The Role of Attention in the Development of Categorization

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Wei Deng, B.S., M.A.

Graduate Program in Psychology

The Ohio State University

2015

Dissertation Committee:

Vladimir M. Sloutsky, Ph.D., Advisor

John E. Opfer, Ph.D.

Stephen A. Petrill, Ph.D.
Abstract

Categorization is a critically important aspect of cognition: it enables recognition and differentiation of objects, people, and events, organizes our existing knowledge, and promotes generalization in new situations. Although this ability appears early in development and it has important ramifications for the acquisition of other cognitive capacities, many questions regarding the development of categorization remain unanswered. How do people learn and represent categories? How does the way categories are learned affect the ways categories are represented? And how do category representations change in the course of development and learning? The current project, consisting of six experiments, attempted to answer these questions by focusing on the role of attention in the development of category learning and category representation.

First, in Experiment 1, 4-year-olds, 6-year-olds, and adults were trained with either a classification task or an inference task and their categorization performance and memory for items were tested. Adults and 6-year-olds exhibited an important asymmetry: they relied on a single deterministic (D) feature during classification training, but not during inference training. In contrast, regardless of the learning regime, 4-year-olds relied on multiple probabilistic (P) features. Second, in Experiments 2-4, 4-year-olds and adults were trained with a classification task and their attention was attracted to the D feature (Experiment 3) or P features (Experiment 4). There was also a baseline (Experiment 2), where no attention manipulation was introduced. There was an important dissociation
between categorization and recognition memory. In terms of categorization responses, children and adults were responsive to manipulations. However, important differences transpired with respect to category representation, as evidenced by recognition memory. Adults, unless told otherwise in Experiment 4, tended to extract the D feature and form a rule-based representation. In contrast, regardless of their categorization performance, young children tended to encode multi-feature information and form a similarity-based representation. These results point to an important developmental difference in the pattern of attention: Whereas adults attend selectively to what they deem to be category-relevant, young children attend diffusely. Importantly, more efficient selective attention in adults was accompanied by worse memory of the to-be-ignored features than of the to-be-attended features, whereas less efficient diffused attention in children was accompanied by equally good memory of both to-be-attended and to-be-ignored features.

Finally, Experiments 5-6 further examined the role of attention in categorization by focusing on the role of words and dynamic visual features in category learning in 8- to 12-month infants. Infants were familiarized with exemplars from one category in a label-defined or motion-defined condition and then tested with prototypes from the studied category and from a novel contrast category. Eye tracking results indicated that infants exhibited better category learning in the motion-defined than in the label-defined condition and their attention was more distributed among different features when there was a dynamic visual feature compared to the label-defined condition. These results are discussed in relation to theories of categorization, the role of selective attention, and the role of linguistic labels in the development of category learning.
Dedicated to my grandfather.
Acknowledgments

First, I would like to thank my advisor Vladimir Sloutsky for his encouragement, guidance, and support over the last six years. He has provided the right balance of mentorship and freedom for my intellectual growth as a critical thinker and an independent researcher.

I would also like to thank John Opfer for challenging me to think deeply about the alternative explanations of my data, the limitations of my approaches, and the developmental story of my findings.

I would like to thank Steve Petrill for being on my committee and for his helpful comments and critics on this project.

I would like to thank Sam Rivera and Keith Apfelbaum for their support and helpful comments on this project. I would also like to thank Sravani Paladugu, Nancy Truong, Priya Patel, and the research assistants in the Cognitive Development Lab. Without their assistance, none of my work would have been realized. I would like to thank all my current and former research assistants who have helped me collect data over the years. Thank you to: Fatima Jawed, Kelsie Smith, Quan Zhou, Becky Plumage, Ashley Fuchs, Danitza Quipuscoa, and Kyle Mundy.

Special thanks to my graduate friends Kevin Darby, Allison O’Leary, Saebiyul Lee, Dan Kim, and Hyungwook Yim, for their support and for sharing many delightful and
insightful conversations over the years. An extra special thanks to Tracey Miser, for her assistance with data collection when she was our lab manager, for her discussions and comments on my work, and for the delicious maple syrup from her farm – all of which were critical to the success of this project. And an extra super special thanks to Xin Yao, for her guidance and help when I started my graduate life, for her continuous support during my graduate study and research, and for sharing great interests in food in the past years.

I would also like to thank my friends Hairong Gu, Yinghao Sun, Xiaoyue Xia, Tao Yang, Yang Liu, and Xin Jiang, for making my life in the States full of joy. Special thanks to Yun Tang, for her help and support from every angle of my graduate life, and for sharing delicious coffee and a taste of sweetness over the past years.

Finally, I would like to express my love and gratitude to my beloved parents, for their understanding, support, and endless love. Special thanks to my husband Kun Wang, for his unbounded support in my academic career, for his infinite love and care, and for his unique sense of humor as a mathematician. If it weren’t for their support, this dissertation may never have gotten finished.
Vita

2005..........................................................No.11 High School, Wuhan, China

2009..........................................................B.S. Psychology, Zhejiang University, Hangzhou, China

2011..........................................................M.A. Psychology, The Ohio State University, Columbus, OH

2011 to 2013 ...........................................Graduate Teaching Associate, Department of Psychology, The Ohio State University

2013 to present ...........................................Graduate Research Associate, Department of Psychology, The Ohio State University

Publications


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.


Fields of Study

Major Field: Psychology

Specialization: Developmental Psychology
Table of Contents

Abstract....................................................................................................................................... ii
Acknowledgments......................................................................................................................... v
Vita.................................................................................................................................................... vii
List of Tables ..................................................................................................................................... x
List of Figures ................................................................................................................................... xi
Chapter 1: Introduction.................................................................................................................... 1
Chapter 2: Overview of the Experiments......................................................................................... 25
Chapter 3: Effects of Classification and Inference Training on Categorization ....................... 30
Chapter 4: Flexibility of Categorization: Evidence from Recognition Memory ..................... 48
Chapter 5: The Role of Words in Early Category Learning ...................................................... 72
Chapter 6: General Discussion....................................................................................................... 86
References......................................................................................................................................... 99
Appendix: Tables and Figures ....................................................................................................... 107
List of Tables

Table 1. Category structure used in Experiments 1-4............................................. 108
Table 2. Training data: Mean proportion of correct responses aggregated in 10-trial blocks across age groups and training conditions in Experiment 1................................. 109
Table 3. Categorization at test: Mean proportions of responses based on old features in new-D, one-new-P, and all-new-P items in Experiment 1............................................. 110
Table 4. Memory at test: Mean proportions of yes responses on different item types in Experiment 1................................................................................................................. 111
Table 5. Training data: Mean proportion of correct responses aggregated in 10-trial blocks across age groups in Experiments 2, 3, and 4....................................................... 112
Table 6. Categorization at test: Mean proportions of responses based on old features in new-D, one-new-P, and all-new-P items in Experiments 2, 3, and 4 ......................... 113
Table 7. Memory at test: Mean proportions of “yes” responses on different item types in Experiments 2, 3, and 4............................................................................................................. 114
Table 8. Category structure used in Experiment 5................................................................................................................................. 115
List of Figures

Figure 1. Examples of stimuli used in Experiments 1-4 ........................................ 116
Figure 2. A schematic representation of the procedure. ........................................ 117
Figure 3. Categorization performance in Experiment 1 ........................................ 118
Figure 4. Recognition performance in Experiment 1 ........................................ 120
Figure 5. Categorization performance in Experiments 2-4 .................................... 122
Figure 6. Recognition performance in Experiments 2-4 .................................... 124
Figure 7. Example stimuli in Experiments 5-6 .................................................. 126
Figure 8. Average looking time at familiarization in Experiment 5 ....................... 128
Figure 9. Novelty preference scores in Experiment 5 ........................................ 129
Figure 10. Individual novelty preference scores in Experiment 5 .......................... 130
Figure 11. Proportion of accumulated looking time to feature at test in Experiment 5.. 131
Figure 12. Proportion of accumulated looking time to feature at familiarization in
Experiment 5 ........................................................................................................ 132
Figure 13. Average number of valid fixation shifts at familiarization in Experiment 5. 133
Figure 14. Average number of valid fixation shifts at each time point within
familiarization trials in Experiment 5 ................................................................. 134
Figure 15. Proportion of accumulated looking time to feature in Experiment 6 .......... 135
CHAPTER 1
INTRODUCTION

The ability to form categories, or equivalence classes, of discriminable entities is a central component of human cognition: categorization enables abstract thought and promotes expansion of knowledge to novel situations. For example, having observed that all 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 and category learning.

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, et al, 2004; Smith, et al, 2012). Second, there is evidence of remarkable development in the ability to form and represent categories (e.g., Younger & Cohen, 1986; Quinn & Johnson, 2000; Kloos & Sloutsky, 2008; L. Smith, 1989; see also Quinn, 2011; Sloutsky, 2010, for reviews). And third, 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 (Heit & Rubinstein, 1994; Gelman & Markman, 1986; Jones, Smith, & Landau, 1991; Macario, 1991; Bulloch & Opfer, 2009; Ross & Murphy, 1999; Sloutsky & Fisher, 2008). For example,
when categorizing toys, participants may rely on shape, but not color, whereas when categorizing foods, they may rely on color, but not shape (Macario, 1991).

It is hardly controversial that adults can acquire exceedingly abstract categories, whereas there is little evidence that infants or even young children (i.e., children younger than 6-years of age) can acquire categories of similar levels of abstraction. Although many agree that categorization does develop, there is less agreement as to what changes and why.

Possible answers to the what question range from (a) profound qualitative (often stage-like) changes in category representations, such as theory change (e.g., Carey, 1991; Inagaki & Hatano, 2002) or characteristic-to-defining shift (Keil & Batterman, 1984; Keil, 1992) to (b) relatively continuous representational change (e.g., Eimas, 1994). According to the shift view, immature representations are replaced by more mature representations, whereas according to the continuity view, the development consists of enrichment rather than replacement of immature representations.

Possible answers to the why question range from the acquisition of domain-specific (or even concept-specific) knowledge (e.g., Carey, 1991; Inagaki & Hatano, 2002; Keil & Batterman, 1984; Keil, 1992) to more domain-general explanations, such the development of selective attention enabling people to focus on relevant information (e.g., Sloutsky, 2010; Smith, 1989). In the former case, development is a function of knowledge acquisition: novices start with more characteristic representations, but shift to defining representations as more knowledge is acquired. In the latter case, development involves changes in basic cognitive processes.
The goal of this dissertation research is to better understand what changes in the course of development and why. In an attempt to answer these questions, I start with three lines of evidence on the role of attention in categorization: (1) effect of learning regime on category representation; (2) flexibility of categorization; and (3) the role of words in early category learning. Then I report three studies examining each issue respectively. I conclude by summarizing the key findings of the reported studies and discussing the implications for understanding the role of attention in the development of categorization, the role of words in categorization, as well as the theories of categorization and category learning.

1.1 Effect of Learning Regime on Category Representation

Adults’ representation of the same category structure may depend on the way the category is learned (Hoffman & Rehder, 2010; Love, Medin, & Gureckis, 2004; Sakamoto & Love, 2010; Yamauchi, Love, & Markman, 2002; see also Markman & Ross, 2003, for a review). Specifically, if they learn the category by classification (i.e., by predicting a label of each item) they tend to represent the structure in a more rule-based (or defining feature based) manner. At the same time, if they learn the category by inference (i.e., by predicting a missing feature of each item) they tend to represent the category in a more similarity-based (or characteristic features based) manner. Therefore, developmental change may not occur in a shift-like manner, with immature representations being replaced by more mature representations. Instead, the development may consist of acquiring the ability to form more defining feature based or rule-based
representations, and forming different category representations under different task conditions.

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

1.1.1 Asymmetry between Learning by Classification and by Inference

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

However, the ways people learn categories are not limited to classification learning. For example, in inference learning participants have to infer a missing feature on the basis of category label and other presented features. On each trial an item is presented and labeled, but one of the features is not revealed to the participant. The participant has to predict whether the non-revealed feature comes from features of category A or category B. In other words, people may learn categories by classification (i.e., by predicting a category label of an item) or by inference (i.e., by predicting an item’s features that are unknown, missing, or unobservable).
There are two lines of evidence that for adults classification and inference learning are not equivalent and they result in different representations. First, in order for classification and inference learning to be equivalent and result in equivalent representations, labels have to be equivalent to other features (see Yamauchi & Markman, 1998; Markman & Ross, 2003, for extensive arguments). This is because classification requires one to predict the category label when features are given, whereas inference requires one to predict a feature, when the label and the rest of the features are given. However, there is much evidence stemming from classification/inference judgment tasks (these tasks do not involve learning) that, at least for adults, labels are not equivalent to features (Yamauchi & Markman, 2000; see also Markman & Ross, 2003, for a review).

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

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

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

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

Do effects of learning regime on category representation emerge in the course of development or are they a developmental default? Two sources of evidence suggest that the former is the case: (1) developmental differences in how labels are treated and (2) developmental differences in selective attention.
First, there is a growing body of developmental work suggesting that, in contrast to the findings with adults, early in development labels may function as features. For example, it has been demonstrated that early in development, labels contribute to similarity of compared entities and the contribution is quantitative, feature-like (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004, 2007; Sloutsky & Fisher, 2004, 2012; Sloutsky & Lo, 1999; but see Waxman & Gelman, 2009 for a review of literature disputing the label-as feature view).

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

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

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

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

1.1.3 The Emergence of Representational Differences and Possible Mechanisms of Change

As discussed above, there is evidence that (1) early in development labels may function as features and (2) infants and young children tend to distribute rather than optimize attention. This evidence suggests that, in contrast to adults, early in development classification and inference learning are equivalent and may result in a similar representation of the learned category. These considerations lead to a number of important hypotheses.

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

1.2 Flexibility of Categorization

While patterns of categorization responses do exhibit differences across the situations, it is less clear whether patterns of representation also exhibit flexibility. In particular, participants can form different representations for different ways of categorizing items, such as having higher attentional weight for color, when foods are categorized. Or they form similar representations, but use different decision weights for different stimulus dimensions across different situations. Does representation of the categories learned from the same input differ across different situations? And, if yes, how does this representational flexibility develop?

1.2.1 Flexibility Starts Early: Evidence from Categorization, Category Learning, and Induction Tasks

As mentioned before, even early in development categorization exhibits remarkable flexibility. 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 an earlier mentioned 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 and similar findings could be interpreted as evidence of children’s reliance on deep conceptual knowledge – understanding of when different features matter (e.g., Booth & Waxman, 2002, 2006; Gelman & Markman, 1986, but see Jones, et al., 1991; Smith & Samuelson, 2006) – it does not have to be the case. There are several studies demonstrating that similar flexibility can be achieved when the deployment of conceptual information is highly unlikely.

For example, Sloutsky and Fisher (2008) conducted a study where children were first trained to make predictions about unambiguous triads. Each triad consisted of a set of items that varied on only one dimension, either color or shape. In one context shape was predictive (such that all items had the same color, but only two had the same shape, whereas the third had a different shape). In another context, color was predictive (such that all items had the same shape, but only two had the same color, whereas the third had a different color). After training participants were presented with ambiguous triads (such
that one test item matched the target on color and another one shape). Researchers found that children relied on shape in the former context and on color in the latter context (see Hayes & Lim, 2013, for replication). Given that the dimension-context correlations were ambiguous and given that children exhibited little awareness of the basis of their choices, flexibility does not necessarily indicate reliance on conceptual information.

Sloutsky and Fisher’s (2008) findings were recently extended to infants (Sloutsky & Robinson, 2013). These researchers demonstrated that even 14 month-olds can learn a color-based category when items were presented in one context and they learn a shape-based category when the items are presented in another context. Similarly, infants can categorize the same set of familiar stimuli in different ways (Ellis & Oakes, 2006; Mareschal & Tan, 2008).

Although these and similar 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 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.
If participants form different representations across the situations, then they should learn that in the context of the origin of a particular food, taxonomic kind is important. Therefore, in this context, bacon and chicken are alike as both are animal products and strawberries and potatoes are alike as both are plants. They also learn that in the context of when the food is eaten, bacon and strawberries are alike as both are breakfast foods and chicken and potatoes are alike as both are dinner foods. In contrast, if participants form the same representations, but use different dimensions when making decision, then they learn that chicken are both animal product and dinner food, but these dimensions play different roles in classification decision across the situations.

Although little is known which possibility is the case, there are several reasons to believe that when learning new categories, children represent all dimensions, whereas adults represent dimensions that are relevant for a given situation, which suggests that representations may differ for adults, but not for young children. I review this evidence in the next section.

1.2.2 The Role of Selective Attention in Category Representation: Possible Mechanisms

A number of models of categorization have proposed to account for how categories are represented (see Murphy, 2002, for a review). For most of these models, selective attention is an integral component. 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, that different stimulus dimensions have on a classification decision. 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 (Smith, Patalano, & Jonides, 1998).

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 stripes become more salient in the context of the latter. Selective attention is key to forming different representation and we are unaware of any model of categorization, in which a categorization decision is not driven by the underlying representation. Thus is it would be difficult for these models to capture a situation, in which all the dimensions of the category are represented with equal attentional weights, yet categorization decision are made on the basis of only one or few diagnostic dimensions.
This attentional selectivity has consequences: while shifting attention to the diagnostic dimension (i.e., presence or absence of stripes), participants learn to ignore non-diagnostic dimensions – the phenomenon known as learned inattention (see Hoffman & Rehder, 2010, for a review). If after learning the two categories, a learner embarks on a new categorization task – differentiating between squirrels and hamsters – the tail that was non-diagnostic for previous learning becomes diagnostic for current learning.

Many theories of categorization predict that, as a result of allocating attention selectively in the first task, participants may have difficulty shifting attention to a previously ignored dimension, and this prediction has been confirmed empirically. For example, Hoffman and Rehder (2010) presented adult participants 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. Using a combination of behavioral and eye tracking methodology the authors found that classification 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).

However, 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). Evidence that older children and adults are generally much better than younger children at selectively attending to one dimension
or property comes from a variety of domains including rule use (e.g. Frye, Zelazo, & Palfai, 1995), discrimination learning (e.g. Kendler & Kendler, 1962), speeded classification (e.g. Smith & Kemler, 1978), and category learning (Robinson et al, 2011; Best et al, 2013).

For example, in a recently published study, Best, Yim, and Sloutsky (2013) presented 6-8-month-old infants and adults with a two-phase category-learning task. In contrast to adults, who exhibited evidence of learned inattention in phase 2, 6-8-old infants did not exhibit evidence of learned inattention, despite the fact that they successfully learned the categories. There is also evidence for a lack of attention optimization in category learning of 4-5 year-olds (Robinson, Best, & Sloutsky, 2011; see also Kloos & Sloutsky, 2008, for related findings). These findings suggest that, given the same task, children and adults exhibit different patterns of attention: adults optimize attention by shifting it to category relevant information, whereas children attend diffusely. As a result, only adults, but not children exhibit evidence of learned inattention.

Therefore, if children and adults allocate attention differently in the course of category learning, they are also likely to form different representations: 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 children should be accompanied by similar representations (i.e., across the situations, both diagnostic and non-diagnostic dimensions should be represented).
If children, form equivalent representations across situations, how is flexible categorization achieved. One possibility is that participants represent the features equivalently, but put different decision weights on some features over others. For example, one could form a representation of squirrels that consists of body size, fur color, tail length, stripe pattern, and so on, but only use the tail when classifying squirrels versus hamsters and only use stripes when classifying squirrels versus chipmunks.

While different patterns of attention allocation across development may result in different category representations, in most developmental studies of categorization, participants’ category representations are inferred from category judgments. At the same time, as discussed before, these judgments can be driven by either representation or decision components. Therefore, it is necessary to access representation more directly to distinguish between representation and decision components. One way of accessing representations is to present participants with a memory task examining what they remember about the studied categories (see Kloos & Sloutsky, 2008, for a discussion). Furthermore, if the representation and decision components of categorization are dissociable, then certain attentional manipulation should change one component without change the other.

In sum, the evidence reviewed above suggests that a number of important hypotheses. Adults, who attend selectively, should optimize attention and extract features that are relevant for a given categorization task. In contrast, across situations, young children should attend diffusely and form equivalent similarity-based representations based on multiple features. In addition, the possibility of dissociation between the representation
and decision components of categorization suggests that we may change one component without changing the other.

1.3 The Role of Words in Categorization

As mentioned earlier, the ability to form categories appears early in development: infants exhibit evidence of category learning during the first months of life (Quinn, et al, 1993; Younger & Cohen, 1985). There is also evidence suggesting that language may affect this process, although the mechanisms underlying the effects of language remain a matter of debate.

1.3.1 Are Words Category-Markers or Part of the Input? Two Theoretical Accounts

Some suggested that words accompanying category members have the special status of category markers and, as such, they guide or supervise category learning in infancy (Waxman & Markow, 1995; see also Westermann & Mareschal, 2014). At the same time, others suggested that early in development words are akin to other features, but they may become category markers in the course of development (Gliozzi, et al., 2009; Sloutsky, 2010; Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Sloutsky, et al., 2001). As I discuss below, distinguishing between these positions has profound consequences for our understanding of the relationships between language and cognition and the nature of learning early in development.

According to the former theory, “infants embark on the task of word learning equipped with a broad, universally shared expectation, linking words to commonalities among objects” (Waxman, 2003, p.220). As a result, words, but not other kinds of auditory input, facilitate infants’ category learning by attracting attention to within-
category commonalities (Waxman & Markow, 1995; Waxman & Booth, 2001), thus effectively supervising category learning. These effects are supervisory because labels guide learning by attracting attention to commonalities.

There is some evidence consistent with this view. First, words may facilitate infants’ categorization above and beyond other kinds of auditory input (Balaban & Waxman, 1997; Fulkerson & Haaf, 2003; Ferry, et al., 2010). Second, facilitative effects of words were reported for basic-level as well as at superordinate or global levels (Balaban & Waxman, 1997; Waxman & Booth, 2003; Waxman & Markow, 1995). Third, there are reports that facilitative effects of labels are specific rather than general in nature: count nouns and adjectives have initially similar effects on category learning, whereas around 14-months of age count nouns are more likely to facilitate category learning than adjectives (Waxman & Booth, 2001). This finding suggests that count nouns may play a special role in category learning. And finally, labels may facilitate property induction above other kinds of input (Keates & Graham, 2008).

There are challenges, however, to the idea that words are category markers in infancy. First, even if words affect category learning in infancy, they do not have to function as category markers supervising learning, but can be instead part of the stimulus input and influence learning in a bottom-up fashion. For example, Plunkett et al (2008) presented 10-month-old infants with a category-learning task, such that the to-be-learned category consisted of two clusters of artificial creatures (i.e., a broad category somewhat analogous to a global category encompassing cats and horses). When the category was presented in silence, participants learned two narrow categories, whereas when one
common label accompanied each item, participants learned the single broad category. Although it is tempting to conclude that these results indicate that labels supervised category learning, this conclusion is unwarranted. Specifically, when Gliozzi et al (2009) modeled data reported by Plunkett et al (2008) using self-organizing maps, a model that assumed that labels are features and function as input rather than top-down supervisory signals was able to account for the reported pattern.

Second, findings that labels facilitate infant category learning are tenuous at best – facilitation transpires in some studies and does not transpire in others. This is because many studies compared the effects of labels with those of unfamiliar sounds, but not with a silent condition. When a silent baseline was introduced (e.g., Robinson & Sloutsky, 2007), labels were not found to facilitate infants’ category learning above the silent baseline.

And finally, even studies demonstrating facilitative effects of labels have generated inconsistent findings regarding the age at which labels facilitate category learning. For example, Booth and Waxman (2002) demonstrated that for an artifact category, words alone facilitated category learning only at 18-months of age, whereas around 14-months of age, words had to be paired with object function to facilitate learning above the baseline. It was argued that both words and functions (a) indicate human agency and (b) highlight commonalities among objects. Results of this study seem to be in a sharp contrast with studies where words were claimed to facilitate category learning in very young infants. In particular, Ferry, et al (2010) found evidence that labels facilitate category learning in 3- to 4-month-olds. Why do labels facilitate category learning in 3-4
months-olds, while failing to facilitate learning in much older infants (e.g., Booth & Waxman, 2002; Robinson & Sloutsky, 2007, 2008)?

According to the label-as-feature proposal, at least early in development, words are part of input – they are features of items rather than category markers (Gliozzi, et al, 2009; Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004; Sloutsky, et al., 2001). For example, there is evidence that early in development auditory input overshadows (or attenuates processing of) corresponding visual input (Lewkowicz, 1988a, 1988b; Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004a, 2004b, 2008; Sloutsky & Napolitano, 2003). Therefore, under most conditions words do not facilitate infant category learning, but, under some conditions, they may interfere with learning.

Of course, even if words start out as features, they may become category markers later in development (e.g., Deng & Sloutsky, 2012). For example, Deng and Sloutsky (2012) used a variant of Yamauchi and Markman’s (1998, 2000) category learning paradigm to teach preschoolers and adults two novel categories. During training, the category label correlated perfectly with a pattern of motion, which is a highly salient visual feature (Egeth & Yantis, 1997, for review), whereas the rest of the category features were probabilistic. At test, the label was pitted against the pattern of motion and participants had to predict one of the features of a test item. The researchers found that unlike many adults, who relied on a category label, children relied on the salient feature, despite the fact that their memory for the label was as good as that of adults. These results raised interesting questions. If a visual feature that is a part of input has a greater effect
on category learning than a label, what makes the label a category marker? And if words are not category markers for preschoolers, how can they be category markers for infants?

In sum, there are two theoretical positions with respect to the role of words in category learning, which differ with respect to the underlying mechanism and a developmental trajectory. Distinguishing between these possibilities and understanding the mechanisms underlying the effect of words on category learning is of critical importance for understanding cognitive development. If from early in development words function as category markers supervising learning, then top-down effects may play a significant role in early cognitive development. Also, given that supervision results in the ability to learn substantially more complex categories than unsupervised learning (Rumelhart, 1989), if words are supervisory signals for infants, our construal of what infants can and cannot learn will be subject to substantial revision.

1.3.2 Effects of Words and Other Features on Category Learning in Infancy: Potential Mechanisms

Although the reviewed evidence is controversial with respect to the idea that labels are category markers, none of the reviewed studies examined the mechanisms hypothesized by each theoretical position. At the same time, each position gives divergent predictions as to how words should affect category learning in comparison with other features. According to the label-as-category-marker position, attracting attention to commonalities is the mechanism via which labels facilitate category learning (e.g., Waxman, 2003). Therefore, when labels are presented, category learning should be accompanied by some form of attention optimization (Blair, et al. 2009; Hoffman &
— diffused attention early in learning and followed by shifting attention to within-category commonalities. Furthermore, attention optimization when labels are present should be greater than in no-label conditions, and this attention optimization should lead to better category learning.

In contrast, according to the label-as-feature position, (a) labels should not attract attention to commonalities and (b) due to possible auditory overshadowing effects, labels should either interfere with infant category learning or not affect it at all. Furthermore, there is evidence that successful category learning in 6-8-months-old infants is accompanied not by increased attention to commonalities (which is often the case in adults), but by broader exploration of features and thus more broadly distributed attention (e.g., Best, Yim, & Sloutsky, 2013). This prediction is based on a substantial body of literature demonstrating that distributed attention (with shorter individual fixations and frequent gaze shifts) is generally associated with better learning in infancy (e.g., Bronson, 1991; Colombo, et al., 1991; Jankowski, et al, 2001; Rose, et al., 2003). There is also evidence that distributed attention is not merely associated, but it leads to better learning. In one study (Jankowski, et al, 2001), researchers introduced a peripheral dynamic cue (a dynamic engager appearing in various parts of the screen) to induce shorter fixations, multiple gaze shifts, and thus a more distributed pattern of attention in 5-month-old infants. Results indicated that training resulted in more distributed attention (compared to no training) and in more efficient learning of presented items. This finding suggests that dynamic visual features (especially those that appear peripherally) may encourage distributed attention, thus facilitating learning. Given that successful category learning in
infancy is also accompanied by distributed attention (e.g., Best, et al., 2013), it is possible that such dynamic visual features may also affect infant category learning.

Why do dynamic visual features encourage distributed attention? In contrast to static visual features that maintain their salience throughout the trial, dynamic visual features become highly salient only during the period when they are dynamic. As a result, if a dynamic visual feature is presented peripherally, at the very minimum participants would need to move their gaze from the point of their initial fixation to the moving part and then, when the motion ceases to the point of their final fixation. Each of these gaze shifts may be accompanied by short fixations (and thus attention to) multiple regions of the stimulus (see Jankowski, et al., 2001, for a discussion).

Therefore, compared to labels, these dynamic visual features may result in more gaze shifts, more distributed attention, and better, more robust learning. This prediction contrasts sharply with that of the label-as-category-marker position, which predicts that labels facilitate category learning by attracting attention to commonalities. The goal of the reported study in Chapter 5 was to test this prediction by using a combination of eye tracking and a more traditional novelty preference paradigm.
CHAPTER 2

OVERVIEW OF THE EXPERIMENTS

The goal of this dissertation is to examine the role of attention in the development of categorization by focusing on three issues: (1) effects of classification and inference training on category representation (Chapter 3); (2) flexibility of categorization (Chapter 4); and (3) the role of words in early category learning (Chapter 5). To address these issues, six experiments were conducted.

Experiment 1 in Chapter 3 examined the effects of training regime on the outcome of category learning and on category representation. Specifically, 4-year-olds, 6-year-olds, and adults were trained on either a classification task (i.e., to predict the category of a given item) or an inference task (i.e., to predict a feature that the item had) and then tested with categorization and recognition tasks. The category structure included a single deterministic feature D (which perfectly distinguished between the two categories) and multiple probabilistic features P (with each providing imperfect probabilistic information about category membership), and both feature types were explicitly mentioned during training. Based on the considerations reviewed above, it was predicted that adults who were shown to optimize attention in classification, but not in inference training, should exhibit asymmetry. They should rely on the D feature in classification, but not in inference training. In addition, they should remember D features better than P features in
classification, but not in inference training. In contrast, 4-year-olds who shown to have difficulty optimizing attention, their categorization performance and recognition memory should be symmetrical across the training conditions. In both conditions, 4-year-olds should rely on multiple probabilistic features rather than on a single deterministic feature. Six-year-olds were included to provide a more detailed account of the developmental transition. In particular, these participants are older than those who typically exhibit difficulty focusing on a single relevant dimension in the presence of distracting dimensions (see, Hanania & Smith, 2010; Plude et al., 1994). Therefore, children of this age may have the capacity to rely on a single feature, which may transpire in Experiment 1.

Having established these developmental differences, it is reasonable to ask: What drives the development? As argued above, there are reasons to believe that these differences are driven by different patterns of attention allocated during category learning: young children attend diffusely regardless of the training condition, whereas older children and adults exhibit focused attention in Classification, but not in Inference training. Experiment 1 presented suggestive evidence supporting this possibility and the goal of Experiments 2-4 in Chapter 4 was to test it directly. Specifically, Experiments 2-4 were based on the following reasoning: there are some conditions under which young children may selectively attend to a single feature, although this selectivity is likely to be exogenous, or driven by characteristics of the stimuli, such as stimulus salience. For example, Deng and Sloutsky (2012) demonstrated that 4-year-olds categorized on the basis of a single salient feature (i.e., pattern of motion) rather than on a combination of
multiple probabilistic features and the label. Therefore, it is possible to affect young children’s attention exogenously.

Another important goal of Experiments 2-4 was to examine the development of category representation and representational flexibility. The basic task for each experiment consisted of three phases: instructions, training and testing. During training, participants (4-year-olds and adults) predicted the category of a given item (similar to the Classification training in Experiment 1) and received corrective feedback. The stimuli were identical to those used in Experiment 1 and the category structure included a deterministic feature D and multiple probabilistic features P. The testing phase consisted of categorization and recognition tasks and was administered immediately after the training phase. Three experiments differed in how participants’ attention was directed to different types of features. In Experiment 2, information about P and D features was not mentioned to participants. The goal of Experiment 2 was to examine the default patterns of categorization and category representation in children and adults. Based on consideration reviewed above, it was predicted that because young children do not optimize attention, their categorization performance and recognition memory should differ from adults’. Specifically, children should rely on multiple P features rather than the deterministic feature in categorization and remember all features equally well. In contrast, adults, who were able to optimize attention in category learning, should rely on the D feature and remember D feature better than P features. The goal of Experiments 3-4 was to examine if changes in categorization outcomes (compared to the default established in Experiment 2) would be accompanied by changes in underlying
representations. Specifically, we directed participants’ attention to the D feature in Experiment 3 and to the P features Experiment 4. 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 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 remain the same across the situations.

Experiments 5-6 in Chapter 5 aimed to further examine the role of attention in category learning in much younger participants, that is, 8- to 12-month-old infants. Specifically, two reported experiments were designed to examine the effects of labels and dynamic visual features on infants’ category learning by using a combination of eye tracking and a more traditional novelty preference paradigm.

In Experiment 5, infants were familiarized with exemplars from one category in a label-defined or motion-defined condition and then tested with prototypes from the studied category and from a novel contrast category. Infants saw the same testing stimuli in both conditions, with neither label nor motion being presented during testing. Experiment 6 was aimed at further comparing the effects of labels with a silent control condition. In both experiments, eye gaze data were collected. Recall, that, according to the label-as-category-marker position, labels should (a) facilitate category learning compared to the other conditions and (b) lead to greater attention to commonalities than in the other two conditions. In contrast, according to the label-as-feature view, (a)
because of auditory overshadowing, labels should not facilitate category learning and (b) more robust category learning may be accompanied by more distributed attention.

The reported results in the chapters followed were either presented as part of a published paper or have been submitted for publication. Specifically, the results of Experiment 1 were presented as part of a recently published paper (Deng & Sloutsky, 2015a) in Developmental Psychology. In addition, the results of Experiments 2-4, as part of a research manuscript (Deng & Sloutsky, in revision), have been submitted to a journal and the manuscript is currently under revision. Finally, the results of Experiments 5-6 were presented in a paper (Deng & Sloutsky, 2015b) that is currently in print in the Journal of Experimental Child Psychology.
CHAPTER 3

EFFECTS OF CLASSIFICATION AND INference training on
category representation

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

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

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

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

In contrast, adults who were shown to optimize attention in classification, but not in inference training, should exhibit asymmetry. They should rely on the D feature in classification, but not in inference training. In addition, they should remember D features better than P features in classification, but not in inference training. Six-year-olds were included to provide a more detailed account of the developmental transition. In particular, these participants are older than those who typically exhibit difficulty focusing on a single relevant dimension in the presence of distracting dimensions (see, Hanania & Smith, 2010; Plude et al., 1994). Therefore, children of this age may have the capacity to rely on a single feature, which may transpire in the current study.
EXPERIMENT 1

Method

Participants

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

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

Materials

Materials were similar to those used previously by Deng and Sloutsky (2012, 2013) and consisted of colorful drawings of artificial creatures. These creatures were accompanied by the novel labels *flurp* (Category F) and *jalet* (Category J). These categories had two prototypes (F0 and J0, respectively) that were distinct in the color and
shape of seven of their features: head, body, hands, feet, antennae, tail, and a body mark (see Figure 1).

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

As shown in Table 1, some of the items were used in training and some in testing. The training stimuli consisted of High-Match items (i.e., P_flurpD_flurp and P_jaletD_jalet). These items had the deterministic feature (D) and four probabilistic features (P) consistent with a given prototype; two other probabilistic features were consistent with the opposite prototype.

The testing stimuli consisted of High-Match items (i.e., P_flurpD_flurp and P_jaletD_jalet), Switch (or critical) items (i.e., P_jaletD_flurp and P_flurpD_jalet), and three additional item types. High-Match items were the items presented during training and they were highly similar to the prototypes of respective categories. Switch items had the D features of one category and most P features of another category, which made these items somewhat analogous to the real life categories of whales or dolphins (these have defining features of
mammals, but the majority of observable features of fish). The three additional item types included: (1) new-D items (i.e., $P_{\text{flurp}} D_{\text{new}}$ and $P_{\text{jalet}} D_{\text{new}}$), which had probabilistic features of the studied categories and a novel feature replacing the deterministic feature; (2) one-new-P items (i.e., $P_{\text{new}} D_{\text{flurp}}$ and $P_{\text{new}} D_{\text{jalet}}$), which had all features of the studied categories but a novel feature replacing one probabilistic feature; and (3) 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 the studied categories and all new features replacing the studied probabilistic features.

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

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

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

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

Design and Procedure

The experiment consisted of instructions, training and testing (see Figure 2). Training was a between-subjects factor, with participants being presented with either Classification or Inference training. Instructions and testing were identical for both training conditions.

The procedures were similar for both adults and children and for all age groups the experiment was presented on the computer and controlled by E-prime software (Version 2.0; Schneider, Eschman, & Zuccolotto, 2002). There were minor differences between children’s and adults’ procedures pertaining to the way the instructions were presented, the questions were asked, and the responses were recorded. Adults read the instructions and questions on the computer screen and pressed the keyboard to make responses, whereas for children, a trained experimenter presented instructions and the questions verbally and recorded children’s responses by pressing the keyboard. The experiment took approximately 10 minutes for adults and approximately 15 minutes for children. Most children and adults finished the experiment and, as evidenced by children’s high recognition accuracy (see below), their response patterns do not stem from confusion or fatigue.
Instructions and Training. In both training conditions, information about P and D features was explicitly given to participants before training. They were told that all flurps (or jalets) had a raindrop-shaped (or a cross-shaped) button and most of the flurps’ (or jalets’) features (at this point, the deterministic and probabilistic features were presented, one at a time). This information was repeated in the corrective feedback on each trial during training using the following script: *This one looks like a flurp (or a jalet) and it has the flurp’s (or the jalet’s) button.* Testing was not mentioned during the training phase. Participants were randomly assigned to one of the two training conditions.

The Classification and Inference training differed in the type of dimensions participants were asked to predict. In Classification training, participants predicted the category label of a stimulus given information about all other features. In Inference training, they predicted a missing feature of a stimulus, given information about the remaining features and the label. The missing feature was randomized across trials (i.e., on one trial the missing feature could be the head and on another trial it could be the feet), it was always one of the four probabilistic features, and the value of the feature was always consistent with the prototype of a given category. Therefore, on each trial, participants were shown an item with one feature covered, the deterministic feature and three probabilistic features from one category, and two probabilistic features from the contrast category. In both conditions, participants were given 30 training trials (15 trials per category) and each trial was accompanied by corrective feedback. The order of the training trials in both conditions was randomized across participants.
Testing. The testing phase was identical for both conditions; it was administered immediately after training and included categorization and recognition tasks. During the testing phase, 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 the 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. As we explain in the results section, the ways participants categorize and remember different item types provide critical information about what they attend to during category learning and thus which features are likely to be represented.

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

For categorization testing, the primary analyses focused on the proportion of responses in accordance with the D feature (i.e., rule-based responses). For recognition memory, the primary analyses focused on the difference between the proportion of hits (i.e., correctly identifying the High-Match items that were presented during training as old) and false alarms (i.e., incorrectly identifying other item types that were not presented during training as old).
If classification and inference learning result in different patterns of attention and in different category representations, then categorization and recognition performance should differ between the Classification and Inference training conditions. In particular, participants should rely on the deterministic feature when categorizing items in the Classification condition, while relying on multiple probabilistic features in the Inference condition. They should also remember the D feature better than P features in the Classification condition, but not in the Inference condition. However, if Classification and Inference training elicit similar patterns of attention and result in similar representations, participants should exhibit symmetric patterns of categorization and recognition performance in the two training conditions.

Based on previous results (e.g. Hoffman & Rehder, 2010; see also Markman & Ross, 2003, for a review), we expected adults to exhibit representational asymmetry between Classification and Inference training, which should transpire in both categorization and recognition performance. In particular, in the Classification condition adults should extract the most diagnostic (or rule) feature, whereas in the Inference condition they should extract within-category information (i.e., the overall similarity). At the same time, given the reviewed above evidence of diffused attention in younger children, we expected them to exhibit representational symmetry between the two training regimes. As a result, in both conditions, 4-year-olds should categorize on the basis of multiple features and should remember multiple features.
Results and Discussion

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

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

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

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

**Testing Phase: Categorization.** Categorization performance of each age group is presented in Figure 3 and Table 3. Preliminary analyzes focused on the ability to correctly categorize trained High-Match items (P₅₅urpD₅₅urp and P₅₅aletD₅₅alet), which was indicative of how well participants learned the categories (see Figure 3).

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

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

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

For 4-year-olds, there was only a significant main effect of trial type, $F(2,76) = 32.75, MSE = 0.66, p < .001, \eta^2 = 0.463$: regardless of the training condition categorization performance on new-D and one-new-P items was above chance ($ps < .001$) and at chance on all-new-P items ($ps > .413$). Recall that all-new-P items had the studied D features and all new P features, which revealed the inability of 4-year-olds to rely exclusively on a single deterministic feature. At the same time, 4-year-olds successfully relied on multiple features that were either all probabilistic (as in new-D items) or a combination of probabilistic and deterministic (as in one-new-P items).
For adults, there was a significant trial type by training condition interaction, $F(2,76) = 3.80$, $MSE = 0.12$, $p = .027$, $\eta^2 = 0.091$. Specifically, adults were able to correctly categorize all three types of items regardless of the training condition (above chance, $ps < .001$), but they exhibited better performance on all-new-P items (reliance on D features) in Classification condition than Inference condition ($p = .042$, Bonferroni adjusted).

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

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

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

The asymmetry between Classification and Inference training in adults is consistent with previous evidence (Yamauchi & Markman, 1998; Hoffman & Rehder, 2010) suggesting differences in representations formed as a result of classification and inference training. Similar to previous findings, adults tended to process and represent categorical information differently, with classification learners being more likely than inference learners to focus on the D feature, which separated the two categories.

At the same time, regardless of the training condition, 4-year-olds relied on the P features. The symmetric performance in 4-year-olds is a novel finding suggesting that unlike adults, they formed similarity-based representation of categories in both
conditions. Finally, equivalent performance in 6-year-olds and adults on switch items coupled with performance differences on other test items suggests that this is a transitional group.

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

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

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

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

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

For 6-year-old children, there was a significant interaction between feature type and training type, \( F (1,38) = 8.17, MSE = 0.41, p = .007, \eta^2 = 0.177 \). Specifically, similar to adults, 6-year-olds exhibited better memory for the D feature than for any single P feature in the Classification condition, paired-samples \( t (19) = 2.99, p = .008, d = 0.72 \), but not in the Inference condition, paired-samples \( t (19) = 0.67, p = .511 \).
Therefore, recognition memory accuracy corroborates findings stemming from categorization performance: Whereas adults and 6-year-olds exhibited attentional (and potentially representational) asymmetry between the Classification and Inference conditions, 4-year-old children attended equivalently in both conditions, and were likely to form similar representations in both conditions.

Overall, the reported results revealed different patterns of representation between adults and 6-year-olds versus 4-year-olds. Specifically, after Classification training, adults and 6-year-olds were more likely to extract the deterministic features than after Inference training. In contrast, 4-year-olds performed symmetrically exhibiting similarity-based representation, regardless of the training condition. Results from 6-year-olds suggest that the developmental change in category representations may begin to occur between 4- and 6-years of age and continue after 6-years of age. These are novel findings pointing to important developmental differences in categorization: young children initially tend to form similarity-based representations, but in the course of development they acquire the ability to form more rule-based representation, without losing the former ability.

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

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

FLEXIBILITY OF CATEGORIZATION: EVIDENCE FROM RECOGNITION MEMORY

The goal of Experiments 2-4 was to examine the development of category representation and representational flexibility. The reported study consisted of three experiments. The basic task for each experiment consisted of three phases: instructions, training and testing. During training, participants (4-year-olds and adults) predicted the category of a given item and received corrective feedback.

There were two family-resemblance categories, with each training item including a single deterministic feature D (which perfectly distinguished between the two categories) and multiple probabilistic features P (with each providing imperfect probabilistic information about category membership). 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 the item was more likely to belong to and (2) whether each item was old or new. No feedback was provided during testing. Both categorization and recognition memory items were created in such a way that participants’ responses would be informative of how participants represent categories. Categorization trials were designed to determine which features participants rely on in their decisions whereas recognition trials were designed to determine what
participants remember from training, which implies 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 2, information about P and D features was not mentioned to participants. The goal of Experiment 2 was to examine the default patterns of categorization and category representation in children and adults. Based on consideration reviewed above, it was predicted that because young children do not optimize attention, their categorization performance and recognition memory should differ from adults’. Specifically, children should rely on multiple P features rather than the D feature in categorization and remember all features equally well. In contrast, adults, who were able to optimize attention in category learning, should rely on the D feature and remember D feature better than P features.

The goal of Experiment 3 and 4 was to examine if changes in categorization outcomes (compared to the default established in Experiment 2) would be accompanied by changes in underlying representations. Specifically, we directed participants’ attention to the D in Experiment 3 and to the P features Experiment 4. 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 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 remain the same across the situations.
EXPERIMENT 2

Method

Participants

Participants were adults (12 women) and 4-year-old children ($M = 53.3$ months, range 48.8 – 59.9 months; 8 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 a female experimenter in a quiet room in their preschool.

Materials

Materials were similar to those used in Experiment 1 and consisted of colorful drawings of artificial creatures. These creatures were accompanied by the novel labels “flurp” (Category F) and “jalet” (Category J). These categories had two prototypes (F0 and J0, respectively) that were distinct in the color and shape of seven of their features: head, body, hands, feet, antennae, tail, and a body mark (see Figure 1).

As shown in Table 1, most of the features were probabilistic and they jointly reflect the overall similarity among the exemplars (we refer to them as the “P features” or as “overall appearance”), whereas one feature was deterministic and it perfectly distinguished the two categories (we refer to 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 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., $P_{\text{flurp}}D_{\text{flurp}}$ and $P_{\text{jalet}}D_{\text{jalet}}$). These items had the deterministic feature (D) and four probabilistic features (P) consistent with a given prototype; two other probabilistic features were consistent with the opposite prototype.

The testing stimuli consisted of High-Match items presented during training (i.e., $P_{\text{flurp}}D_{\text{flurp}}$ and $P_{\text{jalet}}D_{\text{jalet}}$) and four additional types of items. These included: (1) Switch items or critical lures (i.e., $P_{\text{jalet}}D_{\text{flurp}}$ and $P_{\text{flurp}}D_{\text{jalet}}$), 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 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 Figure 1 shows examples of each kind of stimulus.

Design and Procedure

The experiment consisted of instructions, training and testing (see Figure 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 minutes for adults and approximately 15 minutes for
children. Most children and adults finished the experiment and, as evidenced from children’s high recognition accuracy (see below), their response patterns do not stem from confusion or fatigue.

**Instructions and Training.** Before training, participants read a cover story about two categories of creatures from other planets. Information about P and D features was not mentioned to the participants. 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. 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 the 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. 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 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 different patterns of attention result in different category representations, then categorization and recognition performance should differ when attention is allocated differently. In particular, if participants selectively attend to the most diagnostic feature distinguishing two categories, they should rely on the deterministic 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 on recognition performance between D and P features.

Based on previous results (e.g., Hoffman & Rehder, 2010; see also Markman & Ross, 2003, for a review), we expected adults to exhibit attention optimization during category learning (with attention shifting to the deterministic features). As a result, they should extract the most diagnostic (or rule) feature and form a rule-based representation. At the same time, given the reviewed above evidence of diffused attention in young children, we
expected them to categorize on the basis of multiple features and remember multiple features.

Results and Discussion

Analyses below focused on categorization performance during training and testing and on recognition memory during testing.

Training Phase. Training data aggregated into 10-trial blocks across age groups are presented in Table 5.

Overall, children and adults exhibited above-chance training accuracy in the last ten training trials but children’s performance was somewhat lower than adults’: 68.5% in children and 81.5% in adults, $ps < .001$. A 3 (Training Block: 1 vs. 2 vs. 3) by 2 (Age Group: Children vs. Adults) mixed ANOVA revealed a main effect of age, $F (1,38) = 10.29, MSE = 0.69, p = .003, \eta^2 = 0.213$, with adults being more accurate than children. There was also a main effect of training block, $F (2,76) = 11.62, MSE = 0.28, p < .001, \eta^2 = 0.234$, with the accuracy increasing during the course of training for both age groups.

Testing Phase: Categorization. Categorization performance of each age group is presented in Figure 5A and Table 6. Preliminary analyzes focused on the ability to correctly categorize trained High-Match items ($P_{flurpD_{flurp}}$ and $P_{jaletD_{jalet}}$), which was indicative of how well participants learned the categories. As shown in Figure 3, both age groups accurately categorized these test items: adults 82.5% and children 70.6%, both above chance, $ps < .001$.

The second set of preliminary analyses focused on the ability to rely on familiar (i.e., seen during training) features when categorizing new-D, one-new-P, and all-new-P items.
The mean proportions of reliance on old features when categorizing these items are presented in Table 6.

These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 2 (Age Group: Children vs. Adults) mixed ANOVA. There was a significant Trial Type by Age Group interaction, $F (2,76) = 12.62, MSE = 0.55, p < .001, \eta^2 = 0.249$. We further broke down the interaction by conducting a repeated measures ANOVA on Trial Type for each age group.

For adults, there was also a significant main effect of trial type, $F (2,38) = 11.98, MSE = 0.63, p < .001, \eta^2 = 0.387$. However, their pattern of categorization performance on these three trial types was different from children’s. Specifically, adults were able to correctly categorize one-new-P and all-new-P items (81% and 79% respectively, above chance, $ps < .001$), but their categorization performance on new-D items was at chance, (49%, $p = .904$). Therefore, adults clearly relied on deterministic features when categorizing items.

For young children, there was also a significant main effect of trial type, $F (2,38) = 6.47, MSE = 0.22, p = .004, \eta^2 = 0.254$, with the categorization performance on new-D and one-new-P items being above chance (63% and 65% respectively, $ps < .005$) and at chance on all-new-P items (46% $p = .379$). Recall that all-new-P items had the trained D features and all new P features, which revealed the inability of young children to rely exclusively on D features. At the same time, children ably relied on probabilistic features.

Overall, the preliminary analyses indicated that adults exhibited rule-based generalization. They relied on the D feature (when presented with all-new-P items) and...
generalized broadly to items that had an old D feature and all new P feature, whereas they
did not generalize when D features were absent, even when an item had all studied P
features. In contrast, young children exhibited similarity-based generalization: they
generalized successfully when multiple studied P features were present, but failed to
generalize when only studied D features were present.

The primary analyses focused on comparison of categorization of the trained High-
Match items (P\textsubscript{flurp}D\textsubscript{flurp} and P\textsubscript{jalet}D\textsubscript{jalet}) and Switch items (i.e., P\textsubscript{flurp}D\textsubscript{jalet} and P\textsubscript{jalet}D\textsubscript{flurp}
items) between the age groups (see Figure 5A). In the High-Match items both D and P
features come from the same prototype; therefore either rule-based or similarity-based
generalization will result in higher proportions of correct generalizations. In contrast, the
Switch items have D features from one category and P features from another category;
therefore these items allow an unambiguous determination of whether items are
categorized in accordance with D or with P features. Based on preliminary analyses, it
was expected that adults and children exhibit similarly high proportions of correct
categorization for the High-Match items, whereas differences will transpire for the
Switch items. Specifically, it was expected that adults will categorize the Switch items on
the basis of D features, whereas children will categorize on the basis of P features.

Data in Figure 5A were analyzed with a 2 (Trial Type: High-Match vs. Critical Lures)
by 2 (Age Group: Children vs. Adults) mixed ANOVA. As predicted, there was a
significant Trial Type by Age Group interaction, $F (1,38) = 12.01, MSE = 0.57, p = .001,$
$\eta^2 = 0.240$. Specifically, adults exhibited equivalent performance on High-Match items
and Switch items, a paired-samples $t$ test indicated that there was no difference between
two trial types, (83% vs. 84%, $p = .807$. And more importantly, they relied on the D feature to categorize the Switch items, exhibiting primarily rule-based generalization, 84%, above chance, one-sample $t (19) = 6.11, p < .001, d = 1.37$. In contrast, young children exhibited similarity-based generalization, relying on P features to categorize the items. As a result, their proportion of rule-based responses was at 38.1%, below chance, one-sample $t (19) = 2.50, p = .022, d = .56$.

Results of categorization performance indicated that adults tended to spontaneously detect a defining feature and consistently rely on it to categorize items, whereas there was little evidence that young children relied on the deterministic feature in categorization. In contrast to adults, children tended to rely on a pattern of correlated probabilistic features, which reflect overall similarity.

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

*Testing Phase: Recognition Memory.* The proportions of “old” responses on different item types are presented in Table 7 (“old” in response to a High-Match item is a hit, whereas in response to the other items types it is a false alarm). As shown in the table, participants readily distinguished the studied High-Match items from and all-new-P
items: The difference between Hits and False alarms was 0.675 for adults and 0.581 for children, both greater than the chance level of 0, $ps < .001$.

Memory accuracy for the category-inclusion rule (i.e., D feature) and for the overall appearance (i.e., P features) was compared for each age group. Memory accuracy for the rule was obtained by subtracting false alarms on new-D items from hits on High-Match items and memory accuracy for appearance features was obtained by subtracting false alarms on one-new-P items from hits on High-Match items. The main results are presented in Figure 6A and data in the figure indicate that memory accuracy for D and P features was above chance level of 0 for both age groups, $ps < .001$.

Data in Figure 6A were submitted to a 2 (Feature Type: D vs. P) by 2 (Age Group: Children vs. Adults) mixed ANOVA, with feature type as a within-subjects factor and age group as a between-subjects factor. Results revealed a marginally significant interaction, $F(1,38) = 3.75, MSE = 0.18, p = .06, \eta^2 = 0.090$. Specifically, adults exhibited somewhat better memory accuracy for the D feature than for any single P feature (0.563 vs. 0.381), paired-samples $t(19) = 2.05, p = .055, d = .57$, whereas young children exhibited equivalent memory accuracy for any single P feature and for the D feature (0.375 vs. 0.381), paired-samples $t(19) = .16, p = .874$.

Therefore, recognition memory accuracy corroborates findings stemming from categorization performance: Whereas adults exhibited better memory for deterministic feature and consistently relied on it in categorization, young children equally represented all features and relied on the multiple probabilistic features in categorization.
Overall, the reported results revealed different patterns of representation between adults and younger children, pointing to important developmental differences in categorization. Specifically, adults were more likely to extract the deterministic features and formed a rule-based representation. In contrast, young children were more likely to encode multi-feature information and formed a similarity-based representation. As argued above, these developmental differences could be driven by different patterns of attention allocated in the course of category learning. Whereas adults exhibit selective attention, young children attend diffusely. Experiment 2 presents suggestive evidence supporting this possibility and the goal of Experiment 3 and 4 is 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 exogenous, or driven by characteristics of the stimuli, such as stimulus salience. For example, Deng and Sloutsky (2012) demonstrated that 4-year-olds categorized on the basis of a single salient feature (i.e., pattern of motion) rather than on a combination of multiple probabilistic features and the label.

Therefore, it is possible to affect young children’s attention exogenously and we attempt to do that in Experiment 3 and 4 in a more subtle way, without changing the stimuli used in Experiment 2. To achieve this goal, we attempted to direct participants’ attention to the D feature by mentioning this feature on each training trial in Experiment 3 and to the P features by mentioning overall appearance on each training trial in Experiment 4.
EXPERIMENT 3

Method

Participants

Participants were adults (6 women) and 4-year-old children ($M = 55.3$ months, range 48.3 – 62.7 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 a female experimenter in a quiet room in their preschool.

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 didn’t point out the feature type, we directed participants’ attention to the D feature in this experiment. 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: *It has flurp’s (or Jalet’s) button*. The testing phase (both categorization and recognition trials) was identical to Experiment 2 and it was not mentioned during the training phase.

Results and Discussion

Both preliminary and primary analyses were similar to Experiment 2.
Training Phase. Training data aggregated into 10-trial blocks across age groups are presented in Table 5. Overall, children and adults exhibited high training accuracy in the last ten training trials but children’s performance was somewhat lower than adults’: 87.5% in children and 99.5% in adults, both above chance, ps < .001. A 3 (Training Block: 1 vs. 2 vs. 3) by 2 (Age Group: Children vs. Adults) mixed ANOVA revealed a significant interaction, $F(2,76) = 10.18$, $MSE = 0.07$, $p < .001$, $\eta^2 = 0.211$. Specifically, adults were at ceiling throughout the training phase ($p = .167$), whereas children’s training accuracy increased through training blocks ($p < .001$).

Testing Phase: Categorization. Categorization performance of each age group is presented in Figure 5B and Table 6. Preliminary analyzes focused on the ability to correctly categorize 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 Figure 5B, both age groups accurately categorized these test items: adults 98.8% and children 77.5%, both above chance, ps < .001.

The second set of preliminary analyses focused on the ability to rely on familiar (i.e., seen during training) features when categorizing new-D, one-new-P, and all-new-P items. The mean proportions of reliance on old features when categorizing these items are presented in Table 6.

These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 2 (Age Group: Children vs. Adults) mixed ANOVA. There was a significant Trial Type by Age Group interaction, $F(2,76) = 17.04$, $MSE = 0.76$, $p < .001$, $\eta^2 = 0.310$. We
further broke down the interaction by conducting a repeated measures ANOVA on Trial Type for each age group.

For young children, there was a significant main effect of trial type, $F(2,38) = 3.76$, $MSE = 0.21$, $p = .032$, $\eta^2 = 0.165$, with the categorization performance on one-new-P and all-new-P items being 81% and 80% respectively (both above chance, $ps < .001$) and at chance on new-D items ($p = .094$). This pattern was different from that of child participants but similar to adults in Experiment 2.

For adults, there was also a significant main effect of trial type, $F(2,38) = 81.09$, $MSE = 2.84$, $p < .001$, $\eta^2 = 0.810$, and their pattern of categorization performance on these three trial types was similar to children’s. 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, $p = .036$.

Several important findings stem from the preliminary results. Overall, adults exhibited the same pattern of responding in Experiments 2 and 3: whether attention of adults was attracted to the category rule in Experiment 3 or was not in Experiment 2, they exhibit the same pattern of rule based categorization. In contrast, young children exhibited marked differences across the experiments: they relied on the overall similarity in Experiment 2 (when attention was not attracted to the category rule) and on the category rule in Experiment 3 (when attention was attracted to it).

The primary analyses focused on comparison of categorization of the trained High-Match items ($P_{\text{flurp}}D_{\text{flurp}}$ and $P_{\text{jalet}}D_{\text{jalet}}$) and critical lures (i.e., $P_{\text{flurp}}D_{\text{jalet}}$ and $P_{\text{jalet}}D_{\text{flurp}}$ items) between the age groups (see Figure 5B). Recall that in this experiment, we
attempted to direct participants’ attention to the D feature by mentioning this feature on each training trial. If the manipulation is successful as suggested by the preliminary analyses, children should become similar to adults and rely on the D feature in their categorization.

Results presented in Figure 5B indicated that children pattern of categorization was similar that of adults: their categorization was rule-based, which was in sharp contrast to similarity-based categorization exhibited by children in Experiment 2. A 2 (Item Type: High-Match vs. Switch) by 2 (Age Group: Children vs. Adults) mixed ANOVA confirmed these findings. Neither interaction nor main effect of trial type was significant (ps > .824). There was a main effect of age, $F(1,38) = 17.41, MSE = 0.93, p < .001, \eta^2 = 0.314$, with adults making higher proportion of rule-based responses than children. However, both adults and children relied on the D feature to categorize the switch items, with the proportion of rule-based responses being 99% and 77% respectively, above chance, ps < .001, $d > 1.20$.

**Testing Phase: Recognition Memory.** The proportions of “old” responses on different items types are presented in Table 7. Overall, participants readily distinguished the studied High-Match items from all-new-P items. The difference between Hits and False alarms was 0.725 for adults and 0.556 for children, both greater than the chance level of 0, ps < .001.

Similar to Experiment 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 Figure 6B. As shown in the figure, memory accuracy for D and P features was above chance level of 0 for both age groups, ps < .001.

Data in Figure 6B were submitted to a 2 (Feature Type: D vs. P) by 2 (Age Group: Children 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, $F (1,38) = 7.50, MSE = 0.49, p = .009, \eta^2 = 0.165$. Specifically, adults exhibited better memory accuracy for the D feature than for any single P feature (0.669 vs. 0.344), paired-samples $t (19) = 3.92, p = .001, d = 1.28$, whereas young children exhibited equivalent memory accuracy for any single P feature and for the D feature (0.456 vs. 0.469), paired-samples $t (19) = .16, p = .875$.

Overall, results of Experiment 3 substantially expand results of Experiment 2. When participants’ attention was focused on the D feature in Experiment 3, young children’s categorization performance was similar to that of adults, with both age groups exhibiting rule-based categorization. At the same time memory pattern remained the same as that of the child participants in Experiment 2: in both experiments, children exhibited equivalently high memory accuracy for both D and P features.

These findings suggest that in Experiment 3, children continued to distribute their attention across multiple features even though their categorization decision was based on the rule feature. In Experiment 4, we attempted to change adults’ pattern of categorization by directing participants’ attention to the P features.
EXPERIMENT 4

Method

Participants

Participants were adults (10 women) and 4-year-old children \( M = 55.7 \) 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 a female experimenter in a quiet room in their preschool.

Materials, Design, and Procedure

The materials, design, and procedure were similar to those in Experiment 3, with one critical difference. In contrast to Experiment 3 where we directed participants’ attention to the D feature, in Experiment 4 we directed their attention to the P features. In particular, participants were told that most of the flurps’ (or jalets’) features (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: *This one looks like a flurp (or jalet).* The testing phase (both categorization and recognition trials) was identical to Experiment 3 and it was not mentioned during the training phase.

Results and Discussion

Both preliminary and primary analyses were similar to Experiments 2 and 3.

Training Phase. Training data aggregated into 10-trial blocks across age groups are presented in Table 5. Overall, children and adults exhibited above-chance training
accuracy in the last ten training trials but children’s performance was somewhat lower than adults’: 72.0% in children and 75.5% in adults, \( ps < .001 \). A 3 (Training Block: 1 vs. 2 vs. 3) by 2 (Age Group: Children vs. Adults) mixed ANOVA revealed a main effect of age, \( F (1,38) = 4.68, MSE = 0.18, p = .037, \eta^2 = 0.110 \), with adults being more accurate than children. There was also a main effect of training block, \( F (2,76) = 3.30, MSE = 0.10, p = .042, \eta^2 = 0.080 \), with the accuracy increasing during the course of training for both age groups.

Testing Phase: Categorization. Categorization performance of each age group is presented in Figure 5C and Table 6. Preliminary analyzes focused on how well participants learned the categories. As shown in Figure 5C, both age groups accurately categorized the trained High-Match items (PflurpDflurp and PjaletDjalet): adults 70.0% and children 78.1%, both above chance, \( ps < .001 \).

The second set of preliminary analyses focused on the ability to rely on familiar (i.e., seen during training) features when categorizing new-D, one-new-P, and all-new-P items. The mean proportions of reliance on old features when categorizing these items are presented in Table 6.

These data were analyzed with a 3 (Trial Type: new-D vs. one-new-P vs. all-new-P) by 2 (Age Group: Children vs. Adults) mixed ANOVA. There was a significant Trial Type by Age Group interaction, \( F (2,76) = 6.28, MSE = 0.16, p = .003, \eta^2 = 0.142 \). We further broke down the interaction by conducting a repeated measures ANOVA on Trial Type for each age group.
For children, the main effect of trial type was marginally significant, $F(2,38) = 2.77$, $MSE = 0.91$, $p = .075$, $\eta^2 = 0.127$. Similar to the child participants in Experiment 2, children in Experiment 4 correctly categorized one-new-P and new-D items (above chance, $ps < .026$) but their performance on all-new-P items was at chance at chance, $p = .494$. Therefore, in contrast to Experiment 3, children 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 2 and 3, 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 2 and 3. Specifically, adults were able to correctly categorize one-new-P and new-D items (above chance, $ps < .001$), 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 child participants in this experiment as well as in Experiment 2.

The primary analyses focused on comparison of categorization of the trained High-Match items ($P_{flurp}D_{flurp}$ and $P_{jale}D_{jale}$) and switch items (i.e., $P_{flurp}D_{jalet}$ and $P_{jale}D_{flurp}$ items) between the age groups (see Figure 5C). 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 should become similar to children and rely on the overall similarity to categorize switch items.

Results presented in Figure 5C indicated that in this experiment, adults exhibited the same pattern of similarity-based categorization as children. A 2 (Item Type: High-Match vs. Critical Lures) by 2 (Age Group: Children vs. Adults) mixed ANOVA confirmed these findings. There was a main effect of trial type, $F(1,38) = 74.03, MSE = 3.77, p < .001, \eta^2 = 0.661$, with the proportion of rule-based responses on critical lures being significantly lower than that on High-Match items. Neither the interaction ($p = .951$) nor the main effect of age ($p = .060$) were significant. A further one-sample $t$ test indicated that both adults and children relied on the P features to categorize the switch items (27% and 34% respectively), with the proportion of rule-based responses being below chance, $ps < .003, ds > 0.80$.

**Testing Phase: Recognition Memory.** The proportions of “old” responses on different items types are presented in Table 7. Overall, participants readily distinguished the studied High-Match items from all-new-P items. The difference between Hits and False alarms was 0.813 for adults and 0.594 for children, both greater than the chance level of 0, $ps < .001$.

Similar to Experiments 2 and 3, participants’ memory accuracy for the category-inclusion rule (i.e., D feature) and for the overall appearance (i.e., P features) was compared and these data were presented in Figure 6C. As shown in the figure, memory accuracy for D and P features was above chance level of 0 for both age groups, $ps < .001$. 
Data in Figure 6C were submitted to a 2 (Feature Type: D vs. P) by 2 (Age Group: Children 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, $F(1,38) = 22.53, MSE = 0.80, p < .001, \eta^2 = 0.372$. Specifically, adults exhibited better memory accuracy for any single P feature than for the D feature (0.613 vs. 0.206), paired-samples $t(19) = 5.84, p < .001, d = 1.67$, whereas young children exhibited equivalent memory accuracy for any single P feature and for the D feature (0.444 vs. 0.438), paired-samples $t(19) = .13, p = .897$.

Overall, when participants’ attention was directed to the P features in Experiment 4, adults’ categorization performance became similar to children who tended to rely on the overall appearance and their memory accuracy changed accordingly. However, young children’s memory pattern remained the same across three experiments, exhibiting equivalently high memory accuracy for both D and P features.

The reported experiments in Chapter 4 presents several novel findings pointing to important developmental differences in category learning, category representation, and the role of selective attention in these processes. In terms of categorization responses, children and adults were responsive to manipulations. However, important differences transpired with respect to category representation, as evidenced by recognition memory. Adults, unless told otherwise in Experiment 4, tended to extract the deterministic features and form a rule-based representation. In contrast, regardless of their categorization performance, young children tended to encode multi-feature information and form a similarity-based representation. These results point to an important developmental
difference in the pattern attention: Whereas adults attend selectively to what they deem to be category-relevant, young children attend diffusely. Importantly, more efficient selective attention in adults was accompanied by worse memory of the to-be-ignored features than of the to-be-attended features, whereas less efficient diffused attention in children was accompanied by equally good memory of both to-be-attended and to-be-ignored features.

Having established the differences in categorization between young children and adults, it is reasonable to ask: What about infant category learning? Infants are shown to exhibit distributed attention by attending to multiple features when learning novel categories (Best, et al., 2013). Would infants show different patterns of attention when they learn categories accompanied by novel labels? Would there be a relationship between the pattern of attention and the outcome of category learning? Experiments 5-6 in Chapter 5 aimed to further examine the role of attention in category learning in much younger participants, that is, 8- to 12-month-old infants.
CHAPTER 5

THE ROLE OF WORDS IN EARLY CATEGORY LEARNING

Experiments 5-6 were designed to examine the effects of labels and dynamic visual features on category learning in infancy. Experiment 5 included two between-subjects conditions: label-defined condition and motion-defined condition. In both conditions, infants were familiarized with exemplars from one category and then tested with the prototype of this category and that of the contrast category. Infants saw the same testing stimuli in both conditions, with neither label nor motion being presented during testing. Experiment 6 was aimed at further comparing the effects of labels with a silent control condition. In both experiments, eye gaze data were collected. Recall, that, according to the label-as-category-marker position, labels should (a) facilitate category learning compared to the other conditions and (b) lead to greater attention to commonalities than in the other two conditions. In contrast, according to the label-as-feature view, (a) because of auditory overshadowing, labels should not facilitate category learning and (b) more robust category learning may be accompanied by more distributed attention.
EXPERIMENT 5

Method

Participants

Fifty-one infants (30 girls) ranging in age from 8 to 12 months ($M = 10$ months, 11 days; $SD = 1$ month, 20 days) were recruited. Ten infants were excluded from the analyses (two due to fussiness and eight for not looking at a single test trial).

Apparatus

A Tobii T60 eye-tracker with the sampling rate of 60 Hz was used to collect eye gaze data. The eye-tracker was integrated into a 17-inch computer monitor and located on a table inside a booth enclosed by black curtains. A trained experimenter monitored the experiment using Tobii Studio gaze analysis software installed on a 19-inch Dell OptiPlex 755 computer outside the booth. A video stream displaying participants’ activities was projected onto a 9-inch black and white Sony SSM-930 CE television for the experimenter’s online monitoring. Two Dell speakers were located behind a black curtain on each side of the eye-tracker.

Materials and Design

The materials were colorful drawings of artificial creatures and novel labels "flurp" and "jalet" (see Figure 7A). The creatures had five features varying in color and shape and consisted of two categories. As shown in Table 8, the categories had a family-resemblance structure, which was derived from two prototypes (A0 and B0) by modifying the values of one of five features – head, antennae, hands, body, or feet. We used these novel categories to ensure that none of the infants was familiar with these
categories prior to the experiment and all participants had to learn these categories *de novo*. Another advantage of these stimuli is that they have been extensively tested in previous work with preschoolers (Deng & Sloutsky, 2012; 2013).

There were two between-subjects conditions: label-defined and motion-defined. In the label-defined condition, the label presented during training was the same for all the exemplars, whereas in the motion-defined condition the pattern of motion presented during training was the same. To create a dynamic visual feature, the feet were animated using Macromedia Flash MX software. For all “flurps” the feet stretched up and down, whereas for all “jalets” the feet moved sideways. Because the goal was to examine effects of labels and dynamic visual features on category *learning*, neither labels nor patterns of motion were presented at test.

*Procedure*

Infants were seated on parents’ laps approximately 60 cm away from the eye-tracker. Parents were instructed not to interact with infants and to avoid speaking or pointing. Prior to the experiment, infants completed a 5-point calibration sequence. The calibration consisted of dynamic kitten images accompanied by a “bouncing” sound appearing in different locations on the screen.

The experiment proper consisted of 20 familiarization and 4 test trials. The trials were mixed and pseudo-randomly assigned to four blocks, with each block consisting of five familiarization trials followed by one test trial (see Figure 7B for the overall sequence in a block). On each familiarization trial, infants saw a creature generated from the same category (one of the categories shown in Table 8) on a white background lasting for 8000
ms and heard a phrase starting at the onset of each trial. A subset of infants studied Category A whereas the rest of the infants studied Category B. In the label-defined condition, the phrase (e.g., “Look! This is a Flurp”) was presented at the beginning of each trial and lasted for approximately 2800 ms, whereas, in the motion-defined condition, the phrase did not include the label (e.g., “Look at this one!”). The feet of the creature started moving after the phrase ended and the motion lasted for 3000 ms. The onset of motion was approximately the same as that of the label in the label-defined condition (i.e., 2300 ms into the trial).

Each test trial lasted for 8000 ms and presented a pair of items – the prototype of the studied category and the prototype of the non-studied category, with the left-right position of each prototype counterbalanced across test trials. Note that neither prototype was presented during training. These test items (these were the same across the conditions) were different from training trials as they were presented without either label or motion. A dynamic bouncing ball was presented between trials within each block, whereas a short and task-irrelevant cartoon video was presented between blocks to maintain engagement of infant participants. All gaze data were recorded by the computer using Tobii Studio gaze analysis software.

Results and Discussion

Gaze data were exported using Tobii Studio gaze analysis software. For each stimulus seven areas of interest (AOIs) for fixations were defined: ellipses (ranging in visual angle from 2.4° by 2.4° to 3.8° by 5.7°) surrounding each feature of the creatures. Given that hands, feet, and antennae appeared in pairs, each pair was treated as a single AOI, which
resulted in five AOIs used for data analyses: head, body, hands, feet, and antennae (see Figure 7C for AOIs). The gaze durations were weighted by the area of each AOI. The analyses focused on (1) looking time at familiarization; (2) novelty preference score based on the proportion of looking time to the prototype of the novel category as compared to the total looking time to both of the prototypes at test; and (3) patterns of attention during test and familiarization based on (a) the proportion of looking time to different features and (b) the number of gaze shifts between AOIs.

Looking Time at Familiarization

To ascertain that the levels of engagement with the stimuli were comparable across the conditions, we compared accumulated looking time to the stimuli on familiarization trials in the label-defined condition with that in the motion-defined condition. The accumulated looking time data were submitted to a 4 (Block: 1 vs. 2 vs. 3 vs. 4) by 2 (Condition: label-defined vs. motion-defined) mixed ANOVA, with block as a within-subjects factor and condition as a between-subjects factor. Results revealed a main effect of block, $F(3, 117) = 37.26$, $MSE = 0.02$, $p < .01$, $\eta^2_p = 0.296$, with infants’ accumulated looking time decreasing through blocks. However, infants’ accumulated looking time did not differ between the label-defined and motion-defined conditions, $p = .52$. These results indicated that infants exhibited comparable levels of engagement across the two conditions.

To further examine whether infants became more familiarized with stimuli in the course of learning, we compared the average looking time on the first-three and the last-three familiarization trials in the label-defined condition with that in the motion-defined
condition (see Figure 8). These data were submitted to a 2 (Trial: first-three vs. last-three) by 2 (Condition: label-defined vs. motion-defined) mixed ANOVA, with trial as a within-subjects factor and condition as a between-subjects factor. There was a main effect of trial, $F(1, 39) = 57.78, p < .001, \eta^2_p = 0.597$, with infants in both conditions decreasing looking during familiarization, paired-samples $ps < .001$. There was a significant trial by condition interaction, $F(1, 39) = 4.62, p = .038, \eta^2_p = 0.106$. The interaction indicated that infants in the label-defined condition started with longer looking on the first-three trials than those in the motion-defined condition, independent-samples $t(31.7) = 2.44, p = .021, d = 0.75$, whereas there was no difference in the last three trials, $p = .629$. Therefore, infants in the label-defined condition exhibited somewhat greater familiarization than infants in the motion-defined condition.

**Novelty Preference**

To examine how labels or patterns of motion affected infants’ categorization, a novelty preference score was calculated for each test trial. Because there was only one test trial per block, we averaged test trials for blocks 1-2 and for blocks 3-4 (see Figure 9). The data in Figure 9 were submitted to a 2 (Block: 1-2 vs. 3-4) by 2 (Condition: label-defined vs. motion-defined) mixed ANOVA, with block as a within-subjects factor and condition as a between-subjects factor. There was a significant main effect of condition, with infants having higher novelty preference scores in the motion-defined condition, $F(1, 39) = 5.78, MSE = 0.12, p = .021, \eta^2_p = 0.129$. Neither the effect of block ($p = .079$), nor the interaction ($p = .325$) reached significance.
In addition, infants in the motion-defined condition exhibited above-chance novelty preference in Blocks 3-4, one-sample $t(19) = 2.87, p = .010, d = 0.64$, but not in Blocks 1-2, $p = .949$; whereas in the label-defined condition, novelty preference was not different from chance in either block, both $ps > .251$. Therefore, after four blocks of training participants in the motion-defined condition exhibited evidence of category learning, whereas participants in the label-defined condition failed to learn.

Distributions of individual novelty preference scores by condition are presented in Figure 10. As shown in Figure 4, in the first half of training (Blocks 1-2) there were comparable numbers of participants exhibiting novelty preference in either condition, whereas in the second half of training (Blocks 3-4), there were more participants exhibiting novelty preference in the motion-defined condition. To perform statistical analyses, we identified participants with novelty preference scores of 55%\(^1\) or higher as “learners”, while identifying the rest as “non-learners”. For Blocks 1-2, in the label-defined condition there were 4 out 21 learners (19%) and in the motion-defined condition there were 7 out 20 learners (35%), which was not statistically different, $\chi^2 (1, N = 41) = 1.33, p = .249$. However, in Blocks 3-4, there were 13 out 20 learners (65%) in the motion-defined condition, which exceeded the number of learners in the label-defined condition 7 out of 21, (33.3%), $\chi^2 (1, N = 41) = 4.11, p = .043$. Therefore, despite the fact that there was no advantage for the motion-defined condition during familiarization, there

\(^1\) We also performed chi-square analyses by identifying participants with novelty preference scores of 60% or higher and 65% or higher as “learners” (which resulted in a substantial decrease in the number of learners in the label-defined condition) and results remained the same. In Blocks 1-2, there were comparable numbers of participants exhibiting novelty preference in either condition ($ps > .269$), whereas in Blocks 3-4, there were more participants exhibiting novelty preference in the motion-defined condition ($ps < .006$).
was greater evidence of category learning in this condition than in the label-defined condition.

Patterns of Attention

We also examined how attention was distributed among different features of the stimuli on test trials (recall that neither labels nor motion was presented during these trials). Since there was no main effect of block ($p = .54$), the data were collapsed across blocks and results are presented in Figure 11. There was a significant main effect of feature, $F(4, 156) = 112.04, \text{MSE} = 2.39, p < .001, \eta_p^2 = 0.742$. Neither the interaction nor the main effect of condition was significant ($ps > .117$). These results suggest that the difference in novelty preference between the label-defined and the motion-defined conditions did not stem from patterns of attention at test.

Similar analyses were conducted to examine the patterns of attention on familiarization trials. (see Figure 12). The proportion of looking time to each feature was calculated within each trial (8000 ms) and then averaged across five familiarization trials within each block. These data were submitted to a $5 \times 4 \times 2$ mixed ANOVA, with feature and block as within-subjects factors and condition as a between-subjects factor. Because there was no effect of block ($p = .155$), data were collapsed across the four blocks. There was a feature by condition interaction, $F(4, 156) = 30.95, \text{MSE} = 0.391, p < .001, \eta_p^2 = 0.442$. Infants’ accumulated more looking to the head in the label-defined condition compared to the motion-defined condition, independent sample $t(39) = 4.94, p < .001, d = 1.54$, whereas they looked
significantly longer at the feet in motion-defined condition than in label-defined condition, independent sample $t(20.74) = 7.71, p < .001, d = 2.46$.

Although greater attention to the feet in the motion-defined condition may not be surprising, it was nevertheless associated with better learning. Note that feet motion was introduced about 2000 ms into the trial and ended about 3000 ms before the end of the trial. Given that participants spent most of the time looking at the head, it is likely that they moved their gaze at least twice: first from the head to the feet and then back from the feet to the head. Therefore, participants in the motion-defined condition were likely to have more gaze shifts, and this more distributed pattern of attention may have led to better category learning.

In order to examine this possibility, we further analyzed the number of fixation shifts between AOIs after the onset of label or motion during familiarization. To perform statistical analyses, we identified a fixation shift as a valid one if the looking time accumulated in the AOI before the shift and in the AOI after the shift were at least 100 ms respectively. The average number of valid fixation shifts within each block across the conditions is presented in Figure 13. These data were submitted to a 4 (Block: 1 vs. 2 vs. 3 vs. 4) by 2 (Condition: label-defined vs. motion-defined) mixed ANOVA, with block as a within-subjects factor and condition as a between-subjects factor. Results revealed a main effect of block, $F (3, 117) = 3.92, MSE = 79.22, p = .010, \eta^2_p = 0.091$, with the number of valid fixation shifts decreasing through blocks. And more importantly, there was a main effect of condition, $F (1, 39) = 5.09, MSE = 442.85, p = .013, \eta^2_p = 0.149$,
with infants making more shifts after the onset of motion in the motion-defined condition than that after the onset of label in the label-defined condition.

However, one may argue that greater attention to the head in the label condition (see Figure 12) is indicative of the fact that labels attracted attention to commonalities, with participants optimizing attention by shifting it to the head. Another possibility is that participants in the label-defined condition merely exhibited a head bias (see Quinn, et al., 2009, for evidence of head bias in infancy). We therefore deemed it necessary to directly examine these possibilities by examining whether infants exhibited focused attention after the label was introduced and whether infants differentiated the novel head from the familiar head at familiarization.

*Dynamics of attention.* If labels attract attention to commonalities, attention may be diffused early in the trial, but should become more focused on the head after the label was introduced. However, if attention to the head stems from a bias, attention after the introduction of a label should be no more focused than before. To examine the dynamics of infants’ attention within familiarization trials, attention shifts were calculated every 1000 ms, and then averaged across four blocks at each time point for the total duration of 8000 ms. Data between 1000 ms (i.e., after the word “Look” was introduced in both conditions) and 7000 ms (when infants looking started to decrease rapidly) were used for analysis and were shown in Figure 14. Results presented in Figure 14 indicate that labels did not attract attention to commonalities (i.e., if they did, the number of shifts should have dropped rapidly after the onset of the label). In contrast, there was an increase in the number of shifts after the introduction of motion. Therefore, motion resulted in a more
distributed pattern of attention and in better learning. A 6 (Time Point: 1 vs. 2 vs. 3 vs. 4 vs. 5 vs. 6) by 2 (Condition: label vs. motion) mixed ANOVA confirmed these findings. There was a main effect of condition, $F(1, 39) = 7.51, MSE = 34.18, p = .009, \eta_p^2 = 0.161$, with more shifts transpiring in the motion-defined than the label-defined condition. There was also a main effect of time point, $F(5, 195) = 22.28, MSE = 13.57, p < .001, \eta_p^2 = 0.364$, with a quadratic trend showing that shifts increased early in the trial and then decreased later in the trial, $F(1, 39) = 54.94, p < .001, \eta_p^2 = 0.585$. In addition, there was a significant time point by condition interaction, $F(5, 195) = 5.07, MSE = 3.09, p < .001, \eta_p^2 = 0.115$. To examine the interaction, we compared the number of shifts at each time point after the introduction of the label or motion to that before the introduction (i.e., at the first time point). These pair-wise comparisons with Bonferroni correction indicated that in the label-defined condition, the number of shifts at the three consecutive time points after the onset of the label was comparable to that before the label, $ps > .203$, whereas in the motion-defined condition, there were more shifts at the three consecutive time points after the onset of motion than before, $ps < .010$. Therefore, there was no evidence that attention became more focused after the label was introduced, but it became more distributed after motion was introduced.

*Looking time to novel vs. familiar head.* Another way of examining whether longer looking to the head in the label condition stemmed from a head bias or from increased attention to commonalities, we examined novelty preference during familiarization. Recall that because the categories had the family resemblance structure (see Table 8), each familiarization item had one out-of-category (i.e., novel) feature. Therefore, in each
training block, there were four familiarization items with a given head and one familiarization item with a novel head. If labels attract attention to the common head, infants should look longer to the novel head during familiarization. However, if attention to the head stems from a head bias, infants should be interested in the head, whether it is familiar or novel. To examine this, we compared infants’ looking time to the novel head (i.e., the head of item A5 or B5 as shown in Table 1) with the average looking time to the four familiar heads (i.e., the head of items A1-A4 or B1-B4 as shown in Table 1) for each block in the label-defined condition. These data were submitted to a 4 (Block: 1 vs. 2 vs. 3 vs. 4) by 2 (Head Type: novel vs. familiar) within-subjects ANOVA. The results showed that infants did not differentiate the two types of head, with neither the interaction ($p = .467$) nor the main effect of head type ($p = .139$) being significant. These results suggest that looking to the head stemmed from a head bias rather than from labels attracting attention to commonalities. Similarly, their looking at other features did not differ when these features were familiar or novel (all $ps > .1$).

Overall, participants exhibited evidence of category learning in the motion-defined but not in the label-defined condition. This outcome is compatible with two possibilities: (1) label interfering with category learning or (2) motion facilitating category learning. To distinguish between these possibilities, we needed a baseline, in which neither labels nor motion were presented. If participants succeed in the baseline and exhibit patterns of attention comparable to those in the motion-defined condition, then labels interfered with learning. In contrast, if participants fail in the baseline and exhibit patterns of attention
comparable to those in the label-defined condition, then motion facilitated learning. The
goal of Experiment 6 was to distinguish between these possibilities

EXPERIMENT 6

Method

Participants

Twenty-three infants (12 girls) ranging in age from 8 to 12 months ($M = 11$ months, 1
day; $SD = 1$ month, 10 days) were recruited. Five infants were excluded from the
analyses due to fussiness or not looking at a single test trial.

Apparatus, Materials, Design, and Procedure

The apparatus, materials and procedure in Experiment 6 were similar to Experiment
5, with one critical difference: neither labels nor motion patterns were introduced during
training (see Figure 7B).

Results and Discussion

Similar to the data analyses in Experiment 5, five AOIs (i.e., head, body, hands, feet,
and antennae, see Figure 7C) were used and the gaze durations were weighted by the area
of each AOI. The analyses focused on (1) looking time at familiarization; (2) novelty
preference score based on the proportion of looking time to the prototype of the novel
category as compared to the total looking time to both of the prototypes at test; and (3)
patterns of attention based on the proportion of looking time to different features on test
and familiarization trials.

Similar to Experiment 5, infants in Experiment 6 exhibited a drop in looking time in
the last three familiarization trials compared to the first three trials, paired-sample $t(17) =$
2.11, $p = .05$, $d = 0.75$. Also, similar to the label-defined condition, novelty preference score in Blocks 1-2 was not different from that in Blocks 3-4, $p = .574$ and neither was different from chance, $ps > .628$. This was in contrast to the motion-defined condition, in which participants exhibited above-chance novelty preference by the second part of the experiment. Finally, patterns of attention at familiarization and test, as shown in Figure 15, were also similar to the label-defined condition: infants exhibited a head bias, Bonferroni adjusted $ps < .001$, and the proportion of looking to the head in the baseline condition did not differ from that in the label-defined condition ($ps > .144$). However, compared to the motion-defined condition, infants exhibited a stronger head bias in the baseline condition, with the proportion of looking to the head at familiarization being significantly higher ($p = .002$). Taken together these results suggest that labels did little compared to the no-label baseline, whereas patterns of motion changed infants’ patterns of attention which resulted in better category learning.
CHAPTER 6

GENERAL DISCUSSION

The reported studies in Chapters 3-5 present several novel findings pointing to important developmental differences in category learning, category representation, and the role of selective attention in these processes.

First, whereas adults and 6-year-olds exhibited representational asymmetry between classification and inference training, 4-year-olds formed similar representations across these training conditions. Specifically, older participants were more likely to use and represent a rule in classification training than in inference training (cf. Yamauchi & Markman, 1998; Hoffman & Rehder, 2010), whereas 4-year-olds were likely to form similarity-based representations in both conditions (Experiment 1, Chapter 3).

Second, adults, unless told otherwise, tended to extract the deterministic features and form a rule-based representation. In contrast, regardless of their categorization performance, young children tended to encode multi-feature information and form a similarity-based representation. These results point to an important developmental difference in the pattern attention: Whereas adults attend selectively to what they deem to be category-relevant, young children attend diffusely. Importantly, more efficient selective attention in adults was accompanied by worse memory of the to-be-ignored features than of the to-be-attended features, whereas less efficient diffused attention in
children was accompanied by equally good memory of both to-be-attended and to-be-ignored features (Experiments 2-4, Chapter 4).

And third, eye tracking results indicated that better category learning in 8- to 12-month-old infants was associated with more distributed attention among different features. Specifically, the presence of the dynamic visual feature resulted in more robust learning and in more distributed attention than in the other two conditions (cf. Jankowski, et al., 2001), whereas labels failed to either facilitate category learning or attract attention to commonalities (Experiments 5-6, Chapter 5).

These findings elucidate the development of category learning and categorization by linking it to the development of selective attention and they have important implications for theories of categorization and category learning. In addition, these results may contribute to better understanding of the role of linguistic labels in categorization. In what follows, I discuss each of these points.

6.1 Classification and Inference Training and Theories of Categorization

There are a number of models of categorization proposing different accounts of how categories are represented (see Murphy, 2002, for a review). Whereas these models differ in how they construe representations of categories (e.g., prototype or exemplar representations), they typically presume that these representations are independent of the ways categories are learned (but see Love, et al., 2004). Therefore, differential performance of adults in classification and inference training poses challenges to many existing theories of categorization (see Markman & Ross, 2003 for related arguments).
Although both prototype (Smith & Minda, 1998; 2000) and exemplar (Kruschke, 1992; Nosofsky, 1984; 1986) models may account for differential performance by setting different attentional weights for diagnostic features (i.e., by setting higher attentional weights in classification learning and reducing or turning off attentional weights in inference learning), these models do not consider classification and inference to be different tasks and thus have no obvious theoretical machinery capable of explaining differences between classification and inference training or between classification and inference judgment tasks.

There are two models that have separate parameters for category labels and thus can potentially address some but not all the problems. For example SINC (Sloutsky & Fisher, 2004) assumes that the role of category labels may change in the course of development and it has a separate attentional weight for category labels, which allows capturing this developmental change. SINC can also successfully model the asymmetry between categorization performance (where labels has to be predicted) and inference performance (where labels can be used) observed in adults and symmetrical performance observed in children. However, SINC may not have the theoretical machinery to account for the representational differences between classification and inference learning.

SUSTAIN (Love, et al., 2004) has also a separate parameter for the weight of label (i.e., the category focus parameter), which may vary across tasks and training conditions, and, as a model of learning, it can account for outcomes of training conditions as well as categorization performance. At the same time, the observed developmental differences – the asymmetry between classification and inference in adults but not in children – is a
challenge for SUSTAIN as it needs to explain how and why category labels become
different from other features comprising the category. Furthermore, these developmental
differences present a challenge for most existing models of categorization: in addition to
explaining differences between classification and inference learning, these models need
also to explain how these differences come on-line, and how the differences between
labels and other features emerge.

Another theoretical challenge stemming from present findings is the potential
difference between categorization decisions and category representation observed in 4-
year-olds in Experiment 3. Recall that after being instructed to focus on deterministic
features, 4-year-olds exhibited rule-based categorization performance, which was similar
to that of 6-year-olds and adults in Experiment 1. At the same time, similar to 4-year-
olds in Experiment 1, 4-year-olds in Experiment 3 remembered deterministic and
probabilistic features equally well. This pattern suggests that whereas they were likely to
use the deterministic feature in their categorization decisions, they were equally likely to
encode deterministic and probabilistic features in the course of category learning, which
suggests distributed rather than selective attention. However, most models of
categorization assume that decisions are made on the basis of representations, which
implies that similar representations (observed in 4-year-olds in Experiments 1 and 3)
should not result in radically different categorization decisions (observed in 4-year-olds
across the two experiments). Therefore 4-year-olds’ performance in Experiments 1 and 3
may run counter to theoretical assumptions of most models of categorization, thus
challenging these models.
6.2 The Role of Attention in Categorization

6.2.1 Selective Attention and Development of Categorization

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

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

Current results present additional evidence implicating selective attention in the development of categorization: adults and 6-year-olds exhibited more focused attention in
classification than in inference training, whereas 4-year-olds exhibited distributed attention across the training conditions.

Results of Experiment 2 further elucidated the role of selective attention in the development of categorization: focusing 4-year-olds’ attention on D features resulted in adult-like categorization performance, yet their recognition memory performance (i.e., equivalently high memory for D and P features) was similar to that of 4-year-olds in Experiment 1. This dissociation between categorization and recognition memory suggests that (a) categorization may include representation and decision components, (b) both components may depend on attention, (c) it may be easier to externally access the decision component than the representation component, and (d) development of categorization may include both decision and representation components. Therefore, current research presents novel evidence about the role of selective attention in category learning across development, suggesting that (a) important developments occur between 4- and 6-years of age and (b) the decision component of categorization (which is more likely to be affected externally) may develop somewhat earlier than the representation component (which may require more endogenous selective attention).

Taken together, these results advance our understanding of the development of categorization, indicating that (1) category representation undergoes development; (2) the development results in acquiring the ability to form rule-based representations, without losing the ability to form similarity-based representations; and (3) the development of domain-general selective attention contributes to the development of category representations. Although these results do not eliminate the possibility that domain-
specific knowledge contributes to the development of categorization (see Carey, 1991; Inagaki & Hatano, 2002; Keil, 1992, for arguments for domains-specific changes), they clearly emphasize the contribution of domain-general processes.

6.2.2 The Role of Attention in Infant Category Learning

There is much evidence demonstrating the role of selective attention in adult category learning and there is more recent eye-tracking evidence (Blair, et al., 2009; Hoffman & Rehder, 2010) indicating that category learning in adults is accompanied by attention optimization – shifting of attention to within-category commonalities. There is also a related argument pertaining to category learning in infancy: according to the label-as-category-marker position, even in infancy labels should facilitate category learning by attracting attention to commonalities (i.e., by facilitating attention optimization). Therefore, labels are expected to facilitate category learning by affecting selective attention. However, there is evidence that makes this mechanism unlikely.

In particular, there are recent findings that successful category learning in infancy is not accompanied by attention optimization (Best, et al., 2013). This finding is important given previous evidence that distributed attention (with frequent fixation shifts) may result in better learning (e.g., Bronson, 1991; Colombo, et al., 1991; Jankowski, et al, 2001; Rose, et al., 2003). Therefore, a feature that encourages a more distributed pattern may also facilitate learning. One candidate is a dynamic visual feature, especially if it is presented peripherally. Such features may affect attention because at the very minimum participants would need to move their gaze from the point of their initial fixation to the moving part and then, when the motion ceases, to the point of their final fixation. These
ideas, while consistent with the label-as-feature view, run counter to the very core of the label-as-category-marker view.

Many studies have examined the role of labels in infants’ categorization, but this is the first study to demonstrate that effects of dynamic visual features on category learning are greater than those of labels. By comparing the outcome of category learning and examining the patterns of attention in the label-defined and motion-defined conditions, the current study provides novel evidence elucidating how different features may affect category learning in infancy. The results indicate that distributed attention results in successful category learning in infancy and that features that elicit more distributed attention may also lead to better learning.

6.3 The Role of Linguistic Labels in Categorization

6.3.1 What is the Role of Words in Early Category Learning?

Recall that two proposals have been advanced as to the role of words in early category learning: label-as-category-marker and label-as-feature. The label-as-category marker position makes two critical predictions. First, labels should change patterns of attention (compared to no-label baseline) by attracting attention to within-category commonalities. And second, labels should facilitate category learning above the no-label baseline. To our knowledge, the first prediction has not been tested before and the reported study is the first such test. The results clearly indicate that labels failed to attract attention to commonalities.

In contrast to the first prediction, the second prediction has been tested extensively in previous research, and generated conflicting evidence, with some studies finding
facilitative effects of labels and others failing to find such effects. Two sources of supporting evidence are worth considering: (1) differential effects of labels on learning of basic-level and superordinate categories and (2) differential effects of nouns and adjectives on category learning. As we discuss below, many of these effects are inconclusive.

One of the first studies demonstrating such differential effects was the study conducted by Waxman and Markow (1995). In this study with 9-to-20-month-olds two variables were crossed: (1) the category structure (i.e., Basic level vs. Superordinate) and (2) labeling condition (Noun vs. No Word). Results indicated that novelty preference was above chance in all conditions, except for the Superordinate Category-No Word condition. On the basis of these results, it was concluded that words facilitate infants’ attention to superordinate categories, whereas labels have little effect on learning of basic-level categories labels (see also Fulkerson & Haaf, 2003). In contrast, Balaban and Waxman (1997) reported facilitative effects of labels for the basic-level categories in 9-month-olds. Therefore, there is no clear evidence that words have consistently different effects for categories of different levels.

The second source of support has to do with putatively different effects of nouns and adjectives on categorization. For example, in one study (Booth & Waxman, 2009) with 14- and 18-month-olds the category structure (i.e., Basic level vs. Superordinate) was fully crossed with lexical category (i.e., nouns vs. adjectives). The analyses revealed greater novelty preference in the noun condition compared to the other two conditions, but only for a single time window in the third quarter of the trial. In contrast, in a similar
study conducted by the same researchers with a slightly different paradigm (Waxman & Booth, 2001), 14-month-olds exhibited equivalent novelty preference in the noun and in the adjective conditions. Therefore, evidence for different effects of nouns and adjectives on category learning in infancy is rather weak and inconclusive.

Whereas findings used to support the label-as-category-marker position are inconclusive, the reported results in Experiments 5-6 contribute to the growing body of evidence suggesting that labels are similar to other features, in that they are part of input rather than category markers. Data reported here include both negative and positive evidence. Negative evidence indicates that labels fail to facilitate category learning, change pattern of attention compared to the no-label baseline, or attract attention to commonalities. Although this evidence disputes the role of labels in this specific design, it does leave the possibility that perhaps under some other condition(s) labels facilitate category learning by attracting attention to commonalities.

Positive evidence is stronger because it disputes the very core of the label-as-category-marker approach. In particular, if distributed attention results in more successful category learning in infancy, then even if labels are found to attract attention to commonalities, they are unlikely to facilitate category learning. Alternatively, if labels do not attract attention to commonalities, then little is left of the label-as-category-marker position, even if labels are found to facilitate category learning. This is because if labels do not attract attention to commonalities these putative facilitative effects of labels would not uniquely support the label-as-category-marker position. These findings are important because distinguishing between these positions is consequential for our understanding of
the relationships between language and cognition and the nature of learning early in development.

### 6.3.2 The Changing Role of Category Labels

Recall that the standard way of explaining the asymmetry between classification and inference is by assuming a difference between category labels and other features (Love, et al., 2004; Markman & Ross, 2003; Yamauchi & Markman, 2000). However, the fact that 4-year-olds exhibit symmetrical performance suggests that labels may start out as features and their role may change in the course of development.

As discussed earlier, the role of labels in categorization is a matter of considerable debate. Yamauchi and Markman (1998, 2000) presented arguments that in order for the label to be a feature, classification and inference judgment tasks should be functionally equivalent. They presented extensive evidence that the tasks are not functionally equivalent for adults, thus suggesting that for adults labels are category markers. At the same time, there is a growing body of evidence (e.g., Deng & Sloutsky, 2012, 2013) suggesting that classification and inference judgments are equivalent for young children, thus suggesting that labels may function as features early in development. Results presented in Experiment 1 suggest that classification and inference training are also not equivalent for adults and 6-year-olds, while being equivalent for 4-year-olds. These results corroborate and extend previous findings. In particular, these results suggest that the role of category labels changes in the course of development: labels may start out as features, but they may become category markers in the course of development.
addition, results presented in Experiment 1 suggest that important changes in the role of category labels begin to occur between 4- and 6-years of age.

Although there is no conclusive evidence at this point, three developments contributing to the changing role of category labels are worth considering. First, auditory-visual integration becomes more efficient, which results in attenuated auditory interference in visual processing (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003). Second, children are likely to learn that labels are the most reliable cues to the category. Third, the development of selective attention allows them to focus on features most reliably co-occurring with category labels. However, at this point, most of these contentions are preliminary and they have to be addressed in future research. And finally, each of these changes occurs between 4- and 6-year of age (although some of the changes may continue even later in development).

6.4 Conclusion

This dissertation research presents novel evidence pertaining to the development of categorization and the role of attention in this process. First, whereas adults and older children are more likely to attend selectively to deterministic features in classification than in inference training, younger children tend to attend diffusely regardless of the training condition. Second, diffused attention and less efficient category learning in younger children are associated with better memory for specific exemplars, whereas selective attention and more efficient category learning in older children and adults are associated with worse memory for exemplars. Third, by manipulating attention exogenously, we can turn adults’ categorization strategy into a childlike one and increase
their memory for exemplars. In contrast, in young children, we can change only
categorization strategy, whereas their memory accuracy remains uniformly high. And
finally, better category learning in 8- to 12-month-old infants was associated with more
distributed attention among different features. Specifically, the presence of the dynamic
visual feature resulted in more robust learning and in more distributed attention, whereas
labels failed to either facilitate category learning or attract attention to commonalities.

These findings point to important changes in category representation: the
development consists of acquisition of the ability to form rule-based representations,
without losing the ability to form similarity-based representation. Furthermore,
exogenous factors may affect category decisions of younger children, but not their
category representation, which points to a potential distinction between decision and
representation components of categorization and suggests that developmental changes in
decision components may occur prior to developmental changes in representation.
Findings with infants point to an important mechanism by which dynamic features may
affect category learning in early development. These results have important implications
for theories of categorization, understanding the development of categorization, and the
changing role of selective attention in category learning and category representation.
References


Appendix: Tables and Figures

The tables and figures for Experiments 1-6 are provided on the following pages.
Table 1. Category structure used in Experiments 1-4.

<table>
<thead>
<tr>
<th>Category</th>
<th>Head</th>
<th>Body</th>
<th>Hands</th>
<th>Feet</th>
<th>Antennae</th>
<th>Tail</th>
<th>Button</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P_flurpD_flurp</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_jaletD_flurp</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_flurpD_new</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>P_jaletD_flurp</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>N</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P_all_newD_flurp</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Category F

<table>
<thead>
<tr>
<th>Category</th>
<th>Head</th>
<th>Body</th>
<th>Hands</th>
<th>Feet</th>
<th>Antennae</th>
<th>Tail</th>
<th>Button</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P_jaletD_jalet</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P_flurpD_jalet</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P_jaletD_new</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>P_newD_jalet</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>N</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P_all_newD_jalet</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Note: The value 1 = any of seven dimensions identical to Category F (flurp, see Figure 1). The value 0 = any of seven dimensions identical to Category J (jalet, see Figure 1). The value N = new feature which is not presented during training. P = probabilistic feature; D = deterministic feature. F0 is the prototype of Category F and J0 is the prototype of Category J. Items in the first row (i.e., F0 and J0) are prototypes and they were not used in either training or testing. Variants of items in the second row are High-Match items and they were used in both training and testing. Variants of all other item types were used only in testing.
Table 2. Training data: Mean (standard deviation) proportion of correct responses aggregated in 10-trial blocks across age groups and training conditions in Experiment 1.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Training Type</th>
<th>Trials 1-10</th>
<th>Trials 11-20</th>
<th>Trials 21-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>Classification</td>
<td>0.91 (0.11)</td>
<td>0.93 (0.12)</td>
<td>0.97 (0.11)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.85 (0.16)</td>
<td>0.84 (0.18)</td>
<td>0.89 (0.13)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Classification</td>
<td>0.85 (0.14)</td>
<td>0.91 (0.18)</td>
<td>0.95 (0.14)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.60 (0.17)</td>
<td>0.67 (0.18)</td>
<td>0.73 (0.17)</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>Classification</td>
<td>0.62 (0.17)</td>
<td>0.70 (0.20)</td>
<td>0.78 (0.17)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.54 (0.13)</td>
<td>0.61 (0.17)</td>
<td>0.71 (0.13)</td>
</tr>
</tbody>
</table>
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 Experiment 1.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Training Type</th>
<th>new-D</th>
<th>one-new-P</th>
<th>all-new-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>Classification</td>
<td>0.70 (0.23)</td>
<td>0.98 (0.06)</td>
<td>0.94 (0.12)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.73 (0.22)</td>
<td>0.88 (0.20)</td>
<td>0.75 (0.29)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Classification</td>
<td>0.53 (0.22)</td>
<td>0.91 (0.18)</td>
<td>0.94 (0.13)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.54 (0.22)</td>
<td>0.71 (0.26)</td>
<td>0.59 (0.26)</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>Classification</td>
<td>0.74 (0.23)</td>
<td>0.71 (0.21)</td>
<td>0.51 (0.16)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.75 (0.17)</td>
<td>0.79 (0.21)</td>
<td>0.54 (0.23)</td>
</tr>
</tbody>
</table>

Note:

New-D items (i.e., PflurgDnew and PjaletDnew) had probabilistic features of the studied categories and a novel feature replacing the deterministic feature. High proportion of correct responses on new-D items indicates that the participant can categorize items on the basis of old P features, even when the D feature is new.

One-new-P items (i.e., PnewDflurg and PnewDjalet) had all features of the studied categories but a novel feature replacing one probabilistic feature. High proportion of correct responses on one-new-P indicates that the participant can tolerate small distortion in the category prototype when categorizing items.

All-new-P items (i.e., Pall-newDflurg and Pall-newDjalet) had the deterministic features from the studied categories and all new features replacing the studied probabilistic features. High proportion of correct responses on all-new-P items indicates that the participant relies on old D features and generalizes broadly.

The scale effectively ranges from 0.5 to 1, with 0.5 being chance performance.
Table 4. Memory at test: Mean (standard deviation) proportions of yes responses (i.e., old responses) on different item types in Experiment 1.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Training Type</th>
<th>High-Match</th>
<th>new-D</th>
<th>one-new-P</th>
<th>all-new-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>Classification</td>
<td>0.94 (0.11)</td>
<td>0.07 (0.15)</td>
<td>0.23 (0.21)</td>
<td>0.01 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.54 (0.36)</td>
<td>0.32 (0.31)</td>
<td>0.16 (0.17)</td>
<td>0.01 (0.04)</td>
</tr>
<tr>
<td>6-year-olds</td>
<td>Classification</td>
<td>0.88 (0.19)</td>
<td>0.24 (0.32)</td>
<td>0.49 (0.34)</td>
<td>0.22 (0.36)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.89 (0.17)</td>
<td>0.35 (0.39)</td>
<td>0.31 (0.26)</td>
<td>0.11 (0.19)</td>
</tr>
<tr>
<td>4-year-olds</td>
<td>Classification</td>
<td>0.86 (0.19)</td>
<td>0.28 (0.23)</td>
<td>0.31 (0.27)</td>
<td>0.12 (0.21)</td>
</tr>
<tr>
<td></td>
<td>Inference</td>
<td>0.88 (0.16)</td>
<td>0.31 (0.21)</td>
<td>0.25 (0.22)</td>
<td>0.11 (0.17)</td>
</tr>
</tbody>
</table>

Note:

The overall memory accuracy is estimated by the difference in the proportion of yes responses to High-Match items and to all-new-P items.

Memory accuracy for the rule (i.e. the D feature) is estimated by the difference in the proportion of yes responses to High-Match items and to new-D items.

Memory accuracy for the overall appearance (i.e., P features) is estimated by the difference in the proportion of yes responses to High-Match items and to one-new-P items.

The scale ranges from 0 to 1, with 0 being chance performance.
Table 5. Training data: Mean (standard deviation) proportion of correct responses aggregated in 10-trial blocks across age groups in Experiments 2, 3, and 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age Group</th>
<th>Trials 1-10</th>
<th>Trials 11-20</th>
<th>Trials 21-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2</td>
<td>Adults</td>
<td>0.66 (0.17)</td>
<td>0.79 (0.23)</td>
<td>0.82 (0.25)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.52 (0.17)</td>
<td>0.61 (0.20)</td>
<td>0.69 (0.14)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Adults</td>
<td>0.99 (0.02)</td>
<td>0.98 (0.04)</td>
<td>0.99 (0.02)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.73 (0.22)</td>
<td>0.84 (0.24)</td>
<td>0.88 (0.18)</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Adults</td>
<td>0.70 (0.17)</td>
<td>0.73 (0.16)</td>
<td>0.76 (0.16)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.58 (0.15)</td>
<td>0.66 (0.19)</td>
<td>0.72 (0.23)</td>
</tr>
</tbody>
</table>
Table 6. 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 2, 3, and 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age Group</th>
<th>new-D</th>
<th>one-new-P</th>
<th>all-new-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2</td>
<td>Adults</td>
<td>0.49 (0.23)</td>
<td>0.81 (0.28)</td>
<td>0.79 (0.28)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.63 (0.17)</td>
<td>0.66 (0.22)</td>
<td>0.46 (0.19)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Adults</td>
<td>0.34 (0.32)</td>
<td>0.99 (0.04)</td>
<td>0.99 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.63 (0.33)</td>
<td>0.81 (0.24)</td>
<td>0.80 (0.22)</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Adults</td>
<td>0.74 (0.17)</td>
<td>0.80 (0.20)</td>
<td>0.44 (0.14)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.63 (0.19)</td>
<td>0.65 (0.28)</td>
<td>0.53 (0.16)</td>
</tr>
</tbody>
</table>
Table 7. Memory at test: Mean (standard deviation) proportions of “yes” responses (i.e., “old” responses) on different item types in Experiments 2, 3, and 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Age Group</th>
<th>High-Match</th>
<th>new-D</th>
<th>one-new-P</th>
<th>all-new-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2</td>
<td>Adults</td>
<td>0.77 (0.22)</td>
<td>0.21 (0.27)</td>
<td>0.39 (0.25)</td>
<td>0.09 (0.14)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.80 (0.22)</td>
<td>0.42 (0.32)</td>
<td>0.41 (0.28)</td>
<td>0.21 (0.31)</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Adults</td>
<td>0.79 (0.14)</td>
<td>0.13 (0.21)</td>
<td>0.45 (0.26)</td>
<td>0.07 (0.17)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.78 (0.26)</td>
<td>0.31 (0.36)</td>
<td>0.32 (0.32)</td>
<td>0.22 (0.33)</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Adults</td>
<td>0.81 (0.22)</td>
<td>0.61 (0.30)</td>
<td>0.20 (0.20)</td>
<td>0 (0)</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>0.79 (0.24)</td>
<td>0.36 (0.31)</td>
<td>0.35 (0.31)</td>
<td>0.20 (0.27)</td>
</tr>
</tbody>
</table>
Table 8. Category structure used in Experiment 5.

<table>
<thead>
<tr>
<th></th>
<th>Head</th>
<th>Body</th>
<th>Hands</th>
<th>Feet</th>
<th>Antennae</th>
<th>Label/Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Category B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The value 1 = any of five dimensions identical to "Flurp" (see Figure 7). The value 0 = any of five dimensions identical to "Jalet" (see Figure 7). A0 and B0 are prototypes of each category.
<table>
<thead>
<tr>
<th>Category</th>
<th>Prototype</th>
<th>High-Match</th>
<th>switch item</th>
<th>new-D</th>
<th>one-new-P</th>
<th>all-new-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flurp</td>
<td>F0</td>
<td>P_{flurp}D_{flurp}</td>
<td>P_{jalet}D_{flurp}</td>
<td>P_{flurp}D_{new}</td>
<td>P_{new}D_{flurp}</td>
<td>P_{all-new}D_{flurp}</td>
</tr>
<tr>
<td></td>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Jalet</td>
<td>J0</td>
<td>P_{jalet}D_{jalet}</td>
<td>P_{flurp}D_{jalet}</td>
<td>P_{jalet}D_{new}</td>
<td>P_{new}D_{jalet}</td>
<td>P_{all-new}D_{jalet}</td>
</tr>
<tr>
<td></td>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
</tbody>
</table>

*Figure 1.* Examples of stimuli used in Experiments 1-4. Each row depicts items within a category, whereas each column identified an item role (e.g., switch item) and item type (e.g., $P_{jalet}D_{flurp}$). The High-Match items were used in training and testing. The switch items, new-D, one-new-P, and all-new-P items were used only in testing. Neither prototype was shown in training or testing.
Figure 2. A schematic representation of the procedure. In Experiment 1, the phases progressed from A to C. Half of the participants were presented with Classification training in Phase B, whereas the other half were presented with Inference training in Phase B. The procedure of Experiments 2-4 was the same, except that (1) there was only Classification Training condition and (2) participants were given instructions and feedback during training focusing their attention on D features in Experiment 3 and on P features in Experiment 4.
Figure 3. Categorization performance: Proportion of rule-based responses by trial type and training condition for adults (A), 6-year-old children (B), and 4-year-old children (C) in Experiment 1. High-Match items are P_{flurp}D_{flurp} and P_{jalet}D_{jalet}; Switch items are P_{flurp}D_{jalet} and P_{jalet}D_{flurp}.
Figure 3 continued

C. 4-year-olds

![Graph showing the proportion of rule-based response for High-Match Items and Switch Items under different training conditions. The graph compares classification and inference responding.](image)

**Training**
- Classification
- Inference

**Trial Type**
- High-Match Items
- Switch Items

The graph illustrates the proportion of rule-based response for High-Match Items and Switch Items under different training conditions, comparing classification and inference responding.
Figure 4. Recognition performance: Memory accuracy by feature type and training condition for adults (A), 6-year-old children (B), and 4-year-old children (C) in Experiment 1.
Figure 4 continued

C. 4-year-olds

![Bar chart showing hits and false alarms for classification and inference under deterministic (D) and probabilistic (P) feature conditions.]

**Training Condition**
- Classification
- Inference

**Feature**
- Deterministic (D)
- Probabilistic (P)
Figure 5. Categorization performance: Proportion of rule-based responses by trial type and age group in Experiment 2 (A), Experiment 3 (B), and Experiment 4 (C). The chance level is 0.5.
Figure 5 continued

C. Experiment 4

![Figure 5 continued]

- **Proportion of rule-based response**
- **Age Group**
- **High-Match Items**
- **Critical Lures**

<table>
<thead>
<tr>
<th>Age Group</th>
<th>High-Match Items</th>
<th>Critical Lures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Children</td>
<td>0.8</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Rule-based responding

Similarity-based responding
Figure 6. Recognition performance: Memory accuracy by feature type and age group in Experiment 2 (A), Experiment 3 (B), and Experiment 4 (C). The chance level is 0.
Figure 6 continued

C. Experiment 4

![Bar graph showing hits vs. false alarms for adults and children with features labeled as Deterministic (D) and Probabilistic (P).]
Figure 7. Example stimuli. A. Prototypes of stimuli from Categories A and B; B. Procedures used in this study. C. Areas of interest (AOIs) shown as gray ellipses.
Figure 7 continued

C.
Figure 8. Average looking time on the first-three and last-three familiarization trials in the label-defined and motion-defined conditions in Experiment 5.
Figure 9. Novelty preference scores on first two test trials and second two test trials in the label-defined and motion-defined conditions in Experiment 5.
Figure 10. Individual novelty preference scores in the label-defined and motion-defined conditions in Blocks 1-2 (A) and in blocks 3-4 (B) in Experiment 5.
Figure 11. Proportion of accumulated looking time to each feature averaged across 4 blocks at test in the label-defined and motion-defined conditions in Experiment 5.
Figure 12. Proportion of accumulated looking time to each feature averaged across 4 blocks at familiarization in the label-defined and motion-defined conditions in Experiment 5.
Figure 13. Average number of valid fixation shifts within each block during familiarization in the label-defined and motion-defined conditions in Experiment 5.
Figure 14. Average number of valid fixation shifts averaged across four blocks at each time point within familiarization trials in the label-defined and motion-defined conditions in Experiment 5. Each time point represents a 1000 ms time window.
A. Familiarization

![Bar chart showing proportion of accumulated looking time to each feature averaged across 4 blocks at familiarization (A) and test (B) in Experiment 6.](image1)

B. Test

![Bar chart showing proportion of accumulated looking time to each feature averaged across 4 blocks at familiarization (A) and test (B) in Experiment 6.](image2)

*Figure 15.* Proportion of accumulated looking time to each feature averaged across 4 blocks at familiarization (A) and test (B) in Experiment 6.