The Development of Categorization

Vladimir M. Sloutsky and Anna V. Fisher

Please address correspondence to:

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

Anna V. Fisher
Department of Psychology
335-I Baker Hall
5000 Forbes Ave.
Pittsburgh PA 15213
Carnegie Mellon University
Phone: (412) 268-8656
Fax: (412) 268-2798
E-mail: fisher49@andrew.cmu.edu

Abstract

The ability to categorize is a ubiquitous property of human cognition. However, despite much progress in understanding how children learn to categorize, several important issues remain highly debated. The central issues in this debate concern the role of perceptual input and the role of linguistic labels in categorization and category learning. This chapter reviews recent evidence bearing on these key issues. Based on this evidence, we suggest that (1) categorization begins with extracting structure from the input based on overlapping perceptual features and attentional weights, and (2) development of the ability to selectively attend to some features while ignoring others underlies the learning of abstract concepts and categories.

The Development of Categorization

1. Introduction

Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech. Every time we see something as a kind of thing, for example, a tree, we are categorizing (Lakoff, 1987, p. 139).

The ability to categorize is a ubiquitous property of cognition: it enables treating appreciably different entities (e.g., different dogs, different tokens of the same word, different situations, or different problems) as variants of the same thing. This critically important property of intelligence promotes cognitive economy (as different entities are represented the same way) and application of learned knowledge to novel situations (e.g., learning that Dobermans bark would lead one to believe that German Shepherds bark as well).

In many ways categorization in humans is remarkably similar to that in other species. For instance, human and non-human primates, monkeys, rats, birds, fish, and crickets can learn a variety of visual and auditory categories based on the distribution of features in the input (Chase, 2001; Pearce, 1994; Ribar, Oakes, & Spalding, 2004; Wyttenbach, May, & Hoy, 1996; Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008). In the case of human infants, evidence for such perceptual groupings can be exhibited as early as three months of age (Bomba, & Siqueland, 1983; Quinn, Eimas, & Rosenkrantz, 1993).

In some other ways, categorization in humans is remarkably different from that in other species. For example, there is little evidence to suggest that non-human species have the ability to (1) organize categories into multiple levels of abstraction (such as Dalmatian → Dog → Mammal

→ Animal → Living Thing), (2) create ad-hoc categories (i.e., categories that are not established in memory but derived to achieve a novel goal, such as 'activities to do on a vacation in Japan'; Barsalou, 1991), or engage in abstract category-based inductive reasoning (i.e., inferring hidden properties on the basis of common category rather than perceptual similarity).

Therefore, it is not surprising that questions of how people acquire these remarkable abilities in the course of ontogenesis and how category learning interacts with language learning have fascinated generations of researchers. Attempts to answer these questions generated multiple theoretical proposals and a large body of evidence, with surprisingly little agreement with respect to these fundamental questions. The goal of this chapter is to summarize the debate on the development of categorization and category learning and to review some recent evidence bearing on this debate.

In what follows we first summarize two approaches to human category learning that dominate the debate in the current literature. A comprehensive historical overview of the theoretical developments and associated findings – while highly relevant to the issues discussed below – are outside of the scope of this paper. Excellent overviews of these issues are available elsewhere (Murphy, 2002). We then discuss the mechanism of early categorization and possible changes in this mechanism in the course of development.

1.1 Theoretical Approaches to Category Learning

There is a number of theoretical approaches aiming at explaining category learning and the development of mechanisms of categorization. Most researchers agree that even early in development people can learn categories that are based on multiple overlapping perceptual features (Ribar et. al., 2004; Quinn et. al., 1993). Most researchers also agree that adults demonstrate the ability to go beyond perceptual input and to form categories based on abstract unobservable

properties. There is little agreement however on when and how this ability emerges. Some researchers argue that the ability to form categories on the basis of unobservable features emerges from the ability to form perceptual categories (Rakison, & Poulin-Dubois, 2001; Rogers, & McClelland, 2008; Sloutsky & Fisher, 2004). Others argue that this is not the case and the two abilities are relatively independent (Booth, & Waxman, 2002; Gelman, 2003; Keil, Smith, Simons, & Levin, 1998).

Based on certain common theoretical assumptions, it is possible to classify these theories into two broad kinds: the Knowledge-Based Approach and the Coherence-Based Approach. These approaches offer divergent perspectives on the starting point, the mechanism and the development of categorization.

The Knowledge-Based Approach: Poverty of the Stimulus and Representational Constraints

The basic assumption behind this approach is that input is under-constrained and therefore structured knowledge cannot be recovered from input alone. This assumption is based on the Poverty of the Stimulus argument, which was originally proposed for the case of language acquisition (Chomsky, 1980), and was later extended to many other aspects of cognitive development, including category learning. According to this argument, if knowledge can not be recovered from data alone, then some innate constraints are necessary to explain its acquisition. It has been suggested that such constraints may come in the form of skeletal principles (R. Gelman, 1990), core knowledge (e.g., Carey, 2009; Spelke & Kinzler, 2007), or a variety of conceptual assumptions and naïve theories (Gelman, 2003). While there is no explicit agreed-upon definition of *conceptual* knowledge, the term "conceptual" is commonly used in reference to top-down knowledge that is used to structure input information. "…[C]hildren's categories are theory based: They are constructed not merely on the basis of perceptual characteristics and regularities but on the basis of children's beliefs and assumptions about the world and the way language works" (Jaswal,

2004, p. 1872; see also Booth, Waxman, & Huang, 2005**;** Gelman, 2003; Markman, 1991). This meaning of conceptual biases and knowledge as distinctly different from learned associations is assumed in this paper.

The constraints described above are what we will refer to as *representational constraints* (to be contrasted below with other kinds of constraints) – domain-specific knowledge that guides learner's attention towards a small set of features relevant for solving a specific problem in a given domain. The domain-specificity is of critical importance because principles constraining learning in the domain of number (e.g., knowledge that integers never end, cf. R. Gelman, 1990) are not useful in the domain of space.

This argument for representational constraints is based on the assumptions that we can successfully determine which constraints are required to solve a problem and which are available in the input. However, this assumption could be wrong. First, determining the required constraints is a highly non-trivial problem. For example, it has been traditionally accepted that the task of language learning requires the induction of a syntax that has an arbitrary large number of constraints. However, Chater and Christiansen (in press) have recently suggested this construal could be wrong. Specifically, they offered a critical distinction between "natural" and "cultural" induction (i.e., N-induction and C-Induction). N-induction involves the ability to understand the natural world; whereas, C-induction involves the ability to coordinate with other people. While the traditional construal of the problem of language acquisition has conceptualized it as an extremely difficult N-induction problem (i.e., the discovery of abstract syntax), Chater and Christiansen suggested that the problem should be construed as a much easier problem of C-induction. Instead of solving a problem requiring an arbitrary set of constraints (i.e., the problem of N-induction), individuals simply have to make the same guesses as everyone else. In sum, according to the former construal, the number of required constrains is very high, whereas according to the latter

construal, this number is substantially lower. Therefore, different conceptualizations of the same learning problem may result in vastly different estimates of the required constraints. Second, the number of constraints available in the input is also a non-trivial problem that still awaits a theoretical and empirical solution.

Finally, there is much evidence that people possess powerful learning mechanisms allowing them to detect statistical regularities in the input (French, Mareschal, Mermillod, & Quinn, 2004; Gomez & Maye, 2005**;** Pullum & Scholz, 2002; Saffran, Aslin, & Newport, 1996; Sloutsky & Fisher, 2008: Smith & Yu, 2008). In short, if the number of required constrained is not known, if the poverty of the input is questionable and if some structure can be recovered from data, the assumption of innate representational constraints does not seem justified.

The Coherence-Based Approach: Similarity, Statistics and Attentional Weights

According to a different theoretical perspective, category learning early in development is driven by the statistical regularities in the input. The central idea is that items belonging to the same category have multiple overlapping features, which jointly guide category learning. For instance, members of the category "bird": have wings, feathers, and beaks; are referred to as "birds", they lay eggs, and possess a set of typical behaviors (such as flying and nesting). These features cohere and co-vary, providing a learner with richly structured input. Powerful learning mechanisms, such as perceptual and attentional learning, allow children to take advantage of this coherent covariation in the input (Hall, 1991; Smith, 1989; Smith, Jones, & Landau, 1996; French, et. al., 2004; Rogers & McClelland, 2008; Sloutsky, 2003, in press; Sloutsky & Fisher, 2004; Sloutsky & Fisher, 2008).

A common but hardly accurate assumption about this approach is that it is constraint-free and any correlation present in the input gets encoded and learned. If true, this would have made the problem of learning computationally intractable. There are three reasons why this does not happen. First, information that is attended to and learned is limited, in part, by the basic information

processing properties of the organism – what we will refer to as *processing constraints*. It is well known that processing resources available to infants and young children are more modest compared to those available to older children and adults in terms of processing speed, processing capacity, attention capacity, memory, etc. (Baddeley, 1986; Demetriou, Christou, Spanoudis, & Platsidou, 2002). A case has been made that these limitations play a positive role early in development. As Elman (1993, pp. 17-18) put it: "Limited capacity acts like a protective veil, shielding the infant from the stimuli which may either be irrelevant or require prior learning to be interpreted" (for similar arguments see Newport, 1990).

Second, features available in the input differ in salience with salient features being more likely to get through the "protective veil" of limited capacity. In other words, more salient cues have a higher probability of being detected and learned than less salient cues. This learning in turn results in increased attention to these cues in the future, compared to other potentially salient cues (Hall, 1991; Kruschke & Blair, 2000**;** Nosofsky, 1986; Sloutsky & Spino, 2004).

And third, many features in real input are "bundled together" – members of most natural kind categories have much in common and many of those common features distinguish these members of from members of other categories. These bundled features may mutually reinforce each other thus resulting in non-additive effects on learning (see, Rogers & McClelland, 2004; Smith, Colunga, & Yoshida, in press, Sloutsky, in press, for reviews).

Many studies have documented that infants and young children are capable of sophisticated generalizations. For example, studies using a label extension task (e.g., "this is a dax, show me another dax") have demonstrated that 2- and 3-year-olds generalize the novel name to the same-shaped item when the exemplar is a solid object, but to the same material item when the exemplar is a non-solid item (Soja, Carey, & Spelke, 1991). One possible explanation of these findings is that at the start of word learning children have knowledge of ontological distinction between objects and

substances and that this knowledge manifests itself in this task (e.g., Soja, et al., 1991). Another explanation is that the distinction between solids and non-solids is supported by multiple syntactic and perceptual cues systematically co-varying with the distinction. This massive co-variation cues children's attention to the distinction and results in smart generalization behaviors (e.g., Smith et al. in press). Therefore, associative and attentional learning may give rise to smart behaviors. Another example of a smart behavior stemming from mundane mechanisms comes from work of Sloutsky and Fisher (2008) and we describe this work in a greater detail in a section below.

To summarize, according to the Knowledge-based approach early generalization is driven by a few theoretically-relevant features. Importantly, it is impossible to learn from input what the relevant features are – this knowledge comes *a priori* in the form of core knowledge, conceptual assumptions, and naïve theories. According to the Coherence-based approach, category learning is subserved by powerful learning mechanisms (i.e., perceptual and attentional learning) enabling extraction of statistical regularities in the input.

If the end-point of development is a full-fledged system of concepts, it might be beneficial for an organism to start with more abstract ontological kinds and focus on few theoretically-relevant features distinguishing among these kinds (Carey, 2009; R. Gelman, 1990; Spelke & Kinzler, 2007). However, there are several arguments as to why this is a highly implausible starting point. First, reliance on few relevant dimensions requires selective attention whereas, as we argue in the next section, selectivity is a later developing ability. Therefore, selective attention may be insufficiently developed in infants and young children to support conceptual development. The necessity for selectivity, however, could by-passed if (a) categories are based on multiple correlated features (something that is advocated by the coherence-based approach), (b) if category-relevant features are highly salient and capture attention automatically, or (c) if category labels serve as supervisory signals attracting attention to important within-category commonalities and thus

guiding category learning. However, as we argue in the sections *Salience and Categorization* and *Labels and Categorization* respectively, there is very little empirical support for either (b) or (c).

2. Categorization and Selective Attention

Learning of categories based on a few relevant dimensions requires selectivity, whereas learning of categories based on multiple overlapping features could be accomplished without selective attention. At the same time, young children may have difficulty focusing on few relevant features and ignoring multiple irrelevant ones, especially when irrelevant features vary independently of the relevant features. As a result, spontaneous learning of the categories based on a few relevant dimensions could be more challenging than spontaneous learning of the categories based on multiple overlapping features. In what follows, we review two lines of research, one demonstrating that under conditions of spontaneous learning, children are more likely to learn categories based on multiple overlapping features than on few ones. The second line presents evidence explaining the first one and indicating that young children have difficulty focusing on relevant information and ignoring irrelevant.

2.1 Spontaneous Learning of Categories Based on Multiple Overlapping Features

In a set of studies Kloos & Sloutsky (2008) examined the role of category structure in category learning. The measure of structure used in these studies was category density (i.e. the ratio of within category variability to between-category variability). In one study, 5-year-olds and adults were presented with a category learning task where they learned either dense or sparse categories. These categories consisted of artificial bug-like creatures that had a number of dimensions of variation: sizes of tail, wings, and fingers; the shadings of body, antenna, and buttons; and the numbers of fingers and buttons. For dense categories multiple dimensions co-varied and they were

jointly predictive of category membership, whereas for sparse categories there were few category-relevant dimensions, with the rest of the dimensions varying randomly within- and between categories.

Category learning was administered under either an unsupervised, spontaneous learning condition (i.e., participants were merely shown the items) or under a supervised, deliberate learning condition (i.e., participants were told the category inclusion rule). Critical data from this study are presented in Figure 1. The figure presents categorization accuracy (i.e., the proportion of hits, or correct identification of category members minus the proportion of false alarms, or confusion of non-members for members) after the category learning phase. Data presented in Figure 1 clearly indicate that for both children and adults, sparse categories were learned better under the explicit, supervised condition, whereas dense categories were learned better under the implicit, unsupervised condition. Also note that adults learned the sparse category even in the unsupervised condition, whereas young children exhibited no evidence of learning.

In addition, data from Kloos and Sloutsky (2008) indicate that while both children and adults exhibited able unsupervised, spontaneous learning of dense categories, there were marked developmental differences in unsupervised learning of sparse categories. Categorization accuracy in the unsupervised condition by category density and age are presented in Figure 2. Two aspects of these data are worth noting. First, there was no developmental difference in spontaneous learning of the very dense category, whereas there were substantial developmental differences in spontaneous learning of sparser categories, with children exhibiting less evidence for learning than adults.

Why are denser categories easier for young children? And why does learning of sparse but not dense categories undergo developmental progression? One possibility that has been discussed recently (e.g., Sloutsky, in press) is that due to immaturities in the pre-frontal cortex young children

have difficulty selectively focusing on relevant features, while ignoring irrelevant ones. At the same time, this ability seems to be critical for learning sparse categories. As pre-frontal cortex matures, the ability to focus on relevant and ignore irrelevant improves, and so does the ability to learn sparse categories.

This issue has been addressed in a series of recent yet unpublished studies by Yao and Sloutsky. These researchers presented participants with a series of tasks: simple matching, simple generalization, and complex generalization. Each of the tasks was a variant of the match-to-sample task consisting of a target and two test items, with one of the test items being a match for the target. There were two critical modifications to the standard task. First, each item had a less salient component that had to be used in matching and a more salient distracter component that had to be ignored (see Figure 3), with participants instructed to match the items on the basis of the less salient component. There were three within-subject conditions: Supportive (Figure 3a), Conflict (Figure 3b), and Neutral (Figure 3c). In the Supportive condition, the test item that matched the target on the less salient component, also matched the target on the more salient component. In contrast, in the Conflict condition, the test item that matched the target on the less salient component mismatched it on the more salient component, whereas the mismatching test item matched the target on the more salient component. Finally, in the Neutral condition, the more salient component was fixed across the target and test items, so that the task could not be performed on the basis of this component. Therefore, the Neutral condition served as the baseline for task performance. If participants ignore task irrelevant information, their performance should not differ across the conditions. Alternatively, if participants cannot ignore the more salient (yet task-irrelevant) component, their performance should be above the baseline in the Supportive condition and below the baseline in the Conflict condition. In the simple matching task, participants were asked to match the target and one of the test items on the basis of the less salient component. Whereas 4-5-

year-olds performed equally well across the conditions, 3-4 year-olds exhibited significant decrease in matching accuracy in the Conflict condition. Furthermore, increase in task demands resulted in 3-4-year-olds and 4-5-year-olds failing the Conflict condition, while succeeding in the Neutral and Supportive conditions. These findings support the idea that young children have difficulty selectively focusing on some features, while ignoring other features, especially when these other features are more salient than the target features.

While having difficulty attending selectively to few category-relevant features, young children can ably learn multiple dense categories without supervision, and this learning is often implicit. For example, in a recent study Sloutsky and Fisher (2008) presented 4- and 5-year-olds with a contingency learning task. On each trial, participants were presented with triads of items. Sometimes, the triads appeared in Context 1 (i.e., they were presented on a yellow background at the top-right corner of a computer screen), and sometimes the triads appeared in Context 2 (i.e., they were presented on a green background at the bottom-left corner of a computer screen). When the triads appeared in Context 1, shape was predictive of an outcome, whereas when they appeared in Context 2, color was predictive of the outcome. Even though no feedback was provided, children quickly learned the dimension-context contingencies and generalized by shape in Context 1 and by color in Context 2. This finding is remarkable given that children exhibited little awareness of what they had learned (Experiment 2) and failed to exhibit this flexibility when they were explicitly told to focus on a given predictor (Experiment 3). These finding suggests that the learned flexibility was achieved by implicit learning, which is characteristic of the compression-based system.

In sum, young children have no difficulty spontaneously learning dense categories, whereas they have greater difficulty learning categories based on few relevant features. We argue that this difficulty may stem from immaturities of selective attention, with young children having difficulty focusing on few relevant features while ignoring multiple irrelevant features.

2.2 Salience and Categorization

As discussed above, learning categories bound by multiple correlated features can be accomplished in the absence of mature attentional selectivity. Similarly, such selectivity is likely unnecessary for learning a category defined by a single but highly salient feature that captures attention automatically. Could it be that conceptual knowledge can be construed as such a feature? This issue has been explored in recent studies that used a task-switch paradigm. Task-switch paradigms are typically used to investigate executive control by presenting participants with a task in which performance demands change in the middle of the task, such as the Wisconsin Card Sort test (Berg, 1948). In a child-friendly adaptation of this test, the Dimensional Change Card Sort task (DCCS) children are presented with a set of cards depicting familiar objects that differ on two dimensions, such as color and shape (e.g., red and blue flowers, and red and blue boats) (Zelazo, Frye, & Rapus, 1996). Children are first asked to sort cards based on one dimension and then to switch sorting based on the other dimension. Despite understanding the instructions and reminders of the sorting rule given on every trial, children younger than four years of age often perseverate in sorting by the original dimension. Importantly, the pattern of perseveration is symmetric: it does not matter whether children start sorting by color or by shape – post-switch accuracy is reduced regardless of the starting sorting dimension.

Perseveration errors described above are one type of *switch costs* – robust decrease in performance after task switch. The magnitude of switch costs is influenced by a number of factors; the most relevant of these for the study described below is saliency of the post-switch cues. Highly salient post-switch cues reduce switch costs in both children and adults, presumably because salient cues capture attention automatically and thus reduce demands on executive control (Fisher, in press; Koch, 2001). For example, post-switch accuracy in the DCCS task can be nearly as high as pre-

switch accuracy in 3-year-old children when they are switching from sorting by a less salient dimension to sorting by a more salient dimension (Fisher, in press).

Based on the findings that salient post-switch cues reduce switch costs and allow young children to successfully switch to a new sorting task, it is possible to empirically evaluate the possibility that "deep" conceptual features are more salient than "surface" features for young children's category judgments. If conceptual information is central to categorization early in development (and thus highly salient) whereas perceptual information is peripheral (and thus less salient than), then switch costs should be lower when children switch from categorizing by perceptual information to categorizing by conceptual information. However, if perceptual information is more salient than conceptual information in early categorization, the opposite pattern of performance should be observed.

To test these possibilities, Fisher (2009) presented 3-, 4-, and 5-year-old children with a novel task, in which participants were asked to categorize the same set of objects twice and categorization instructions were changed in the middle of the task. Stimuli in this task consisted of iconic images of well-known objects (a familiarity check administered after the experiment proper confirmed that all children were indeed familiar with the materials). These images were organized into triads that put category information in conflict with appearance similarity. For example, an open red umbrella was paired with a folded blue-and-yellow umbrella to create a category match and with a red mushroom to create an appearance match (see Figure 4). Children were asked to categorize the objects twice: once by grouping together items that "look similar" and once by grouping together items that are "the same kind of thing" (the order of tasks was randomized for each child). Proportion of correct responses in each task was then used to calculate perceptual and conceptual costs of switching.

Results of this study indicated that for every age group tested in the study the costs of switching to categorizing by kind were higher than costs of switching to categorizing by appearance. Across all three age groups, costs of switching to categorizing by object kind were nearly 25% higher than costs of switching to categorizing by appearance. Furthermore, in 4- and 5-year-old children switching to categorizing by appearance resulted in negligibly low switch costs (8%, not different from zero), whereas switching to categorizing by kind produced high switch costs (39%, significantly above zero). These findings indicate that perceptual information is more salient than conceptual information.

3. The Role of Labels in the Development of Categorization

Although young children might have difficulty spontaneously focusing on few category-relevant features, their attention to these features could be guided by words (Gelman, 2003; Jaswal, 2004). In particular, category labels (especially when presented as count nouns) may serve as invitations to form categories and they attract attention to within-category commonalities (Gelman & Markman, 1986; Markman, 1991). As we argue in this section, there is little evidence that this is the case.

3.1 Labels as Features Contributing to Similarity

There is much evidence that category labels guide category learning in adults (Yamauchi & A. Markman, 2000). The overall reasoning behind this work is that if labels are category markers, they should be treated differently from the rest of features (such shape, color, size, etc). However, this may not be the case if labels are features. Therefore, inferring a label when features are given (i.e., a classification task) should elicit different performance from a task of inferring a feature when the label is given (i.e., a feature induction task). To test these ideas, Yamauchi and A. Markman (2000,

see also A. Markman & Ross, 2003, for a review) a category learning task that was presented under either classification or feature induction learning condition. There were two categories, C_1 and C_2 denoted by two labels, L_1 and L_2. In the classification task, they were presented with a creature that was not labeled, and the task was to predict the category label. The critical condition was the case when an item was a member of C_1, but was similar to C_2, with the dependent variable being the proportion of C_1 responses. The results indicated that there were significantly more category-based responses in the induction condition (where participants could rely on the category label) than in the categorization condition (where participants had to infer the category label). It was concluded therefore that category labels differed from other features in that participants treated labels as category markers.

At the same time, there is much evidence that, in contrast to adults, early in development labels are features of objects, with similarity of compared entities computed over both appearance and labeling attributes (Sloutsky, Lo, & Fisher, 1999; Sloutsky, 2003; Sloutsky & Fisher, 2004). These intuitions were implemented in a similarity-based model of early generalization abbreviated as SINC (for Similarity-Induction-Naming-Categorization).

According to SINC, early in development induction is a function of the overall perceptual similarity computed over weighted visual and auditory feature (Sloutsky & Fisher, 2004). The theory underlying SINC assumes that auditory information (including linguistic labels) often has greater attentional weights than visual information early in development (support for this assumption is reviewed in the next section). Therefore, when two entities A and B look equally similar to the target entity T, but only one of these entities (e.g., B) is referred to by the same label as the target, *perceived similarity* of test B to the target will be greater than of test A to the target. In this situation children should be highly likely to rely on matching labels to perform tasks such as property induction and categorization. However, when test A looks markedly more similar to the

target than test B but test B shares the name with the target, children may perceive both A and B as equally similar to the target. In this situation children should be equally likely to rely on matching labels and on matching appearances when performing property induction and categorization tasks.

The above predictions were supported in a series of studies in which 4- to 5-year-old children were presented with similarity judgment, property induction, and categorization tasks, with SINC accounting for over 90% of observed variance in performance on these tasks (Sloutsky & Fisher, 2004). Furthermore, SINC accurately predicted a bi-modal distribution of responses on the same property induction task and stimuli used in prior research to argue that children are more likely to rely on matching labels than on matching appearances in the course of induction (Gelman & Markman, 1986) (see Figure 5 panel A). In particular, as predicted by SINC, preschool-age children were above chance in relying on matching labels when none of the test items were overwhelmingly perceptually similar to the target (e.g., when one test item had the same shape as the target and the other test item had the same texture and color as the target; see Figure 5 panel C). In contrast, when one of the test items looked overwhelmingly similar to the target and another test item had the same name as the target (see Figure 5 panel B), children's reliance on matching labels did not exceed chance level.

As stated above, SINC assumes that auditory features may have a greater attentional weight than visual features early in development. The reasons as to why this might be the case are reviewed in the next section.

3.2 Labels Overshadow Visual Input

There is evidence that auditory input may affect attention allocated to corresponding visual input (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008), and these effects may change in the course of learning and

development. In particular, linguistic labels may strongly interfere with visual processing in prelinguisitic infants and young children, although these interference effects may somewhat weaken with age (Sloutsky & Robinson, 2008, see also Robinson & Sloutsky, 2007a; 2007b).

These issues have been examined in depth in a series of experiments by Sloutsky and colleagues (e.g., Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). In these experiments, 4-year-olds and adults were presented with a compound Target stimulus, consisting of simultaneously presented auditory and visual components ($AUD_{Target}VIS_{Target}$). These experiments were similar to the described above experiment, except that no learning was involved. Participants were presented with a Target item, which was followed immediately by a Test item (the Test item could be either the same as the Target or different) and participants had to determine whether the Test item was exactly the same as the Target.

There were four types of test items: (1) $AUD_{Target}VIS_{Target}$, which was the Old Target item, (2) $AUD_{Target}VIS_{New}$, which had the target auditory component and a new visual component, (3) $AUD_{New}VIS_{Target}$, which had the target visual component and a new auditory component, or (4) $AUD_{New}VIS_{New}$, which had a new visual component and a new auditory component. The task was to determine whether each presented test item was exactly the same as the Target (i.e., both the same auditory and visual components) or a new item (i.e., differed on one or both components).

It was reasoned that if participants process both auditory and visual stimuli, they should correctly respond to all items by accepting Old Target items and rejecting all other test items. Alternatively, if they fail to process the visual component, they should falsely accept $AUD_{Target}VIS_{New}$ items, while correctly responding to other items. Finally, if they fail to process the auditory component, they should falsely accept $AUD_{New}VIS_{Target}$ items, while correctly responding to other items. In one experiment (Napolitano & Sloutsky, 2004), speech sounds were paired with either geometric shapes or pictures of unfamiliar animals. Results indicated that while children ably

processed either stimulus in the uni-modal condition, they failed to process visual input in the cross-modal condition. Furthermore, a yet unpublished study by Napolitano and Sloutsky indicates that interference effects attenuate gradually in the course of development, with very little evidence of interference in adults.

There is also evidence that this dominance of auditory input is not under strategic control: even when instructed to focus on visual input young children had difficulties doing so (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004). In one of the experiments described in Napolitano and Sloutsky (2004), 4-year-olds were explicitly instructed to attend to visual stimuli, with instructions repeated before each trial. However, despite the repeated explicit instruction to attend to visual stimuli, 4-year-olds continued to exhibit auditory dominance. These results suggest that auditory dominance is unlikely to stem from deliberate selective attention to a particular modality, but it is more likely to stem from automatic pulls on attention.

If linguistic labels attenuate visual processing, such that children ably process a label, but they do so to a lesser extent the corresponding visual input, then these findings can explain the role of labels in categorization tasks. In particular, items that share a label may appear more similar than the same items presented without a label. In other words, early in development, labels may function as features contributing to similarity, and their role may change in the course of development. In fact, there is evidence supporting this possibility (e.g., Sloutsky & Fisher, 2004a; Sloutsky & Lo, 1999).

3.3 Effect of Semantic Similarity and Co-Occurrence Probability of Labels

If labels are merely features contributing to similarity, then why do semantically similar labels (as opposed to identical labels) affect young children's generalization? Specifically, Gelman and Markman (1986; Experiment 2) demonstrated that 4- to 5-year-old children generalize unobservable

properties from one object to another not only when the objects are referred to by identical labels (which can be readily explained by the label overshadowing hypothesis), but also when objects are referred to by semantically similar labels. For example, in this study children could be told that a 'rabbit eats bugs' whereas a 'squirrel eats grass', and asked whether the target item (called a 'rabbit' in the Identical Labels condition and a 'bunny' in the Synonyms condition) 'eats bugs like the rabbit' or 'eats grass like the squirrel.' Gelman and Markman found that children generalized properties to categorically similar items at a level above that expected by chance in both the Identical Labels and the Synonyms conditions. Importantly, children's performance with synonyms was statistically equivalent to their performance with identical labels (63% and 67% of category-based responses, respectively). These findings suggested that labels not merely promote label-based inferences but truly point to commonalities in kind.

However, Fisher (2010a) recently suggested that some label pairs in the Gelman and Markman (1986) study included pairs of labels that were not only semantically similar but also co-occurred in child-directed speech (e.g., *bunny-rabbit, puppy-dog*) according to the CHILDES database (MacWhinney, 2000; see Fisher 2010a for details of the co-occurrence analyses). Co-occurrence of words in natural language has been argued to give rise to strong lexical associations (Brown & Berko, 1960; McKoon & Ratcliff, 1992). For instance, when people are presented with the words *puppy* and *bunny*, the probability of obtaining the words *dog* and *rabbit* in response is 71% and 74% respectively. In comparison, the probability of producing the word *stone* in response to *rock* (a pair that was also used in Gelman and Markman's Synonyms condition) is only 3% (Nelson, McEvoy, & Schreiber, 1998)[1].

[1] Of course, words can become associated by means other than co-occurrence (e.g., thematic relatedness); however, co-occurrence is argued to play an important role in the formation of lexical associations (Brown & Berko, 1960; McKoon & Ratcliff, 1992).

It is possible that effects reported by Gelman and Markman's (1986) stemmed from labels pointing to commonalities in kind and promoting category-based reasoning. However, it is also possible that responses (at least to some of the items) stemmed from associative priming: children's responses could be based not on the reasoning that bunnies and rabbits are the same kind of animal, but on the fact that the word *bunny* primed the word *rabbit*, but not the word *squirrel*.

Fisher and colleagues (Matlen, Godwin, & Fisher, 2010) recently obtained evidence in support of the latter possibility. They presented a group of 4-year-old children and adults with an induction task in either an Identical Labels condition or a Semantically Similar Labels condition. Within each labeling condition participants were presented either with labels that were likely to co-occur in the CHILDES database (e.g., bunny-rabbit, puppy-dog, kitty-cat) or with labels that were unlikely to co-occur (e.g., alligator-crocodile, mouse-rat, toad-frog). All labels referred to animal names and all properties consisted of blank predicates. For example, in the Semantically Similar Labels condition participants could be told the "alligator has matlen inside" and asked whether the "crocodile" (a semantically-related test item) or the "hippo" (an unrelated test item) was likely to share this property.

Adults were at ceiling in performing category-based induction in all conditions. In the Identical Labels condition children's performance was above chance. However, in the Semantically Similar Labels condition children performed above chance only with co-occurring labels, but their performance was at chance when non-co-occurring semantically similar labels were used. A Picture Identification task administered immediately after the experiment proper confirmed that children were well familiar with all of the labels used in this study (accuracy on the Picture Identification task was 99%).

Results of a follow-up experiment indicated that less than 20% of 4-year-old children spontaneously utilize semantic similarity when labels are unlikely to co-occur in natural language and approximately 50% of 5-year-old children can do so; however, the majority of children do not spontaneously utilize semantic similarity of non-co-occurring labels in reasoning tasks until 6 years of age (a similar developmental pattern was reported by Ramesh, Matlen, and Fisher (2010) in a study that utilized an analogical reasoning task). These findings pose a serious challenge to the notion that labels guide attention to commonalities in kind from as early as two years of age (Gelman and Coley, 1990).

3.4 Flexible Adjustment of Attention Allocated to Labels and Appearances

Another challenge to the notion that labels are centrally important for categorization and appearances are peripheral from early in development, comes from a series of studies that examined the flexibility with which children can rely on label and appearance attributes in the course of induction. If reliance on labels and appearances in induction and categorization tasks stems from automatic allocation of attention to these attributes, then it should be possible to modify children's reliance on these attributes by manipulating the amount of attention directed to labels and appearances. If however, children believe that labels are more theoretically central than appearances, then such a change should be difficult (if not impossible), because beliefs are notoriously resistant to change. When children and adults hold strong beliefs they tend to disregard evidence that conflicts with these beliefs. For example, when children and adults come to a task with pre-existing beliefs, they tend to incorrectly encode information in such a way that is consistent with their beliefs – phenomenon known as *illusory correlation* (Hamilton & Rose, 1980; Johnston & Jacobs, 2003; Lord, Ross, Lepper, 1979; Meehan and Janik, 1990).

The Knowledge-Based approach predicts that the status of conceptually-central attributes (i.e., labels) and conceptually-peripheral attributes (i.e., appearances) is relatively fixed and resistant to change, even when new data become available. The Coherence-Based approach predicts that children's reliance on various attributes in the course of induction is shaped by data – for instance data on how well a particular attribute correlates with a particular outcome and thus how useful this attribute might be in predicting the outcome. Therefore, this approach predicts a high degree of flexibility in relying on labeling and appearance attributes in young children who are yet to form beliefs about theoretical importance of labels.

There is ample evidence suggesting that attention allocated to different perceptual attributes can be flexibly changed in both animals and humans by manipulating the predictive value of an attribute: when a particular cue is consistently predictive, attention allocated to this cue increases automatically, whereas when a cue is consistently non-predictive, attention allocated to this cue decreases (see Hall, 1991 for a review). Therefore, manipulating predictive values of labels and appearances in the course of induction should change children's reliance on these attributes.

Evidence in support of this hypothesis comes from the studies in which 4- to 5-year-old children were given evidence that a particular cue – such as a shared label, similar appearance, or shared inheritance (i.e., the items were introduced as *mother* and *baby*) – does not correlate with the outcome (Fisher and Sloutsky 2006; Fisher, 2010b; Sloutsky & Spino, 2004). During the training phase in these studies, children's task was to predict an unobservable property of a target object (e.g., whether it has thick blood or thin blood) based on the information about the test objects. Children were presented with one target object and up to three test objects, each of which shared only one attribute with the target (e.g., each test object could have the same label as the target, look similar to the target, or have given birth to the target). After each response children were provided with feedback indicating whether or not the response was correct. The correctness of the response

varied based on the condition to which the children were assigned. For example, in the Label Training condition children were provided with positive feedback for making label-based predictions and in the Appearance Training condition children were provided with positive feedback for making appearance-based predictions. In some of the studies the feedback was explicit (i.e., children were told whether their response was correct or not; Sloutsky & Spino, 2004) and in other studies the feedback was implicit (i.e., children were shown a short cartoon when they made correct predictions; Fisher and Sloutsky 2006; Fisher, 2010b).

The training phase was immediately followed by a transfer phase in which children were tested on an induction task in the absence of feedback. The transfer phase included novel stimuli on which children were not trained (see Figure 6 for examples). A consistent pattern of results emerged from these studies suggesting that the pattern of responses in the transfer phase was (1) different from the baseline pattern of induction and (2) consistent with the training condition to which children were assigned. In other words, children who were assigned to the Label Training condition relied predominantly on labels during the transfer phase, children who were assigned to the Appearance Training condition relied predominantly on appearances, and children assigned to the Inheritance Training relied predominantly on inheritance. In all of these conditions children's performance was different from the baseline pattern of induction, which typically reflects children's integration of information from multiple cues (Sloutsky & Fisher, 2004).

Furthermore, Sloutsky and Spino (2004) showed that the trained pattern of performance persisted when preschool-age children were tested in a delayed transfer task, which was administered approximately three months after the training phase, by a different experimenter, in a different location, and used a novel set of stimuli. Overall, results described above suggest that reliance on labeling and appearance attributes in the course of induction is flexible: when either attribute becomes non-predictive in the course of training, reliance on this attribute decreases

markedly during testing. These findings challenge the position that effects of labels on induction stem from young children's belief in their conceptual importance. It is not clear how this hypothesis can account for the fact that a relatively modest amount of training – 10 trials in Sloutsky & Spino (2004) and 16 trials in Fisher & Sloutsky (2006) resulted in a flexible shift of attention away from predictors that are supposed to be theoretically central (i.e., linguistic labels) to those that are supposed to be theoretically peripheral (i.e., appearances). At the same time, these findings support the idea that reliance on labels and appearances in the course of induction stems from allocation of attention to more predictive (and hence more salient) features and not from theoretical beliefs about centrality of certain attributes over others.

In sum, categorization on the basis of few theoretically relevant features may hinge on selective attention that exhibits a great degree of immaturity early in development. In principle, this immaturity could have been compensated for if early categorization were assisted by language with words attracting attention to important within-category commonalities or across-category differences. However, this is not likely due to immaturities in cross-modal processing, with auditory input overshadowing visual input early in development. We presented arguments that instead of being based on few category-relevant features, early categorization is based on automatic attention to bundles of category-relevant features or to few highly salient ones. Although much research is needed to flesh out the details, there is sufficient evidence to consider this mechanism a plausible candidate for how categories are learned early in development.

4. Early Categorization: What Develops?

If categorization starts out as implicit learning of perceptual categories, how do people acquire the ability to learn highly abstract categories as well as concepts of mathematics, science, morality or law? Although we do not have a precise answer to this question, we believe that there are several

critical steps in this process. Perhaps, the most important step is the ability to selectively attend to category relevant information while ignoring category-irrelevant information. This ability is critical for learning categories that are based on few relevant features, and it may be even more critical when these features are unobservable, as is the case of highly abstract concepts. For example, many abstract categories are bound by unobservable relational characteristics, while having very few (if any) perceptual features in common. Consider such concepts as *fairness* and *reciprocity*. While these concepts are lexicalized categories, one would be hard pressed to find perceptual commonalities in instances of reciprocity. Therefore, learning of such a category puts high demands on selective attention as one needs to ignore much irrelevant perceptual information. As we discussed in this chapter, the ability to do so has a protracted developmental course and is likely immature long after children make great strides in concept acquisition. As a result of this immaturity of selective attention, some of the abstract concepts are beyond children's reach. These developmental limitations could be highly adaptive as they prevent young category learners from entertaining one commonality at a time. As a result, young learners have to automatically focus on bundles of overlapping features. Unlike isolated commonalities, that may not generate a clear distinction among categories, bundled commonalities are more likely to do so. Therefore, there is little surprise that the first categories that children acquire are based on multiple overlapping features (e.g., cats, dogs, birds, or people) and that even children of scientists and lawyers do not start their learning with concepts of *gravity* or *litigation*.

References

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Barsalou, L. W. (1991). Deriving categories to achieve goals. in G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, (Vol. 27). New York: Academic Press.

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 39, 15-22.

Bomba, P. C. & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, 35, 294-328.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, 84, B11-B22.

Booth, A., Waxman, S. R., & Huang, Y.T. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology*, 41, 491-505.

Brown, R. & Berko, J. (1960). Word association and the acquisition of grammar. *Child Development*, 31, 1-14.

Brown, A, M, Lindsey, D, T, McSweeney, E, M, & Walters, M M. (1995). Infant luminance and chromatic contrast sensitivity: Optokinetic nystagmus data on 3-month-olds. *Vision Research*, 35, 3145-60.

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Chase, A. (2001). Music discriminations by carp (Cyprinus carpio). *Animal Learning & Behavior*, 29, 336-353.

Chater, N. & Christiansen, M. H. (in press). Language acquisition meets language evolution. *Cognitive Science*.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Demetriou, A., Christou, C., Spanoudis, G., & Platsidou, M. (2002). The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society of Research in Child Development*, 67, SN 268.

Elman, J. L. (1993) Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71-99.

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

Fisher, A.V. (2010b). Mechanisms of induction early in development. In M. Banich & D. Caccamise (Eds.) *Generalization of Knowledge: Multidisciplinary Perspectives* (pp. 89-112). New York: Psychology Press.

Fisher, A.V. (2009). Does conceptual information take precedence over perceptual information early in development? Evidence from perseveration errors. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the XXXI Annual Conference of the Cognitive Science Society*, 1330-1335.

Fisher, A.V., & Sloutsky, V.M. (2006). Flexible Attention to Labels and Appearances in Early Induction. In R. Sun and N. Miyake (Eds.) *Proceedings of the XXVIII Annual Conference of the Cognitive Science Society*, 220-225.

French R. M., Mareschal D., Mermillod, M., & Quinn P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, 133, 382-397.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, 14: 79–106.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S.A. & Coley, J.D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796-804.

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

Gómez, R. & Maye, J. (2005). The developmental trajectory of non-adjacent dependency learning. *Infancy*, 7, 183-206.

Hall, G. (1991). *Perceptual and associative learning*. New York: Oxford University Press.

Hamilton, D.L., & Rose, T.L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Personality and Social Psychology,* 39, 832-845.

Jaswal, V.K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*, 3, 279-300.

Johnston, K.E. & Jacobs, J.E. (2003). Children's Illusory Correlations: The Role of Attentional Bias in Group Impression Formation. *Journal of Cognition and Development*, 4, 129-160.

Keil, F.C., Smith, C.S., Simons, D., & Levin, D. (1998). Two dogmas of conceptual empiricism. *Cognition*, 65, 103-135.

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.

Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1474-1486.

Kruschke, J. K. & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, 7, 636-645.

Lakoff, G. (187). *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press.

Lord, C.G., Ross, L., & Lepper, M.R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Markman, E.M. (1991). *Categorization and Naming in Children: Problems of Induction*. MIT Press.

Matlen, B., Fisher, A.V., & Godwin, K.E. (2010). Children's Inductive Inference with Synonymous Labels. In S. Ohlsson and R. Catrambone (Eds.) *Proceedings of the XXXII Annual Conference of the Cognitive Science Society*, 1493-1498.

Meehan, A. M. & Janik, L. M. (1980). Illusory correlation and the maintenance of sex role stereotypes in children. *Sex Roles*, 22, 83-95.

McKoon, G. & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1155 – 1172.

Murphy, G. L. (2002). *The Big Book of Concepts*, MIT Press.

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11-28.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.

Pearce, J.M. (1994). Discrimination and Categorization. In N. J. Mackintosh (Ed.), *Animal Learning and Cognition*, 110-134.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month odl infants. *Child Development*, 80, 674-685.

Pullum, G. K. & Scholz, B. (2002) Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 19, 9-50.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3- and 4-month-old infants. *Perception*, 22, 463-475.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, 127, 209-228.

Ramesh, S., Matlen, B., & Fisher, A.V. (2010). Development of Relational Reasoning with Semantically Similar Labels. In S. Ohlsson and R. Catrambone (Eds.) *Proceedings of the XXXII Annual Conference of the Cognitive Science Society*, 1499-1504.

Ribar, R. J., Oakes, L. M., & Spalding, T. L. (2004). Infants can rapidly form new categorical representations. *Psychonomic Bulletin & Review*, 11, 536-541.

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. (2007a). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, 10, 734-740.

Robinson, C. W., & Sloutsky, V. M. (2007b). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, 11, 223-253.

Rogers, T. T., & McClelland, J. L. (2008). Precis of Semantic Cognition, a Parallel Distributed Processing Approach. *Behavioral and Brain Sciences*, 31, 689-749.

Rogers, T. T. & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.

Sloutsky, V.M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, 7, 246-251.

Sloutsky, V. M. (in press). From perceptual categories to concepts: What develops? *Cognitive Science.*

Sloutsky, V. M, & Napolitano, A.C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, 74 (3), 822-833.

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

Sloutsky, V. M., & Spino, M. A. (2004). Naive theory and transfer of learning: When less is more and more is less. *Psychonomic Bulletin and Review*, 11, 528-535.

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.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in visual processing: From overshadowing to attentional tuning. *Cognitive Science*, 32, 354-377.

Smith, L. B. (1989) A model of perceptual classification in children and adults. *Psychological Review*, 96, 125-144.

Smith, L. B., Colunga, E., & Yoshida, H. (in press). Contextually-cued attention as a mechanism in development. *Cognitive Science*.

Smith, L.B., Jones, S.S., Landau, B., Gershkoff-Stowe, L. & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological Science*, *13*, 13-19.

Smith, L.B., & Yu, C. (2008). Infants Rapidly Learn Word-Referent Mappings via cross-situational Statistics. *Cognition*, *106*, 333-338.

Soja, N.N., Carey, S., & Spelke, E.S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38, 179-211.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10, 89-96.

Wyttenbach, R, May, M, &Hoy, R. (1996). Categorical perception of sound frequency by crickets. *Science*, 273, 1542-1544.

Yao, X., & Sloutsky, V. M. (2010). Selective attention and development of categorization: An Eye tracking study. In S. Ohlsson and R. Catrambone (Eds.) *Proceedings of the XXXII Annual Conference of the Cognitive Science Society*, 1980-1985.

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

Zelazo, P. D., Frye, D. & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11, 37-63.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews*, 3, 13-45.

Figure Captions.

Figure 1. Figure 1. Mean accuracy scores by category type and learning condition in adults (Panel A) and children (Panel B). Error bars represent standard errors of the mean. After Kloos and Sloutsky (2008).

Figure 2. Unsupervised category learning by density and age group in Kloos and Sloutsky (2008).

Figure 3. Example of stimuli used by Yao and Sloutsky (in press). Small circles with symbols inside are task-relevant, whereas the large colorful stimuli are task-irrelevant. The item below is the Target, whereas two items above are Test stimuli. Participants are asked to match small circles of the Target item with the small circles of one of the Test items. A. Supportive condition; B. Conflict condition; C. Neutral condition.

Figure 4. Example of a triad used in Fisher (2009). Target: open red umbrella; Appearance Match: red mushroom; Category Match: folded multi-color umbrella.

Figure 5. Summary of results from Sloutsky and Fisher (2004; Experiment 4). Panel A presents responses observed in preschool-age children by Sloutsky & Fisher (2004) and predicted by SINC on the property induction task and materials originally used by Gelman & Markman (1986). Examples of triads used originally in Gelman and Markman (1986) and then in Sloutsky and Fisher (2004) are presented in Panels B-C. In Panel B the test item that was referred to by the label different from the target (i.e., *bat*) looks overwhelmingly more similar to the target than the test item that was referred to by the same label as the target (i.e., *bird*). This triad corresponds to triad #1 on the graph in Panel A. In Panel C both test items look equally similar to the target item (for

results of the similarity calibration with children see Fisher & Sloutsky, 2004, Experiment 4). This triad corresponds to triad #2 on the graph in Panel A.

Figure 6. Example of training stimuli (Panel A) and testing stimuli (Panel B) from Fisher and Sloutsky (2006). Children were trained that either similar appearance of matching labels were a predictive cue in a property induction task (Experiment 1) or a similarity judgment task (Experiment 2) using stimuli similar to those in Panel A. Transfer of this learning was tested in a property induction task using stimuli similar to those in Panel B.

Figure 1.

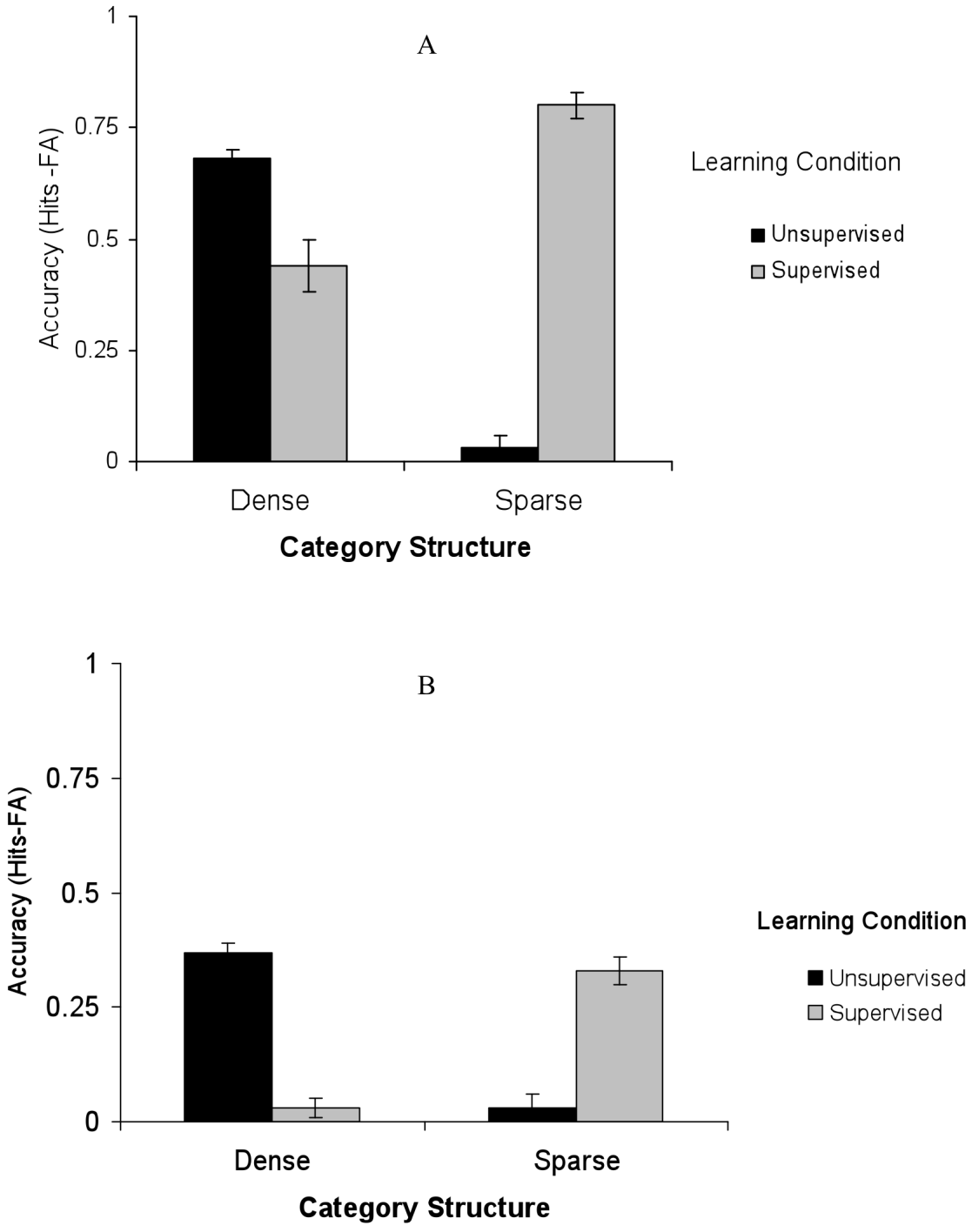

A
1
0.75
0.5
0.25
0
Accuracy (Hits -FA)
Dense
Sparse
Category Structure
Learning Condition
Unsupervised
Supervised
B
1
0.75
0.5
0.25
0
Accuracy (Hits-FA)
Dense
Sparse
Category Structure
Learning Condition
Unsupervised
Supervised

Figure 2.

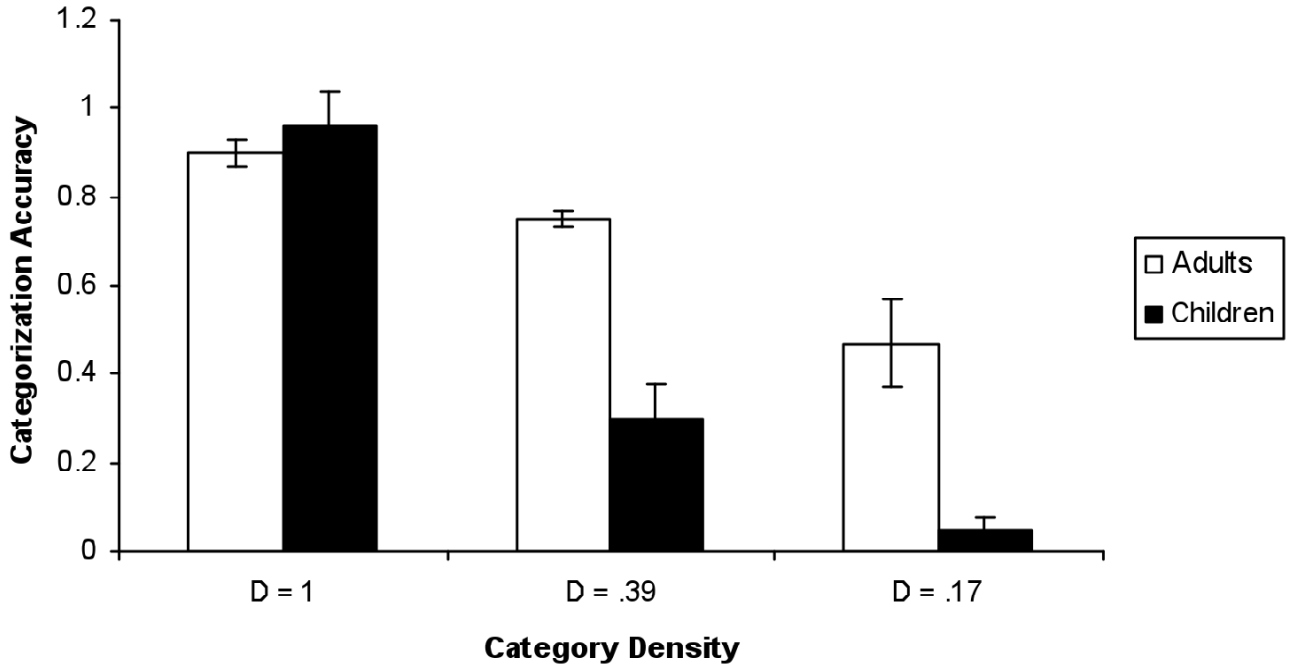

1.2
1
0.8
0.6
0.4
0.2
0
Categorization Accuracy
D = 1
D = .39
D = .17
Category Density
Adults
Children

Figure 3.

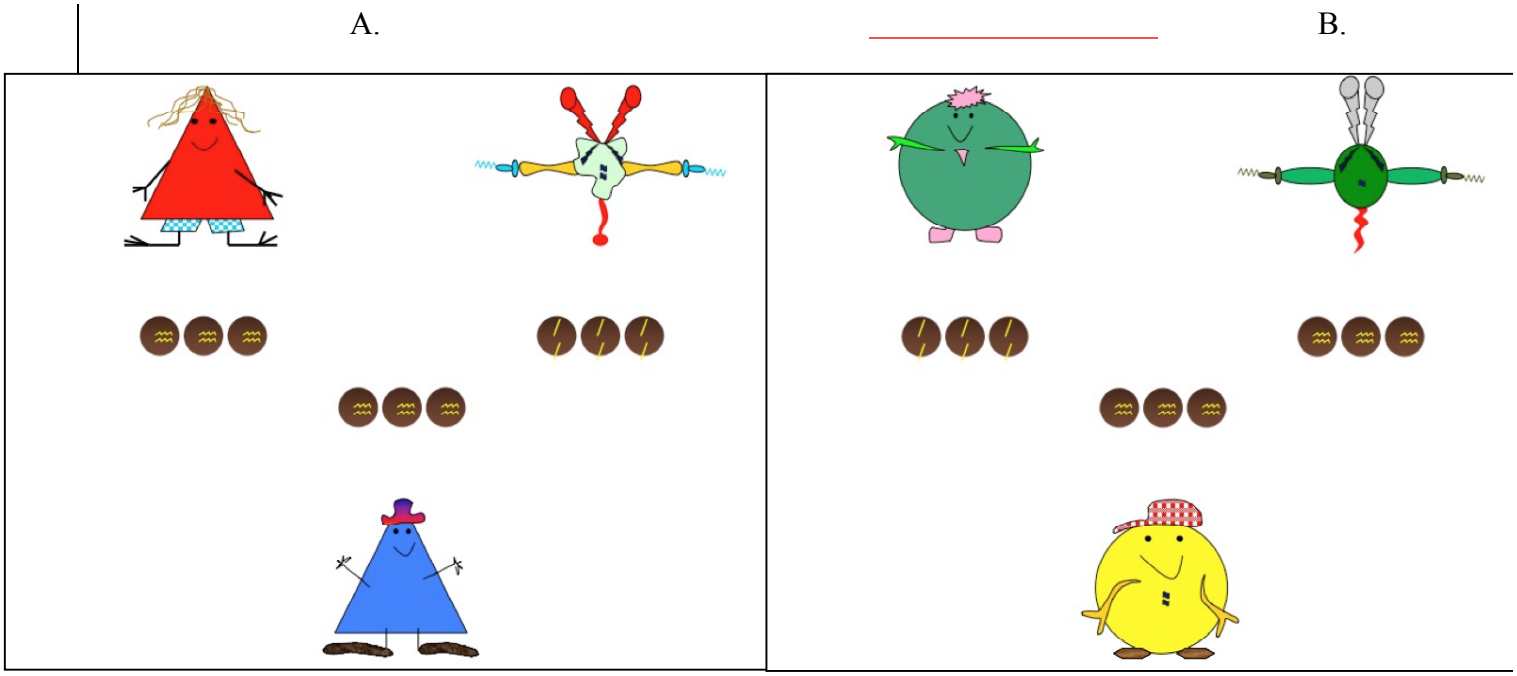
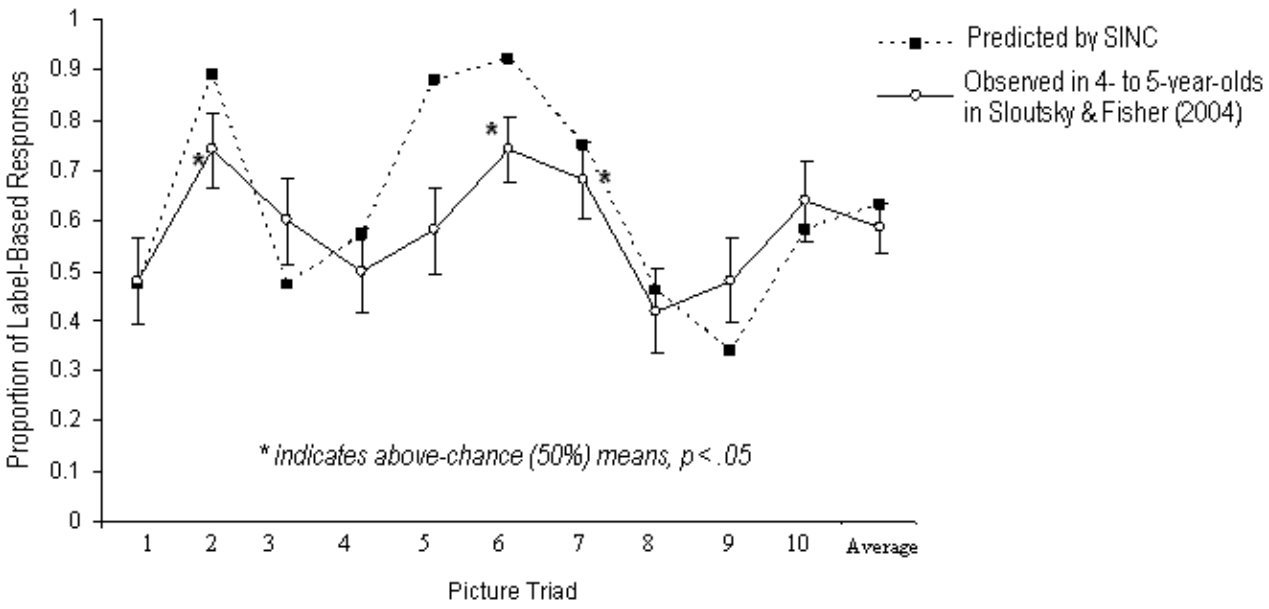


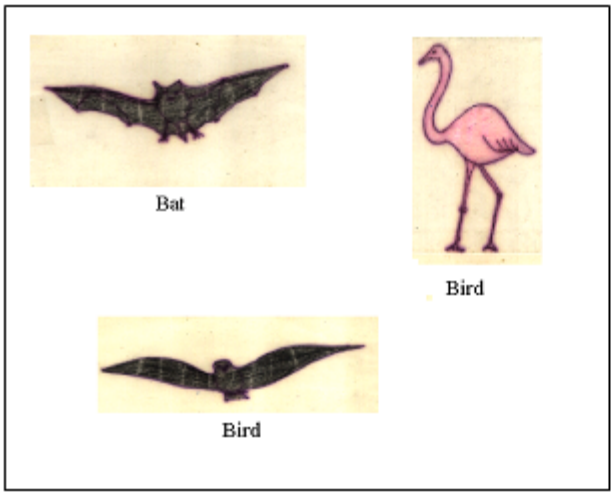
Figure 4.

Figure 5.

A.



B (Triad 1 on the graph above).



C (Triad 2 on the graph above).

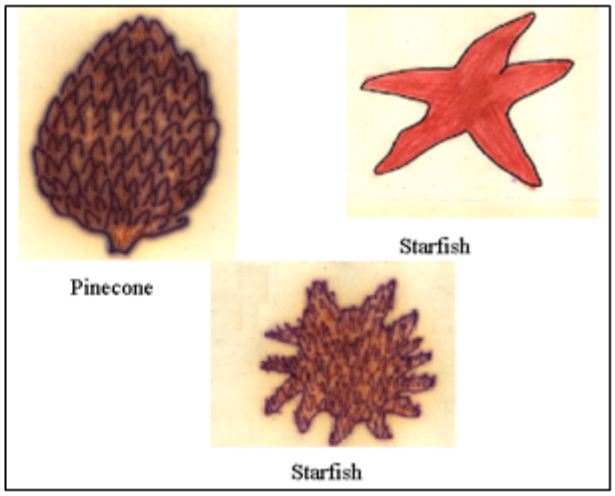
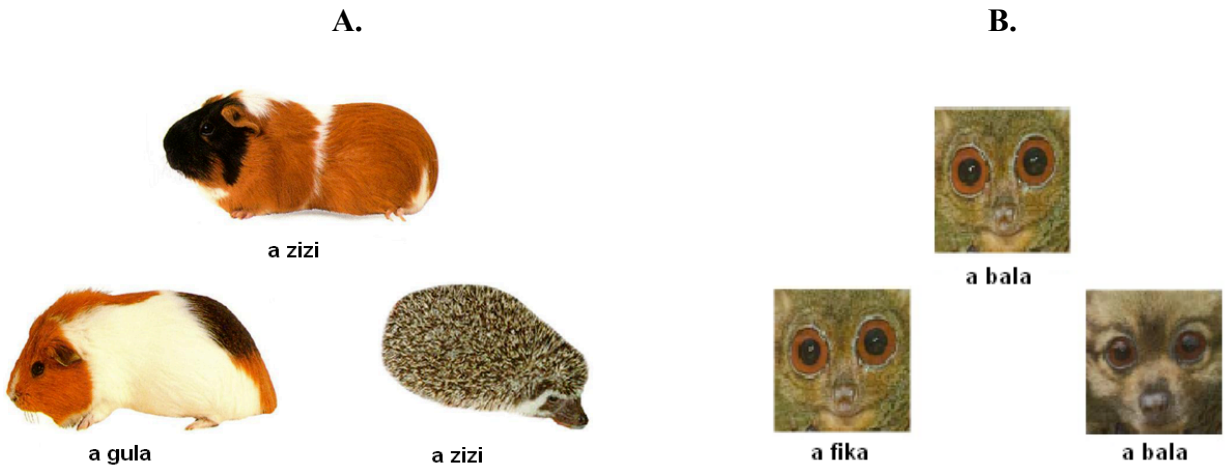



Figure 6.