Development of Category-Based Reasoning in Preschool-Age Children: Preliminary Results of a Longitudinal Study

Karrie E. Godwin (kegodwin@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Bryan J. Matlen (bmatlen@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Anna V. Fisher (fisher49@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Abstract

Category-based reasoning is central to mature cognition; yet, the developmental course of this fundamental ability remains unclear. We designed a longitudinal study to investigate the development of category-based reasoning. We also took an individual differences approach to identify possible cognitive factors that may facilitate category-based reasoning. In this paper we report preliminary results of our longitudinal investigation into the development of category-based reasoning.

Keywords: Induction. Reasoning. Categories. Cognitive Development

Introduction

A great deal of prior research has investigated the development of category-based reasoning. This work suggests that the fundamental ability to make inferences on the basis of category labels (i.e., category-based reasoning) is early developing (Gelman & Coley, 1990; Gelman & Markman, 1986; Jaswal, 2004; Jaswal & Markman, 2007; Welder & Graham, 2001). In a simple test of this skill, Jaswal and Markman (2007, Experiment 1) presented 24-month-old children with pairs of familiar animals (e.g., *dog* and *cat*). The children watched as the animals engaged in specific activities (e.g., the cat drinks milk and the dog chews on a bone). A third hybrid-animal was then presented (e.g., a cross between a cat and dog) and children were asked to use the props to demonstrate which action the hybrid animal would make (e.g., drink milk or chew on a bone). Importantly, the hybrid animal was designed to look more similar to one of the targets (e.g., The hybrid animal was designed to look more similar to the cat). In the no-label condition the hybrid was referred to generically (e.g., the experimenter labeled the cat-like animal as "*this one*"). In the label condition the hybrid animal was labeled counter-intuitively (e.g., the experimenter labeled the cat-like animal a "*dog"*). In the no-label condition Jaswal and Markman found that 24-month-olds generalized based on perceptual similarity 69% of the time. However, in the label condition, when perceptual similarity was pitted against category information, perceptually-based generalizations dropped to 37%. These results suggest children as young as 24-months of age can utilize labels to infer category membership.

In a seminal study, Gelman and Markman (1986) examined children's ability to make inductive inferences using category information that was conveyed by synonymous labels. In this experiment, preschool-aged children were presented with triads of objects and provided with respective labels. The children were told that two of the objects possessed particular properties, and the children were asked to infer which property the third object possessed. For example, children were presented with a bunny and a squirrel and told that the bunny eats grass, and the squirrel eats bugs. Subsequently, the children were asked to determine whether the rabbit ate grass like the bunny or bugs like the squirrel. Gelman and Markman found that children made category-based inductions 63% of the time, which is slightly above chance, and posited that preschool children are sensitive to the cues synonymous labels provide about category membership.

Despite these intriguing findings, there is mounting evidence demonstrating that the course of category-based reasoning follows a more protracted developmental course (Fisher, 2010; Fisher, Matlen, & Godwin, 2011). For example, Fisher et al. (2011) found that children's ability to make inductive inferences using synonyms is limited to a small set of semantically-similar words that co-occur in child-directed speech according to the CHILDES database (MacWhinney, 2000). In particular, Fisher et al. found that most 4-year-old children were able to perform category-based inferences with synonyms that are likely to co-occur in child-directed speech (e.g., *bunny-rabbit, puppy-dog*); however, they were unlikely to make category-based inferences with non co-occurring synonyms (e.g., *alligator-crocodile*, *rock-stone*). This pattern of results was found with both natural kinds and artifacts. Additionally, children's reliance on category information was found to improve gradually with age. Although 5-year-olds evidenced improvement in their reliance on category information compared to 4-year-olds, the majority of children did not reliably utilize category information

conveyed by non-co-occurring semantically-similar labels until six years of age.

If category-based reasoning has a protracted developmental course, an important question to be addressed is identifying what actually develops that enables children to utilize labels as windows into categories and reliably use this information in the course of induction.

One possibility is that advances in category-based reasoning are facilitated by changes in how children organize knowledge. There is evidence that children begin to organize concepts into networks by 21 months (Arias-Trejo & Plunkett, 2009). There is also evidence that conceptual organization changes over the course of development, with associative networks emerging prior to semantic networks (McCauley, Weil, & Sperber, 1976; Plaut & Booth, 2000). It is also possible that development of executive functioning may facilitate category-based reasoning by allowing children to disengage their attention from – often misleading or irrelevant – surface similarities and consider deeper relational similarities (Sloutsky & Fisher, 2005; Sloutsky, 2010).

The goal of the present research is to examine possible cognitive factors contributing to the development of category-based reasoning. Towards this goal we designed a longitudinal investigation taking an individual differences approach. Specifically, we collected measures of children's category-based reasoning at Time 1, verbal working memory, IQ, and semantic knowledge organization. Collection of additional measurements (i.e., inhibitory control, non-verbal working memory, semantic priming, and category-based reasoning at Time 2) is currently in progress. In what follows we report the preliminary results of this study.

Method

Participants

Participants were 43 four-year-old children from a local preschool (M_{age}=4.32 years, SD=0.28 years, 20 females, 23 males).

Materials & Procedure

Children were tested individually in a quiet room adjacent to their classroom by a trained research assistant. The tasks were administered across 6 sessions over the course of approximately 2 weeks. A detailed description of each task is provided below.

Category-Based Reasoning Task

The category-based reasoning task consisted of a triad induction task. Visual Stimuli were sets of three identical doors which were presented on a computer; see Figure 1. Verbal stimuli included 9 label triads: 3 triads referring to animate natural kinds, 3 triads referring to inanimate natural kinds, and 3 triads referring to artifacts (see Table 1). The properties participants were asked to generalize consisted of two-syllable blank predicates. Each trial was comprised of a target item, a category-choice, and an unrelated lure (e.g., *rock-stone-grass*). The children were told that the objects were hiding behind doors. This design was employed to encourage children's reliance on category information conveyed via labels. This procedure has been successfully used in prior research (Fisher et al., 2011). On every trial children were told what was hiding behind each door.

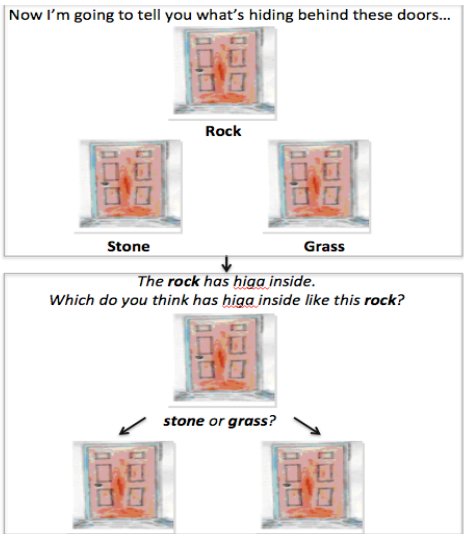

Figure 1: Schematic depiction of the category-based reasoning task. All instructions were given verbally by the experimenter.

Then, children were told that the target had a novel-property and they were asked to generalize the property to one of the test items (the category-choice or lure).

Table 1: Linguistic Stimuli for the Category-Based Reasoning Task

Target	Category Choice	Lure	Property
Rock	Stone	Grass	Higa
Alligator	Crocodile	Butterfly	Omat
Rug	Carpet	Window	Koski
Rat	Mouse	Fish	Lignin
Hill	Mountain	Flower	Erwin
Sea	Ocean	Apple	Manchin
Sofa	Couch	Cup	Creighan
Shoe	Boot	Car	Troxel
Lamb	Sheep	Frog	Matlen

The trials were presented in one of two orders: all trials were randomized for order 1, and for order 2 the presentation was reversed. Presentation order was counterbalanced across participants. The reasoning task was administered during session 2 and again in session 4 to assess stability in children's generalization performance.

Picture Identification Task

The picture identification task served to assess children's familiarity with the labels used in the reasoning task. Verbal

stimuli included 27 labels (the target, category-choice, and lure from the reasoning task). Visual stimuli consisted of a set of 108 pictures presented on a computer. All instructions and labels were given by hypothesis-blind experimenters. The picture identification task is similar to the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). On each trial children were asked to point to the object labeled by the experimenter from 4 pictorial response options (the target and 3 lures). The trials were presented in one of two orders. Presentation order was counterbalanced across participants.

Intelligence Test

IQ materials consisted of a commercially purchased intelligence test, the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI). The IQ test was administered in order to assess if children's reasoning performance was related to their general intelligence and/or a particular intelligence component. Eight of the WPPSI subscales were administered over 3 testing sessions in order to obtain an index of children's Verbal IQ, Performance IQ, Processing Speed Quotient, and Full Scale IQ. The WPPSI was administered by the first author of this paper and two trained research assistants.

Verbal Working Memory Tasks

Children's verbal working memory capacity was assessed using a simple and complex word-span task. Verbal stimuli entailed 60 words that were arranged into 6 sets. Set length ranged in size from a list length of 2 words to 6 words. Each set was comprised of 3 lists of the same length.

In the simple word-span task, children listened to the experimenter read a series of familiar count nouns, as judged by the MacArthur Communicative Development Inventory (Dale & Fenson, 1996). Then, children were asked to recite the words in the same order in which they were presented. The number of words in each set increased monotonically after children correctly completed two out of three trials within a given set size. For example, if children correctly completed 2 trials with set size 2, they then moved on to set size 3 (for a minimum of 2 trials or a maximum of 3 trials), and then set size 4 (for a minimum of 2 trials or a maximum of 3 trials), and so on until children made two errors within a set at which point testing stopped. The child's score is the longest list length he or she recited successfully. The complex word-span task was identical to the simple word-span task, except that children were asked to repeat items in the reverse order in which they were presented (e.g., If children were given the string, "*duck, house, chair*", the correct response would be "*chair, house, duck*"). The word lists were presented in one of two orders. Presentation orders were counterbalanced across participants.

The simple word-span task was included in the assessment battery in order to assess if children's performance on the category-based reasoning task was related to their general working memory capacity. The complex word-span task was included as it was hypothesized that the complex word-span task more closely resembled the demands of the reasoning task itself (e.g., both the category-based reasoning task and the complex word-span task contain a memory component as well as a transformation/processing component).

Semantic Space Task

Visual stimuli included a game board consisting of a 9x9 grid. Two 1" wooden cubes served as the game pieces. Verbal stimuli consisted of 24 animal pairs. The list of linguistic stimuli is provided in Table 2. In the semantic space task, children are asked to help *Zibbo the Zookeeper* organize his zoo. Children are told that Zibbo wants to put animals of the same kind close together. Children are presented with 24 animal pairs (i.e., a target animal and a test item). Of the 24 animal pairs, 6 dyads were semantically-similar (e.g., *lamb-sheep*), 6 dyads shared a common habitat or setting (e.g., *lamb-horse*), 6 dyads were physically similar – according to size and/or color (e.g., *lamb-swan*), and 6 dyads served as filler trials. Note that the target animal was paired with 3 different animals throughout the game - the category-choice, a physically-similar item, and the habitat match. On each trial, the experimenter shows the child where Zibbo put the target animal (e.g., the experimenter places the game piece on a designated space on the board and tells the child, *"The zookeeper put the **crocodile** here"*). Then, the experimenter hands the child the second game piece and asks the child where the test item should go (e.g., *"Where do you think the **grasshopper** should go?"*). The board is then cleared and the experimenter presents the next dyad. The child's response on each trial is recorded so the distance between the target animal and test item can be calculated.

Placement of the 18 critical trials (i.e., semantically-similar dyads, physically similar dyads, or similar habitat dyads) was pseudo randomized to eight potential squares; see Figure 2. Each square was utilized at least twice and no more than three times. The 6 filler trials were randomly assigned to one of the remaining 24 squares in order to encourage participants to use the entire game board. Trials were presented in one of two orders and presentation orders were counterbalanced across participants.

Table 2: List of Stimuli for the Semantic Space Task

Critical Trials			
Target	**Category - Choice**	**Physical Similarity**	**Habitat**
Crocodile	Alligator	Grasshopper	Fish
Chick	Hen	Goldfish	Goat
Lamb	Sheep	Swan	Horse
Whale	Dolphin	Elephant	Octopus
Monkey	Gorilla	Chipmunk	Parrot
Mouse	Rat	Hippo	Pig
Filler Pairs			
1. Zebra/Turkey; 2. Bear/Snake; 3. Panther/ Turtle; 4. Tiger/Butterfly; 5. Frog/Lion; 6. Giraffe/Seal			

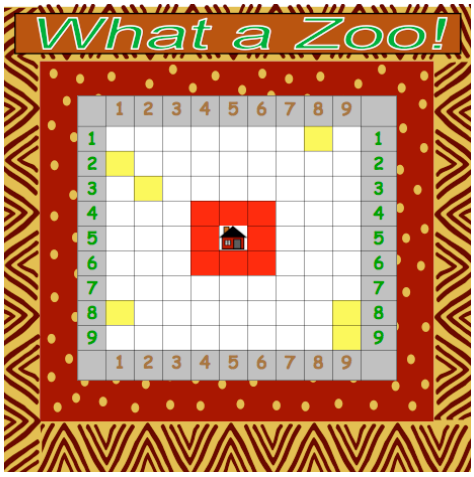

Figure 2: Depiction of the game board for the semantic space task. Red squares indicate the location of the critical trials and the yellow squares mark the location of the filler trials.

The semantic space task was included in the assessment battery to assess whether the organization of children's semantic space was related to their performance on the category-based reasoning task. Specifically, we were interested in identifying how children weight different dimensions (e.g., semantic-similarity, physical similarity, and habitat) and whether the distribution of weights to various dimensions enhances or hinders children's ability to successfully make category-based inductions.

Results

Picture Identification

The results of the picture identification task suggest that children possessed the prerequisite knowledge to perform category-based induction as children were highly familiar with the labels used in the reasoning task (M=0.92, SD=0.14). Additionally, the correlation between children's performance on the picture identification task and children's average reasoning score was only marginally significant (r=.28, p=0.07).

Category-Based Reasoning Task

As stated previously, performance on the category-based reasoning task was measured twice over the course of 1 week in order to examine the stability of this measure. Mean category-based reasoning at Time 1a and 1b were very similar (M=0.62, SD=0.22; M=0.63, SD=0.26 respectively) and these measures were significantly correlated (r=.483, p=0.001). Proportions of category-based responses were compared to chance level (0.5) using single-sample t-tests. All mean scores (scores at Time 1a & 1b and average reasoning score) were significantly above chance; all t's>3.30, all p's<0.0022.

To investigate individual patterns of responses, participants were classified as either category-based or non-category-based responders. A category-based responder was defined as a participant who gave a category-based response on at least 7 out of 9 (78%) trials (binomial probability = 0.09). At Time 1a and 1b only a small percentage of children were classified as category-based responders (33% and 37% respectively).

To further investigate stability in children's category-based reasoning performance we also examined whether children's classification remained stable across Time 1a and 1b. We found that 67% (29 out of 43) of children were categorized as stable across Time 1a and 1b. Of these children only 19% (8 out 43) were classified as consistently category-based responders, 49% (21 out of 43) were consistently non category-based, and 33% (14 out of 43) were considered unstable responders; See Figure 3. For the purposes of the remaining analyses the average reasoning score was utilized (M=0.63, SD=0.21).

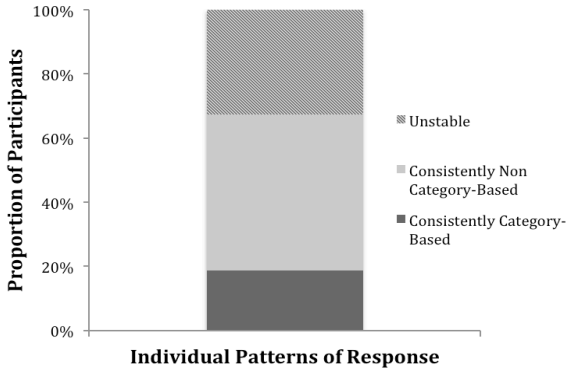

Figure 3: Proportion of children classified as consistently category-based, consistently non-category-based, or unstable responders.

Intelligence Test

Children's mean composite IQ scores and their Full Scale IQ were in the average range (M_{VIQ}=107.36, SD=24.30; M_{PIQ}= 107.26, SD=14.78; M_{PSQ}=92.78, SD=23.56; M_{FSIQ}=107.05, SD=15.79). Children's Verbal IQ and Performance IQ composite scores were significantly correlated with their average performance on the category-based reasoning task (r=.33, p=0.03; r=.31, p=0.05 respectively). However, Processing Quotient was not significantly correlated with children's average reasoning score (r=.15, p=0.35). Children's Full Scale IQ was also significantly correlated with their average reasoning score (r=.50, p=0.001).

Verbal Working Memory Tasks

Children's performance on the simple word-span task was better than their performance on the complex word-span task (M=3.07, SD=1.32; M=1.28, SD=1.14 respectively). A

mean score of 3.07 on the simple word-span task indicates that on average children were able to successfully recall a list length of 3 words. Children's score on the simple word-span task was found to be correlated with their average performance on the category-based reasoning task (r=.35, p=0.02). A mean score of 1.28 on the complex word-span task suggests that many children obtained a score of 0 on the task as the smallest list length was 2 words. Performance on the complex word-span task was not significantly correlated with children's average induction performance (r=.08, p=0.60), possibly due to floor effects on the complex word-span task.

Semantic Space Task

Children's semantic space score was calculated in the following way: First, for each child an average score for each category was calculated (i.e., an average score for semantically-similar dyads, an average score for similar habitat dyads, and an average score for physically similar dyads). Children's scores for similar habitat dyads and physically similar dyads were averaged together to create an average score for non-semantically-similar dyads. This score was subtracted from the average score for semantically-similar dyads to obtain a difference score. Larger difference scores indicate that children placed semantically-similar dyads closer together and non-semantically-similar dyads farther apart. Smaller difference scores indicate that children did not reliably discriminate between semantically-similar dyads and non-semantically-similar dyads.

Table 3: Correlation Matrix

(**Note**: CB Reasoning = Category-based reasoning, Pic ID = picture identification, SWS = simple word-span, CWS = complex word-span, SS = semantic space, VIQ = verbal IQ, PIQ = performance IQ, PSQ = processing speed, FSIQ = full scale IQ)

	1	2	3	4	5	6	7	8	9
1. CB Reasoning	-	.283	.345*	.082	.460**	.332*	.305*	.149	.496**
2. Pic ID	-	-	.706**	.275	.228	.723**	.100	-.157	.584**
3. SWS	-	-	-	.415**	.313*	.671**	.182	-.038	.578**
4. CWS	-	-	-	-	.126	.170	.188	.009	.202
5. SS	-	-	-	-	-	.384*	.432**	.145	.526**
6. VIQ	-	-	-	-	-	-	.049	-.314*	.800**
7. PIQ	-	-	-	-	-	-	-	.598**	.540**
8. PSQ	-	-	-	-	-	-	-	-	.123
9. FSIQ	-	-	-	-	-	-	-	-	-

Note. *. Correlation is significant at the 0.05 level. **. Correlation is significant at the 0.01 level

Children's mean score for semantically-similar dyads was 4.37 (SD=1.35). Children's score for physically similar dyads and similar habitats was 5.64 (SD=1.76) and 5.83 (SD=1.60) respectively. Children's mean score for non semantically-similar dyads was 5.74 (SD=1.51). Difference scores ranged from -2.58 to 5.67 suggesting considerable variability in children's performance on this task. The average difference score was 1.37 (SD=1.88). Children's performance on the semantic space task was found to be significantly correlated with their average performance on the category-based reasoning task (r=.46; p=0.002).

Predicting Category-Based Reasoning Performance

There were a total of 8 possible predictors of children's reasoning performance. As can be seen in Table 3, several of the predictors were significantly correlated with each other. Concerns regarding collinearity are allayed as tolerance values for predictors entered into the regression model were within the acceptable range. Children's scores on the simple word-span task, semantic space task, and FSIQ were entered into the model as predictors. Children's average score on the category-based reasoning task was the dependent variable.

The regression model significantly predicted children's average reasoning score, $F(3, 38)=5.47$, p=0.003. The R squared value indicates that 30% of the variance in children's performance on the category-based reasoning task was explained by the model. Simple word-span was not found to be a significant predictor (β=0.08, p=0.62); however, semantic space (β=0.26, p=0.11) and FSIQ (β=0.31, p=0.10) were marginally significant predictors of children's reasoning performance.

Discussion

Overall, the results from this study, although preliminary, point to several findings. First, the analysis of individual patterns of response on the reasoning task replicate previous work (Fisher et al., 2011; Godwin, Matlen, & Fisher, 2011). Specifically, we found that when young children are presented with non-co-occurring semantically-similar labels only a small percentage of children spontaneously engage in category-based reasoning.

Second, the present findings suggest that several factors are related to children's induction performance. We found that children's Full Scale IQ, Verbal IQ, Performance IQ, simple working memory, and semantic space performance were all significantly correlated with children's average reasoning score. Additionally, children's Full scale IQ and performance on the semantic space task were identified as unique predictors of children's performance on the category-based reasoning task according to the regression model. The correlation between semantic organization and category-based reasoning is also consistent with related research on the development of analogical reasoning, which has suggested that children's shift from focusing primarily on perceptual similarity to relational similarity is mediated by increases in domain knowledge (Rattermann & Gentner,

1998). On-going research will examine whether other cognitive factors (e.g., inhibitory control, non-verbal working memory, etc.) are also related to the development of category-based reasoning.

Third, the present study provides novel information on the stability in children's induction performance. This study is the first to our knowledge to look at the stability in children's performance on a conceptual development measure. The findings from this study suggest that children's category-based reasoning performance between Time 1a and 1b is correlated; however, there is still a great deal of variability in children's performance as indicated by the small percentage of children who were classified as consistently category-based across the two time points. Additionally, the longitudinal component of this study will enrich our understanding of the stability in children's inductive reasoning. Once data collection is complete, we will be able to examine whether the percentage of children who are classified as consistently category-based increases with age.

In conclusion, the present study contributes to our understanding of children's emerging ability to engage in category-based reasoning. The contributions of this work include identifying potential factors that may be predictive of children's induction performance as well as the opportunity to investigate the stability of children's category-based reasoning. Future research is needed to extend these findings and disentangle the different hypotheses put forth to explain this fundamental aspect of conceptual development.

Acknowledgements

We thank Malika Sinha, Laura Pacilio, Alyssa Montanaro, Anna Loiterstein, Like Li, and Kayoung Joung for their help collecting data. We also thank the children, parents, and teachers who made this project possible. This work was supported in part by a Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education (R305B090023 and R305B040063).

References

Arias-Trejo, N. & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society Biological Sciences,* 364, 3633-3647.

Dale, P. S., and Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, & Computers, 28*, 125-127.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition.* Circle Pines, MN: AGS Publishing.

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

Fisher, A.V., Matlen, B.J., & Godwin, K.E. (2011). Semantic similarity of labels and inductive generalization: Taking a second look. *Cognition*, 118, 432-438.

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

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

Godwin, K., Matlen, B., & Fisher, A. (2011). The influence of co-occurrence and inheritance information on children's inductive generalization. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (2989-2994). Austin, TX: Cognitive Science Society.

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

Jaswal, V. K. & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*, 8(1) 93-111.

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

McCauley, C., Weil, C. M., & Sperber, R. D. (1976). The development of memory structures as reflected by semantic-priming effects. *Journal of Experimental Child Psychology*, 22, 511-518.

Plaut, D.C. & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*(4), 786-823.

Rattermann, M. J. & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development, 13*, 453 – 478.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, 34, 1244–1286.

Sloutsky, V. M., & Fisher, A. V. (2005). Similarity, Induction, Naming, and Categorization (SINC): Generalization or verbal inductive reasoning? Response to Heit and Hayes. *Journal of Experimental Psychology: General*, 134, 606-611.

Welder, A.N., & Graham, S.A. (2001). The influences of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, *72*, 1653-1673.