see Waxman & Gelman, 2009 for a review of literature disputing the *label-as feature* view).

Additional evidence suggesting that early in development labels may function as features stems from more recent work by Deng and Sloutsky (2012, 2013). These researchers adapted a variant of Yamauchi and Markman's (2000) paradigm to 4- to 5-year-olds children. It was found that, in contrast to adults, young children treat labels no differently than other features.

The second source of evidence pertains to developmental differences in selective attention. More generally, children younger than 5 years of age often have difficulty focusing on a single relevant dimension, while ignoring multiple distracting dimensions (see, Hanania & Smith, 2010; Plude, Enns, & Brodeur, 1994, for reviews; see also Rabi & Minda, 2014, for recent category learning findings).

There is also more specific evidence that when learning categories by classification, adults and young children allocate attention differently (Best, Yim, & Sloutsky, 2013; Robinson, Best, & Sloutsky, 2011). As discussed above, adults tend to optimize attention by shifting it to the most diagnostic feature (or features) that separates the categories. In contrast, infants and young children tend to learn categories while attending diffusely and extracting within-category statistics.

Therefore, if young children achieve learning by distributing rather than optimizing attention, then they should not optimize attention in classification learning and thus form similar representations and exhibit symmetrical performance in classification and inference learning. As we discuss in the next section, these findings offer answers to the *why* question, pointing to a possible mechanism of developmental change.

The Emergence of Representational Differences and Possible Mechanisms of Change

As discussed above, there is evidence that (a) early in development labels may function as features, and (b) infants and young children tend to distribute rather than optimize attention. This evidence suggests that, in contrast to adults, early in development classification and inference learning are equivalent and may result in a similar representation of the learned category. These considerations lead to a number of important hypotheses.

Because early in development classification and inference learning could be equivalent, young children in both learning regimes should (a) exhibit a similar pattern of diffused attention, and (b) form similar representations (based on multiple within-category features). In contrast, for adults (for whom classification and inference learning are not equivalent), representations formed in the course of classification learning would differ from those formed in the course of inference learning. Specifically, adults may optimize attention and extract deterministic features when learning by classification and they may attend diffusely and extract multiple within-category features when learning by inference.

Present Study

The reported study consisted of two experiments. Experiment 1 was designed to examine the developmental differences in category representation in classification and inference learning, whereas Experiment 2 attempted to further examine the mechanisms of developmental change. The basic task of Experiment 1 consisted of three phases: instructions, training, and testing. During training, participants (4-year-olds, 6-year-olds, and adults) had to predict either the category of a given item (in classification training) or a feature that the item had (in inference training) and they were provided with corrective feedback. There were two family resemblance categories, with each training item including a single deterministic feature D (which perfectly distinguished between the two categories) and multiple probabilistic features P (with each providing imperfect probabilistic information about category membership).

Participants were then tested on how they categorized items and represented categories. The testing phase (which was identical for the two training conditions) was administered immediately after the training phase and no feedback was provided during testing.

Testing consisted of categorization and recognition tasks. On categorization trials participants were asked to determine which category the item was more likely to belong to, whereas on recognition trials they were asked whether or not each item was presented during training. The goal of categorization trials was to determine which features participants rely on in their decisions. The goal of recognition trials was to determine what participants remember from training, which may shed light on how they allocate attention during training. In addition, memory for features may be informative with respect to how this feature is used in category representation: greater memory for a given feature makes it more likely that this feature is included in category representation.

Based on the considerations reviewed above, it was predicted that because 4-year-olds do not optimize attention, their categorization performance and recognition memory should be symmetrical across the training conditions. In both conditions, participants should rely on multiple probabilistic features rather than on a single deterministic feature.

In contrast, adults who were shown to optimize attention in classification, but not in inference training, should exhibit asymmetry. They should rely on the D feature in classification, but not in inference training. In addition, they should remember D features better than P features in classification, but not in inference training. Six-year-olds were included to provide a more detailed account of the developmental transition. In particular, these participants are older than those who typically exhibit difficulty focusing on a single relevant dimension in the presence of distracting dimensions (see, Hanania & Smith, 2010; Plude et al., 1994). Therefore, children of this age may have the capacity to rely on a single feature, which may transpire in the current study.

The goal of Experiment 2 was to test the proposed attentional account of the development of category learning. In particular, we attempted to exogenously direct attention of 4-year-olds to D features in order to elicit changes in their categorization performance and category representation. If our attentional manipulation is successful in affecting 4-year-olds' categorization performance (and category representation), this finding would link the development of categorization with the development of selective attention.

Experiment 1

Method

Participants. The sample consisted of 40 adults (19 women), 40 4-year-old children (M = 54.5 months, range 47.5–60.1 month; 19 girls), and 40 6-year-old children (M = 71.5 months, range 66.1–78.3 months; 15 girls). There were two between-subjects conditions (classification and inference training), with 20 participants of each age group per condition. Data from one additional adult were excluded from analyses because of extremely poor performance in training. Data from two additional 4-year-olds and one additional 6-year-old were also excluded from analyses because of the experiment being disrupted by school activities.

Adults were The Ohio State University undergraduate students participating for course credit and they were tested in a quiet room in the laboratory on campus. Child participants were recruited from childcare centers and preschools, located in middle-class suburbs of Columbus, Ohio, and were tested by a female experimenter in a quiet room in their childcare center or preschool.

Materials. Materials were similar to those used previously by Deng and Sloutsky (2012, 2013) and consisted of colorful drawings of artificial creatures. These creatures were accompanied by the novel labels *flurp* (Category F) and *jalet* (Category J). These categories had two prototypes (F0 and J0, respectively) that were

distinct in the color and shape of seven of their features: head, body, hands, feet, antennae, tail, and a body mark (see Figure 1).

As shown in Table 1, most of the features were probabilistic and they jointly reflected the overall similarity among the exemplars (we refer to them as the *P features* or as *overall appearance*), whereas one feature was deterministic and it perfectly separated the two categories (we refer to as the *D feature* or as a *category-inclusion rule*). The body mark (introduced as a *body button*) was the deterministic feature: all members of Category F had a raindrop-shaped button with the value of 1, whereas all members of Category J had cross-shaped button with the value of 0. All the other features—the head, body, hands, feet, antennae, and tail—varied within each category, thus constituting the probabilistic features.

As shown in Table 1, some of the items were used in training and some in testing. The training stimuli consisted of high-match items (i.e., $P_{flurp}D_{flurp}$ and $P_{jalet}D_{jalet}$). These items had the deterministic feature (D) and four probabilistic features (P) consistent with a given prototype; two other probabilistic features were consistent with the opposite prototype.

The testing stimuli consisted of high-match items (i.e., $P_{flurp}D_{flurp}$ and $P_{jalet}D_{jalet}$), Switch (or critical) items (i.e., $P_{jalet}D_{flurp}$ and $P_{flurp}D_{jalet}$), and three additional item types. High-match items were the items presented during training and they were highly similar to the prototypes of respective categories. Switch items had the D features of one category and most P features of another category, which made these items somewhat analogous to the real life categories of whales or dolphins (these have defining features of mammals, but the majority of observable features of fish). The three additional item types included: (a) new-D items (i.e., $P_{flurp}D_{new}$ and $P_{jalet}D_{new}$), which had probabilistic features of the studied categories and a novel feature replacing the deterministic feature; (b) one-new-P items (i.e., $P_{new}D_{flurp}$ and $P_{new}D_{jalet}$), which had all features of the studied categories but a novel feature replacing one probabilistic feature; and (c) all-new-P items (i.e., $P_{all\text{-}new}D_{flurp}$ and $P_{all\text{-}new}D_{jalet}$), which had the deterministic features from the studied categories and all new features replacing the studied probabilistic features.

The high-match items were used to examine how well the participants learned the categories and to assess their recognition accuracy on the old items. The switch items had most of the P features from one category and the D feature from another, thus allowing determining whether participants in their categorization decisions relied on the overall similarity (i.e., P features) or on the deterministic rule (i.e., D feature).

The new-D items were used to assess whether participants could rely in their categorization on old P features when the old D feature was not available. These items were also used to examine whether participants encoded the deterministic feature, in which case they should judge these items as new during the memory test.

The one-new-P items were used to assess whether participants could categorize items when one P feature was new. These items were also used to examine whether participants encoded all individual P features, in which case they should judge these items as new during memory test.

And finally, the all-new-P items were used to assess whether participants could perform rule-based categorization (i.e., rely on the old D feature when none of the old P features was available). In addition, these items were used to assess participants' overall memory accuracy for probabilistic features: if they encoded at least one such feature, they should judge these items as new. Table 1 presents an example of category structure with P and D being combined to create five types of stimuli, and Figure 1 shows examples of each kind of stimulus.

Design and procedure. The experiment consisted of instructions, training, and testing (see Figure 2). Training was a between-

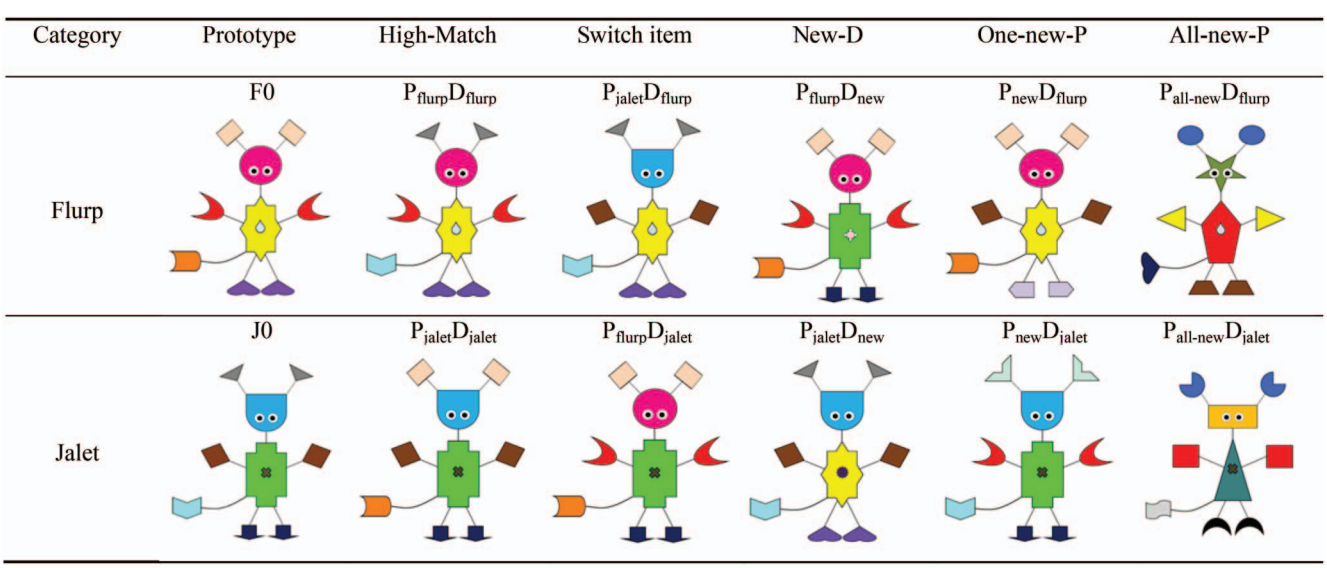

Figure 1. Examples of stimuli used in this study. Each row depicts items within a category, whereas each column identified an item role (e.g., switch item) and item type (e.g., $P_{jalet}D_{flurp}$). The high-match items were used in training and testing. The switch items, new-D, one-new-P, and all-new-P items were used only in testing. Neither prototype was shown in training or testing. Note that Figures 1–2 were presented to participants in color. See the online article for the color version of this figure.

Table 1
Category Structure Used in Experiments 1 and 2

Category F	Head	Body	Hands	Feet	Antenna	Tail	Button	Category J	Head	Body	Hands	Feet	Antenna	Tail	Button
F0	1	1	1	1	1	1	1	J0	0	0	0	0	0	0	0
$P_{flurp}D_{flurp}$	1	1	1	1	0	0	1	$P_{jalet}D_{jalet}$	0	0	0	0	1	1	0
$P_{jalet}D_{flurp}$	0	1	0	1	0	0	1	$P_{flurp}D_{jalet}$	1	0	1	0	1	1	0
$P_{flurp}D_{new}$	1	0	1	0	1	1	N	$P_{jalet}D_{new}$	0	1	0	1	0	0	N
$P_{new}D_{flurp}$	1	1	0	N	1	0	1	$P_{new}D_{jalet}$	0	0	1	0	N	0	0
$P_{all\text{-}new}D_{flurp}$	N	N	N	N	N	N	1	$P_{all\text{-}new}D_{jalet}$	N	N	N	N	N	N	0

Note. The value 1 = any of seven dimensions identical to Category *F*(flurp, see Figure 1). The value 0 = any of seven dimensions identical to Category J (jalet, see Figure 1). The value N = new feature which is not presented during training. p = probabilistic feature; D = deterministic feature. F0 is the prototype of Category F and J0 is the prototype of Category J. Items in the first row (i.e., F0 and J0) are prototypes and they were not used in either training or testing. Variants of items in the second row are *High-Match* items and they were used in both training and testing. Variants of all other item types were used only in testing.

subjects factor, with participants being presented with either classification or inference training. Instructions and testing were identical for both training conditions.

The procedures were similar for both adults and children and for all age groups the experiment was presented on the computer and controlled by E-prime software (Version 2.0; Schneider, Eschman, & Zuccolotto, 2002). There were minor differences between children's and adults' procedures pertaining to the way the instructions were presented, the questions were asked, and the responses were recorded. Adults read the instructions and questions on the computer screen and pressed the keyboard to make responses, whereas for children, a trained experimenter presented instructions and the questions verbally and recorded children's responses by pressing the keyboard. The experiment took approximately 10 min for adults and approximately 15 min for children. Most children and adults finished the experiment and, as evidenced by children's high recognition accuracy (see below), their response patterns do not stem from confusion or fatigue.

Instructions and training. In both training conditions, information about P and D features was explicitly given to participants before training. They were told that all flurps (or jalets) had a raindrop-shaped (or a cross-shaped) button and most of the flurps' (or jalets') features (at this point, the deterministic and probabilistic features were presented, one at a time). This information was repeated in the corrective feedback on each trial during training using the following script: *This one looks like a flurp (or a jalet) and it has the flurp's (or the jalet's) button*. Testing was not mentioned during the training phase. Participants were randomly assigned to one of the two training conditions.

The classification and inference training differed in the type of dimensions participants were asked to predict. In classification training, participants predicted the category label of each item given information about all other features. In Inference training, they predicted a missing feature of each item, given information about the remaining features and the label. The missing feature was randomized across trials (i.e., on one trial the missing feature could be the head and on another trial it could be the feet), it was always one of the four probabilistic features, and the value of the feature was always consistent with the prototype of a given category. Therefore, on each trial, participants were shown an item with one feature covered, the deterministic feature and three probabilistic features from one category, and two probabilistic features from the contrast category. In both conditions, participants were given 30 training trials (15 trials per category) and each trial was accompanied by corrective feedback. The order of the training trials in both conditions was randomized across participants.

Testing. The testing phase was identical for both conditions; it was administered immediately after training and included categorization and recognition tasks. During the testing phase, participants were presented with 40 test trials (eight trials per item-type; with equal number coming from each of the two categories) and were asked to determine (a) which category the creature was more likely to belong to, and (b) whether each creature was old (i.e., exactly the one presented during the training phase) or new. As we explain in the results section, the ways participants categorize and remember different item types provide critical information about what they attend to during category learning and thus which features are likely to be represented.

Each trial included a categorization and recognition question and the order of the questions was counterbalanced between participants and the order of the 40 test items was randomized across participants. All recognition questions referred to the *first part of the game* (i.e., the training phase), with participants being asked whether an item in question was presented during the first part of the game or was a new item. No feedback was provided during testing.

For categorization testing, the primary analyses focused on the proportion of responses in accordance with the D feature (i.e., rule-based responses). For recognition memory, the primary analyses focused on the difference between the proportion of hits (i.e., correctly identifying the high-match items that were presented during training as *old*) and false alarms (i.e., erroneously identifying other item types that were not presented during training as *old*).

If classification and inference learning result in different patterns of attention and in different category representations, then categorization and recognition performance should differ between the classification and inference training conditions. In particular, participants should rely on the deterministic feature when categorizing items in the classification condition, while relying on multiple probabilistic features in the Inference condition. They should also remember the D feature better than P features in the classification condition, but not in the inference condition. However, if classification and inference training elicit similar patterns of atten-

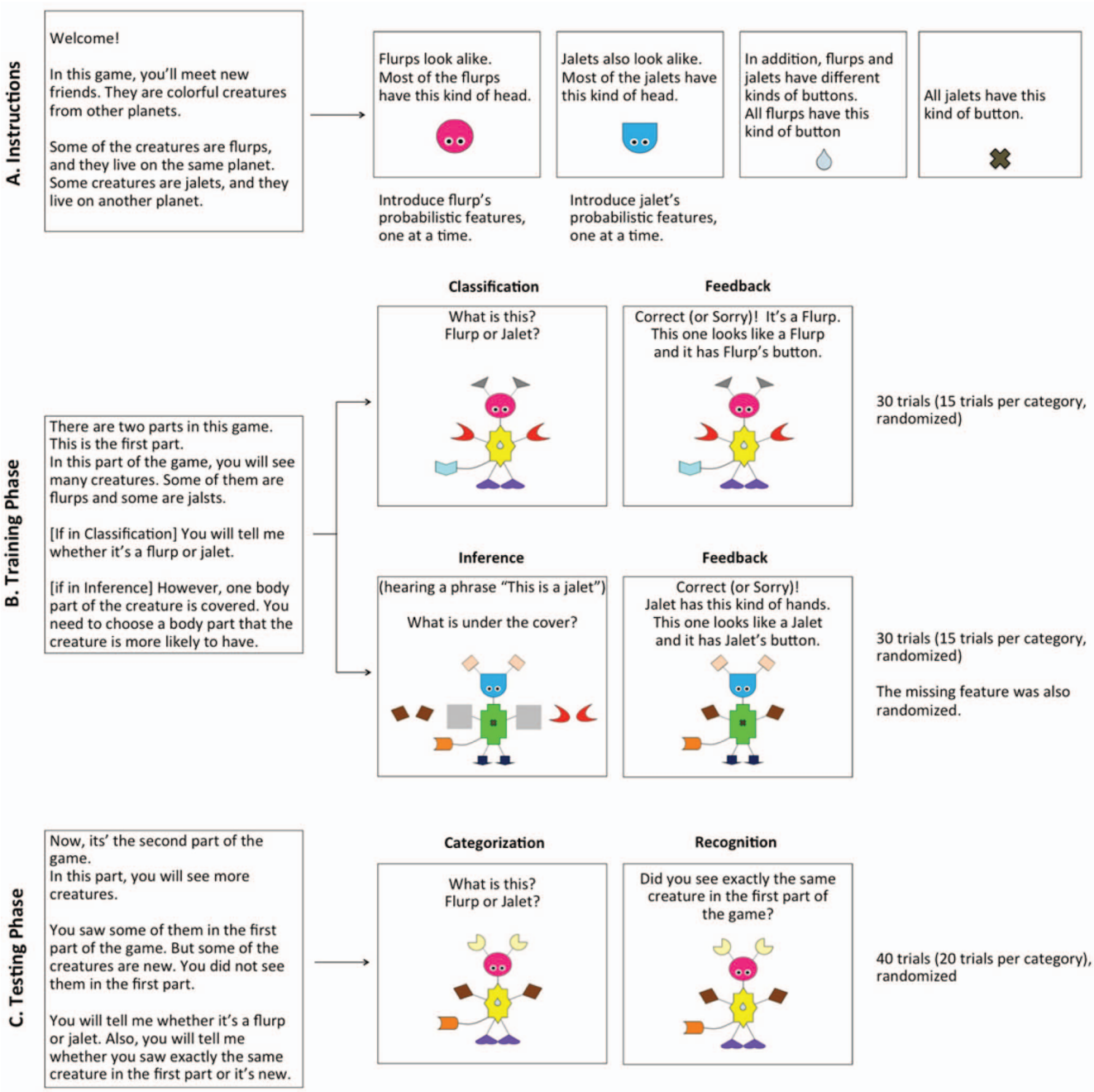

Figure 2. Overview of the procedure. In Experiment 1, the phases progressed from A to C. Half of the participants were presented with classification training in Phase B, whereas the other half were presented with Inference training in Phase B. The procedure of Experiment 2 was the same, except that (a) participants were given instructions and feedback during training focusing their attention on D features, and (b) there was only classification training condition. A. Adults; B. 6-year-olds; C. 4-year-olds. See the online article for the color version of this figure.

tion and result in similar representations, participants should exhibit symmetric patterns of categorization and recognition performance in the two training conditions.

Based on previous results (e.g., Hoffman & Rehder, 2010; see also Markman & Ross, 2003, for a review), we expected adults to exhibit representational asymmetry between classification and inference training, which should transpire in both categorization and recognition performance. In particular, in the classification condition adults should extract the most diagnostic (or rule) feature, whereas in the inference condition they should extract within-

category information (i.e., the overall similarity). At the same time, given the reviewed above evidence of diffused attention in younger children, we expected them to exhibit representational symmetry between the two training regimes. As a result, in both conditions, 4-year-olds should categorize on the basis of multiple features and should remember multiple features.

Results and Discussion

Analyses below focused on performance during training and testing. Note that testing performance is of primary importance because, in contrast to the training phase, all participants were presented with the same task.

Training phase. One adult in the Inference training was two standard deviations below the mean of accuracy in the last 10 training trials and data from this participant were excluded from the following analyzes. Training data aggregated into three 10-trial blocks across age groups and training conditions are presented in Table 2.

Overall, children and adults exhibited high training accuracy in the last 10 training trials in the classification training condition: 77.5% in 4-year-olds (above chance, $p < .001$), 94.5% in 6-year-olds (above chance, $p < .001$), and 96.5% in adults (above chance, $p < .001$). Performance was somewhat lower in the Inference training condition: 71.0% in 4-year-olds (above chance, $p < .001$), 72.5% in 6-year-olds (above chance, $p < .001$), and 88.5% in adults (above chance, $p < .001$).

A 2 (Training Type: Classification vs. Inference) × 3 (Age Group: 4-year-olds vs. 6-year-olds vs. Adults) between-subjects ANOVA revealed a main effect of age, $F(2, 114) = 16.27$, $MSE = 0.33$, $p < .001$, $\eta^2 = 0.222$, with adults being the most accurate whereas the 4-year-olds were the least accurate. There was also a main effect of condition, $F(1, 114) = 21.70$, $MSE = 0.44$, $p < .001$, $\eta^2 = 0.160$, with all age groups being less accurate in the inference training condition. Inference training was more difficult for both children and adults: In classification training they needed to remember the assignment of only two possible labels to two categories, whereas in Inference training, they had to remember the assignment of 12 possible features to two categories. Given these differences in difficulty, the differences between the training conditions are not surprising. In addition, inference training did not test category learning (just the participants' ability to infer the feature in question), and, as we demonstrate in the section on testing, in both training conditions participants learned categories well. Age differences are potentially informative and we return to this issue after the analyses of the testing phase.

Testing phase: Categorization. Categorization performance of each age group is presented in Figure 3 and Table 3. Preliminary analyzes focused on the ability to correctly categorize trained high-match items ($P_{flurp}D_{flurp}$ and $P_{jalet}D_{jalet}$), which was indicative of how well participants learned the categories (see Figure 3).

Across the training conditions, participants accurately categorized these test items (Adults: 96.3% in classification and 75.6% in inference, above chance, $ps < .001$; 6-year-olds: 90.0% in classification and 73.8% in inference, above chance, $ps < .001$; and 4-year-olds: 86.9% in classification and 75.6% in inference, above chance, $ps < .001$). A 3 (Age Group: 4-year-olds vs. 6-year-olds vs. Adults) × 2 (Training Condition: Classification vs. Inference) between-subjects ANOVA reveled a significant main effect of training condition, $F(1, 114) = 17.71$, $MSE = 0.77$, $p < .001$, $\eta^2 = 0.134$, with no main effect of age or an interaction, both $ps > .554$. Therefore, participants of all age groups learned both categories well, exhibiting somewhat better learning in classification training.

The second set of preliminary analyses focused on the ability to rely on familiar (i.e., seen during training) features when categorizing new-D, one-new-P, and all-new-P items. The mean proportions of reliance on old features when categorizing these items are presented in Table 3. High proportion of correct responses on all-new-P items indicates that the participant relies on old D features and generalizes broadly. High proportion of correct responses on new-D items indicates that the participant can categorize items on the basis of old P features, even when the D feature is new. And finally, high proportion of correct responses on one-new-P indicates that the participant can tolerate small distortion in the category prototype when categorizing items.

Data in Table 3 were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) × 3 (Age Group: 4-year-olds vs. 6-year-olds vs. Adults) × 2 (Training Condition: Classification vs. Inference) mixed ANOVA. There was a significant three-way interaction, $F(4, 228) = 3.49$, $MSE = 0.10$, $p = .009$, $\eta^2 = 0.058$. We broke down the interaction by conducting a mixed ANOVA on trial type and training condition for each age group.

For 4-year-olds, there was only a significant main effect of trial type, $F(2, 76) = 32.75$, $MSE = 0.66$, $p < .001$, $\eta^2 = 0.463$: regardless of the training condition categorization performance on new-D and one-new-P items was above chance ($ps < .001$) and at chance on all-new-P items ($ps > .413$). Recall that all-new-P items had the studied D features and all new P features, which revealed the inability of 4-year-olds to rely exclusively on a single deterministic feature. At the same time, 4-year-olds successfully relied

Table 2
Training Data: Mean (Standard Deviation) Proportion of Correct Responses Aggregated in 10-Trial Blocks Across Age Groups and Training Conditions in Experiments 1 and 2

Experiment	Age group	Training type	Trials 1–10	Trials 11–20	Trials 21–30
Experiment 1	Adults	Classification	0.91 (0.11)	0.93 (0.12)	0.97 (0.11)
		Inference	0.85 (0.16)	0.84 (0.18)	0.89 (0.13)
	6-year-olds	Classification	0.85 (0.14)	0.91 (0.18)	0.95 (0.14)
		Inference	0.60 (0.17)	0.67 (0.18)	0.73 (0.17)
	4-year-olds	Classification	0.62 (0.17)	0.70 (0.20)	0.78 (0.17)
		Inference	0.54 (0.13)	0.61 (0.17)	0.71 (0.13)
Experiment 2	4-year-olds	Classification	0.73 (0.19)	0.85 (0.17)	0.89 (0.12)

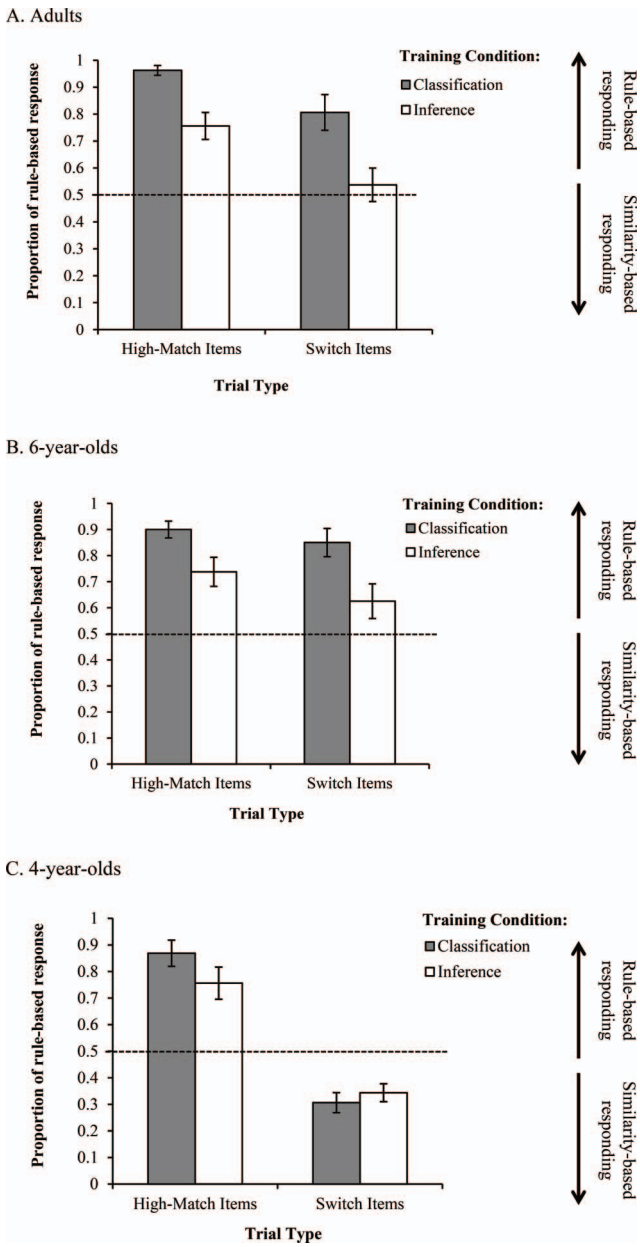

Figure 3. Categorization performance: Proportion of rule-based responses by trial type and training condition for adults (A), 6-year-old children (B), and 4-year-old children (C) in Experiment 1. High-match items are $P_{flurp}D_{flurp}$ and $P_{jalet}D_{jalet}$; switch items are $P_{flurp}D_{jalet}$ and $P_{jalet}D_{flurp}$. A. Adults; B. 6-year-olds; C. 4-year-olds.

on multiple features that were either all probabilistic (as in new-D items) or a combination of probabilistic and deterministic (as in one-new-P items).

For adults, there was a significant trial type by training condition interaction, $F(2, 76) = 3.80$, $MSE = 0.12$, $p = .027$, $\eta^2 = 0.091$. Specifically, adults were able to correctly categorize all three types of items regardless of the training condition (above chance, $ps < .001$), but they exhibited better performance on all-new-P items (reliance on D features) in classification condition than inference condition ($p = .042$, Bonferronni adjusted).

For 6-year-olds (similar to adults) there was a significant interaction, $F(2, 76) = 8.43$, $MSE = 0.32$, $p < .001$, $\eta^2 = 0.182$. Specifically, they ably relied on old features when categorizing one-new-P (these had old D and most of old P features) and all-new-P items (these had only old D features) in the classification condition, $ps < .001$. In contrast, in the inference condition they could correctly categorize only one-new-P items, $p = .002$. Therefore, 6-year-olds' performance was more similar to adults in the classification condition, but more similar to 4-year-olds in the inference condition, which suggests that this is a transitional group.

Overall, adults could rely on either old D features (when presented with all-new-P items) or old P features (when presented with new-D items), with somewhat higher reliance on old D features in the classification condition. Four-year-olds, regardless of the condition, relied on multiple features, but failed to rely on a single D feature. Finally, 6-year-olds could rely on the old D feature only in classification, but not in inference condition. These results point to the predicted asymmetry in adults: although adults could rely on either feature type, they were more likely to rely on a single D feature in classification than in inference training. In contrast, 4-year-olds exhibited symmetric performance relying on multiple features regardless of the training condition. At the same time, 6-year-olds appear to be a transitional group.

The primary analyses focused on comparison of categorization of the switch items (i.e., $P_{flurp}D_{jalet}$ and $P_{jalet}D_{flurp}$) across the training conditions (see Figure 3). These data were analyzed with a 3 (Age Group: 4-year-olds vs. 6-year-olds vs. Adults) × 2 (Training Condition: Classification vs. Inference) between-subjects ANOVA. As predicted, there was a significant Training Condition × Age Group interaction, $F(2, 114) = 4.50$, $MSE = 0.27$, $p = .013$, $\eta^2 = 0.073$. Specifically, adults and 6-year-olds exhibited asymmetry between classification and inference training by relying on the D features following classification training (above chance, both $ps < .001$, both $ds > 1.03$), but not inference training, both $ps > .079$. In contrast, 4-year-olds exhibited symmetry relying on P features in both conditions, both $ps < .001$, both $ds > 1.03$, with no difference between the training conditions, $p = .465$.

The asymmetry between classification and inference training in adults is consistent with previous evidence (Hoffman & Rehder, 2010; Yamauchi & Markman, 1998) suggesting differences in representations formed as a result of classification and inference training. Similar to previous findings, adults tended to process and represent categorical information differently, with classification learners being more likely than inference learners to focus on the D feature, which separated the two categories. At the same time, regardless of the training condition, 4-year-olds relied on the P features. The symmetric performance in 4-year-olds is a novel finding suggesting that unlike adults and 6-year-olds), they formed similarity-based representation of categories in both conditions.

However, although categorization performance points to differences in representation, this evidence is suggestive because categorization performance may not distinguish between the representation and decision processes. For example, participants could represent all the features equivalently, but put different decision weights on some features over others. Alternatively, they could represent only some features, but not the others (see Kloos & Sloutsky, 2008, for a discussion of these issues). These issues

Table 3
Categorization at Test: Mean (Standard Deviation) Proportions of Responses Based on Old Features in New-D, One-New-P, and All-New-P Items in Experiments 1 and 2

Experiment	Age group	Training type	New-D	One-new-P	All-new-P
Experiment 1	Adults	Classification	0.70 (0.23)	0.98 (0.06)	0.94 (0.12)
		Inference	0.73 (0.22)	0.88 (0.20)	0.75 (0.29)
	6-year-olds	Classification	0.53 (0.22)	0.91 (0.18)	0.94 (0.13)
		Inference	0.54 (0.22)	0.71 (0.26)	0.59 (0.26)
	4-year-olds	Classification	0.74 (0.23)	0.71 (0.21)	0.51 (0.16)
		Inference	0.75 (0.17)	0.79 (0.21)	0.54 (0.23)
Experiment 2	4-year-olds	Classification	0.59 (0.23)	0.80 (0.23)	0.84 (0.16)

Note. *New-D items* (i.e., $P_{flurp}D_{new}$ and $P_{jalet}D_{new}$) had probabilistic features of the studied categories and a novel feature replacing the deterministic feature. High proportion of correct responses on new-D items indicates that the participant can categorize items on the basis of old P features, even when the D feature is new. *One-new-P items* (i.e., $P_{new}D_{flurp}$ and $P_{new}D_{jalet}$) had all features of the studied categories but a novel feature replacing one probabilistic feature. High proportion of correct responses on one-new-P indicates that the participant can tolerate small distortion in the category prototype when categorizing items. *All-new-P items* (i.e., $P_{all\text{-}new}D_{flurp}$ and $P_{all\text{-}new}D_{jalet}$) had the deterministic features from the studied categories and all new features replacing the studied probabilistic features. High proportion of correct responses on all-new-P items indicates that the participant relies on old D features and generalizes broadly. The scale effectively ranges from 0.5 to 1, with 0.5 being chance performance.

could be addressed by analyzing participants' memory for the studied categories.

Testing phase: Recognition memory. The proportions of *old* responses on different item types are presented in Table 4 (*old* in response to a high-match item is a hit, whereas in response to the other items types it is a false alarm). As shown in the table, participants readily distinguished the studied high-match items from all-new-P items (all differences between hits and false alarms were above 0.53, which was greater than the chance level of 0, $ps < .001$).

Memory accuracy for the category-inclusion rule (i.e., D feature) and for the overall appearance (i.e., P features) was compared for each age group. Memory accuracy for the rule was obtained by subtracting false alarms on new-D items from hits on high-match items and memory accuracy for appearance features was obtained by subtracting false alarms on one-new-P items from hits on high-match items. The main results are presented in Figure 4 and data in the figure indicate that memory accuracy for D and P features was above chance level of 0 for all age groups, $ps < .001$.

To determine effects of training condition on representation of D and P features, data in Figure 4 were submitted to a 2 (Feature Type: D vs. P) × 2 (Training Condition: Classification vs. Inference) mixed ANOVA, with feature type as a within-subjects factor and training condition as a between-subjects factor. For adults, there was a significant feature type by training condition interaction, $F(1, 38) = 15.08$, $MSE = 0.53$, $p < .001$, $\eta^2 = 0.284$. Specifically, in the classification condition participants exhibited better memory for the D feature than for any single P feature, paired-samples $t(19) = 2.70$, $p = .014$, $d = 0.75$, whereas in the inference condition, participants exhibited better memory for any single P feature than for the D feature, paired-samples $t(19) = 2.80$, $p = .012$, $d = 0.53$.

For 4-year-olds, neither the main effects ($ps > .729$) nor the interaction ($p = .318$) were significant. Specifically, participants exhibited equivalent memory accuracy for any single P feature and for the D feature in classification condition ($p = .658$) and in inference condition ($p = .320$). Furthermore, as shown in Figure 4, their memory accuracy was uniformly high.

For 6-year-old children, there was a significant interaction between feature type and training type, $F(1, 38) = 8.17$, $MSE = 0.41$, $p = .007$, $\eta^2 = 0.177$. Specifically, similar to adults, 6-year-olds exhibited better memory for the D feature than for any single P feature in the classification condition, paired-samples $t(19) =$

Table 4
Memory at Test: Mean (Standard Deviation) Proportions of Yes Responses (i.e., Old Responses) on Different Item Types in Experiments 1 and 2

Experiment	Age group	Training type	High-match	New-D	One-new-P	All-new-P
Experiment 1	Adults	Classification	0.94 (0.11)	0.07 (0.15)	0.23 (0.21)	0.01 (0.04)
		Inference	0.54 (0.36)	0.32 (0.31)	0.16 (0.17)	0.01 (0.04)
	6-year-olds	Classification	0.88 (0.19)	0.24 (0.32)	0.49 (0.34)	0.22 (0.36)
		Inference	0.89 (0.17)	0.35 (0.39)	0.31 (0.26)	0.11 (0.19)
	4-year-olds	Classification	0.86 (0.19)	0.28 (0.23)	0.31 (0.27)	0.12 (0.21)
		Inference	0.88 (0.16)	0.31 (0.21)	0.25 (0.22)	0.11 (0.17)
Experiment 2	4-year-olds	Classification	0.88 (0.14)	0.36 (0.35)	0.38 (0.41)	0.30 (0.37)

Note. The *overall memory accuracy* is estimated by the difference in the proportion of *yes* responses to High-Match items and to all-new-P items. *Memory accuracy for the rule (i.e. the D feature)* is estimated by the difference in the proportion of *yes* responses to High-Match items and to new-D items. *Memory accuracy for the overall appearance (i.e., P features)* is estimated by the difference in the proportion of *yes* responses to High-Match items and to one-new-P items. The scale ranges from 0 to 1, with 0 being chance performance.

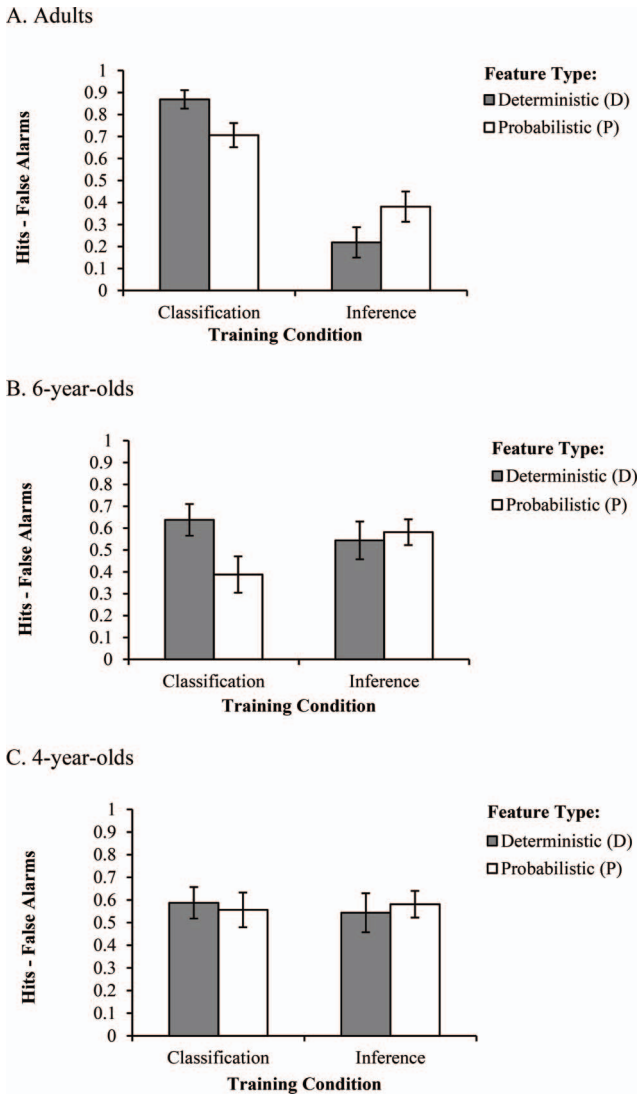

Figure 4. Recognition performance: Memory accuracy by feature type and training condition for adults (A), 6-year-old children (B), and 4-year-old children (C) in Experiment 1.

2.99, $p = .008$, $d = 0.72$, but not in the inference condition, paired-samples $t(19) = 0.67$, $p = .511$.

Therefore, recognition memory accuracy corroborates findings stemming from categorization performance: Whereas adults and 6-year-olds exhibited attentional (and potentially representational) asymmetry between the classification and inference conditions, 4-year-old children attended equivalently in both conditions, and were likely to form similar representations in both conditions.

Overall, the reported results revealed different patterns of representation between adults and 6-year-olds versus 4-year-olds. Specifically, after classification training, adults and 6-year-olds were more likely to extract the deterministic features than after inference training. In contrast, 4-year-olds performed symmetrically exhibiting similarity-based representation, regardless of the training condition. Results from 6-year-olds suggest that the developmental change in category representations may begin to occur between 4 and 6 years of age and continue after 6 years of age. These are novel findings pointing to important developmental differences in categorization: young children initially tend to form similarity-based representations, but in the course of development they acquire the ability to form more rule-based representation, and form, depending on the task, either rule-based or similarity based representations.

Although these findings are important, one potential concern is the fact that 4-year-olds were significantly lower than 6-year-old or adults in classification training (see Table 2). Specifically, as discussed in the section on the results of training, categorization performance of 4-year-olds on the last 10 trials was 78%, which was significantly lower than 95% and 97% accuracy exhibited by 6-year-olds and adults. It could be argued therefore, that the differences between 4-year-olds and the other groups can be explained by this somewhat lower learning. Although, this explanation is unlikely, given that no differences in category learning transpired at testing, we deemed it necessary to address this issue directly. It turned out that training performance in 4-year-olds yielded enough variability to separate a group of high learners ($N = 10$, $M_{accuracy} = 91\%$), whose training performance did not differ from that of 6-year-olds and adults (both ps $> .204$, Cohen ds < 0.50). The analyses indicated that high learners exhibited the same pattern as the entire sample: they relied on P features on switch items ($M_{Deterministic} = 0.26$, below chance, $p < .001$) and they exhibited exceedingly high memory for both D and P features (0.75 and 0.79, respectively), with no difference between the two feature types. These analyses strongly indicate age differences in training cannot explain age differences in categorization and recognition memory performance.

Having found these developmental differences, it is reasonable to ask: What drives the development? As argued above, there are reasons to believe that these differences are driven by different patterns of attention allocated during category learning: 4-year-olds attend diffusely regardless of the training condition, whereas 6-year-olds and adults exhibit focused attention in classification, but not in inference training. Experiment 1 presents suggestive evidence supporting this possibility and the goal of Experiment 2 is to test it directly.

Experiment 2 was based on the following reasoning: there are some conditions under which 4-year-olds may selectively attend to a single feature, although this selectivity is likely to be exogenous, or driven by characteristics of the stimuli, such as stimulus salience. For example, Deng and Sloutsky (2012) demonstrated that 4-year-olds categorized on the basis of a single salient feature (i.e., pattern of motion) rather than on a combination of multiple probabilistic features and the label.

Therefore, under some conditions, it is possible to affect 4-year-olds' attention exogenously and we attempt to do that in Experiment 2 in a more subtle way, without changing the stimuli used in Experiment 1. To achieve this goal, we attempted to direct participants' attention to the D feature by mentioning only this feature on each training trial. Given that adults and 6-year-olds exhibited evidence of relying on the D feature only in the classification training condition, we presented 4-year-olds in Experiment 2 with only classification training. If our manipulation is successful and 4-year-olds' performance will become more similar to adults, we would implicate attention as an important factor driving development.

Experiment 2

Method

Participants. There were 20 preschool children (M = 55.3 months, range 48.3–62.7 months; 11 girls) participating in this experiment. Child participants were recruited from childcare centers and preschools, located in middle-class suburbs of Columbus and were tested in a quiet room in their childcare center or preschool by a female experimenter.

Materials, design, and procedure. The materials, design, and procedure were similar to those in Experiment 1, with two critical differences. First, in contrast to Experiment 1, participants were only given classification training. Second, in contrast to Experiment 1, we directed participants' attention only to the D feature, both during instructions and during training. In particular, participants were told that all flurps (or jalets) had a raindrop-shaped (or a cross-shaped) button (at this point, the D feature of each category was presented, one at a time). In addition, this information was repeated in the corrective feedback to each response during training (i.e., *This one has the flurp's [or the jalet's] button*). The testing phase (both categorization and recognition trials) and the logic of both preliminary and primary analyses were the same as in Experiment 1.

Results and Discussion

Training phase. Training data aggregated into three 10-trial blocks across age groups and training conditions are presented in Table 2. Overall, children exhibited high training accuracy (88.5%, above chance, $p < .001$) in the last 10 training trials.

Testing phase: Categorization. Preliminary analyzes of categorization performance focused on the ability to correctly categorize high-match items. As shown in Figure 5A, child participants accurately categorized these test items (75.0%, above chance, $p = .001$), which indicated that they learned both categories well.

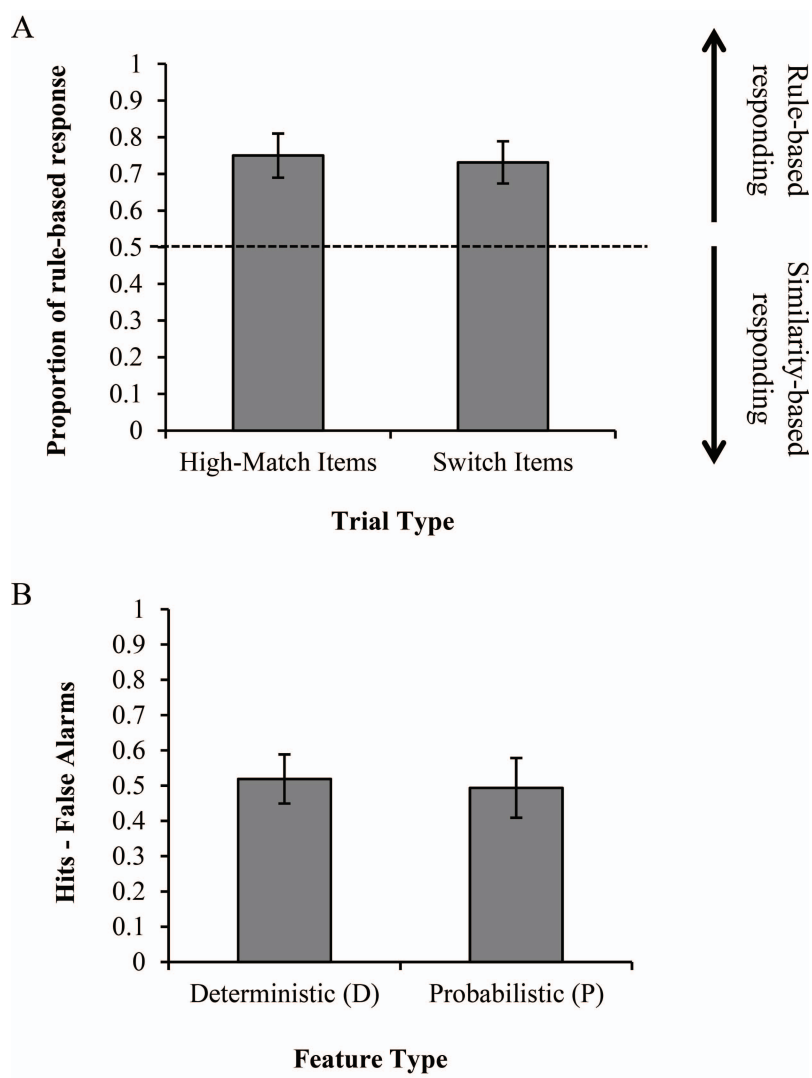

Figure 5. Categorization performance (A) and recognition performance (B) in Experiment 2.

The second set of preliminary analyses focused on new-D, one-new-P, and all-new-P items and the mean proportions of reliance on old features when categorizing these items are presented in Table 3. Data in the table indicate that in Experiment 2, 4-year-olds (who were presented only with classification training) were similar to 6-year-old classification learners in Experiment 1. In particular, 4-year-olds successfully categorized on the basis of old D-features (categorization of all-new-P items was 84%, above chance, $p < .001$), but not on the basis of old P-features (performance on new-D items was 59%, not different from chance, $p = .083$).

The primary analyses focused on categorization of the switch items (i.e., $P_{flurp}D_{jalet}$ and $P_{jalet}D_{flurp}$ items). The main results are shown in Figure 5A. As shown in the figure, unlike the 4-year-old participants in Experiment 1, who relied on the P features to categorize switch items ($M_{switch} = 0.31$), 4-year-olds in Experiment 2 relied on the D feature ($M_{switch} = 0.73$), with the proportion of rule-based responses on switch items being significantly above chance, one-sample $t(19) = 4.02$, $p = .001$, $d = 0.90$. Therefore, attentional manipulation was successful in affecting participants' patterns of categorization response.

Testing phase: Recognition memory. The proportions of *old* responses on different item types are presented in Table 4. Overall, participants readily distinguished the studied high-match items from and all-new-P items (the difference between hits and false alarms was 0.58, which was greater than the chance level of 0, $p < .001$).

Similar to Experiment 1, participants' memory accuracy for the category-inclusion rule (i.e., D feature) and for the overall appearance (i.e., P features) was compared and these data were presented in Figure 5B. As shown in the figure, participants exhibited equivalently high memory accuracy for D feature ($M = 0.52$) and for any single P feature ($M = 0.49$), $p = .811$, both above the chance level of 0, $ps < .001$, $ds > 1.30$.

Overall, when their attention was focused on the D feature in Experiment 2, 4-year-olds categorization performance was similar to the adults and older children in Experiment 1 (who tended to rely on the rule to categorize items). At the same time, the memory pattern of 4-year-olds in Experiment 2 was similar to that of 4-year-olds in Experiment 1: In both Experiments 4-year-olds exhibited equivalently high memory accuracy for both D and P features. The latter finding suggests that in Experiment 2, 4-year-olds continued to distribute their attention across multiple features.

The results are both important and surprising: the fact that the attentional manipulation affected categorization responses but not memory points to an important dissociation between (a) response or decision strategies that transpire in the pattern of categorization responses, and (b) pattern of attention during category learning and category representation that transpire in the pattern of recognition memory for items. In other words, attention seems be engaged both during item encoding (i.e., when representations are formed) and during decision (when categorization is performed) and we succeeded in affecting the latter, but not the former. This finding is important because it suggests that (a) categorization is affected by both representation and decision factors, (b) these factors may be independent, (c) young children may differ from adults with respect to both factors, and (d) developmental change may affect categorization *decisions* prior to affecting category *representations*. We return to each of these issues in the General Discussion.

General Discussion

The reported study presents several novel findings pointing to important developmental differences in category learning, category representation, and the role of selective attention in these processes. First, whereas adults and 6-year-olds exhibited representational asymmetry between classification and inference training, 4-year-olds formed similar representations across these training conditions. Specifically, older participants were more likely to use and represent a rule in classification training than in inference training (cf. Hoffman & Rehder, 2010; Yamauchi & Markman, 1998), whereas 4-year-olds were likely to form similarity-based representations in both conditions. In addition, in Experiment 2 attracting attention of 4-year-olds to only D features affected their categorization performance, but not their representation of categories, implicating representation and decision factors in categorization. These findings elucidate the development of category learning and categorization by linking it to the development of selective attention and they have important implications for theories of categorization and category learning. In addition, these results may contribute to better understanding of the role of linguistic labels in categorization. In what follows, we discuss each of these points.

Selective Attention and Development of Categorization

Representational differences between classification and inference training in adults have been inferred from different allocation of attention in the two tasks (e.g., Hoffman & Rehder, 2010). In classification training, adults shifted attention to the category boundary (which is best marked by a deterministic feature separating the two categories), whereas in inference training they distributed attention among features within a category in an attempt to find how features are related to one another within a category. Therefore, adults allocate attention flexibly and in a task-dependent manner. Results of Experiment 1 (both categorization and recognition memory findings) suggest that similar to adults, 6-year-olds also tend to optimize attention in classification training, but not in inference training.

In contrast to adults, selective attention in young children is immature (see Plude et al., 1994, for a review) and they may have difficulty shifting attention to a single feature (unless, this feature is highly salient and it captures attention automatically). In particular, they tend to distribute attention among multiple features within a category and to form representations that are based on multiple features (which transpires in their categorization and memory performance). There is recent evidence supporting this possibility and indicating that in contrast to adults, successful category learning in infants was accompanied by distributed attention (Best et al., 2013).

Current results present additional evidence implicating selective attention in the development of categorization: Adults and 6-year-olds exhibited more focused attention in classification than in inference training, whereas 4-year-olds exhibited distributed attention across the training conditions. Results of Experiment 2 further elucidated the role of selective attention in the development of categorization: attracting attention of 4-year-olds' to D features resulted in adult-like categorization performance, yet their recognition memory performance (i.e., equivalently high memory for D and P features) was similar to that of 4-year-olds in Experiment 1. This dissociation between categorization and recognition memory

suggests that (a) categorization may include representation and decision components, (b) both components may depend on attention, (c) it may be easier to externally access the decision component than the representation component, and (d) development of categorization may include both decision and representation components. Therefore, current research presents novel evidence about the role of selective attention in category learning across development, suggesting that (a) important developments occur between 4 and 6 years of age, and (b) the decision component of categorization (which is more likely to be affected externally) may develop somewhat earlier than the representation component (which may require more endogenous selective attention).

Taken together, these results advance our understanding of the development of categorization, indicating that (a) category representation undergoes development; (b) the development results in acquiring the ability to form rule-based representations, without losing the ability to form similarity-based representations; and (c) the development of domain-general selective attention contributes to the development of category representations. Although these results do not eliminate the possibility that domain-specific knowledge contributes to the development of categorization (see Carey, 1999; Inagaki & Hatano, 2002; Keil, 1992, for arguments for domains-specific changes), they clearly emphasize the contribution of domain-general processes.

Implications for Theories of Categorization

The reported results may have important implications for theories of categorization. This is particularly true with respect to the reported differences between categorization decisions and category representation observed in 4-year-olds in Experiment 2. Recall that after being instructed to focus on deterministic features, 4-year-olds exhibited rule-based categorization performance, which was similar to that of 6-year-olds and adults in Experiment 1. At the same time, similar to 4-year-olds in Experiment 1, 4-year-olds in Experiment 2 remembered deterministic and probabilistic features equally well. This pattern suggests that whereas they were likely to use the deterministic feature in their categorization decisions, they were equally likely to encode deterministic and probabilistic features in the course of category learning, which suggests distributed rather than selective attention. However, most models of categorization assume that decisions are made on the basis of representations, which implies that similar representations (observed in 4-year-olds in Experiments 1 and 2) should not result in radically different categorization decisions (observed in 4-year-olds across the two experiments). This dissociation between categorization decisions and category representation found in young children requires additional theory development.

Limitations and Future Directions

There are several important issues that are not addressed by current research and will require further research. For example, it could be argued that the current results may not generalize readily to more realistic categories or to categories in knowledge domains, such as biology. It could also be argued that the saliency of the deterministic feature may affect category learning and category representation in 4-year-olds. This is not impossible given that 4- to 5-year-olds can rely on a single feature if the feature is highly salient, such as a pattern of motion (e.g., Deng & Sloutsky, 2012, 2013). However, if a more salient feature is used, it is likely that young children will rely on this feature in both classification and inference learning, whereas adults are likely to continue exhibiting differences between classification and inference training. Finally, it could be argued that the lack of conceptual status of the deterministic feature may make young children less likely to use it in categorization (e.g., Gelman & Davidson, 2013). Although we have preliminary data demonstrating that this is not the case, more evidence is needed to fully address this issue.

Conclusion

Despite these limitations and need for future research, current research presents novel evidence pertaining to the development of categorization and the role of attention in this process. Whereas adults and 6-year-olds are more likely to attend selectively to diagnostic features in classification than in inference training, 4-year-olds tend to attend diffusely regardless of the training condition. These findings point to important changes in category representation between 4 and 6 years of age: the development consists of acquisition of the ability to form rule-based representations, without losing the ability to form similarity-based representation.

Furthermore, exogenous factors may affect category decisions of younger children, but not their category representation, which points to a potential distinction between decision and representation components of categorization and suggests that developmental changes in decision components may occur prior to developmental changes in representation. These results have important implications for theories of categorization, understanding the development of categorization, and the changing role of selective attention in category learning.

References

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116,* 105–119. http://dx.doi.org/10.1016/j.jecp.2013.05.002

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change. In S. Carey & R. Gelman (Eds.), *The Epigenesis of mind: Essays on biology and cognition* (pp. 257–291). Hillsdale, NJ: Erlbaum. http://dx.doi.org/10.1002/edp.2430040307

Deng, W., & Sloutsky, V. M. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science, 23,* 178–186. http://dx.doi.org/10.1177/0956797611429133

Deng, W., & Sloutsky, V. M. (2013). The role of linguistic labels in inductive generalization. *Journal of Experimental Child Psychology, 114,* 432–455. http://dx.doi.org/10.1016/j.jecp.2012.10.011

Eimas, P. D. (1994). Categorization in early infancy and the continuity of development. *Cognition, 50,* 83–93. http://dx.doi.org/10.1016/0010-0277(94)90022-1

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development, 65,* 903–917. http://dx.doi.org/10.2307/1131427

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology, 66,* 327–353. http://dx.doi.org/10.1016/j.cogpsych.2013.02.001

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science, 13,* 622–635. http://dx.doi.org/10.1111/j.1467-7687.2009.00921.x

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. http://dx.doi.org/10.1037/a0019042

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York, NY: Psychology Press.

Keil, F. C. (1992). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23,* 221–236. http://dx.doi.org/10.1016/S0022-5371(84)90148-8

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137,* 52–72. http://dx.doi.org/10.1037/0096-3445.137.1.52

Lazareva, O. F., Freiburger, K. L., & Wasserman, E. A. (2004). Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychonomic Bulletin & Review, 11,* 1111–1117. http://dx.doi.org/10.3758/BF03196745

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332. http://dx.doi.org/10.1037/0033-295X.111.2.309

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129,* 592–613. http://dx.doi.org/10.1037/0033-2909.129.4.592

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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. http://dx.doi.org/10.1111/j.1467-8624.2004.00821.x

Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Object examining: Habituation and categorization. *Cognitive Development, 6,* 377–392. http://dx.doi.org/10.1016/0885-2014(91)90045-F

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica, 86,* 227–272. http://dx.doi.org/10.1016/0001-6918(94)90004-3

Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: The role of age and executive functioning. *PLoS ONE, 9,* e85316. http://dx.doi.org/10.1371/journal.pone.0085316

Robinson, C. W., Best, C. B., & Sloutsky, V. M. (2011, March). *Auditory dominance and category learning*. Paper presented at Biennial Meeting of Society for Research in Child Development, Montreal, Canada. http://dx.doi.org/10.1037/e516372011-002

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75,* 1387–1401. http://dx.doi.org/10.1111/j.1467-8624.2004.00747.x

Robinson, C. W., & Sloutsky, V. M. (2007). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy, 11,* 233–253. http://dx.doi.org/10.1111/j.1532-7078.2007.tb00225.x

Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *Journal of Experimental Psychology: Applied, 16,* 361–377. http://dx.doi.org/10.1037/a0021610

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

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science, 34,* 1244–1286. http://dx.doi.org/10.1111/j.1551-6709.2010.01129.x

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. http://dx.doi.org/10.1037/0096-3445.133.2.166

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology, 111,* 65–86. http://dx.doi.org/10.1016/j.jecp.2011.07.007

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Pt. 1. Linguistic labels and the development of similarity judgment. *Developmental Psychology, 35,* 1478–1492. http://dx.doi.org/10.1037/0012-1649.35.6.1478

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., . . . Grace, R. C. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience and Biobehavioral Reviews, 36,* 2355–2369. http://dx.doi.org/10.1016/j.neubiorev.2012.09.003

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96,* 125–144. http://dx.doi.org/10.1037/0033-295X.96.1.125

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13,* 258–263. http://dx.doi.org/10.1016/j.tics.2009.03.006

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 585–593. http://dx.doi.org/10.1037/0278-7393.28.3.585

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–148. http://dx.doi.org/10.1006/jmla.1998.2566

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 776–795. http://dx.doi.org/10.1037/0278-7393.26.3.776

Received October 4, 2013
Revision received December 8, 2014
Accepted December 15, 2014 ■