



Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych



Selective attention, diffused attention, and the development of categorization



Wei (Sophia) Deng^{a,*}, Vladimir M. Sloutsky^{b,*}

^a Department of Psychology, University of Macau, China.

^b Department of Psychology, The Ohio State University, United States

ARTICLE INFO

Article history:

Accepted 24 September 2016

Keywords:

Categorization

Attention

Learning

Cognitive development

ABSTRACT

How do people learn categories and what changes with development? The current study attempts to address these questions by focusing on the role of attention in the development of categorization. In Experiment 1, participants (adults, 7-year-olds, and 4-year-olds) were trained with novel categories consisting of deterministic and probabilistic features, and their categorization and memory for features were tested. In Experiment 2, participants' attention was directed to the deterministic feature, and in Experiment 3 it was directed to the probabilistic features. Attentional cueing affected categorization and memory in adults and 7-year-olds: these participants relied on the cued features in their categorization and exhibited better memory of cued than of non-cued features. In contrast, in 4-year-olds attentional cueing affected only categorization, but not memory: these participants exhibited equally good memory for both cued and non-cued features. Furthermore, across the experiments, 4-year-olds remembered non-cued features better than adults. These results coupled with computational simulations provide novel evidence (1) pointing to differences in category representation and mechanisms of categorization across development, (2) elucidating the role of attention in the development of categorization, and (3) suggesting an important distinction between representation and decision factors in categorization early in development. These issues are discussed with respect to theories of categorization and its development.

© 2016 Published by Elsevier Inc.

* Corresponding authors at: Department of Psychology, 247 Psychology Bldg., 1835 Neil Avenue, The Ohio State University, Columbus, OH 43210, United States (V.M. Sloutsky) and Department of Psychology, FSS, University of Macau, E21-3037, Avenida da Universidade, Taipa, Macau S.A.R., China (W. (Sophia) Deng).

E-mail addresses: wdeng@umac.mo (W. (Sophia) Deng), sloutsky.1@osu.edu (V.M. Sloutsky).

1. Introduction

Categorization, or the ability to form equivalence classes of discriminable entities, is an essential component of human cognition. This ability to assign *sameness* to appreciably different stimuli was identified by William James (1883/1890) as the key property of the mind. Categories enable recognition and differentiation of objects, people, and events, help organizing our existing knowledge, and promote generalization of knowledge to new situations. For example, having observed that all of the previously encountered birds have been able to fly, one might infer that a newly encountered bird can fly as well. There are a number of key (and relatively uncontroversial) findings pertaining to categorization.

First, 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). And second, there is evidence of remarkable development in the ability to form and represent categories (e.g., Huang-Pollock, Maddox, & Karalunas, 2011; Kloos & Sloutsky, 2008; Minda, Desroches, & Church, 2008; Quinn & Johnson, 2000; Rabi, Miles, & Minda, 2015; Rabi & Minda, 2014; Smith, 1989; Younger & Cohen, 1986; see also Quinn, 2011; Sloutsky, 2010, for reviews). It is hardly controversial that adults can acquire exceedingly abstract (often non-perceptual) categories, whereas there is little evidence that infants or even young children can acquire categories of similar levels of abstraction. Although many agree that categorization does develop, there is less agreement as to *what* changes with development and *why*.

According to some explanations, the development is driven by acquisition of domain-specific (or even concept-specific) knowledge (e.g., Carey, 1999; Inagaki & Hatano, 2002; Keil, 1992; Keil & Batterman, 1984). According to other (more domain-general) explanations, the development is driven by changes in basic cognitive processes, such as selective attention, working memory, and cognitive control (see Fisher, Godwin, & Matlen, 2015; Rabi & Minda, 2014; Sloutsky, 2010; Sloutsky, Deng, Fisher, & Kloos, 2015; Smith, 1989). Selective attention is a particularly promising candidate because, as discussed in detail below, it (a) has been identified as an important component of adult category learning and (b) clearly undergoes development.

Although these possibilities provide radically different developmental accounts, they do not have to be mutually exclusive. Perhaps the former account explains developmental changes in how familiar categories are interconnected and used (see Fisher, Godwin, Matlen, & Unger, 2015), whereas the latter explains developmental changes in acquisition and representation of novel categories.

The goal of current research is to better understand developmental changes in how novel categories are learned and represented. We propose that changes in selective attention may drive this development, derive specific hypotheses from this general proposal, and test these hypotheses in the reported experiments. In the remainder of this section, we first review the role of selective attention in category learning and category representation. We then discuss the development of selective attention and its implication for category learning.

1.1. Selective attention, category learning, and category representation

Since the pioneering research on category learning by Shepard, Hovland, and Jenkins (1961), selective attention¹ has been an important component of models of categorization and category learning. Exemplar models (Hampton, 1995; Medin & Schaffer, 1978; Nosofsky, 1986), prototype models (Smith & Minda, 1998), clustering models (Love, Medin, & Gureckis, 2004), and dual process models (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) all include some form of selective attention as a factor

¹ Note that in the attention literature (e.g., Egeth & Yantis, 1997; Pashler, Johnston, & Ruthruff, 2001; Posner & Petersen, 1990) selective attention has been conceptualized as either involuntary, bottom-up, and stimulus-driven (when it is captured automatically by a highly salient stimulus) or as voluntary, top-down, and goal-driven (when the goal is to find a red object in a pile of things of different colors). In the categorization literature, selective attention has been conceptualized in the latter sense. For the purpose of consistency, in this paper we will use the conceptualization adopted in the categorization literature. However, we will discuss limitations of this conceptualization in Section 6.

determining the influence (or weight) of stimulus dimensions on categorization. According to some of these models, as learning progresses, these weights change, with changes occurring gradually in associative models and abruptly in rule-based models (see [Rehder & Hoffman, 2005](#) for a discussion).

There are three important sources of evidence of how selective attention affects category learning and under what conditions: (1) selective attention has consequences in the form of attention optimization and learned inattention, and these consequences are often observed after adults learn categories; (2) attention optimization is specific to learning categories of particular structures; (3) attention optimization is specific to learning categories under certain task conditions.

1.1.1. Consequences of selective attention: attention optimization and learned inattention

If category learning is accompanied by attending selectively to features distinguishing the to-be-learned categories (i.e., *diagnostic features*), such selectivity may result in a number of testable consequences (see [Hoffman & Rehder, 2010](#), for a review). The most important of these consequences is shifting attention to the diagnostic features (i.e., *attention optimization*) and learning to ignore less diagnostic or irrelevant features (i.e., *learned inattention*).

Consider a situation in which one learns two categories, such as squirrels versus chipmunks. As learning progresses, the learner's attention may shift to stripes (which is a diagnostic feature), which would indicate attention optimization. At the same time, the learner would attend less to or even ignore the tail, which is a non-diagnostic feature. This phenomenon of learned inattention to non-diagnostic features would transpire if, after learning the categories of squirrels and chipmunks, a learner embarks on a new categorization task – differentiating between squirrels and hamsters. In this case, the tail that was non-diagnostic for previous learning becomes diagnostic for current learning. Importantly, prior history of ignoring the tail would make it more difficult to shift attention to the tail than it would have been without learning of the first set of categories.

To examine this issue, [Hoffman and Rehder \(2010\)](#) presented adults with a multi-phase category learning task, such that dimensions that were diagnostic in phase 1 of category learning became non-diagnostic in phase 2, whereas dimensions that were non-diagnostic in phase 1 became diagnostic in phase 2. Using a combination of behavioral and eye tracking methodologies, the authors found that adult learners optimized attention in phase 1 by shifting it to the category-relevant (or diagnostic) dimension and exhibited learned inattention in phase 2. These findings suggest that in the course of category learning, adults tend to attend selectively, trying to extract the most diagnostic (or rule) dimension(s).² Attention optimization accompanying category learning in adults has been also reported in an eye-tracking study by [Blair, Watson, and Meier \(2009\)](#).

1.1.2. Attention is allocated differently to categories of different structures

In their seminal study of category learning, [Shepard et al. \(1961\)](#) identified a number of category structures that elicited different learning profiles in human adults. For example, type I category is the easiest and the most basic category structure: categorization decision can be made on the basis of a single dimension (e.g., blue items vs. red items). In contrast, type VI category is the most difficult one, as no dimension or their combination supports categorization: the assignment of each item to a category should be learned by rote and memorized. Therefore, whereas it is adaptive to attend selectively when learning type I category, it is hardly useful when learning type VI category. To examine this issue, [Rehder and Hoffman \(2005\)](#) recorded adult participants' eye movements during learning of categories of different structures. Their results indicated that participants examined primarily the diagnostic dimension in the case of type I category, thus presenting evidence of attention optimization. However, it could be argued that attention optimization is a consequence of any learning, not just selective attention to diagnostic features. This possibility was rejected when researchers examined attention allocation in the case of type VI category, when none of the dimensions was diagnostic. Upon observing that participants examined all dimensions, the researchers concluded that participants allo-

² Note that in most research discussed here participants learned visual categories. However, when categories were presented as lists of features that participants read, less selectivity/optimization was observed (e.g., [Bott, Hoffman, & Murphy, 2007](#)).

cate attention differently, depending on the structure of the to-be-learned category: selectively when there are diagnostic features and diffusely when there are no diagnostic features.

1.1.3. Attention is allocated differently across different categorization tasks

Multiple tasks can be used to elicit category learning. The tasks most frequently used in the lab studies are *classification* learning and *inference* learning. In the former task, 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 whether the item is labeled A or B. 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. A participant has to predict whether the non-revealed feature comes from features of category A or category B. There is evidence that classification and inference learning are not equivalent for adults and these tasks may result in different representations in adults (see Markman & Ross, 2003; Yamauchi & Markman, 1998, for extensive arguments).

If classification and inference tasks result in different representations, perhaps adults allocate attention differently in these tasks. This issue was addressed in the Hoffman and Rehder's (2010) study reviewed above. Recall that these authors presented adult participants with a multi-phase category-learning task and recorded participants' eye movements during learning. Hoffman and Rehder (2010) found evidence that classification learning, but not inference learning, resulted in optimized attention in phase 1 and learned inattention in phase 2. The authors concluded therefore that whereas selective attention transpires in some category learning tasks, it does not transpire in other category learning tasks: 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 they interrelate.

These findings also suggest that classification and inference learning lead to differences in allocation of attention and subsequently to differences in category representation. In classification learning, adults are likely to extract the most diagnostic (or rule) feature, whereas in inference learning they are more likely to extract within-category similarity. Note that, as we discuss below, differences between classification and inference learning do not transpire until 6-to-7 years of age, with younger children exhibiting similar performance in both types of tasks (Deng & Sloutsky, 2013, 2015a). These findings suggest that young children attend similarly in both types of tasks and subsequently form similar representations across these tasks.

Therefore, the reviewed evidence suggests that depending on the task and category structure, adults attend either selectively or diffusely and form representations that reflect their pattern of attention. In contrast, young children tend to distribute attention and form similar representations across different conditions. These findings are theoretically consequential for understanding of mature category learning and of developmental changes in categorization and category learning.

1.2. Diffused attention and early categorization and category learning

If adults attend selectively, at least under some category structure and learning task conditions, it is reasonable to ask: how do children learn categories under these conditions? The question is potentially informative because 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). Note that these difficulties transpire when no dimension captures attention automatically and top-down attention is required. In contrast, when a single highly salient feature or dimension captures attention automatically, young children focus on this dimension (Deng & Sloutsky, 2012), and they often have difficulty ignoring this feature or dimension (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; see also Robinson & Sloutsky, 2007, for similar tendency in infancy.).

There is much evidence supporting the idea of developmental differences in top-down selective attention, with older children and adults being generally better than younger children at selectively attending to a single dimension. These findings transpire in a variety of tasks, including rule use (e.g. Frye, Zelazo, & Palfai, 1995), discrimination learning (e.g. Kendler & Kendler, 1962), free

classification (Smith, 1989; Smith & Kemler, 1977), speeded classification (e.g. Smith & Kemler, 1978), and category learning (Best, Yim, & Sloutsky, 2013).

For example, in Smith and Kemler (1977, see also Smith, 1989) participants were presented with triads of two-dimensional stimuli. One of the stimuli matched the target on a single dimension, but had a very different value on the second dimension (e.g., the two could have the same color, but different shapes). At the same time, the other stimulus was similar to the target on both dimension, with neither dimension value being the same (e.g., the two had somewhat similar color and shape). When asked to select two out of three items that would go together, 5-year-olds opted for the overall similarity, whereas older children preferred dimensional matches. These findings suggest that young children tend to distribute attention across multiple dimensions, rather than focusing on a single dimension.

More recently, Best et al. (2013) presented 6–8-month-old infants and adults with a two-phase category learning task, such that the dimensions that were relevant in the first phase became irrelevant in the second phase. Results indicated that although both groups learned categories, their patterns of allocating attention differed. Adults optimized attention to category relevant dimensions in phase 1 and continued attending to these dimensions in phase 2 (when these dimensions were no longer relevant). In contrast, infants allocated attention to all dimensions in both phases. Therefore, whereas adults attended selectively when learning categories, infants attended diffusely. There is also some evidence suggesting a lack of attention optimization in category learning of 4–5 year-olds (Robinson, Best, & Sloutsky, 2011): even when children exhibited robust learning of a rule-based category, they failed to exhibit evidence of attention optimization.

Taken together, these studies support the idea that early in development children tend to distribute attention across multiple dimensions, unless a single dimension captures attention automatically. This pattern of attention allocation may have important consequences for how people learn and represent categories and what they remember after learning. Current research examines these consequences across development, with the goal of better understanding of developmental differences in the mechanism of categorization. We consider these general consequences and more specific predictions in the next two sections.

Importantly, diffused attention seems to be more than just a limitation in focusing attention early in development, it could be also an important mechanism, sub-serving early category learning. In one study, Deng and Sloutsky (2015b) presented infants with a category-learning task. While category learning was established using a traditional novelty preference procedure, attention during category learning was examined by recording participants' eye movements. Results indicated that more successful learning was accompanied by more distributed attention evidenced by a greater number of gaze shifts across different features of presented objects.

1.3. Attention and category representation across development

If young children and adults allocate attention differently in the course of category learning, they are also likely to form different representations: young children should represent all or most dimensions, whereas adults should represent primarily category-relevant dimensions. If this is the case, then categorization in adults should be accompanied by different representations across different situations (i.e., depending on a situation, they should represent different diagnostic dimensions), whereas categorization in young children should be accompanied by similar representations (i.e., across the situations, both diagnostic and non-diagnostic dimensions should be represented). In a recently published study, Deng and Sloutsky (2015a) trained 4-year-olds, 6-year-olds, and adults with either a classification task or an inference task and tested their categorization performance and memory for items. Adults and 6-year-olds exhibited an asymmetry: they relied on a single deterministic feature and formed rule-based representations during classification training, but not during inference training. In contrast, regardless of the learning regime, 4-year-olds relied on multiple probabilistic features and formed similarity-based representations. These findings suggest that whereas older children and adults attend selectively to a subset of features that were particularly useful for a given task, younger children tended to attend diffusely across the tasks. These developmental differences in attention allocation during category learning may have important consequences for what is remembered about the

category and what is not and for how these memories differ across development. We discuss these consequences in the next section.

1.4. How to infer category representation?

Attention is important for category representation because it affects which information is encoded into long-term memory and which is not (Chun, Golomb, & Turk-Browne, 2011). Therefore, analyses of memory data following learning can provide useful information about attention in the course of learning and about category representation. Specifically, if a dimension is remembered after learning, it was likely to be attended to in the course of learning. In addition, if a dimension is not remembered (or remembered poorly) it is unlikely to be part of category representation.

Furthermore, categorization and generalization data (i.e., how participants categorize items consisting of old and new features) may provide further evidence about representations being formed in the course of learning. For example, if a given feature is not used in categorization and generalization, it is unlikely to be a part of category representation, whereas if the feature is used in categorization and generalization, it is likely to be a part of category representation.

In addition to these relatively straightforward cases, it is also possible that features that are not used in categorization and/or generalization are still remembered. Such cases may highlight differences between representation and decision components in categorization. For example, learning a category of red objects versus blue objects may be achieved by color being encoded (because it is an attended dimension), and shape and texture not being encoded (because these are ignored dimensions). Alternatively, it is possible that all dimensions are encoded, but categorization decisions are made only on the basis of the diagnostic dimension. Therefore, cases when features are not used in categorization and/or generalization but are still remembered will require a more detailed analysis and we provide these analyses when such cases occur. If adults attend selectively, they should better encode and remember the features that control their categorization. In contrast, young children (i.e., those who are younger than 6 years of age) attend diffusely and they should remember well all the features.

1.5. Current study

The reviewed above theoretical considerations and evidence suggest a number of important hypotheses. First, adults, whose attention allocation may differ across tasks and conditions, should optimize attention in some conditions and distribute attention in other conditions. In contrast, across conditions, young children should attend diffusely. Second, depending on attention allocation, adults should extract different features and form different representations. In contrast, across conditions, young children should extract multiple features and form equivalent representations. And third, these differences in attention allocation should transpire in what is remembered: while young children should remember all or most features well, adults should have better memory for features that determine their categorization decisions.

The goal of the current study was to test these hypotheses, thus advancing our understanding of the development of categorization and the role of selective attention in this process. The reported study consisted of three experiments conducted with 4-year-olds, 7-year-olds, and adults. The basic task for each experiment consisted of three phases: instructions, training, and testing. As explained below, all attentional manipulations were introduced during the instructions and training phases, whereas the testing phase was identical across the experiments.

During training, participants predicted the category of a given item and received corrective feedback. There were two family-resemblance categories, with each training item having a single deterministic feature D (which perfectly distinguished the two categories) and multiple probabilistic features P (with each providing imperfect probabilistic information about category membership).

The testing phase consisted of categorization and recognition tasks and was administered immediately after the training phase. During testing participants were asked to determine (1) which category each item was more likely to belong to and (2) whether each item was old or new. No feedback was provided during testing. Categorization trials were designed to determine which features participants

rely on in their decisions, whereas recognition trials were designed to determine what participants remembered from training, which provides information about how they allocate attention during training and how they represent the learned categories.

Three experiments differed in how participants' attention was directed to different types of features. In Experiment 1, information about P and D features was explicitly mentioned to participants. The goal of Experiment 1 was to replicate and further extend [Deng and Sloutsky's \(2015a\)](#) findings with children and adults, with the goal of establishing a baseline for Experiments 2 and 3.

Based on the considerations reviewed above, it was predicted that because young children do not optimize attention, their categorization performance and recognition memory should differ from adults'. Specifically, young children should rely on multiple P features rather than the D feature in categorization and remember well all or most features. In contrast, adults, who ably optimize attention in category learning, should rely on the D feature and remember D feature better than P features. Performance of 7-year-olds (in comparison with the other two age groups) will help better understand the development of categorization.

In Experiments 2 and 3, we cued participants' attention, with the goal of examining whether attentional cueing results in changes in categorization performance (compared to the pattern established in Experiment 1) and in changes in underlying representations. Specifically, we directed participants' attention to the D feature in Experiment 2 and to the P features in Experiment 3. If we observe changes in both categorization performance and memory for features, then different ways of categorizing are driven by different underlying representations.

In contrast, if we observe changes only in categorization performance, but not in memory for features, then different ways of categorizing are driven by different decision weights of different features in different situations, whereas underlying representations are likely to remain the same across the situations. As we discuss in Section 6, this information is consequential for understanding the development of categorization, and for theories and models of categorization.

2. Experiment 1: establishing a baseline

2.1. Method

2.1.1. Participants

Participants were adults (15 women), 7-year-old³ children ($M_{\text{age}} = 83.2$ months, range 73.2–89.4 months; 10 girls), and 4-year-old children ($M_{\text{age}} = 54.1$ months, range 48.3–59.6 months; 7 girls), with 20 participants per age group. Adult participants 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 and were tested by an experimenter in a quiet room in their preschool. Data from two additional adults, one additional 7-year-old, and one additional 4-year-old were excluded from analyses because of extremely poor performance in training (their categorization performance was two standard deviations below the mean of accuracy in the last ten training trials). Data from one additional 4-year-old were also excluded from analyses because of the experiment being disrupted.

2.1.2. Materials

Materials were similar to those used previously by [Deng and Sloutsky \(2015a\)](#) 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 [Fig. 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”),

³ In this and all other experiments reported here, the older child participants consisted of 6-year-olds and 7-year-olds, with the mean ages being just under seven. For purposes of brevity, we refer to this group as “7-year-olds”.

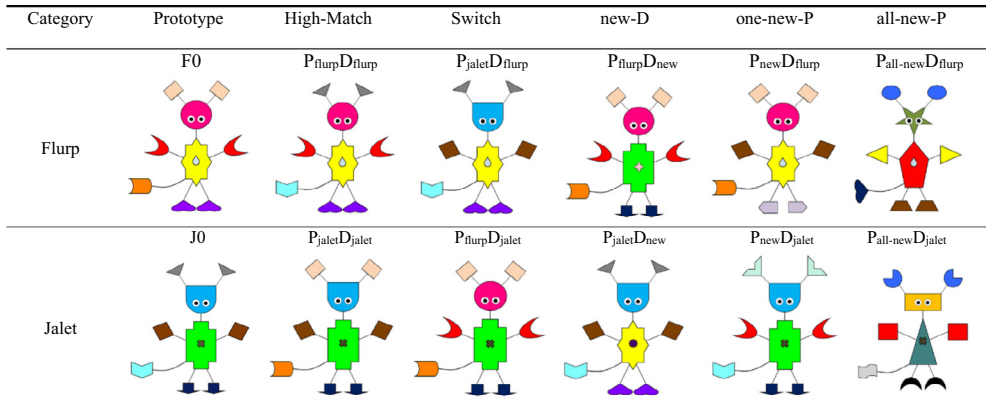


Fig. 1. Examples of stimuli used in this study. Each row depicts items within a category, whereas each column identified an item role (e.g., switch) and item type (e.g., $P_{jaletDflurp}$). The High-Match items were used in training and testing. The switch, new-D, one-new-P, and all-new-P items were used only in testing. Neither prototype was shown in training or testing.

Table 1
Category structure used in Experiments 1–3.

Item type	Stimulus	Probabilistic feature						Deterministic feature
		Head	Body	Hands	Feet	Antenna	Tail	Button
High-Match	$P_{flurpDflurp}$	1	1	1	1	0	0	1
	$P_{jaletDjalet}$	0	0	0	0	1	1	0
Switch	$P_{jaletDflurp}$	0	1	0	1	0	0	1
	$P_{flurpDjalet}$	1	0	1	0	1	1	0
New-D	$P_{flurpDnew}$	1	0	1	0	1	1	N
	$P_{jaletDnew}$	0	1	0	1	0	0	N
One-new-P	$P_{newDflurp}$	1	1	0	N	1	1	1
	$P_{newDjalet}$	0	0	1	0	N	0	0
All-new-P	$P_{all-newDflurp}$	N	N	N	N	N	N	1
	$P_{all-newDjalet}$	N	N	N	N	N	N	0

Note. The value 1 = any of seven dimensions identical to the prototype of Category F (flurp, see Fig. 1). The value 0 = any of seven dimensions identical to the prototype of Category J (jalet, see Fig. 1). The value N = new feature which is not presented during training. P = probabilistic feature; D = deterministic feature. This table presents an example of stimulus structure for each item type. See Appendix A for full tables of category structure of all the variants. Variants of High-Match items were used in both training and testing. Variants of all other item types were used only in testing.

whereas one feature was deterministic and it perfectly distinguished the two categories (we refer to it as “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 a 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., $P_{flurpDflurp}$ and $P_{jaletDjalet}$). 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 presented during training (i.e., $P_{flurpDflurp}$ and $P_{jaletDjalet}$) and four additional types of items. These included: (1) Switch items (i.e., $P_{jaletDflurp}$ and $P_{flurpDjalet}$), which had the deterministic feature of a studied category but most probabilistic features

consistent with the opposite prototype; (2) new-D items (i.e., $P_{\text{flurp}}D_{\text{new}}$ and $P_{\text{jalet}}D_{\text{new}}$), which had probabilistic features of a studied category and a novel feature replacing the deterministic feature; (3) one-new-P items (i.e., $P_{\text{new}}D_{\text{flurp}}$ and $P_{\text{new}}D_{\text{jalet}}$), which had all features of a studied category but a novel feature replacing one probabilistic feature; and (4) all-new-P items (i.e., $P_{\text{all-new}}D_{\text{flurp}}$ and $P_{\text{all-new}}D_{\text{jalet}}$), which had the deterministic features from a studied category 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 category-inclusion rule, in which case they should reject these items as new during the memory test. The one-new-P items were used to assess whether participants could rely in their categorization on the old D feature and remaining P features when a single old P feature was not available. These items were also used to examine whether participants encoded all individual P features, in which case they should reject these items as new during memory test. And finally, the all-new-P items were used to assess whether participants could rely in their categorization 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 reject these items as new. Table 1 presents example of category structure with P and D being combined to create five types of stimuli, and Fig. 1 shows examples of each kind of stimulus.

2.1.3. Design and procedure

The experiment consisted of instructions, training, and testing (see Fig. 2). The procedures were similar for both adults and children, except 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 from children's high recognition accuracy (see below), their response patterns did not stem from confusion or fatigue.

Instructions and training. Before training, participants were presented with a cover story about two categories of creatures from other planets. Information about P and D features was explicitly given to the participants (see Fig. 2A). 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 flurp's (or jalet's) button.* After instructions, participants were given 30 training trials (15 trials per category). On each trial, participants predicted the category label of a stimulus given information about all other features and each trial was accompanied by corrective feedback (see Fig. 2B). The order of the training trials was randomized across participants. Testing was not mentioned until the end of the training phase.

Testing. The testing phase was administered immediately after training and included categorization and recognition tasks. During the testing phase, adults and children were presented with 40 test trials (8 trials per item-type; with equal number coming from each of the two categories) and were asked to determine (1) which category each creature was more likely to belong to and (2) whether each creature was old (i.e., exactly the one presented during the training phase) or new (see Fig. 2C). As we explain in the results section, the ways participants categorize and remember different item types provide critical information about how they represented the categories.

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 par-

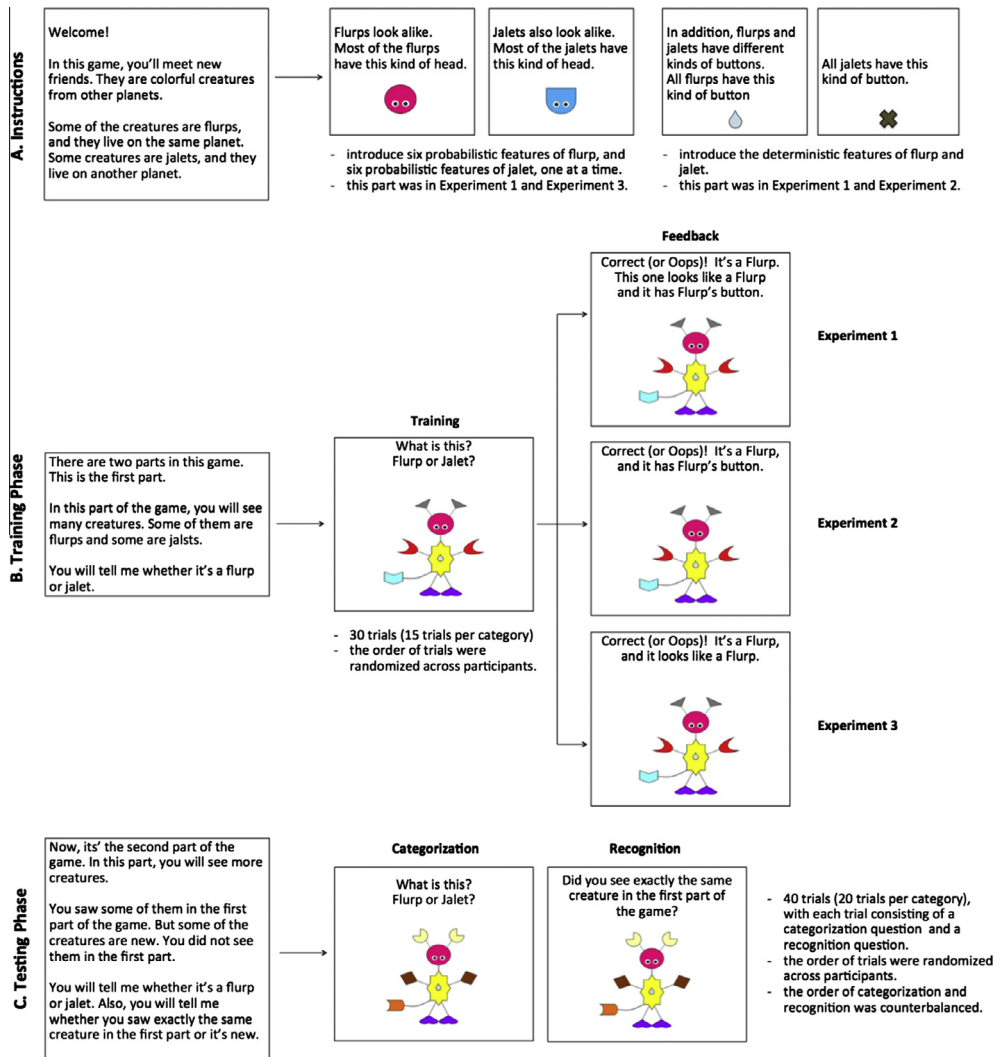


Fig. 2. Schematic representation of the procedure in Experiments 1–3.

Participants. All recognition questions referred to the “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., incorrectly identifying other item types that were not presented during training as “old”).

If there are developmental differences in attention allocation, these differences should transpire in both categorization and recognition performance. In particular, if participants selectively attend to the most diagnostic feature distinguishing two categories, they should rely on the D feature in categorization and remember the D feature better than P features. However, if participants attend diffusely to

multiple features within each category, they should rely on the overall similarity in categorization and exhibit no difference in recognition performance between D and P features. These developmental differences in categorization and memory, if found, would suggest developmental differences in representations of the studied categories.

Based on previous results (e.g., Deng & Sloutsky, 2015a; Hoffman & Rehder, 2010; see also Markman & Ross, 2003, for a review), we expected adults (and perhaps 7-year-olds) to exhibit attention optimization during category learning (with attention shifting to the deterministic features). As a result, they should use the most diagnostic (or rule) feature and remember this rule well. These findings would suggest a rule-based representation of categories in adults. At the same time, given the reviewed above evidence of diffused attention in young children, we expected 4-year-olds to categorize on the basis of multiple features and remember multiple features. These findings would suggest multi-feature-based (or similarity-based) representation in 4-year-olds. Therefore, analyses below focused on categorization and recognition memory performance.

2.2. Results and discussion

2.2.1. Preliminary analyses

Preliminary analyses focused on categorization performance during training and testing.

Training phase. Two adults, one 7-year-old, and one 4-year-old were two standard deviations below the mean of accuracy in the last ten training trials, and data from these participants were excluded from the following analyses. Training data were aggregated into three 10-trial blocks across age groups and they are presented in Table 2.

Overall, children and adults exhibited above-chance training accuracy in the last ten training trials: 85.0% in 4-year-olds, 96.5% in 7-year-olds, and 96.5% in adults, $ps < .001$. A one-way ANOVA on Age Group (4-year-olds vs. 7-year-olds vs. Adults) revealed significant differences between age groups, $F(2, 59) = 12.53$, $MSE = 0.09$, $p < .001$, $\eta^2 = 0.305$, with adults and 7-year-olds being more accurate than 4-year-olds.

Testing phase. Categorization performance of each age group is presented in Fig. 3A and Table 3. Preliminary analyses focused on the ability to correctly categorize the trained High-Match items ($P_{flurp}D_{flurp}$ and $P_{jalet}D_{jalet}$), which was indicative of how well participants learned the categories. As shown in Fig. 3A, all age groups accurately categorized these test items: 83.8% in 4-year-olds, 96.9% in 7-year-olds, and 96.3% in adults, all above chance, one-sample $ts > 9.28$, $ps < .001$, $ds > 2.07$.

The second set of preliminary analyses focused on generalization performance—the ability to rely on familiar (i.e., seen during training) features when some of the features changed as was the case for 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. These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with trial type as a within-subjects factor and age group as a between-subjects factor. There was a significant Trial Type by Age Group interaction, $F(4, 114) = 12.61$, $MSE = 0.39$, $p < .001$, $\eta^2 = 0.307$. We further broke down the interaction by conducting a repeated measures ANOVA on Trial Type for each age group.

For 4-year-olds, there was a significant main effect of trial type, $F(2, 38) = 7.22$, $MSE = 0.23$, $p = .002$, $\eta^2 = 0.275$, with the categorization performance on new-D and one-new-P items being above chance (69% and 77% respectively, $ps < .001$) and at chance on all-new-P items (56%, $p = .324$). Recall that all-new-P items had the trained D features and all new P features, which revealed the inability of 4-year-olds to rely exclusively on D features and to generalize broadly. At the same time, 4-year-olds ably relied on P features in their generalization.

For adults, there was also a significant main effect of trial type, $F(2, 38) = 17.19$, $MSE = 0.46$, $p < .001$, $\eta^2 = 0.475$, with categorization performance on new-D items being lower than that on one-new-P ($p < .001$) and all-new-P items ($p = .016$). However, their pattern of categorization performance on these three trial types was different from 4-year-olds'. Specifically, adults were able to correctly categorize all three types of items (above chance, $ps < .001$). Therefore, in contrast to 4-year-olds who could rely on only P features, adults could rely on either D or P features, with greater reliance on the D features.

Table 2

Training data: Mean (standard deviation) proportion of correct responses aggregated in 10-trial blocks across age groups in Experiments 1–3.

Experiment	Age group	Trials 1–10	Trials 11–20	Trials 21–30
Experiment 1	Adults	0.930 (0.098)	0.910 (0.207)	0.965 (0.081)
	7-year-olds	0.870 (0.108)	0.955 (0.089)	0.965 (0.067)
	4-year-olds	0.605 (0.188)	0.760 (0.160)	0.850 (0.100)
Experiment 2	Adults	0.995 (0.022)	0.980 (0.041)	0.995 (0.022)
	7-year-olds	0.945 (0.083)	0.980 (0.052)	0.990 (0.031)
	4-year-olds	0.780 (0.120)	0.900 (0.149)	0.900 (0.130)
Experiment 3	Adults	0.700 (0.169)	0.730 (0.156)	0.755 (0.161)
	7-year-olds	0.600 (0.149)	0.590 (0.207)	0.630 (0.142)
	4-year-olds	0.580 (0.154)	0.655 (0.193)	0.720 (0.226)

For 7-year-olds, there was a significant main effect of trial type, $F(2,38) = 32.53$, $MSE = 1.13$, $p < .001$, $\eta^2 = 0.631$, with the categorization performance on one-new-P and all-new-P items being above chance (93% and 93% respectively, $ps < .001$) and at chance on new-D items (51%, $p = .843$). Therefore, 7-year-olds were able to generalize broadly by relying on D features but failed to rely exclusively on P features.

Overall, the preliminary analyses indicated that adults exhibited flexible generalization: they could rely on either the old D features (when presented with all-new-P items) or the old P features (when presented with new-D items). Seven-year-olds exhibited rule-based generalization: they could generalize on the basis of the old D features, whereas they failed to generalize on the basis of the old P features. In contrast, 4-year-olds exhibited multi-feature-based (or similarity-based) generalization: they generalized successfully when multiple studied P features were present, but failed to generalize when only studied D features were present.

2.2.2. Primary analyses

The primary analyses focused on categorization of switch items and recognition memory performance.

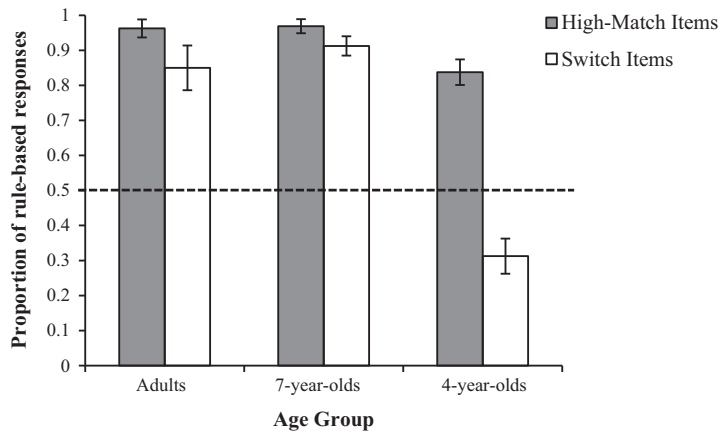
Categorization. To examine the pattern of categorization, we analyzed performance on Switch items (i.e., $P_{\text{flurp}}D_{\text{jalet}}$ and $P_{\text{jalet}}D_{\text{flurp}}$) across the age groups (see Fig. 3A). The Switch items had D features from one category and most of the P features from another category; therefore, these items allowed an unambiguous determination of whether D or P features controlled categorization. Based on preliminary analyses, it was expected that adults and 7-year-olds would categorize the Switch items on the basis of D features, whereas 4-year-olds would categorize on the basis of P features.

Data for Switch items in Fig. 3A were analyzed with a one-way ANOVA, with Age Group (4-year-olds vs. 7-year-olds vs. Adults) as the factor. As predicted, there were significant differences between age groups, $F(2,59) = 44.50$, $MSE = 2.18$, $p < .001$, $\eta^2 = 0.610$, with adults and 7-year-olds exhibiting higher proportion of rule-based responses than 4-year-olds. More importantly, both adults and 7-year-olds relied on the D feature to categorize the Switch items, exhibiting primarily rule-based generalization, 85% and 91%, respectively, both above chance, one-sample $ts > 5.48$, $ps < .001$, $ds > 1.23$. In contrast, 4-year-olds exhibited similarity-based generalization, relying on the P features to categorize the Switch items. As a result, their proportion of rule-based responses was at 31%, below chance, one-sample $t(19) = 3.74$, $p = .001$, $d = 0.84$.

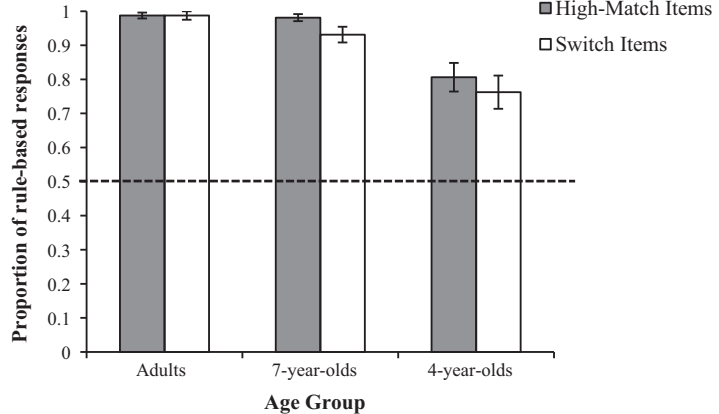
Results of categorization performance indicated that adults and 7-year-olds tended to rely on the deterministic feature (i.e., the category-inclusion rule) to categorize items, whereas 4-year-olds relied on multiple probabilistic features (i.e., the overall similarity).

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 other item types it is a false alarm). As shown in the table, participants readily distinguished the studied High-Match items from all-new-P items: the differences between hits and false alarms were 0.781 for adults, 0.675 for 7-year-olds, and 0.519 for 4-year-olds, all greater than the chance level of 0, $ps < .001$.

A. Experiment 1



B. Experiment 2



C. Experiment 3

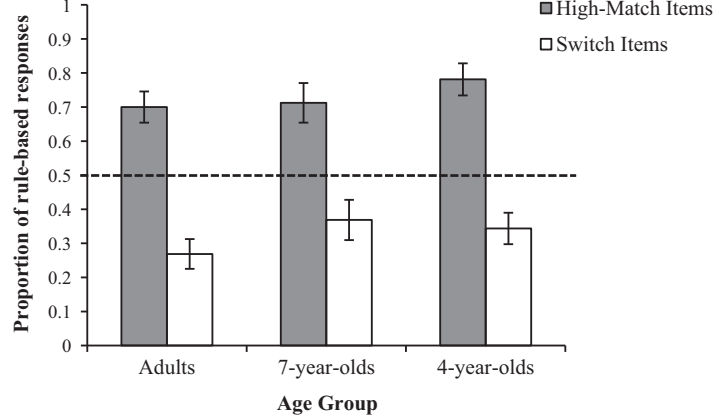


Fig. 3. Categorization performance: proportion of rule-based responses by trial type and age group in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C). The chance level is 0.5. Error bars represent standard errors of mean.

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 1–3.

Experiment	Age group	New-D	One-new-P	All-new-P
Experiment 1	Adults	0.681 (0.170)	0.975 (0.051)	0.894 (0.219)
	7-year-olds	0.513 (0.278)	0.925 (0.159)	0.925 (0.118)
	4-year-olds	0.694 (0.149)	0.769 (0.216)	0.556 (0.248)
Experiment 2	Adults	0.338 (0.322)	0.988 (0.038)	0.994 (0.028)
	7-year-olds	0.600 (0.328)	0.975 (0.051)	1 (0)
	4-year-olds	0.644 (0.328)	0.806 (0.238)	0.794 (0.216)
Experiment 3	Adults	0.744 (0.174)	0.800 (0.200)	0.438 (0.143)
	7-year-olds	0.613 (0.246)	0.675 (0.258)	0.525 (0.197)
	4-year-olds	0.631 (0.192)	0.706 (0.251)	0.519 (0.164)

Table 4

Memory at test: Mean (standard deviation) proportions of “yes” responses (i.e., “old” responses) on different item types in Experiments 1–3.

Experiment	Age group	High-Match	New-D	One-new-P	All-new-P
Experiment 1	Adults	0.844 (0.181)	0 (0)	0.444 (0.277)	0.063 (0.224)
	7-year-olds	0.744 (0.288)	0.213 (0.314)	0.350 (0.316)	0.069 (0.174)
	4-year-olds	0.913 (0.152)	0.363 (0.303)	0.394 (0.200)	0.263 (0.289)
Experiment 2	Adults	0.794 (0.136)	0.125 (0.211)	0.450 (0.264)	0.069 (0.174)
	7-year-olds	0.800 (0.325)	0.225 (0.288)	0.481 (0.345)	0.263 (0.413)
	4-year-olds	0.769 (0.254)	0.306 (0.355)	0.319 (0.321)	0.213 (0.327)
Experiment 3	Adults	0.813 (0.217)	0.606 (0.302)	0.200 (0.204)	0 (0)
	7-year-olds	0.863 (0.198)	0.544 (0.385)	0.244 (0.273)	0.025 (0.051)
	4-year-olds	0.794 (0.244)	0.350 (0.316)	0.356 (0.312)	0.200 (0.270)

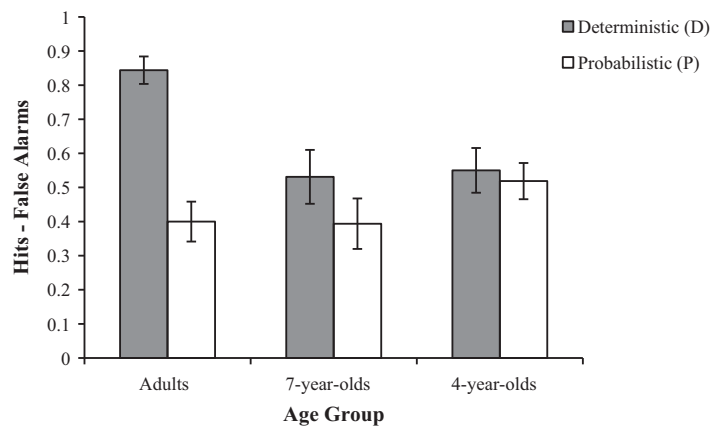
Memory accuracy for the category-inclusion rule (i.e., D feature) and for the overall appearance (i.e., P features) was compared for each age group. Memory accuracy for the rule was obtained by subtracting false alarms on new-D items from hits on High-Match items, and memory accuracy for appearance features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

Data in Fig. 4A were submitted to a 2 (Feature Type: D vs. P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with feature type as a within-subjects factor and age group as a between-subjects factor. Results revealed a significant interaction, $F(2,57) = 10.85$, $MSE = 0.46$, $p < .001$, $\eta^2 = 0.276$. Specifically, adults exhibited better memory accuracy for the D feature than for a P feature (0.844 vs. 0.400), paired-samples $t(19) = 7.18$, $p < .001$, $d = 1.97$. Similar to adults, 7-year-olds also exhibited better memory accuracy for the D feature than for a P feature (0.531 vs. 0.394), paired-samples $t(19) = 2.24$, $p = .037$, $d = 0.40$. In contrast, 4-year-olds exhibited equivalent memory accuracy for the D feature and for a P feature (0.550 vs. 0.519), paired-samples $t(19) = 0.44$, $p = .666$. Furthermore, 4-year-olds exhibited numerically better memory for P features than older participants, although these differences did not reach significance.

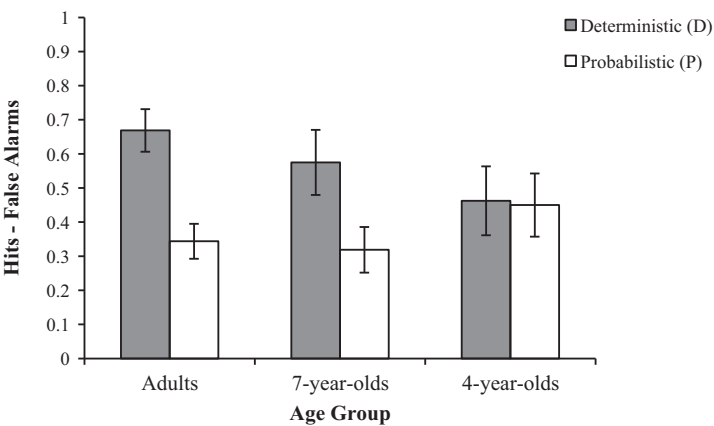
Therefore, recognition memory accuracy corroborates findings stemming from categorization performance: whereas adults and 7-year-olds exhibited better memory for the deterministic feature and consistently relied on it in categorization, 4-year-olds remembered will all features and relied on multiple probabilistic features in categorization.

One may argue, however, that 4-year-olds' reliance on probabilistic features instead of the deterministic feature might stem from the difficulty of focusing on this particular deterministic feature (i.e., the body button that was relatively small in size). According to this argument, 4-year-olds might fail to notice this least salient feature in the training phase, thus failing to rely on it in the testing phase. Although we do not think that this was the case because 4-year-olds exhibited high memory accuracy

A. Experiment 1



B. Experiment 2



C. Experiment 3

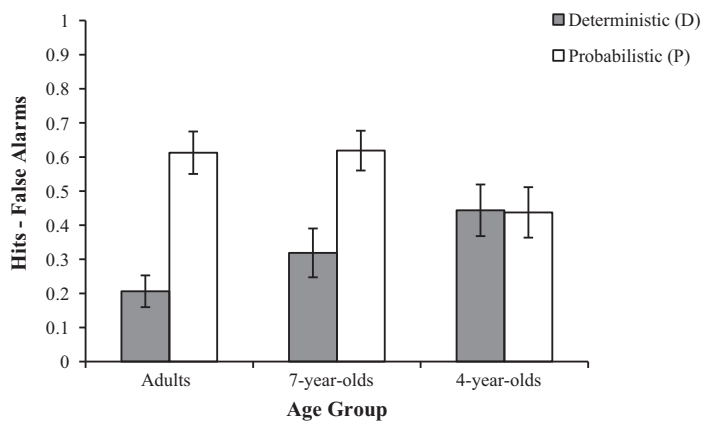


Fig. 4. Recognition performance: memory accuracy by feature type and age group in Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C). The chance level is 0. Error bars represent standard errors of mean.

for the deterministic feature, we deemed it necessary to add counterbalancing conditions using other features as deterministic ones. The head, which is presumably highly salient, may not be appropriate, and there is evidence showing that 4-year-old children are able to rely on a single highly salient feature in categorization (Deng & Sloutsky, 2012, 2013). Therefore, we chose the hands and the feet, which seem to be, in terms of salience, somewhere between the body button and the head. Data from twenty-four 4-year-olds (twelve participants in each condition) indicated that upon learning the categories, similar to the 4-year-old participants in Experiment 1, 4-year-olds in both hands and feet conditions relied on the multiple features in categorization. The proportions of rule-based responses on Switch items were $M = 0.365$ in the hands condition, and $M = 0.385$ in the feet condition, both below chance, $ps < .001$.

Overall, Experiment 1 replicated and extended Deng and Sloutsky's findings (2015a; see also Rabi et al., 2015, for similar findings) and revealed important developmental differences in categorization. Specifically, adults and older children were more likely to rely on the deterministic features in their categorization and they better remembered these features. Therefore, it is likely that older participants formed rule-based representations of the categories. In contrast, younger children were more likely to rely on multiple probabilistic features in their categorization and remember all the features. These findings suggest that, in contrast to older participants, 4-year-olds formed similarity-based representations.

As argued above, these developmental differences in categorization and memory are likely to be driven by different patterns of attention allocated in the course of category learning. Whereas adults and older children exhibit selective attention, younger children attend diffusely. Experiment 1 presented suggestive evidence supporting this possibility and the goal of Experiments 2 and 3 was to test it directly. There are some conditions under which young children may selectively attend to a single feature, although this selectivity is likely to be either exogenous (i.e., driven by characteristics of the stimuli, such as stimulus salience) or based on the selection history (Awh, Belopolsky, & Theeuwes, 2012). For example, Deng and Sloutsky (2012, 2013) demonstrated that 4-year-olds categorized on the basis of a single salient feature (i.e., pattern of head motion). As we discuss in Section 6, these variants of selectivity differ sharply from endogenous (top-down, goal-driven) selective attention.

In Experiments 2 and 3 we attempted to cue participants' attention, without changing the salience of the stimuli (Deng & Sloutsky, 2015a presented evidence that this could be done with 4-year-olds). To achieve this goal, we directed participants' attention to the D feature (by only mentioning this feature in instructions and in feedback on each training trial) in Experiment 2 and to the P features (by only mentioning overall appearance in instructions and in feedback on each training trial) in Experiment 3.

3. Experiment 2: cueing deterministic features

The goal of Experiment 2 was to cue participants' attention to the D feature and examine changes in (a) categorization performance (compared to the pattern established in Experiment 1) and (b) underlying category representations. If participants attend selectively to the cued features, this manipulation should affect categorization decisions and memory for cued and non-cued features. In particular, memory for D features should increase, whereas memory for P features should decrease.

3.1. Method

3.1.1. Participants

Participants were adults (6 women), 7-year-old children ($M_{\text{age}} = 78.6$ months, range 70.8–89.7 months; 13 girls), and 4-year-old children ($M_{\text{age}} = 54.5$ months, range 49.8–59.7 months; 11 girls), with 20 participants per age group. Adult participants 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 and were tested by an experimenter in a quiet room in their pre-

school. Data from one additional 4-year-old were excluded from analyses because of extremely poor performance in training.

3.1.2. Materials, design, and procedure

The materials, design, and procedure were similar to those in Experiment 1, with one critical difference. In contrast to Experiment 1, where attention was attracted to both D and P features, in Experiment 2, we directed participants' attention to the D feature (see Fig. 2). 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 using the following script: *correct (or oops), it is a flurp (or jalet); it has flurp's (or jalet's) button*. The testing phase (both categorization and recognition trials) was identical to Experiment 1 and it was not mentioned during the training phase.

3.2. Results and discussion

3.2.1. Preliminary analyses

Preliminary analyses were similar to those in Experiment 1.

Training phase. One 4-year-old was two standard deviations below the mean of accuracy in the last ten training trials, and data from this participant were excluded from the following analyses. Training data were aggregated into three 10-trial blocks across age groups and they are presented in Table 2. Overall, children and adults exhibited high training accuracy in the last ten training trials: 90.0% in 4-year-olds, 99.0% in 7-year-olds, and 99.5% in adults, all above chance, $ps < .001$. A one-way ANOVA on Age Group (4-year-olds vs. 7-year-olds vs. Adults) revealed significant differences between age groups, $F(2, 59) = 9.377$, $MSE = 0.06$, $p < .001$, $\eta^2 = 0.248$, with adults and 7-year-olds being more accurate than 4-year-olds.

Testing phase. Categorization performance of each age group is presented in Fig. 3B and Table 3. Preliminary analyses focused on the ability to correctly categorize the trained High-Match items ($P_{\text{flurp-D}_{\text{flurp}}}$ and $P_{\text{jalet-D}_{\text{jalet}}}$), which was indicative of how well participants learned the categories. As shown in Fig. 3B, all age groups accurately categorized these test items: 80.6% in 4-year-olds, 98.1% in 7-year-olds, and 98.8% in adults, all above chance, one-sample $ts > 7.29$, $ps < .001$, $ds > 1.63$.

The second set of preliminary analyses focused on generalization performance—the ability to rely on familiar (i.e., seen during training) features when some of the features changed as was the case for 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.

These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with trial type as a within-subjects factor and age group as a between-subjects factor. There was a significant trial type by age group interaction, $F(4, 114) = 10.07$, $MSE = 0.41$, $p < .001$, $\eta^2 = 0.261$. We further broke down the interaction by conducting a repeated measures ANOVA on Trial Type for each age group.

For 4-year-olds, there was a marginal effect of trial type, $F(2, 38) = 3.23$, $MSE = 0.16$, $p = .051$, $\eta^2 = 0.145$, with the categorization performance on one-new-P and all-new-P items (81% and 80% respectively) being above chance ($ps < .001$) and marginally above chance on new-D items ($p = .065$). This pattern was different from that of 4-year-old participants in Experiment 1: in Experiment 1, 4-year-olds could rely in their generalization on only P features, whereas in the current experiment they could rely on either P or D features.

For adults, there was a significant main effect of trial type, $F(2, 38) = 81.09$, $MSE = 2.84$, $p < .001$, $\eta^2 = 0.810$. Specifically, adults were able to correctly categorize one-new-P and all-new-P items (98% and 99% respectively, both above chance, $ps < .001$). Their performance on new-D items was below chance, $p = .036$. Therefore, adults exhibited reliance on D features, but not on P features. This pattern was different from that of adults' in Experiment 1: in Experiment 1, adults could rely in their generalization on either P or D features, whereas in the current experiment they could rely on only D features.

For 7-year-olds, there was also a significant main effect of trial type, $F(2, 38) = 26.7$, $MSE = 1.00$, $p < .001$, $\eta^2 = 0.584$, with the categorization performance on one-new-P and all-new-P items (98%

and 100% respectively) being above chance ($ps < .001$) and at chance on new-D items ($p = .189$). Therefore, similar to adults, 7-year-olds exhibited reliance on D features, but not on P features. This pattern was similar to that of 7-year-olds' in Experiment 1.

Several important findings stem from the preliminary results. When attention was attracted to the D feature (i.e., the rule) in Experiment 2, adults and 7-year-olds exhibited rule-based generalization. In contrast, 4-year-olds could rely in their generalization on either D or P features, thus exhibiting evidence of both rule-based and similarity-based generalization.

3.2.2. Primary analyses

Similar to Experiment 1, the primary analyses focused on categorization of switch items and recognition memory performance.

Categorization. To examine the pattern of categorization, we analyzed performance on the Switch items (i.e., $P_{\text{flurp}}D_{\text{jale}}$ and $P_{\text{jale}}D_{\text{flurp}}$) across the age groups (see Fig. 3B). Recall that in this experiment, we attempted to direct participants' attention to the D feature by only mentioning this feature in instructions and on each training trial. If the manipulation is successful as suggested by the preliminary analyses, 4-year-olds should become similar to adults and 7-year-olds and rely on the D feature in their categorization.

Results presented in Fig. 3B indicated that 4-year-olds' pattern of categorization was similar to that of adults and 7-year-olds: their categorization was rule-based, which was in sharp contrast to similarity-based categorization exhibited by 4-year-olds in Experiment 1. A one-way ANOVA on Age Group (4-year-olds vs. 7-year-olds vs. Adults) confirmed these findings. There were significant differences between age groups, $F(2, 59) = 44.50$, $MSE = 2.18$, $p < .001$, $\eta^2 = 0.610$, with adults and 7-year-olds exhibiting higher proportion of rule-based responses than 4-year-olds. However, all age groups relied on the D feature to categorize the switch items, with the proportion of rule-based responses being 76.3% in 4-year-olds, 93.1% in 7-year-olds, and 98.8% in adults, above chance, all $ps < .001$, all $ds > 1.20$.

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 all-new-P items. The difference between hits and false alarms was 0.556 for 4-year-olds, 0.538 for 7-year-olds, and 0.725 for adults, all greater than the chance level of 0, $ps < .001$.

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 (see Fig. 4B). As shown in the figure, memory accuracy for D and P features was above chance level of 0 for all age groups, $ps < .001$.

Data in Fig. 4B were submitted to a 2 (Feature Type: D vs. P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with feature type as a within-subjects factor and age group as a between-subjects factor. There was a significant interaction, $F(2, 57) = 4.46$, $MSE = 0.27$, $p = .016$, $\eta^2 = 0.135$. Specifically, adults exhibited better memory accuracy for the D feature than for a P feature (0.669 vs. 0.344), paired-samples $t(19) = 3.92$, $p = .001$, $d = 1.27$. Similar to adults, 7-year-olds also exhibited better memory accuracy for the D feature than for a P feature (0.575 vs. .319), paired-samples $t(19) = 3.58$, $p = .002$, $d = 0.70$. In contrast, 4-year-olds exhibited equivalent memory accuracy for a P feature and for the D feature (0.463 vs. 0.450), paired-samples $t(19) = 0.16$, $p = .875$. Similar to Experiment 1, 4-year-olds exhibit numerically better memory for P features than older participants, although these differences did not reach statistical significance.

Overall, results of Experiment 2 substantially expand results of Experiment 1. When participants' attention was focused on the D feature in Experiment 2, 4-year-olds' categorization performance (in contrast to Experiment 1) was rule-based and was similar to that of older children and adults. At the same time, their memory pattern was similar to that of 4-year-olds in Experiment 1 and different from those of older children and adults in this experiment. Whereas older children and adults exhibited better memory for D features (that controlled their categorization) than for P features, 4-year-olds exhibited equivalently high memory accuracy for both D and P features.

Patterns of categorization, generalization and memory performance suggest that adults and 7-year-olds formed representations based primarily on deterministic features. At the same time, 4-year-olds were able to generalize on the basis of both D and P features and remembered both features equally well. Therefore, it seems likely that these participants formed similarity-based representations. How-

ever, testing this possibility will require additional analyses and we will return to this issue and provide these analyses when we discuss results of all three experiments.

In addition, memory performance of 4-year-olds in Experiment 2 suggest that these participants continued to distribute their attention across multiple features even though their categorization decision was based on the rule feature. One interesting consequence of such distributed attention was that 4-year-old remembered non-cued features numerically better than 7-year-olds and adults. In Experiment 3, we attempted to change adults and 7-year-olds' pattern of categorization by attracting participants' attention to the P features.

4. Experiment 3: cueing probabilistic features

The goal of Experiment 3 was to cue participants' attention to the P features and examine changes in categorization (compared to the pattern established in Experiment 1) and memory for features. If participants attend selectively to the cued features, this manipulation should affect categorization decisions and memory for cued and non-cued features. In particular, memory for P features should increase, whereas memory for D features should decrease.

4.1. Method

4.1.1. Participants

Participants were adults (12 women), 7-year-old children ($M_{\text{age}} = 80.2$ months, range 72.3–90.1 months; 10 girls), and 4-year-old children ($M_{\text{age}} = 55.7$ months, range 49.3–60.2 months; 10 girls), with 20 participants per age group. Adult participants 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 and were tested by an experimenter in a quiet room in their preschool. Data from one additional 7-year-old were excluded from analyses because of extremely poor performance in training.

4.1.2. Materials, design, and procedure

The materials, design, and procedure were similar to those in Experiment 2, with one critical difference. In contrast to Experiment 2 where we directed participants' attention to the D feature, in Experiment 3 we directed their attention to the P features (see Fig. 2). In particular, participants were told that most of the flurps' (or jalets') had a certain body part (at this point, 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: *correct (or oops), this is a flurp (or jalet); it looks like a flurp (or jalet)*. The testing phase (both categorization and recognition trials) was identical to Experiment 2 and it was not mentioned during the training phase.

4.2. Results and discussion

4.2.1. Preliminary analyses

Preliminary analyses were identical to those in Experiments 1 and 2.

Training phase. One 7-year-old was two standard deviations below the mean of accuracy in the last ten training trials, and data from this participant were excluded from the following analyses. Training data were aggregated into three 10-trial blocks across age groups and they are presented in Table 2. Overall, children and adults exhibited above-chance training accuracy in the last ten training trials: 72.0% in 4-year-olds, 66.0% in 7-year-olds, and 75.5% in adults, $ps < .001$. A one-way ANOVA on Age Group (4-year-olds vs. 7-year-olds vs. Adults) revealed no differences between age groups, $F(2,59) = 1.47$, $MSE = 0.05$, $p = .238$, $\eta^2 = 0.049$.

Testing phase. Categorization performance of each age group is presented in Fig. 3C and Table 3. Preliminary analyses focused on how well participants learned the categories. As shown in Fig. 3C, all age groups accurately categorized the trained High-Match items ($P_{\text{flurpDflurp}}$ and $P_{\text{jaletDjalet}}$): 78.1% in 4-

year-olds, 71.3% in 7-year-olds, and 70.0% in adults, all above chance, all one-sample t s > 3.66 , p s $< .002$, all d s > 0.82 .

The second set of preliminary analyses focused on generalization performance—the ability to rely on familiar (i.e., seen during training) features when some of the features changed as was the case for 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.

These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA. There was a significant trial type by age group interaction, $F(4, 114) = 3.53$, $MSE = 0.91$, $p = .009$, $\eta^2 = 0.110$. We further broke down the interaction by conducting a repeated ANOVA on Trial Type for each age group.

For 4-year-olds, the main effect of trial type was significant, $F(2, 38) = 5.66$, $MSE = 0.18$, $p = .007$, $\eta^2 = 0.230$. Similar to the 4-year-old participants in Experiment 1, 4-year-olds in Experiment 3 correctly categorized one-new-P and new-D items (above chance, p s $< .007$) but their performance on all-new-P items was at chance, $p = .614$. Therefore, in contrast to Experiment 2, 4-year-olds successfully relied on the overall similarity (when presented with one-new-P and new-D items), but failed to rely on the D features (when presented with all-new-P items).

For adults, similar to Experiments 1 and 2, there was a significant main effect of trial type, $F(2, 38) = 39.75$, $MSE = 0.76$, $p < .001$, $\eta^2 = 0.677$. However, their pattern of categorization was different from that in Experiments 1 and 2. Specifically, adults were able to correctly categorize one-new-P and new-D items (above chance, p s $< .001$), whereas their performance on all-new-P items was at chance, $p = .066$. Recall that all-new-P items had the trained D features and all new P features, which revealed the inability of adults to rely exclusively on D features after their attention was directed to P features. This pattern was similar to that of the 4-year-old participants in this experiment as well as in Experiment 1.

For 7-year-olds, there was also a significant main effect of trial type, $F(2, 38) = 4.26$, $MSE = 0.11$, $p = .021$, $\eta^2 = 0.183$. Their pattern of categorization was different from that in Experiments 1 and 2. Similar to adults and 4-year-olds, 7-year-olds were able to correctly categorize one-new-P items (above chance, $p = .007$) and new-D items (marginally above chance, $p = .055$), whereas their performance on all-new-P items was at chance, $p = .577$.

4.2.2. Primary analyses

Primary analyses were identical to those in Experiments 1–2: similar to Experiments 1–2, these analyses focused on categorization of switch items and recognition memory performance.

Categorization. To examine the pattern of categorization, we analyzed performance on the Switch items (i.e., $P_{\text{flurp}D_{\text{jalet}}}$ and $P_{\text{jalet}D_{\text{flurp}}}$) between the age groups (see Fig. 3C). Recall that in this experiment, we attempted to direct participants' attention to the P features by mentioning overall appearance on each training trial. If the manipulation is successful, as suggested by the preliminary analyses, adults and 7-year-olds should become similar to 4-year-olds and rely on the overall similarity to categorize the Switch items.

Results presented in Fig. 3C indicated that in this experiment, adults and 7-year-olds exhibited the same pattern of similarity-based categorization as 4-year-olds. A one-way ANOVA on Age Group (4-year-olds vs. 7-year-olds vs. Adults) for the Switch items confirmed these findings. There were no significant differences between age groups, $F(2, 59) = 1.08$, $MSE = 0.05$, $p = .347$, $\eta^2 = 0.036$. A further one-sample t test indicated that all age groups relied on the P features to categorize the Switch items, with the proportion of rule-based responses being 34.4% in 4-year-olds, 36.9% in 7-year-olds, and 26.9% in adults, all below chance, all p s $< .039$, all d s > 0.50 .

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 all-new-P items. The difference between hits and false alarms was 0.594 for 4-year-olds, 0.838 for 7-year-olds, and 0.813 for adults, all greater than the chance level of 0, p s $< .001$.

Similar to Experiments 1 and 2, 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 Fig. 4C. As shown in the figure, memory accuracy for D and P features was above chance level of 0 for all age groups, p s $< .001$.

Data in Fig. 4C were submitted to a 2 (Feature Type: D vs. P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with feature type as a within-subjects factor and age group as a between-subjects factor. There was a significant interaction, $F(2,57) = 9.74$, $MSE = 0.46$, $p < .001$, $\eta^2 = 0.255$. Specifically, adults exhibited better memory accuracy for a P feature than for the D feature (0.613 vs. 0.206), paired-samples $t(19) = 5.84$, $p < .001$, $d = 1.66$. Similar to adults, 7-year-olds also exhibited better memory accuracy for a P feature than for the D feature (0.619 vs. 0.319), paired-samples $t(19) = 3.62$, $p = .002$, $d = 1.03$. In contrast, 4-year-olds exhibited equivalent memory accuracy for a P feature and for the D feature (0.438 vs. 0.444), paired-samples $t(19) = 0.13$, $p = .900$. Furthermore, 4-year-olds' memory for P features (that were non-cued ones in this experiment) was significantly better than that in adults (Bonferroni adjusted $p = .040$, $d = 0.85$), and numerically better than that of 7-year-olds.

Overall, when participants' attention was directed to the P features in Experiment 3, adults' and 7-year-olds' categorization performance became similar to 4-year-olds who tended to rely on the overall appearance, and their memory accuracy changed accordingly. However, 4-year-olds' memory pattern remained the same across three experiments, exhibiting equivalently high memory accuracy for both D and P features. These results suggest that, in contrast to adults and 7-year-olds who tended to attend selectively to the cued features, 4-year-olds tended to distribute their attention across multiple features, regardless of attentional manipulation.

Results of categorization, generalization, and memory suggested that when P features were cued in Experiment 3, adults, 7-year-olds, and 4-year-olds formed similarity-based representations. However, whereas memory of older participants was better for the cued P features, 4-year-olds remembered both P and D features equally well. These results indicated that, in contrast to older participants, 4-year-olds tended to distribute attention across both cued and non-cued features.

To further examine the effect of attentional manipulation on memory, we compared participants' memory accuracy for cued features (i.e., the D features in Experiment 2 and the P features in Experiment 3) and non-cued features (i.e., the P features in Experiment 2 and the D features in Experiment 3) across experimental conditions (i.e., directing attention to the D features in Experiment 2 and to the P features in Experiment 3). These data were submitted to a 2 (Feature Type: Cued vs. Non-cued) by 2 (Experimental Condition: cueing D vs. cueing P) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA, with feature type as a within-subjects factor and experimental condition and age group as between-subjects factors. There was neither main effect of experimental condition ($p = .562$) nor interaction involving experimental condition ($ps > .555$). Therefore, these data were collapsed across two experimental conditions and were submitted to a 2 (Feature Type: Cued vs. Non-cued) by 3 (Age Group: 4-year-olds vs. 7-year-olds vs. Adults) mixed ANOVA. There was a significant feature type by age group interaction, $F(2,117) = 13.55$, $MSE = 0.72$, $p < .001$, $\eta^2 = 0.188$. We further broke down the interaction by performing two separate one-way ANOVAs on Age Group (4-year-olds vs. 7-year-olds vs. Adults) for the memory accuracy of the cued features and for the memory accuracy of the non-cued features. For the cued features, there were significant differences between age groups, $F(2,119) = 3.41$, $MSE = 0.40$, $p = .037$, $\eta^2 = 0.055$, with adults exhibiting better memory for the cued features than 4-year-olds, $p = .042$, $d = 0.563$. For the non-cued features, there were also significant differences between age groups, $F(2,119) = 3.37$, $MSE = 0.32$, $p = .038$, $\eta^2 = 0.054$, but in the opposite direction: 4-year-olds exhibited better memory for non-cued features than adults, $p = .042$, $d = 0.557$.

These findings point to an important developmental difference in the pattern of attention: whereas adults and older children attended selectively to what they deemed to be category-relevant, younger children attended diffusely. Importantly, more efficient selective attention in adults and older children was accompanied by worse memory of the non-cued (i.e., to-be-ignored features) than of the to-be-attended features, whereas less efficient diffused attention in younger children was accompanied by equally good memory of both to-be-attended and to-be-ignored features.

Taken together, categorization and memory performance suggest that across the conditions, adults and older children encoded condition-specific information and formed a condition-specific representation (with rule-based representation being a default). In contrast, 4-year-olds tended to encode multi-feature information and form similarity-based representations (although their representations in Experiment 2 requires additional analyses and we provide these analyses below).

5. Interrelationships between categorization and memory across Experiments 1–3 and developmental changes in the mechanism of categorization

In the three reported experiments we examined participants' categorization judgments and memory for features, with the goal of better understanding developmental changes in the mechanism of category learning and in category representation. In Experiment 1, we attracted participants' attention to both D and P features and this experiment served as a baseline. As evidenced by performance on switch trials (see Fig. 3A), adults and 7-year-olds relied predominately on D features in their categorization, whereas 4-year-olds relied predominately on P features. In terms of memory accuracy, adults and 7-year-olds exhibited better memory for D features than for P features, whereas 4-year-olds exhibited comparable memory accuracy. Therefore, in older participants, memory performance was consistent with their categorization performance – they had better memory for features that controlled their categorization. However, this was not the case in 4-year-olds: these participants remembered all the features, regardless of whether these features controlled their categorization.

In Experiment 2, participants' attention was attracted to D features. Whereas categorization and memory performance of older participants were similar to those in Experiment 1, there were some changes in 4-year-olds. Specifically, in contrast to Experiment 1, in this experiment 4-year-olds' categorization was rule-based (see Fig. 3B), whereas similar to Experiment 1, they exhibited comparably high memory for D and P features. Again, categorization and memory were consistent in older participants, but not in 4-year-olds.

Finally, in Experiment 3, participants' attention was attracted to P features. In contrast to Experiments 1–2, 7-year-olds and adults relied on P features in their categorization (see Fig. 3C) and they remembered P features better than D features. Four-year-olds also relied on P features in their categorization. However, in contrast to older participants, they again remembered P and D features equally well.

Therefore, across experiments, in 7-year-olds and adults, memory for features was consistent with categorization performance: if they performed rule-based categorization, they exhibited better memory for D features, whereas if they performed similarity-based categorization, they exhibited better memory for P features. In contrast, in 4-year-olds, categorization and memory performance were independent: regardless of their pattern of categorization, they remembered P and D features equally well. In other words, categorization and memory performance were interdependent (or coupled) in 7-year-olds and adults, but independent (or decoupled) in 4-year-olds.

To further examine these relationships between categorization and memory, we calculated correlations between the proportion of rule-based categorization and difference in memory accuracy for D features compared to P features for each age group across the three experiments. If reliance on a feature and memory for this feature are interdependent, then greater reliance on D features in categorization should be accompanied by better memory for D features than for P features, thus resulting in a positive correlation. The results (see Fig. 5) indicate that there was high positive correlation in adults ($r = .66, p < .001$) and in 7-year-olds ($r = .60, p < .001$), whereas there was no correlation in 4-year-olds ($r = .07, p = .620$).

These results are important as they further indicate that categorization and memory are coupled in 7-year-olds and adults, but decoupled in 4-year-olds. In addition, 4-year-olds tend to remember all the features and form similarity-based representations (as evidenced by their successful generalization with new-D items across the experiments), even when they performed rule-based categorization in Experiment 2. The latter finding is of critical importance as it suggests that there is a distinction between category representation and decision weights that can be put on different features when categorizing the items. These findings also suggest potential developmental differences in the mechanisms of categorization and we examine this possibility in the next section.

5.1. Attentional weights, categorization, and recognition: implications for mechanisms of categorization

The results of the experiments reported here point to decoupling between categorization and memory in younger children, but not in older children and adults. This decoupling is important,

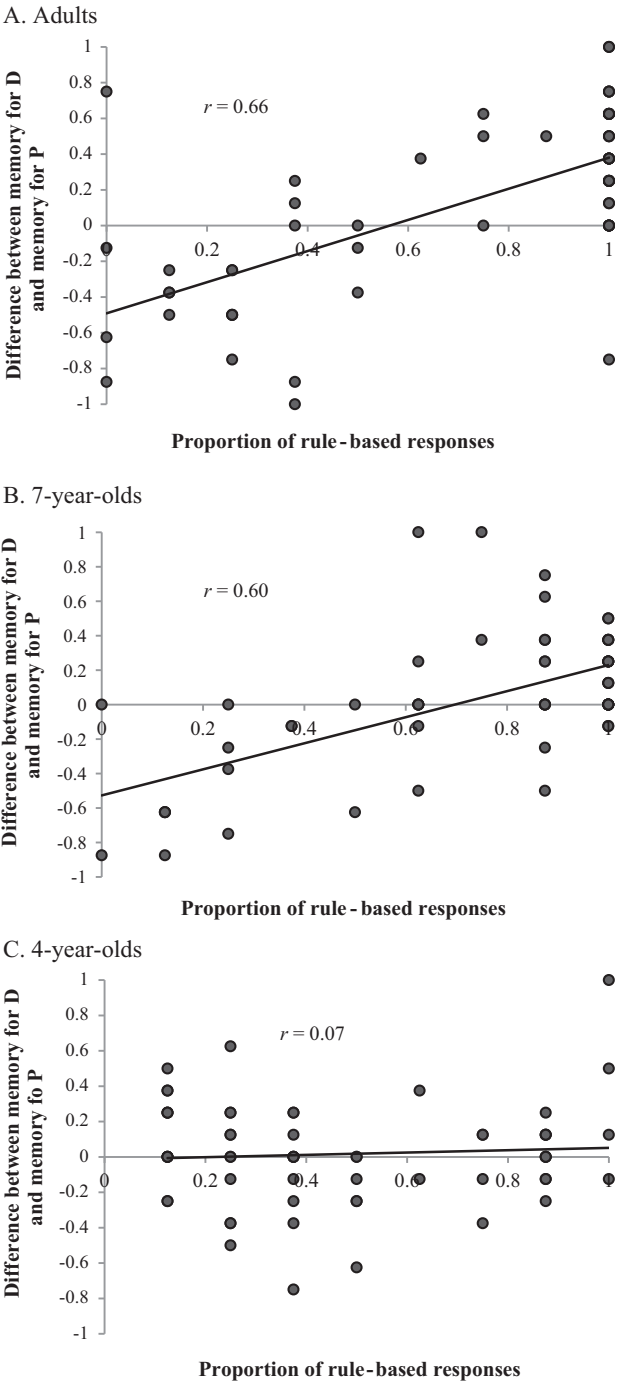


Fig. 5. Correlations between the proportion of rule-based categorization and difference in memory accuracy for D features compared to P features in adults (A), 7-year-olds (B), and 4-year-olds (C) across the experiments.

and it may reveal developmental differences in the mechanism of categorization. We consider two possible mechanisms that are consistent with the reported data. The first mechanism (Mechanism 1) assumes attention optimization coupled with representational change in the course of category learning: Category learning is accompanied by shifting attention to category-relevant features. As a result, representation of categories is based primarily on the category-relevant features, and categorization and generalization are performed on the basis of these representations. An alternative possibility (Mechanism 2) assumes changes only in decision, but not representation. In this case, people may distribute attention, encode all features, form similarity-based representation, but put different decision weights on different features when performing rule-based categorization and generalization. The reported results suggest that Mechanism 1 transpires in adults and older children, whereas Mechanism 2 transpires in younger children.

To further examine potential developmental differences in the mechanism of categorization, we simulated the data using the model that embodies Mechanism 1 and compared it to the observed data. If Mechanism 1 is the case, then simulated data should be consistent with the observed data. In contrast, if Mechanism 2 is the case, the simulated data should be inconsistent with the observed data.

One such model is the generalized context model (GCM, [Nosofsky, 1986, 1988](#)) – a member of the family of categorization models – which has been successful in accounting for patterns of categorization and recognition data stemming from the same stimulus sets. The model assumes that categorization and recognition are based on the common representational substrate, with categorization and recognition responses being a positive function of similarity between the stimulus and studied items (although categorization and recognition decision rules may differ).

According to the GCM, in the case of two mutually exclusive categories A and B, the probability that Stimulus S_i is classified in Category C_A , $P(R_A|S_i)$, is given by the following equation:

$$P(R_A|S_i) = \frac{b_A \sum_{j \in C_A} sim_{ij}}{b_A \sum_{j \in C_A} sim_{ij} + (1 - b_A) \sum_{k \in C_B} sim_{ik}} \quad (1)$$

where b_A ($0 \leq b_A \leq 1$) is Category A response bias and sim_{ij} and sim_{ik} are similarities between a given exemplar i and exemplars belonging to categories A and B, respectively.

The similarity between items is an exponential decay function of psychological distance d depicted in Eq. (2):

$$sim_{ij} = \exp[-cd_{ij}] \quad (2)$$

where c ($0 \leq c < \infty$) is a scaling parameter reflecting the overall discriminability in the psychological space. This parameter can be affected by a variety of task and individual variables. For example, mixing stimuli with noise, presenting them for a short period of time, or requiring to simultaneously perform more than one task may make the stimuli less discriminable and reduce the value of c . Similarly, low perceptual or memory acuity may also result in smaller values of c (cf. [Nosofsky & Zaki, 1998](#)).

Psychological distance d is calculated according to the standard Euclidean formula, as shown in Eq. (3):

$$d_{ij} = \left[\sum w_m (x_{im} - x_{jm})^2 \right]^{1/2} \quad (3)$$

where x_{im} is the psychological value of exemplar i on dimension m , and w_m ($0 \leq w_m \leq 1$, $\sum w_m = 1$) is the attentional weight given to dimension m . These weights are free parameters and are interpreted as reflecting the attention allocated to each dimension during categorization ([Nosofsky, 1984, 1986](#)). Attentional weights (w_m) are of critical importance because they are highly malleable and may change as a result of categorization (GCM, [Nosofsky, 1986](#)) and category learning (ALCOVE, [Kruschke, 1992](#)). In contrast to c that can be interpreted as the overall attention to stimuli or even to the task, w reflects the amount of attention allocated to a particular dimension. Increased attention to a particular dimension makes this dimension more discriminable (and perhaps more memorable), whereas reduced attention make dimension less discriminable (and perhaps less memorable).

According to GCM ([Nosofsky, 1986](#), see also ALCOVE, [Kruschke, 1992](#)), category learning is often accompanied by attention optimization – increased attention to predictive dimensions and decreased

attention to non-predictive dimensions. Attention optimization is an important mechanism of category learning and has potential consequences for memory for dimensions and their values (i.e., features). Decreased attention to a given dimension should result in lower memory for values on this dimension and in memory confusions between old and new values. In contrast, increased attention to a given dimension should result in higher memory.

According to the GCM, the recognition judgment for an item is based on the summed similarity of the item to all the exemplars of all the categories, fam , as shown in Eq. (4):

$$fam_i = \sum_{i=1}^k sim_{ik} \quad (4)$$

The response rule for recognition judgments is shown in Eq. (5). The probability of a positive (“Old”) response to Stimulus S_i , $P(Old|S_i)$, is a monotonic function of an item’s similarity to all studied items, regardless of their category. The response rule has a single scaling parameter β , and it is interpreted as a criterion for making “old” recognition judgments. A lower value of β corresponds to a greater overall tendency to respond positively (i.e., “yes” bias).

$$P(Old|S_i) = \frac{fam_i}{fam_i + \beta} \quad (5)$$

The goal of this section is to examine the ability of the family of models embodying Mechanism 1 and assuming attention optimization to predict recognition performance from categorization performance across the age groups. Experiments 2 and 3 offer an opportunity for such an examination because all age groups exhibited the same pattern of categorization (i.e., D feature categorization in Experiment 2 and P feature categorization in Experiment 3).

5.1.1. Predicting memory from categorization based on D features (Experiment 2)

To predict memory for different features from categorization performance, we first needed to estimate attentional weights w for different features. To obtain such estimates for the D feature, w_D , we first estimated the probability of a rule-based response on switch items (recall that these had D features from one category and the majority of P features from another category) using Eqs. (1)–(3) under every possible value of w_D between 0 and 1 in 0.01 increments, assuming the response biases to two categories are equal⁴ and setting the value of c at 3.50.⁵ Then, we selected the attentional weight that best predicted the probability of rule-based responses (for adults, $w_D = 0.99$, for 7-year-olds, $w_D = 0.81$, and for 4-year-olds, $w_D = 0.50$).⁶ These attentional weights derived from categorization suggested selective attention to D features in all age groups, with perhaps greater selectivity in adults and older children than younger children.

To simulate recognition performance under the selected attentional weight for the D feature, we first estimated the probability of a “yes” response to the new-D item and the one-new-P item respectively using Eqs. (2)–(5) under every possible value of β between 20 and 0.1 in 0.1 increments. Recall that a lower value of β corresponds to a greater overall tendency to respond “old”. The estimated probabilities of “old” responses are effectively familiarity judgments for the new-D items and the one-new-P items. Given that these items were new, familiarity judgments predict the rate of false alarms. To convert these values into memory accuracy data (i.e., correct rejections), we inverted the estimated probabilities. We also simulated recognition performance under the assumption of attention distributed between D and P features, such that attentional weights for all P features combined equaling to the D feature. The resulting curve corresponds to attentional weights that would generate chance categorization performance on switch items. This was important because this curve presents a bound-

⁴ In fact, there was no evidence in the data suggesting unequal response bias.

⁵ In addition to this value, we also obtained estimates by setting different values of C parameter. Although the estimates changed under different C values, the overall patterns remained the same across the age groups.

⁶ In addition to group estimates, we also obtained individual estimates and used them in the subsequent analyses. However, the overall fit remained essentially the same.

ary: recognition data on or above this curve cannot be predicted by the observed categorization performance.

Simulated and observed memory values are presented in Fig. 6A. Simulated values are presented in black (each point corresponds to a different value of β), whereas observed data are shown in color (each point represents individual participant). Age groups are presented in different colors in separate plots: 4-year-olds are shown in blue, 7-year-olds are shown in orange, and adults are shown in green. The curve in yellow presents the above described boundary.

There are two regions of space on each plot. First, individual data points that are around the black curve but below the yellow curve indicate qualitatively good prediction of memory data from categorization data with no free parameters. Note that these predictions could be even more accurate with addition of a free parameter.⁷

The second region of space consists of individual values falling on or above the boundary (i.e., on or above the yellow curve). These values indicate no relationship between categorization and recognition: this recognition performance (i.e., equally good memory for D and P features or even better memory for P features than for D features) cannot be predicted from the observed categorization performance (i.e., participants basing their responses on D features). This is because recognition performance suggests no advantage for the D-features (or even an advantage for P-features), whereas categorization performance on switch trials suggests an advantage for D-features.

As shown in Fig. 6A, the vast majority of adults and 7-year-olds (i.e., all but two adults and all but four 7-year-olds) fall below the yellow curve. Therefore, these participants exhibit evidence of coupling between categorization and memory: their recognition could be predicted from their categorization. The fact that the model embodying Mechanism 1 can predict well recognition performance from categorization performance of adults and 7-year-olds strongly support the possibility of Mechanism 1 of categorization in these participants.

At the same time, the majority of 4-year-olds fall at or above the yellow curve: in these cases, recognition could not be predicted from categorization performance. Therefore, 4-year-olds' data are inconsistent with the model embodying Mechanism 1 and categorization in these participants is more likely to be based on Mechanism 2 – distributed attention, the formation of similarity-based representation, and applying different decision weights to different features when performing rule-based categorization.

In addition, decoupling of categorization and memory in 4-year-olds presents an interesting challenge to any model that assumes attention optimization and representational change in the course of category learning (including GCM and ALCOVE): children's categorization data suggests attention optimization, but recognition data do not. At present, we do not see any obvious solution to this problem within the models assuming attention optimization in the course of learning. Turning attentional learning off and assuming comparable attentional weights for all features does not solve the problem either: this strategy would successfully simulate recognition performance of 4-year-olds, but would fail to simulate their categorization performance.

These findings point to potentially different mechanisms of category learning in younger and older participants. Whereas older participants optimize attention in the course of category learning (which in turn affects their representation of dimensions, and possibly exemplars and categories), younger participants do not. Instead they seem to allocate attention to all dimensions (as evidenced by their memory performance) but put different decision weights on different dimensions when categorizing items.

Although this conclusion is important, one potential counterargument is that perhaps young children's data are too noisy to be fit by any model. This possibility is however undermined by the analyses of Experiment 3. Recall that in Experiment 3, participants' attention was attracted to the P features and their categorization was based on P features. Participants also performed similarity-based categorization and generalization and remembered mostly probabilistic features (older children

⁷ For example, the numerator and each addend in the denominator of Eq. (1) could be raised to the power of γ (Ashby & Maddox, 1993; Rouder & Ratcliff, 2004), with $\gamma > 0$. Larger values of γ may transform even small advantage in the attentional weight of the D feature into a consistent rule-based response.

A. Predicting memory from categorization based on D features (Experiment 2)

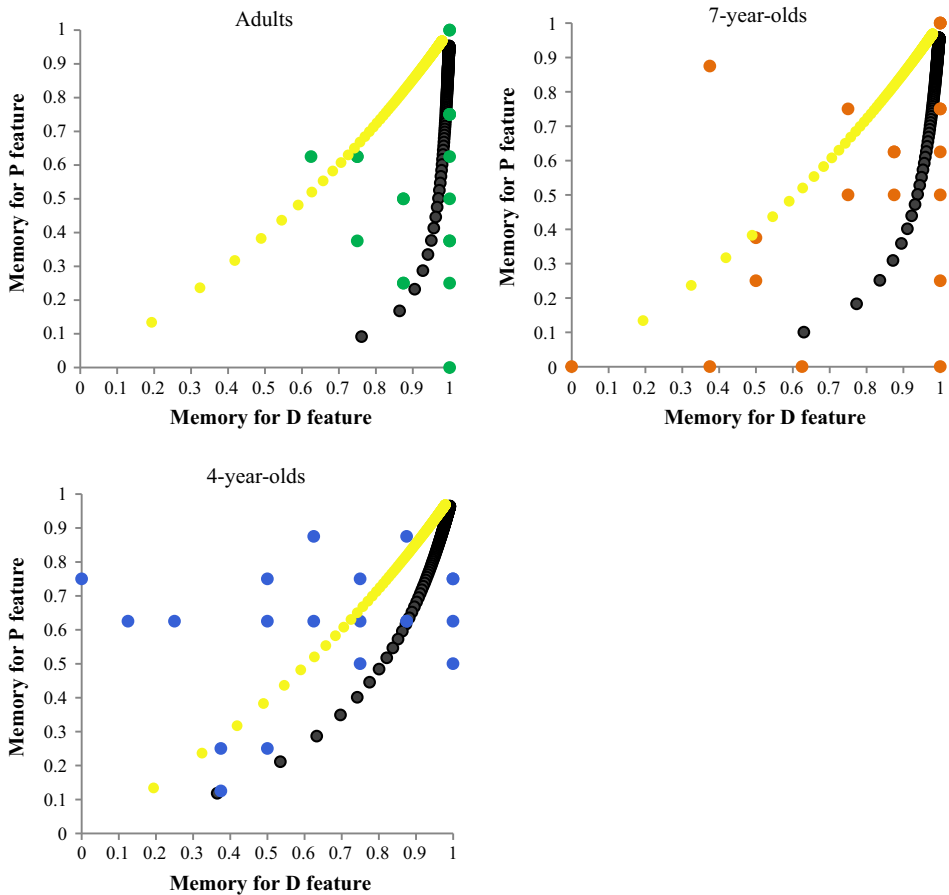


Fig. 6. GCM simulated memory values and observed memory values in Experiment 2 (A) and in Experiment 3 (B). Simulated values are shown in black, with each point corresponding to a different value of β ; observed memory accuracies are shown in green for adults, in orange for 7-year-olds, and in blue for 4-year-olds, with each point corresponding to an individual participant. The yellow curve presents a boundary that would generate chance categorization performance on switch items. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and adults) or all features (younger children). These results could be readily captured by assuming distributed attention. We therefore used this assumption in the analyses presented below.

5.1.2. Predicting memory from categorization based on P features (Experiment 3)

To simulate recognition performance, we used the same strategy as was used to simulate results of Experiment 2. To implement an assumption of distributed attention, we estimated attentional weight of each feature by dividing the total attentional weight ($\sum w = 1$) by the number of features ($n = 7$), $w_p = 0.14$. This attentional weight under the assumption of distributed attention was identical across all age groups.

Similar to the estimations under the assumption of selective attention to the D feature, we examined the pattern of recognition by estimating the probability of an “Old” response to the new-D item and the one-new-P item respectively using Eqs. (2)–(5) under the distributed attentional weight ($w_p = 0.14$) assumption and under every possible value of β between 20 and 0.1 in 0.1 decrements. Then, we inverted the estimated probability into memory accuracy. Also, similar to simulating results

B. Predicting memory from categorization based on P features (Experiment 3)

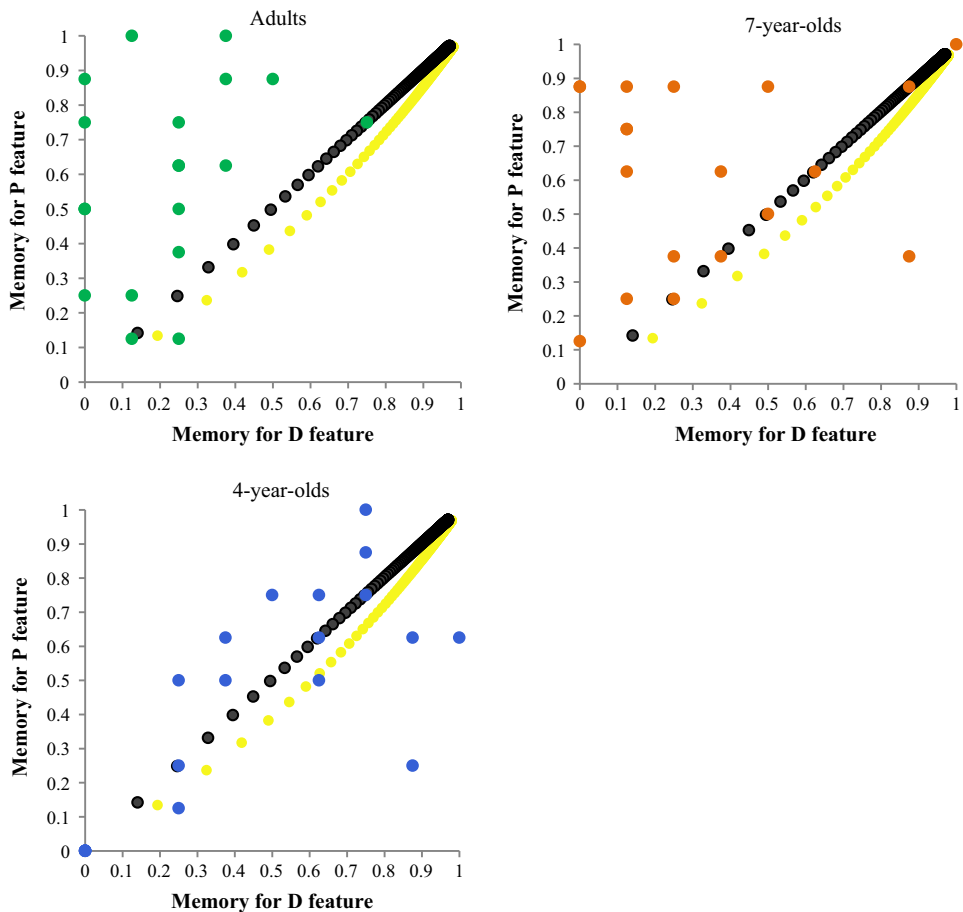


Fig. 6 (continued)

of Experiment 2, we simulated the boundary (presented as the yellow curve). This curve corresponds to attentional weights that would generate chance categorization performance on switch items. Observed recognition data that are on or below this boundary cannot be predicted from the observed P-based categorization performance.

Simulated and observed memory values are presented in Fig. 6B. Simulated values are presented in black (each point corresponds to a different value of β), whereas observed data are shown in color (each point represents individual participant). Age groups are presented in different colors in separate plots: 4-year-olds are shown in blue, 7-year-olds are shown in orange, and adults are shown in green. The curve in yellow presents the above-described boundary. There are several issues worth discussing.

First, there are few data points that are at or below the yellow curve (i.e., one adult, two 7-year-olds, and five 4-year-olds): recall that these data points cannot be captured by the model. At the same time, for the majority of participants across the age groups recognition data could be predicted from their categorization performance. This finding is important because it undermines the possibility that the failure to predict 4-year-olds' memory from their categorization performance in Experiment 2 stemmed from 4-year-olds' data being noisier than that of older participants.

And second, data points that are close to the black curve indicate comparable attentional weight of D and P features, whereas data points that are well above the black curve indicate higher attentional

weights of P features than of the D feature. The former pattern indicates that attention was distributed among all the features, whereas the latter pattern points to selective attention albeit to multiple features – participants focus on P features, while ignoring the D features. As shown in Fig. 6B, the majority of adults and many 7-year-olds exhibit such a pattern, whereas relatively few 4-year-olds do. These results were confirmed by the quantitative analyses: the model of attention distributed across all the features poorly predicts memory of adults ($RMSE = 0.467$) and 7-year-olds ($RMSE = 0.460$), while predicting well memory of 4-year-olds ($RMSE = 0.215$). Therefore, memory data suggest that 4-year-olds were likely to distribute attention across all features, whereas older participants were likely to focus mostly on the cued (probabilistic) features.

Overall, several important findings stem from simulating of memory performance using categorization performance. Most importantly, the model that assumes attention optimization in the course of category learning can capture categorization performance of all participants but fails to capture memory data of 4-year-olds in Experiment 2. These findings suggest developmental differences in the mechanism of categorization, indicating that early in development rule-based categorization is achieved by putting greater decision (rather than attention) weights on the cued features. Therefore, early in development, rule-based categorization can be performed on the basis of similarity-based representations. This presents interesting challenges to the models that presume stable mechanism of categorization across development.

6. General discussion

The reported study presents several novel findings pointing to important developmental differences in attention allocation during category learning, mechanisms of categorization, and category representation. In the three reported experiments, we examined categorization and memory in 4-year-olds, 7-year-olds, and adults. Across all experiments, categories had a single deterministic (D) feature and multiple probabilistic (P) features. In Experiment 1, both D and P features were cued during training. In Experiment 2, only D features were cued, and in Experiment 3 only P features were cued. In terms of categorization responses, both children and adults were responsive to attentional manipulations introduced in Experiments 2 and 3. However, important differences transpired with respect to recognition memory. Adults and 7-year-olds tended to remember better features that they used in categorization, whereas 4-year-olds tended to remember all features equally well. These results coupled with computational simulations point to an important developmental difference in (1) the pattern attention and (2) the mechanism of categorization. First, whereas adults and 7-year-olds attend selectively to what they deem to be category-relevant, 4-year-olds attend diffusely. And second, whereas selective attention and subsequent attention optimization lead to representational change in older participants (i.e., some features were more likely to be included in category representation than others as a result of learning), no representational change occurred in 4-year-olds.

Importantly, selective attention in adults and 7-year-olds (which presumably sub-serves more efficient learning) was accompanied by worse memory of the to-be-ignored features than of the to-be-attended features, whereas diffused attention in 4-year-olds was accompanied by equally good memory of both to-be-attended and to-be-ignored features. In addition, in Experiment 2 attracting attention of 4-year-olds to only D features affected their categorization performance, but not their memory, suggesting an important distinction between representation and decision factors in early categorization. These findings have important implications for (a) the role of attention in categorization and its development, (b) flexibility of categorization and its development, and (c) theories of categorization. In what follows, we discuss each of these points in greater detail.

6.1. Selective attention, diffused attention, and categorization

Selective attention is an integral component for most of the models of categorization. For example, in both exemplar models (Hampton, 1995; Medin & Schaffer, 1978; Nosofsky, 1986) and prototype models (Nosofsky, 1992; Smith & Minda, 1998), selective attention is formalized in terms of the influence, or weight, of each stimulus dimensions on categorization. In rule-based models, it is implicitly

assumed that the operation of selective attention to the stimulus dimension(s) referred to by the current hypothesis (i.e., rule) being tested (Ashby et al., 1998; Smith, Patalano, & Jonides, 1998, see also Rehder & Hoffman, 2005). Most of the models agree that categorization decisions are sub-served by underlying representations of stimulus dimensions. A given representation is formed in the course of category learning, with learning of a category resulting in increased attention to the dimension (s) that distinguish the studied categories and decreased attention to those that do not (e.g., ALCOVE, Kruschke, 1992; GCM, Nosofsky, 1986). For example, if one learns two categories, such as squirrels versus chipmunks, the learner's attention may shift to stripes (which is a diagnostic feature) and away from the tail (which is not diagnostic). At the same time, if one learns two other categories, such as squirrels versus hamsters, the tail is a diagnostic feature, whereas stripes are not. Thus learning of different ways of categorizing items should result in different attentional weights of stimulus dimensions and subsequently in different representations of these dimensions. Therefore, as a result of this attentional selectivity, stripes become more salient in the context of the former categorization task, whereas the tail becomes more salient in the context of the latter. Many theories of categorization predict that, as a result of allocating attention selectively in the first category learning task, participants may have difficulty shifting attention to a previously ignored dimension and exhibit learned inattention in the second category learning task, and this prediction has been confirmed empirically (e.g., Hoffman & Rehder, 2010).

However, young children may exhibit different pattern of attention from adults, thus resulting in different category representations. For example, Best et al. (2013) found that, in contrast to adults who attended selectively to relevant dimensions in category learning and exhibited evidence of learned inattention, 6- to 8-month-old infants attended to both relevant and irrelevant dimensions, and they did not exhibit learned inattention. Furthermore, in a recently published study, Deng and Sloutsky (2015b) demonstrated that diffused attention is an important property in early category learning: more successful learning in 8- to 12-month-old infants was accompanied by more distributed attention among different features of presented objects. There is also evidence showing that 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 et al., 1994, for reviews). Current findings present additional evidence implicating the role of selective attention in the development of categorization: whereas adults and 7-year-olds exhibited selective attention to relevant information, 4-year-olds exhibited diffused attention to both relevant and irrelevant information across the attentional cueing conditions. An important challenge for future research is to understand when and why this important development takes place.

Note that in the attention literature (e.g., Egeth & Yantis, 1997; Pashler et al., 2001; Posner & Petersen, 1990) selective attention has been conceptualized as either involuntary, bottom-up, and stimulus-driven (when it is captured automatically by a highly salient stimulus) or as voluntary, top-down, and goal-driven (when the goal is to find a red object in a pile of things of different colors). Top-down selective attention is considered to be intentional, deliberate, person-controlled, and goal-directed; whereas, bottom-up selective attention is considered to be autonomous and stimulus-controlled. When selectivity is top-down and goal-driven, attention is shifted to particular stimulus dimensions that are important for task goals or when a person is instructed to do so. For example, in classical shadowing experiments where different auditory information is presented to different ears (see Pashler, 1999, for a review), people can attend selectively to a predetermined auditory channel.

When selectivity is bottom-up and stimulus-controlled, attention is shifted to particularly salient or novel stimuli, regardless of how relevant these stimuli are for a task at hand. For example, novelty preferences in infants, “pop out” effects, or shifting attention automatically to highly salient stimuli. Given that selective attention to particular stimulus dimensions in categorization is a result of learning and not of inherent differences in salience among stimulus dimensions, it is likely to be a variant of top-down selectivity. Although, some argue for a more complex taxonomy of selective attention (e.g., Awh et al., 2012), it is hardly controversial that bottom-up selective attention exhibits an early onset (e.g., Posner & Petersen, 1990). At the same time, the fact that 4-year-old children deploy bottom-up selective attention in categorization, categorizing on the basis of a single highly salient feature (Deng & Sloutsky, 2012, 2013), raises an important question: what are the consequences of bottom-up selective attention for memory? Are these effects similar to those of top-down selectivity, with children

remembering primarily the highly salient feature? Or do these effects differ from those of top-down selectivity, with children remembering multiple features? This is an important question for understanding the links among attention, categorization and memory, and it has to be addressed in future research.

6.2. Flexibility of categorization

Even early in development, people's categorization is remarkably flexible – when presented with a given input, people, depending on a situation, may rely on different aspects of this input (Bulloch & Opfer, 2009; Gelman & Markman, 1986; Heit & Rubinstein, 1994; Jones, Smith, & Landau, 1991; Macario, 1991; Ross & Murphy, 1999; Sloutsky & Fisher, 2008). This flexibility has been observed in a variety of categorization, category learning, and property induction tasks. For example, in one study (Jones et al., 1991), 2–3 year-olds were presented with a target item, which was named (i.e., “this is a dax”), and asked to find another dax among test items. When the target and test objects were presented with eyes, children relied on both shape and texture, whereas when the objects were presented without eyes children tended to rely on shape alone. In a categorization study, 3–4 year-olds were more likely to group novel items differing in color and shape on the basis of color, if the items were introduced as food, but on the basis of shape, if the items were introduced as toys (Macario, 1991).

In another categorization study (Nguyen & Murphy, 2003), 4-year-olds were presented with triads of food items, consisting of a target and two test items. In all triads, one test item was unrelated to the other two, but in some triads one test item matched the target taxonomically (i.e., both were the same kinds of foods, such as meats), whereas in other triads one test item matched the target thematically (i.e., both could be eaten during the same time of the day, such as breakfast). For example, a taxonomic triad could consist of bacon (target), chicken (taxonomic choice), and lemon (unrelated choice), whereas a thematic triad could consist of bacon (target), pancakes (thematic choice), and carrot (unrelated choice). Researchers found that 4-year-olds could cross-classify items by selecting either a taxonomic or thematic test item.

In a property induction study (Gelman & Markman, 1986), 4- to 5-year-olds were presented with a target and two test items, such that one test item shared the label with the target and the other looked similar to the target. Participants were then told that the target had a particular property and asked which of the test items had the same property. Participants were more likely to rely on linguistic labels when inferring a biological property than when inferring a physical property (see also Heit & Rubinstein, 1994, for similar findings in adults).

More recently, Bulloch and Opfer (2009) presented 4- to 5-year-olds with another variant of property induction task. Participants were shown triads of items, with one item being the target and two others being test stimuli. The target stimulus and the test stimuli each consisted of a set of three items. Two larger members of each set were identical and looked like bugs, whereas the smaller member of the set was different and looked like larvae. One of the test items had the same bugs as the target (but different larvae), whereas the other had the same larvae (but different bugs). Researchers introduced a property of the target larvae and asked which of the test larvae had the same property. It was found that participants relied on the similar looking bugs when the items were introduced as “parents and offspring,” whereas they relied on the similar looking larvae, when items were introduced as “predators and prey”.

Although these findings demonstrate that, depending on the situation, children can categorize the same items in different ways or rely on different dimensions when making predictions, none of the reviewed studies examined underlying representations. As a result, little is known as to whether children's representations of the categories also differ across the situations. In particular, it is possible that such flexible behaviors are based on different representations of the same items, at least when participants learn new categories. Alternatively, it is possible that representations are the same, with participants making different decisions on the basis of the same representations. In the former case participants represent primarily the predictive dimension and ignore the non-predictive, whereas in the latter case, they represent both dimensions, but decide later which one to rely on.

Current results provide evidence indicating that younger children can be as flexible as older children and adults in categorization responses and they are able to use different dimensions in different

Table A1

Category structure of High-Match items, used in both training and testing in Experiments 1–3.

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
$P_{flurp}D_{flurp}$							
1	1	1	1	1	0	0	1
2	1	1	1	0	1	0	1
3	1	1	0	1	1	0	1
4	1	0	1	1	1	0	1
5	0	1	1	1	1	0	1
6	1	1	1	0	0	1	1
7	1	1	0	1	0	1	1
8	1	0	1	1	0	1	1
9	0	1	1	1	0	1	1
10	1	1	0	0	1	1	1
11	1	0	1	0	1	1	1
12	0	1	1	0	1	1	1
13	1	0	0	1	1	1	1
14	0	1	0	1	1	1	1
15	0	0	1	1	1	1	1
$P_{jalet}D_{jalet}$							
1	0	0	0	0	1	1	0
2	0	0	0	1	0	1	0
3	0	0	1	0	0	1	0
4	0	1	0	0	0	1	0
5	1	0	0	0	0	1	0
6	0	0	0	1	1	0	0
7	0	0	1	0	1	0	0
8	0	1	0	0	1	0	0
9	1	0	0	0	1	0	0
10	0	0	1	1	0	0	0
11	0	1	0	1	0	0	0
12	1	0	0	1	0	0	0
13	0	1	1	0	0	0	0
14	1	0	1	0	0	0	0
15	1	1	0	0	0	0	0

situations: categorization of younger children, older children and adults was responsive to attentional manipulations. However, across the experiments younger children tended to represent all dimensions and form similarity-based representations, whereas adults and older children tended to represent dimensions that are relevant for a given situation. At the very minimum, these results point to a distinction between category representation and categorization decision and indicate that decision flexibility develops before representational flexibility. We discuss some of these issues in the next section.

6.3. Distinction between representation and decision and theories of categorization

While different patterns of attention allocation may result in different category representations, in most of the behavioral studies on categorization, participants' category representations are inferred from participants' category judgments, or decisions. And in most of the cases, adults' categorization decisions change according to different category representations formed during learning (Chin-Parker & Ross, 2002; Hoffman & Rehder, 2010; Sakamoto & Love, 2010; Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 2000). For example, adult participants exhibited different patterns of categorization performance after they were trained by classification, where they predicted category label of a given item, compared to inference, where they predicted feature of a given item. Specifically, their categorization was based on the most diagnostic dimension distinguishing between categories after classification learning but on the within-category featural relation after inference learning (Hoffman & Rehder, 2010; Yamauchi & Markman, 2000). These findings suggest that, for

Table A2
Category structure of Switch items, used in testing in Experiments 1–3.

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
<i>P_{jallet}D_{flurp}</i>							
1	0	0	0	0	1	1	1
2	0	0	0	1	0	1	1
3	0	0	1	0	0	1	1
4	0	1	0	0	0	1	1
5	1	0	0	0	0	1	1
6	0	0	0	1	1	0	1
7	0	0	1	0	1	0	1
8	0	1	0	0	1	0	1
9	1	0	0	0	1	0	1
10	0	0	1	1	0	0	1
11	0	1	0	1	0	0	1
12	1	0	0	1	0	0	1
13	0	1	1	0	0	0	1
14	1	0	1	0	0	0	1
15	1	1	0	0	0	0	1
<i>P_{flurp}D_{jallet}</i>							
1	1	1	1	1	0	0	0
2	1	1	1	0	1	0	0
3	1	1	0	1	1	0	0
4	1	0	1	1	1	0	0
5	0	1	1	1	1	0	0
6	1	1	1	0	0	1	0
7	1	1	0	1	0	1	0
8	1	0	1	1	0	1	0
9	0	1	1	1	0	1	0
10	1	1	0	0	1	1	0
11	1	0	1	0	1	1	0
12	0	1	1	0	1	1	0
13	1	0	0	1	1	1	0
14	0	1	0	1	1	1	0
15	0	0	1	1	1	1	0

adults, different learning regimes may result in different representations of a category, which gives rise to different categorization decisions.

However, the empirical evidence of the consistency between representation and decision does not necessarily mean that these two components completely overlap. In category learning, participants can represent all the features equivalently, but put different decision weights on some features over others. For example, one could form a representation of squirrel category that consists of body size, fur color, tail length, stripe pattern, and so on, but only use the dimension of tail when classifying squirrels and hamsters and only use the dimension of stripes when classifying squirrels and chipmunks. Alternatively, they could represent only some features, but not the others (see Kloos & Sloutsky, 2008, for a discussion of these issues).

Current findings provide novel evidence supporting these ideas. Across experiments, in adults and 7-year-olds, memory for features was consistent with categorization performance: if they performed rule-based categorization, they exhibited better memory for the D feature, whereas if they performed similarity-based categorization, they exhibited better memory for the P features. In contrast, in 4-year-olds, categorization and memory performance were independent: regardless of their pattern of categorization, they remembered P and D features equally well. Correlation analyses on the interrelationships between categorization and memory further support the idea of decoupling between categorization and memory in 4-year-olds. Positive correlations were found in adults and 7-year-olds between the proportion of rule-based categorization and difference in memory accuracy for the D feature compared to the P features, whereas no correlation was found in 4-year-olds.

Table A3

Category structure of new-D items, used in testing in Experiments 1–3.

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
$P_{flurp}D_{new}$							
1	1	1	1	1	0	0	N
2	1	1	1	0	1	0	N
3	1	1	0	1	1	0	N
4	1	0	1	1	1	0	N
5	0	1	1	1	1	0	N
6	1	1	1	0	0	1	N
7	1	1	0	1	0	1	N
8	1	0	1	1	0	1	N
9	0	1	1	1	0	1	N
10	1	1	0	0	1	1	N
11	1	0	1	0	1	1	N
12	0	1	1	0	1	1	N
13	1	0	0	1	1	1	N
14	0	1	0	1	1	1	N
15	0	0	1	1	1	1	N
$P_{jalet}D_{new}$							
1	0	0	0	0	1	1	N
2	0	0	0	1	0	1	N
3	0	0	1	0	0	1	N
4	0	1	0	0	0	1	N
5	1	0	0	0	0	1	N
6	0	0	0	1	1	0	N
7	0	0	1	0	1	0	N
8	0	1	0	0	1	0	N
9	1	0	0	0	1	0	N
10	0	0	1	1	0	0	N
11	0	1	0	1	0	0	N
12	1	0	0	1	0	0	N
13	0	1	1	0	0	0	N
14	1	0	1	0	0	0	N
15	1	1	0	0	0	0	N

Results of modeling also support this decoupling. When the model assumes selective attention to the D feature, it well captures the recognition performance in adults and 7-year-olds, predicting better memory for the D feature than the P features; but it fails to capture 4-year-olds' performance, who exhibited comparably high memory accuracy for both D and P features.

Taken together, these findings point to developmental differences in the mechanism of rule-based category learning. Whereas older children and adults tend to attend selectively, form rule-based representations, and perform categorization on the basis of these representations (Mechanism 1), young children attend diffusely, form similarity-based presentations, but put greater decision weights on D features when making categorization judgments (Mechanism 2).

These differences in the mechanism of category learning may have important implications for early learning in and outside of academic settings. If young children attend to and process information that is part of the to-be-learned concept and that is extraneous to it, the latter information may become a part of their representations following learning. Therefore, distributed attention may result in an exceedingly rich representation of a to-be-learned concept, thus impeding its generalization and transfer to novel situations (see Kaminski & Sloutsky, 2013; Son, Smith, & Goldstone, 2011, for examples pertaining to a mathematical concept and to properties of a set, respectively).

7. Conclusions

Current research examined the role of attention in categorization across development and possible developmental differences in the mechanisms of categorization. The results present novel evidence

Table A4
Category structure of one-new-P items, used in testing in Experiments 1–3.

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
<i>P_{new}D_{flurp}</i>							
1-1	1	1	1	1	0	N	1
1-2	1	1	1	0	1	N	1
1-3	1	1	0	1	1	N	1
1-4	1	0	1	1	1	N	1
1-5	0	1	1	1	1	N	1
2-1	1	1	1	1	N	0	1
2-2	1	1	1	0	N	1	1
2-3	1	1	0	1	N	1	1
2-4	1	0	1	1	N	1	1
2-5	0	1	1	1	N	1	1
3-1	1	1	1	N	1	0	1
3-2	1	1	1	N	0	1	1
3-3	1	1	0	N	1	1	1
3-4	1	0	1	N	1	1	1
3-5	0	1	1	N	1	1	1
4-1	1	1	N	1	1	0	1
4-2	1	1	N	1	0	1	1
4-3	1	1	N	0	1	1	1
4-4	1	0	N	1	1	1	1
4-5	0	1	N	1	1	1	1
5-1	1	N	1	1	1	0	1
5-2	1	N	1	1	0	1	1
5-3	1	N	1	0	1	1	1
5-4	1	N	0	1	1	1	1
5-5	0	N	1	1	1	1	1
6-1	N	1	1	1	1	0	1
6-2	N	1	1	1	0	1	1
6-3	N	1	1	0	1	1	1
6-4	N	1	0	1	1	1	1
6-5	N	0	1	1	1	1	1
<i>P_{new}D_{jalet}</i>							
1-1	0	0	0	0	1	N	0
1-2	0	0	0	1	0	N	0
1-3	0	0	1	0	0	N	0
1-4	0	1	0	0	0	N	0
1-5	1	0	0	0	0	N	0
2-1	0	0	0	0	N	1	0
2-2	0	0	0	1	N	0	0
2-3	0	0	1	0	N	0	0
2-4	0	1	0	0	N	0	0
2-5	1	0	0	0	N	0	0
3-1	0	0	0	N	0	1	0
3-2	0	0	0	N	1	0	0
3-3	0	0	1	N	0	0	0
3-4	0	1	0	N	0	0	0
3-5	1	0	0	N	0	0	0
4-1	0	0	N	0	0	1	0
4-2	0	0	N	0	1	0	0
4-3	0	0	N	1	0	0	0
4-4	0	1	N	0	0	0	0
4-5	1	0	N	0	0	0	0
5-1	0	N	0	0	0	1	0
5-2	0	N	0	0	1	0	0
5-3	0	N	0	1	0	0	0
5-4	0	N	1	0	0	0	0
5-5	1	N	0	0	0	0	0
6-1	N	0	0	0	0	1	0

Table A4 (continued)

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
6-2	N	0	0	0	1	0	0
6-3	N	0	0	1	0	0	0
6-4	N	0	1	0	0	0	0
6-5	N	1	0	0	0	0	0

Table A5

Category structure of all-new-P items, used in testing in Experiments 1–3.

Stimulus	Probabilistic feature						Deterministic feature
	Head	Body	Hands	Feet	Antenna	Tail	Button
<i>P_{all-new}D_{flurp}</i>							
1	N1	N1	N1	N1	N1	N1	1
2	N2	N2	N2	N2	N2	N2	1
3	N3	N3	N3	N3	N3	N3	1
4	N4	N4	N4	N4	N4	N4	1
<i>P_{all-new}D_{jalet}</i>							
1	N5	N5	N5	N5	N5	N5	0
2	N6	N6	N6	N6	N6	N6	0
3	N7	N7	N7	N7	N7	N7	0
4	N8	N8	N8	N8	N8	N8	0

that diffused attention and perhaps less efficient category learning in 4-year-olds are associated with better memory for specific exemplars, whereas selective attention and more efficient category learning in 7-year-olds and adults are associated with worse memory for exemplars. Furthermore, by externally cueing attention, we can turn adults' categorization strategy into a childlike one and increase their memory for exemplars. In contrast, in 4-year-olds, we can change only categorization strategy, whereas their memory accuracy remains uniformly high. These results coupled with computational simulations suggest (1) important decoupling between categorization and memory early in development and coupling of these processes later in development, (2) distinction between representation and decision in early categorization, and (3) potentially different mechanisms of category learning in younger and older participants. The reported results have important implications for understanding the role of attention in the development of categorization. They may also pose interesting challenges to theories and models of categorization that presume developmentally invariant mechanisms of categorization.

Author note

This research is supported by IES Grant R305A140214 and NIH grants R01HD078545 and P01HD080679 to Vladimir Sloutsky. We thank members of the Cognitive Development Lab for helpful comments.

Appendix A

Tables A1–A5 present the category structure of variants of all item types used in Experiments 1–3. The value 1 = any of seven dimensions identical to the prototype of Category F (flurp, see Fig. 1). The value 0 = any of seven dimensions identical to the prototype of Category J (jalet, see Fig. 1). The values N, N1, N2, N3, and N4 = new features which are not presented during training. P = probabilistic feature; D = deterministic feature. Variants of High-Match items (A1) were used in both training and test-

ing. Variants of Switch items (A2), new-D items (A3), one-new-P items (A4), and all-new-P items (A5) were used only in testing.

References

- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological model of multiple systems in category learning. *Psychological Review*, 105, 442–481.
- Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372–400.
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16, 437–443.
- 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, 105–199.
- Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, 112, 330–336.
- Bott, L., Hoffman, A. B., & Murphy, G. L. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, 136, 685–699.
- Bullock, M. J., & Opfer, J. E. (2009). What makes relational reasoning smart? Revisiting the relational shift in cognitive development. *Developmental Science*, 12, 114–122.
- 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.
- Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory and Cognition*, 3, 353–362.
- Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, 62, 73–101.
- 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.
- Deng, W., & Sloutsky, V. M. (2013). The role of linguistic labels in inductive generalization. *Journal of Experimental Child Psychology*, 114, 432–455.
- Deng, W., & Sloutsky, V. M. (2015a). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, 51, 392–405.
- Deng, W., & Sloutsky, V. M. (2015b). Labels, dynamic visual features, and distribution of attention in infant category learning. *Journal of Experimental Child Psychology*, 134, 62–77.
- Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48, 269–297.
- 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.
- Fisher, A. V., Godwin, E. K., & Matlen, B. (2015). Development of inductive generalization with familiar categories. *Psychonomic Bulletin & Review*, 22, 149–1173.
- Fisher, A. V., Godwin, E. K., Matlen, B., & Unger, L. (2015). Development of category-based induction and semantic knowledge. *Child Development*, 86, 48–62.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183–209.
- Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, 34, 686–708.
- Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, 13, 622–635.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 411–422.
- 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.
- Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. *Journal of Experimental Child Psychology*, 109, 321–335.
- Inagaki, K., & Hatano, G. (2002). *Young children's naïve about the biological world*. New York: Psychology Press.
- James, W. (1983/1890). *The principles of psychology* (originally published in 1890). Cambridge, MA: Harvard University Press.
- Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, 62, 499–512.
- Kaminski, J. A., & Sloutsky, V. M. (2013). Extraneous perceptual information interferes with children's acquisition of mathematical knowledge. *Journal of Educational Psychology*, 105, 351–363. <http://dx.doi.org/10.1037/a0031040>.
- 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.
- Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem solving. *Psychological Review*, 69, 1–16.
- 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.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- 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.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309–332.

- Macario, J. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, 6, 17–46.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129, 592–613.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1518–1533.
- 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.
- Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development*, 74, 1783–1806.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104–114.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 700–708.
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25–53.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, 9, 247–255.
- Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Object examining: Habituation and categorization. *Cognitive Development*, 6, 377–392.
- Pashler, H. (1999). *The Psychology of Attention*. Cambridge, MA: MIT Press.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, 52(1), 629–651.
- Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, 86, 227–272.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Quinn, P. C. (2011). Born to categorize. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (2nd ed., pp. 129–152). Oxford, UK: Blackwell Publishers.
- Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy*, 1, 31–46.
- Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: Comparing adults and children. *Journal of Experimental Child Psychology*, 131, 149–169.
- Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: The role of age and executive functioning. *PLoS ONE*, 9, e85316.
- Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51, 1–41.
- Robinson, C. W., Best, C. B., & Sloutsky, V. M. (2011). Auditory dominance and category learning. In *Paper presented at biennial meeting of society for research in child development, Montreal, Canada*.
- 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). Visual processing speed: effects of auditory input on visual processing. *Developmental Science*, 10, 734–740. <http://dx.doi.org/10.1111/j.1467-7687.2007.00627.x>.
- Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, 38, 495–553.
- Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, 133, 63–82.
- Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *Journal of Experimental Psychology: Applied*, 16, 361–377.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (Whole No. 517).
- Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, 34, 1244–1286.
- Sloutsky, V. M., Deng, W. S., Fisher, A. V., & Kloos, H. (2015). Conceptual influences on induction: A case for a late onset. *Cognitive Psychology*, 82, 1–31.
- Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, 79, 639–651.
- Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, 96, 125–144.
- Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., et al. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience and Biobehavioral Reviews*, 36, 2355–2369.
- Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, 24, 279–298.
- Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology*, 10, 502–532.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1411–1430.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, 65, 167–196.
- Son, J. Y., Smith, L. B., & Goldstone, R. L. (2011). Connecting instances to promote children's relational reasoning. *Journal of Experimental Child Psychology*, 108, 260–277. <http://dx.doi.org/10.1016/j.jecp.2010.08.011>.
- 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.
- 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.
- Younger, B., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, 57, 803–815.