Title
Memory Dynamics in Cross-Situational Statistical Learning

Author
Vlach, Haley Amelia Heublein

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Memory Dynamics in Cross-Situational Statistical Learning

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychology

by

Haley Amelia Heublein Vlach

2012

ABSTRACT OF THE DISSERTATION

Memory Dynamics in Cross-Situational Statistical Learning

by

Haley Amelia Heublein Vlach

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2012

Professor Catherine Sandhofer, Chair

A central pursuit in cognitive science and developmental psychology has been to characterize how humans encode a seemingly infinite amount of information, interpret this information, and use the information at later points in time. For example, in the domain of language learning and development, researchers have long sought to determine how infants and adults are able to determine word-referent pairings, despite the infinite number of possibilities (Quine, 1960). More recent research has suggested that learners are able to determine word-referent pairings by tracking co-occurrence probabilities across time, a behavior commonly termed *cross-situational statistical learning*.

The current series of experiments built upon research on cross-situational statistical learning by examining learning over varying timescales, from a matter of seconds to up to one week. In particular, this work examined the mechanisms underlying learning that promoted and/or deterred the learning and long-term retention of cross-situational statistics. Experiments 1

– 3 presented adult learners with a cross-situational learning task and then presented learners with a forced-choice inference test immediately or one week later. Experiments 2-3 examined how retrieval dynamics, the ease and/or difficulty of retrieving information while learning, was related to long-term retention and inference performance. Experiment 4 presented infant learners with a cross-situational learning task that manipulated the timing at which word-referent pairings were presented, requiring infants to retrieve prior potential pairings from long-term memory.

The results of the these studies indicate that adult learners are able to retain cross-situational statistics for up to one week later and that the amount of retention is related to ease and/or difficulty of retrieval during the learning process. However, young infants demonstrated constraints on their ability to retrieve information over short timescales, indicating that memory development in retrieval abilities may be critical to acquiring word-referent pairings. This work challenges broad theories of cognition and development that rest on retrieval processes as successful and automatic. Indeed, retrieving the past is a dynamic process, which undergoes dramatic developments over the lifespan, and should be incorporated into theoretical and computational accounts of learning.

The dissertation of Haley Amelia Heublein Vlach is approved.

Robert A. Bjork

Noel Enyedy

Scott P. Johnson

Catherine M. Sandhofer, Chair

University of California, Los Angeles

2012

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Portions of the research reported in my dissertation are currently published, under review, or in prep for publication. Specifically, research in Chapter 2 will be published with co-author Catherine M. Sandhofer and research in Chapter 4 will be published with co-author Scott P. Johnson. My dissertation research was supported by NIH Grant NICHD R03 HD064909-01 and NIH Grant R01-HD40432.

I would like to acknowledge my entire committee (Bob Bjork, Noel Enyedy, Scott Johnson, & Cathy Sandhofer) for their support and feedback on this work. A special thanks to my committee chair, Catherine M. Sandhofer—thank you for being an incredible mentor.

VITA / BIOGRAPHICAL SKETCH

ACADEMIC POSITIONS

Starting Fall 2012	Assistant Professor Human Development Area, Department of Educational Psychology University of Wisconsin, Madison

EDUCATION

2007-2012	Ph.D., University of California, Los Angeles Developmental Psychology	Los Angeles, CA
2006-2007	M.A., University of California, Los Angeles Developmental Psychology	
2002-2006	B.S., Carnegie Mellon University Major: Psychology; Phi Beta Kappa, Highest Honors	Pittsburgh, PA

PUBLICATIONS

Journal Articles (published and in press):

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*. Published online via EarlyView.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 246-254. doi: 10.1037/a0025260

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory mechanisms support children's ability to retain words. *Frontiers in Psychology, 3:46*. doi: 10.3389/fpsyg.2012.00046

Vlach, H. A. & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology, 108*, 394-401. doi: 10.1016/j.jecp.2010.09.011

Vlach, H. A., Krogh, L., Thom, E. E. & Sandhofer, C. M. (2010). Doing with development: Moving toward a complete theory. *Behavioral and Brain Sciences, 33*, 227-228. doi: 10.1017/S0140525X1000049X

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in young children's memory and category induction. *Cognition, 109*, 163-167. doi: 10.1016/j.cognition.2008.07.013

Vlach, H. A. & Carver, S. M. (2008). The effects of observation coaching on children's graphic representations. *Early Childhood Research & Practice, 10*, 1-15.

Vlach, H. A. & Carver, S. M. (2006). The impact of observation on children's drawings. *National Association of Laboratory Schools Journal, 30*, 6-15.

Conference Proceedings:

Vlach, H. A. & Johnson, S. P. (2011). *Memory constraints on infants' statistical word learning* [Abstract]. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 829). Boston, MA: Cognitive Science Society.

Vlach, H.A., & Sandhofer, C. M. (2011). *Retrieval dynamics of in-the-moment and long-term statistical word learning*. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 789-794). Boston, MA: Cognitive Science Society.

Vlach, H. A., & Sandhofer, C. M. (2010). *Desirable difficulties in cross-situational word learning*. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 2470-2476). Austin, TX: Cognitive Science Society.

Vlach, H. A., & Sandhofer, C. M. (2008). *Context and induction: The impact of background context on children's category learning* [Abstract]. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (p. 1573). Austin, TX: Cognitive Science Society.

RESEARCH GRANTS

2011-2013	NICHD R03 HD064909-01 Desirable Difficulties in Children's Word Learning PIs: Catherine Sandhofer, Robert Bjork, & Haley Vlach Priority Score: 12, Percentile: 1.0
2009-2010 & 2008-2009	Excellence in Research Grant University of California, Los Angeles, PI: Haley Vlach
2007-2008	CONNECT Research Grant University of California, Los Angeles, PI: Haley Vlach
2007-2008	Rapid Start Research Grant, University of California, Los Angeles, PI: Haley Vlach

CHAPTER 1: An Introduction to Cross-Situational Statistical Learning

How Do We Learn Words?

In one moment in time, the world presents learners with a seemingly infinite amount of information - including perceptual, linguistic, conceptual, social, and cultural information. In cognitive science and developmental psychology, a large research pursuit has been to characterize how it is that learners acquire, store, and later retrieve such a large data set of information. For example, in the sub-domain of language learning and development, consider a scenario first proposed by Quine (1960). In this scenario, a linguist is sent to a new culture in order to learn a new language. Upon arriving, a native comes up to the linguist, points out into a field at a rabbit, and says “gavagai”. The question raised by Quine (1960) is: How does the linguist determine what “gavagai” refers to? Is it the rabbit? Is it the field? Is it the color of the rabbit? Is it the action of hopping? Quine (1960) argued that, because the world offers learners an infinite number of word-to-world mappings, choosing the correct referent for a single word should be impossible.

How do young children and adults learn words despite the inherent difficulty of the task? Historically, research has sought to understand how learners are able to reduce ambiguity of potential word-referent mappings in a single moment of time. The results of this research are three main classes of theories: the Constraints/Principles theories, the Social-Pragmatic theories, and the Domain-General theories, described below. All three of these theories propose tools that young learners use to reduce ambiguity.

The Constraints/Principles theories suggest that word learning is made easier and more feasible by constraints that narrow the search space for possible word-to-world mappings, such as mutual-exclusivity (e.g., Markman, 1989) and the novel-name nameless-category assumption (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994). These constraints guide learners' interpretations of new words and thus reduce the degree of indeterminacy. For example, the mutual-exclusivity principle (Markman, 1989) proposes words have mutually-exclusive meaning—one object can have only one referent. Consequently, when learners hear an unfamiliar label, they will assign the unfamiliar label to an unfamiliar object rather than an object that has already been named. In sum, according to these theories, even young infants possess heuristics that are used to constrain the number of possible referents for one word, easing the ability to learn words.

A second class of theories, the Social-Pragmatic theories, propose that word learning is simplified because learners are embedded in a social world in which they are guided by expert word learners (e.g., Bloom, 1993; Tomasello & Barton, 1994). Adults, as expert word learners, resolve the ambiguity of the word-learning scenario by guiding children's attention and thus make the task of word learning easier. For example, adults commonly talk about objects, events, and actions that young learners are already focused on and consequently make it easier for learners to make word-to-world mappings (Bloom, 1993). Moreover, young children look to adults for subtle social cues to guide learning (Tomasello & Barton, 1994). The Social-Pragmatic theories suggest that there is a powerful bi-directional relationship between children and adults that results in a narrow space and small number of referents for mapping words.

Finally, a third class of theories, the Domain-General theories, assert that general cognitive mechanisms such as perceptual saliency, association, and frequency make word learning straightforward (e.g., Samuelson & Smith, 1998; Smith, 1995)—learners notice objects and actions that are most salient in their environment and pair them with the most frequently associated label. For example, in one study (Samuelson & Smith, 1998), children were able to learn a novel word-novel object link by using saliency cues in the absence of other cues, such as social cues, suggesting that saliency cues alone guided children's word learning. In sum, the Domain-General theories argue that the basic processes of attention, perception, and memory are powerful enough to reduce the ambiguity of mapping words to referents in the world.

Moving Beyond One Moment in Time: Cross-Situational Statistical Learning

In recent years, researchers have moved toward examining how it is that learners resolve ambiguity across moments in time, rather than in one moment in time. This research pursuit was born out computer science (e.g., Siskind, 1996) and is most commonly termed *cross-situational statistical learning* (e.g., Blythe, Smith, & Smith, 2010; Fazly, Alishahi, & Stevenson, 2010; Fitneva & Christiansen, 2011; Frank, Goodman, Tenenbaum, 2009; Scott & Fisher, 2012; Smith, Smith, & Blythe, 2010; Smith & Yu, 2008; Yu & Smith, 2007, 2011, 2012). This research has sought to determine if learners can resolve ambiguity across learning events by tracking co-occurrence of words and referents.

In a typical experiment of cross-situational statistical learning, learners are presented with a series of learning events which include multiple words and multiple objects. After learning, participants are presented with a test in which they infer object-label pairings. For example, in Yu & Smith (2007), adult participants were presented with a series of learning events that consisted of either two words and two objects, three words and three objects, or four words and four objects. Participants' ability to track co-occurrence of words and objects was tested at a final test, where learners were asked to infer which word(s) corresponded to the objects presented during learning. If learners used cross-situational co-occurrence statistics to guide word mapping, then they should have selected object-label pairings based upon words and objects that were always presented in the same learning event. Indeed, this is what the results of this study revealed – the participants were able track co-occurrence information to guide word mapping. This body of work has revealed that both adults (e.g., Yu & Smith, 2007) and young infants (e.g., 12- and 14-month-olds; Smith & Yu, 2008) are able to learn to learn word mappings by tracking co-occurrence probabilities across learning events.

To date, research on cross-situational statistical learning has focused on *what* questions – what can be learned, what degree of ambiguity can be resolved, what equations/mathematical functions can be used to model learning processes, etc. Because research on cross-situational statistical learning is in its infancy, the majority of studies have been demonstrative rather than mechanistic and explanatory (for a review of this issue, see Yu & Smith, 2012). Thus, although research has demonstrated that infants and adults are able to track co-occurrence probabilities in order to guide word mapping, very little is known about the mechanisms underlying cross-situational statistical learning.

The current set of studies builds upon current research on cross-situational learning by identifying mechanisms underlying the acquisition and retention of statistics. This research examines *how* questions – how does the way that information is encoded affect learning outcomes, how are cross-situational statistics retained, how do memory processes facilitate/limit learners ability to acquire cross-situational statistics, etc. These questions seek to build a mechanistic account for how cross-situational statistics are acquired, stored, and retrieved over time.

Chapter 2 describes two experiments that examined if and how cross-situational statistics are retained over extended periods of time. Moreover, the experiments in this chapter identified a component of learning (the retrieval dynamics during the acquisition of cross-situational statistics) that may be a key mechanism that supported acquiring and retaining statistics over extended periods of time. Chapter 3 describes an experiment that expanded upon the research in Chapter 2. In particular, the experiment in Chapter 3 built upon work in Chapter 2 by experimental manipulating the timing at which learners were presented with object-label pairings. Chapter 4 extends research in Chapters 2 & 3 by examining cross-situational statistical learning during infancy. Finally, Chapter 5 provides a discussion of the work by outlining the theoretical implications of this work.

References for CHAPTER 1

Bloom, L. (1993). *The transition from infancy to language: Acquiring the power of expression*. Cambridge, England: Cambridge University press.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science, 34*, 620-642. doi: 10.1111/j.1551-6709.2009.01089.x

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *6*, 1017-1063. doi: 10.1111/j.1551-6709.2010.01104.x

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*, 367-380. doi: 10.1111/j.1551-6709.2010.01156.x

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*, 578-585. doi: 10.1111/j.1467-9280.2009.02335.x

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*, 125-155.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94–104. doi: 10.2307/1132073

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122*, 163-180. doi: 10.1016/j.cognition.2011.10.010

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science, 35*, 480-498. doi: 10.1111/j.1551-6709.2010.01158.x

Smith, L. B. (1995). Self-organizing processes in learning to learn words: Development is not induction. In C. A. Nelson (Ed.), *Basic and applied perspectives on learning, cognition, and development: The Minnesota Symposia on Child Psychology* (Vol. 28, pp. 1-32). Mahwah, NJ: Erlbaum.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568. doi: 10.1016/j.cognition.2007.06.010

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-91. doi: 10.1016/S0010-0277(96)00728-7

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology, 30*, 639-650. doi: 10.1037/0012-1649.30.5.639

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414-420. doi: 10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165-180. doi: 10.1111/j.1467-7687.2010.00958.x

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21-39. doi: 10.1037/a0026182

CHAPTER 2: Retrieval Dynamics and Retention in Cross-Situational Statistical Word Learning

In any single moment in time, the world presents a seemingly infinite number of possible referents for just one word (Quine, 1960). However, despite the ambiguity and inherent difficulty of mapping words to referents, children and adults appear to learn words with great ease. In fact, by age 6 children typically know approximately 14,000 words (Templin, 1957). Thus, a central research question has been: how do we learn words despite the ambiguity and difficulty of the task?

Historically, word learning research has focused on identifying the processes involved in resolving ambiguity in one moment in time. This work has revealed that young children and adults use several mechanisms, such as basic cognitive processes (e.g., Samuelson & Smith, 1998; Smith 2000), social/cultural cues and dynamics (e.g., Akhtar, Carpenter, & Tomasello, 1996; Baldwin, 1993; Tomasello & Barton, 1994), and heuristics/constraints (e.g., Gleitman, 1990; Markman, 1989). These processes reduce the number of potential referents for a word and, in turn, support the ability to map words to referents.

More recent research has begun to examine how learners resolve ambiguity across several moments in time. This research has revealed that learners track co-occurrence of words and referents across multiple learning events. Learners then use the co-occurrence statistics to guide the inference of word-referent pairings. This behavior is commonly termed *cross-situational statistical word learning* (e.g., Blythe, Smith, & Smith, 2010; Fazly, Alishahi, & Stevenson, 2010; Fitneva & Christiansen, 2011; Frank, Goodman, Tenenbaum, 2009; Scott & Fisher, 2012; Siskind, 1996; Smith, Smith, & Blythe, 2010; Smith & Yu, 2008; Yu & Smith, 2007, 2011, 2012). This work has revealed that adult learners can track co-occurrence of word-referent pairings with varying numbers of words and referents (e.g., Yu & Smith, 2007) and under conditions of high uncertainty (e.g., Smith et al., 2010).

The vast majority of research on cross-situational word learning has focused on learners' immediate acquisition and inference of word-referent pairings (e.g., Fitneva & Christiansen, 2011; Scott & Fisher, 2012; Smith & Yu, 2008; Yu & Smith, 2007, 2011). That is, most paradigms present participants with a series of ambiguous learning trials and then have participants infer the word-referent pairings at an immediate test. Consequently, very little is known about the long-term retention of cross-situational statistics.

Do learners retain cross-situational statistics over extended periods of time? In real-world word learning, learners are likely to experience a delay between learning events and situations in which they infer the meanings of words. Thus, a complete theory of cross-situational learning (and broader theories of word learning) must account for how word-referent pairings are retained across time. This work takes an important first step in examining whether learners can retain cross-situational statistics for extended periods of time and, if they are able to retain statistics, how low-level memory processes support the ability to do so.

In this paper, we report a series of experiments that were designed to examine learners' long-term retention of cross-situational statistics. In both Experiment 1 and 2, learners' acquisition and retention of word-referent (i.e., object-label) pairings was tested at an immediate or one week delayed forced-choice test. Experiment 2 was also designed to reveal how in-the-moment learning dynamics may be supporting and/or deterring the ability to retain cross-situational statistics. Specifically, we examined the retrieval dynamics occurring during learning. Previous research has indicated that struggling to retrieve information/engaging in deeper retrieval (e.g., Carpenter & DeLosh, 2006; Halamish & Bjork, 2011; Kornell, Hays, &

Bjork, 2009; Pyc & Rawson, 2009; Richland, Kornell, & Kao, 2009; Vlach, Ankwoski, & Sandhofer, 2012) and retrieval practice (e.g., Karpicke & Roediger, 2007; Roediger & Butler, 2011) can support to the long-term retention of information. We examined whether these dynamics occur during cross-situational word learning and how they may be related to long-term retention. In sum, these experiments took the important first steps in elucidating the mechanisms that support the long-term ability to retain cross-situational statistics.

Experiment 1

In this experiment, we started by examining whether or not learners would be able to retain cross-situational statistics over extended periods of time. Learners were presented with a cross-situational word learning task, with varying learning conditions, and tested immediately or one week later. If learners are able to retain cross-situational statistics, we predicted that performance would be above chance at the one week delayed test. If participants are not able to retain these statistics over extended periods of time, we predicted that performance would be at chance at the one week delayed test.

Method

Participants. Seventy-two undergraduate students in the department participant pool participated in this study. Participants were randomly assigned to one of the six between-subjects conditions of the experiment, resulting in 12 participants in each of the conditions. Participants received course credit for their participation.

Apparatus and Stimuli. Participants were presented with a cross-situational word learning task using a laptop computer. Pictures of objects were presented on the 15-inch computer screen and the sound for the labels was presented by the computer's speakers. As Figure 1 shows, the objects were pictures of novel objects. There were a total of 18 objects and

18 labels. The labels were novel words following the phonotactic probabilities of English (e.g., “blicket”, “dax”), presented in the same woman’s voice. Objects and labels were randomly paired together, for a total of 18 object-label pairs. In all conditions (2 x 2, 3 x 3, and 4 x 4), there were a total of six presentations of each of the 18 object-label pairs during the learning phase.

Design. This study used a 3 (Learning Condition) x 2 (Testing Delay) design. Learning Condition (2 x 2, 3 x 3, and 4 x 4) and Testing Delay (immediate and 1-week delay) were both between-subjects factors.

Procedure. The cross-situational word learning task consisted of three phases: a training phase, a learning phase, and a testing phase.

Training Phase. The training phase was designed to introduce participants to what the experiment would be like and how it would be ambiguous as to which words went with which objects during one learning trial. Participants were seated in front of the computer and told that they would be shown a series of children’s toys and hear novel words. After providing instructions, the experimenter presented the training learning trials. There were four novel objects and four novel labels presented; these object-label pairings were presented in three learning trials, each with two objects and two labels, immediately followed by one forced-choice testing trial. These objects and labels were not used during the learning or testing phases of the experiment.

Learning Phase. Following the training phase, the learning phase began and participants were randomly assigned to one of three learning conditions, 2 x 2, 3 x 3, and 4 x 4. In the 2 x 2 condition, two objects and two words were presented in each learning trial (see Figure 1). In the 3 x 3 condition, three objects and three labels were presented. In the 4 x 4 condition, four objects and four labels were presented. Because the same number of object-label pairs (18 pairs) were presented in each condition, the same number of times (6 presentations each), other presentation factors varied across conditions in order to ensure equivalent exposure to the object-label pairs. Table 1 outlines these variations, which were adapted from Yu and Smith (2007). Although the number of trials and time per trial varied, the total exposure time remained constant across the conditions (see Table 1).

Participants were presented with learning trials according to the condition in which they had been randomly assigned (2 x 2, 3 x 3, or 4 x 4). After viewing all of the trials in the learning phase, participants were presented with the testing phase according to the condition in which they had been assigned; participants in the immediate testing condition received the test trials immediately following the learning phase and participants in the 1-week delayed condition were asked to return one week later and complete the testing phase.

Testing Phase. The testing phase consisted of four four-choice trials (see Figure 1). Each testing trial presented one label over the computer's speakers and asked participants to identify the corresponding object among four objects. Participants were instructed to record their answers on a piece of paper. The three foil objects were other objects used in the experiment. No one object was repeated across testing trials. Hence, 16 of the 18 objects were presented during the test. The labels and objects used during the testing phase were randomly assigned.

Results and Discussion

We were interested in whether learners would be able to learn and retain cross-situational statistics over extended periods of time. In order to examine this question, we first conducted a 3 (Learning Condition) x 2 (Testing Delay) ANOVA, with the number of correct responses as the dependent measure. Results of this test revealed a significant main effect of learning condition,

$F(2, 66) = 19.086$, $p < .001$, $\eta_p^2 = .366$, a significant main effect of testing delay, $F(1, 66) = 31.641$, $p < .001$, $\eta_p^2 = .324$, and a significant interaction of learning and testing delay, $F(2, 66) = 6.070$, $p = .004$, $\eta_p^2 = .155$.

In order to characterize the nature of the interaction, we conducted two univariate ANOVAs within each testing condition. We then computed three planned comparisons using t-tests with Bonferroni corrections to determine the nature of the differences between the learning conditions within each testing delay condition. If there were in-the-moment learning dynamics that affected the long-term ability to retain cross-situational statistics, we would expect there to be differences in performance between the learning conditions and across the testing conditions. As can be seen in Figure 2, there appeared to be significant changes in the pattern of performance over time.

In the immediate testing condition, there was a main effect of learning condition, $F(2, 33) = 14.741$, $p < .001$, $\eta_p^2 = .472$ (see Figure 2). Participants in the 2 x 2 condition had significantly higher performance than participants in the 3 x 3, $p = .043$, and 4 x 4 conditions, $p < .001$. Moreover, performance in the 3 x 3 condition was significantly higher than the 4 x 4 condition, $p = .023$. Thus, the greater the number of object-label pairings in each learning trial, the lower the performance on an immediate test. This finding replicates that of Yu & Smith (2007).

However, in the one week delayed testing condition, there was a strikingly different pattern of performance (see Figure 2). There was a main effect of learning condition, $F(2, 33) = 10.482$, $p < .001$, $\eta_p^2 = .388$. Participants in the 3 x 3 condition had significantly higher performance than participants in the 2 x 2, $p = .048$, and 4 x 4 conditions, $p < .001$. Moreover, performance in the 2 x 2 condition was marginally significantly different than performance in the 4 x 4 condition, $p = .086$. Hence, although initially participants in the 3 x 3 condition had lower performance than participants in the 2 x 2 condition, one week later participants in the 3 x 3 condition had higher performance than participants in both the 2 x 2 and 4 x 4 conditions.

Why did we observe an interaction of retention across timescales? One possibility is that the in-the-moment dynamics of the three learning conditions engendered differences in the ability to retain cross-situational statistics over time. The fact that there were significant differences between each learning condition at the immediate test suggests that there could be differences of in-the-moment dynamics occurring during learning.

How did the in-the-moment dynamics differ across the three learning conditions? There has been a long history of research in memory tasks that has identified several processes underlying the ability to retain information over time (starting with Ebbinghaus, 1885/1964; also see Estes, 1955a, 1955b; Shiffrin & Atkinson, 1969; Tulving & Thomson, 1973). One such process that has often been shown to be related to long-term retention is the ability to retrieve information during learning (e.g., Kornell et al., 2009; Pyc & Rawson, 2009; Richland et al., 2009; Vlach et al., 2012). Interestingly, it appears that if learners struggle, but are eventually successful at retrieving information, there is a detriment to initial performance but stronger long-term retention. However, learners that engage in easier retrieval often have higher initial performance, but demonstrate poorer performance on a retention test. This pattern of performance is often termed the retrieval effort hypothesis (for a review, see Pyc & Rawson, 2009). Indeed, this pattern of performance parallels the interaction of test performance differences of the 2 x 2 and 3 x 3 conditions in Experiment 1.

Were there different retrieval dynamics occurring across the three learning conditions? We predicted that participants in the 2 x 2 condition were experiencing the greatest ease of retrieving object-label pairings during learning, compared to participants in the 3 x 3 and 4 x 4 conditions. For participants in the 3 x 3 condition, we predicted that they were experiencing an intermediate degree of difficulty retrieving object-label pairings. Because participants in the 3 x 3 condition demonstrated the highest long-term test performance, we predicted that they may have struggled to retrieve pairings initially, but obtained more success in retrieving pairings over the course of the learning phase. Finally, we predicted that participants in the 4 x 4 condition experienced the greatest degree of difficulty retrieving object-label pairings. We predicted this pattern of results because it would be consistent with the test performance results obtained in Experiment 1 and principles of memory processes and retention.

Experiment 2

In this experiment, we examined whether there were in-the-moment retrieval dynamics that differed across the three learning conditions. We hypothesized that participants in the 3 x 3 condition were experiencing more optimal retrieval processes than participants in the 2 x 2 and 4 x 4 conditions. Thus, in this experiment we used the same protocol as Experiment 1 but included a self-report retrieval task designed to capture the in-the-moment dynamics occurring during learning.

Method

Participants. Seventy-eight undergraduate students in the department participant pool participated in this study. Participants were randomly assigned to one of the six between-subjects conditions of the experiment, resulting in 13 participants in each of the conditions. Participants received course credit for their participation and had not participated in Experiment 1.

Apparatus and Stimuli. The computer and stimuli used in Experiment 1 were also used in Experiment 2. Additionally, the participants in Experiment 2 were provided with a worksheet in which to record their perceived ability to retrieve object-label pairings. As shown in Figure 3, the worksheet was a list of trial numbers and letters. This worksheet was used during the training and learning phases of the experiment. The experimenter collected the worksheet before the testing phase.

Design. Same as Experiment 1.

Procedure. The procedure used in Experiment 1 was also used in Experiment 2, with one exception. During the training and learning phases, participants were asked to record their perceived ability to successfully retrieve object-label pairings on a worksheet (for an example, see Figure 3). In the training phase, the experimenter provided instructions for how to record the retrieval successes on the worksheet in the training trials section. The experimenter demonstrated that the participants could circle one and/or multiple of the letters on the worksheet (e.g., in the 4 x 4 condition, ‘A’, ‘B’, ‘C’, and/or ‘D’) if they knew the word that corresponded to a particular object. If the participant was not able to successfully retrieve any object-label pairings, they were instructed to circle ‘None’. The participants recorded their responses on the worksheet during the three training trials.

After the training trials, the experimenter asked the participants if they had any clarification questions for how to record information on the worksheet. If a question was asked, the experimenter would repeat information provided during the learning phase, without extra elaboration. Following the training phase, the learning phase began and the participants recorded their retrieval successes for each trial of the learning phase on the worksheet. The worksheets contained the appropriate number of letters and trials according to the condition in which the participant was assigned (see Table 1).

Results and Discussion

Final Test Performance. We first examined the final test performance in order to see if the findings from Experiment 1 were replicated in Experiment 2. Experiment 2 included the additional demand of the retrieval task during training and learning, which could have resulted in differences on the final test performance. We conducted a 3 (Learning Condition) x 2 (Testing Delay) ANOVA, with the number of correct responses at the final test as the dependent measure (see Figure 4). Results of this test revealed a significant main effect of learning condition, $F(2, 72) = 12.582$, $p < .001$, $\eta_p^2 = .259$, a significant main effect of testing delay, $F(1, 72) = 9.573$, $p = .003$, $\eta_p^2 = .117$, and a significant interaction between learning condition and testing delay, $F(2, 72) = 5.808$, $p = .001$, $\eta_p^2 = .173$.

In order to examine the interaction, we conducted two univariate ANOVAs, one in each testing condition. We then computed three planned comparisons using t-tests with Bonferroni corrections to determine the nature of the differences between learning conditions within each testing delay condition. If the results were similar to Experiment 1, we expected there to be differences in performance between learning conditions across the testing conditions.

In the immediate testing condition, there was a main effect of learning condition, $F(2, 36) = 13.930$, $p < .001$, $\eta_p^2 = .436$. Participants in the 2 x 2 condition had significantly higher performance than in the 4 x 4 condition, $p < .001$. Participants' performance was also higher in the 2 x 2 condition than the 3 x 3 condition, $p = .049$. Finally, participants' performance in the 3 x 3 condition was significantly higher than the 4 x 4 condition, $p = .028$. Thus, the greater the number of object-label pairings in each learning trial, the lower the performance at an immediate test.

However, there was a different pattern of results in the one week delay condition. There was a main effect of learning condition, $F(2, 36) = 6.568$, $p = .004$, $\eta_p^2 = .267$. Participants in the 3 x 3 condition had higher performance than both the 2 x 2 condition, $p = .039$, and 4 x 4 condition, $p = .004$. Participants in the 4 x 4 condition did not have significantly different performance than participants in the 2 x 2 condition, $p > .05$. In sum, the pattern of final test performance seen in Experiment 1 was replicated in Experiment 2 (compare Figures 2 and 4).

Retrieval Task Performance. After examining the final test performance, we examined participants' performance on the self-report retrieval task during the learning phase of the experiment. We hypothesized that there may be differences of in-the-moment learning processes that could be contributing to differences in long-term performance. Specifically, we predicted that retrieval dynamics during learning could be a mechanism underlying in-the-moment and long-term performance. To explore this possibility, we analyzed participants' self-report of what they were successfully retrieving during learning. If there were differences in the number and timing of retrieval successes, this could be contributing to differences in immediate and long-term performance.

We started by examining the overall number of reported retrieval successes by learning condition. We conducted a univariate ANOVA with the overall number of reported retrieval successes as the outcome variable. We found a significant main effect of condition, $F(2, 75) = 19.769$, $p < .001$, $\eta_p^2 = .345$. We then computed three planned comparisons using t-tests with Bonferroni corrections to determine the nature of the differences between the learning conditions. Participants in the 2 x 2 condition reported a significantly higher number of retrieval successes ($M = 53.81$, $SD = 21.83$) than participants in the 3 x 3 condition ($M = 42.23$, $SD = 14.02$), $p = .058$, and participants in the 4 x 4 condition ($M = 23.65$, $SD = 15.48$), $p < .001$.

Moreover, participants in the 3 x 3 condition reported a significantly higher number of retrieval successes than participants in the 4 x 4 condition, $p = .001$. Thus, there were striking differences in the overall number of retrieval successes across the three learning conditions; the greater the number of objects and labels in each learning trial, the smaller the number of retrieval successes during learning.

In addition to the overall number of retrieval successes, we were also interested in the pattern of self-reported retrieval performance across the learning phase. In order to examine the in-the-moment ability to successfully retrieve object-label pairings, we started by dividing the learning phase into nine blocks of time, 36 seconds each. We chose this timescale because, over 36 seconds, participants in all of the conditions were exposed to the same number of object-label pairings. For example, in the 2 x 2 condition, there were 6 trials with 2 object-label pairings, for a total of 12 object-label pairings. In the 3 x 3 condition, there were 4 trials with 3 object-label pairings, for a total of 12 object-label pairings. Finally, in the 4 x 4 condition, there were 3 trials with 4 object-label pairings each, for a total of 12 object-label pairings.

After dividing the learning phase into nine timescales, we then computed the mean number of reported retrieval successes during each timescale, for each learning condition. Each time point (Time1 – Time9) represents the mean number of reported of retrieval successes between the previous time point and the current time point. For example, Time1 represents the mean number of retrieval success between Time0 (i.e., the beginning of the experiment) and Time1, Time 2 represents the mean number of retrieval successes between Time1 and Time2, and so forth. The descriptive results can be seen in Figure 5.

We then conducted a mixed 3 (Learning Condition) x 9 (Time Point) ANOVA, with learning condition as a between-subjects variable and time point as a within-subjects variable.

Results of this test revealed a significant main effect of learning condition, $F(2, 75) = 19.680$, $p < .001$, $\eta_p^2 = .344$, a significant main effect of time point, Wilks' Lambda = .221, $F(8, 68) = 30.009$, $p < .001$, $\eta_p^2 = .779$, and a significant interaction of learning condition and time point, Wilks' Lambda = .509, $F(16, 136) = 3.412$, $p < .001$, $\eta_p^2 = .286$.

In order to examine the interaction between learning condition and time point, we conducted a post-hoc analysis using planned comparisons between the three learning conditions, at each time point. To correct for all 27 comparisons, we computed a corrected alpha using Bonferroni standards ($\alpha = .05/27$, corrected $\alpha = .00185$). This alpha level was used for all of the planned comparisons.

The results of the post-hoc analyses revealed many differences between the three learning conditions, across time points, ps < .00185. We have categorized the nature of these differences into three distinct periods of the learning phase (see Figure 5). In the first period, the Beginning Period (Time0 – Time2), participants in the 2 x 2 condition reported significantly more retrieval successes than participants in the 3 x 3 and 4 x 4 conditions, at Time1 and Time2, ps < . 00185. There were no significant differences between the number of reported retrieval successes in the 3 x 3 and 4 x 4 conditions, at each time point, ps > .00185. Thus, during the early part of the learning phase, termed the Beginning Period, participants in the 2 x 2 condition were experiencing a greater ease in retrieval compared to participants in the 3 x 3 and 4 x 4 conditions.

During the next period of the learning phase, the Transition Period (Time3 – Time6), there were significant differences between all three learning conditions at Time3 – Time6, ps < .00185. That is, participants in the 2 x 2 condition reported significantly more retrieval successes than participants in the 3 x 3 and 4 x 4 conditions. Moreover, participants in the 3 x 3 condition reported significantly more retrieval successes than participants in the 4 x 4 condition. Hence, in the Transition Period, participants in the three learning conditions were experiencing three different degrees of ease in retrieving information, with participants in the 3 x 3 condition experiencing an intermediate degree of difficulty compared to participants in the 2 x 2 and 4 x 4 conditions.

Finally, in the last period of the learning phase, the End Period (Time7 – Time 9), there were also significant differences between the learning conditions, but differences which followed a strikingly different pattern than in the earlier periods. First, there were no significant differences in the number of reported retrieval successes between participants in the 2 x 2 and 3 x 3 conditions at all time points, Time7 – Time9, ps > .00185. Moreover, participants in the 2 x 2 and 3 x 3 conditions reported significantly more retrieval successes than participants in the 4 x 4 condition, at all time points, ps < .00185. Thus, in the last period of the experiment, there were no differences in the degree of retrieval difficulty experienced by participants in the 2 x 2 and 3 x 3 conditions.

In sum, there were significant changes in the pattern of reported retrieval successes across the learning phase of the 3 x 3 condition. Initially (during the Beginning Period), participants in the 3 x 3 condition appeared to struggle to retrieve object-label pairings; retrieval performance was significantly lower than retrieval performance in the 2 x 2 condition and not significantly different than that of retrieval performance in the 4 x 4 condition. In the middle part of the learning phase (during the Transition Period), participants in the 3 x 3 condition were reporting an intermediate degree of retrieval success. Finally, in the last part of the learning Phase (during the End Period), performance did not differ across the 2 x 2 and 3 x 3 conditions, suggesting that by the end of the learning phase, participants in the 3 x 3 condition were experiencing a greater degree of retrieval success.

These findings confirm our hypothesis that the in-the-moment retrieval dynamics differed across the three learning conditions. There were differences in the overall number and pattern of retrieval successes across the three learning conditions. Participants in the 3 x 3 condition had a significantly different pattern of retrieval successes than participants in the other conditions, suggesting that this could be contributing to the stronger performance at the one week delayed test. Indeed, the pattern of retrieval successes in the 3 x 3 condition is consistent with the retrieval effort hypothesis—when learners engage in difficult but eventually successful retrieval, this deters initial performance but supports long-term performance (e.g., Pyc & Rawson, 2009). The implications of these findings are discussed in the General Discussion.

Accuracy of Self-Report Retrieval Task. The retrieval dynamics self-reported by participants during the learning phase demonstrated that there were differences in participants' experience in retrieving object-label pairings during learning. Learners often have difficulty monitoring their own ability to retrieve and remember information (e.g., Benjamin, Bjork, & Schwartz, 1998; Kornell, Rhodes, Castel, & Tauber, 2011). Thus, we wanted to verify that participants were able to do so in the current retrieval task.

If participants were accurately reporting their own retrieval dynamics, participants that reported successfully retrieving more object-label pairings should have had higher final test performance at the immediate test. Conversely, participants that reported not being able to retrieve object-label pairings should have had lower final test performance at the immediate test. Initial retrieval difficulty can lower the overall number of retrieval successes but promote performance on a delayed test (e.g., Pyc & Rawson, 2009; Vlach et al., 2012), resulting in potential interactions of performance over time. Consequently, we did not examine the results for participants in the one week delay testing condition.

We analyzed the relationship between the total number of retrieval successes during learning and immediate test performance using Pearson's *r*. Results of this test revealed that, for participants in the immediate testing condition, there was a significant relationship between participants' reported number of retrieval successes and their overall test performance, $r(39) = .506$, $p = .001$. In sum, participants were able to monitor whether or not they were successfully retrieving correct object-label pairings.

General Discussion

The experiments in this paper were designed to examine if cross-situational statistics are retained over extended periods of time. To our knowledge, this is the first work to demonstrate that learners are able to retain cross-situational statistics up to one week later. Moreover, the ability to retain word-referent pairings was related to the retrieval dynamics that occurred during learning. Interestingly, participants that were initially struggling to successfully retrieve correct object-label pairings, but were eventually relatively more successful by the end of the learning phase (the 3 x 3 condition), demonstrated the strongest retention. This finding is consistent with studies of long-term memory and retention—struggling to retrieve information during learning often leads to stronger retention and performance (e.g., Kornell et al., 2009; Pyc & Rawson, 2009; Richland et al., 2009; Vlach et al., 2012). We discuss the implications of these results below and suggest several important future directions for research on cross-situational statistical word learning.

Memory Processes and Cross-Situational Word Learning

This work demonstrates that different conditions of cross-situational word learning engender varying retrieval dynamics during learning and over extended periods of time, such as at a delayed test. In the current study, there were striking differences in the number and pattern of retrieval successes across the 2 x 2, 3 x 3, and 4 x 4 learning conditions. Why did the 3 x 3 condition engender more favorable retrieval dynamics for long-term retention?

We do not hypothesize that there is something unique about the 3 x 3 learning condition vs. the other learning conditions. What is more likely is that working memory and/or short-term memory moderated the in-the-moment ability to hold enough information in memory in order to successfully map words to objects. That is, in one learning event, the different learning conditions presented learners with varying numbers of items to hold in working memory. For example, in the 2 x 2 condition, there were two words and two objects presented in each presentation, for a total of four items. In the 3 x 3 condition there were three words and three objects, for a total of six items. Finally, in the 4 x 4 condition there were four words and four objects, for a total of eight items. In addition to the information presented during the learning trials, participants in all of these conditions would need to retrieve prior pairings as well, adding to the number of items held in working memory. Based upon previous research on working memory, there are limits on the number of items that can be held in working memory at any one moment, such as the often cited "magic number 7 ± 2" (Miller, 1956; for a more recent review, see Baddeley, 1994). These limits and capacities may have been critical in moderating the degree to which retrieval difficulty was beneficial for long-term retention.

In sum, the 3 x 3 condition may have been more taxing of working memory than the 2 x 2 condition, but still within the limits of short-term/working memory capacities. Given the low retrieval and test performance of participants in the 4 x 4 condition, it may have been that this learning condition engendered retrieval dynamics that crossed the bounds of short-term/working memory capacities. Future work should explore how learners' individual short-term/working memory capacities are related to their ability to acquire, retrieve and retain cross-situational statistics in order to explore this possibility.

It is also important to note that, given that retrieval dynamics are related to word learning performance, memory development is also likely to be a critical factor moderating cross-situational word learning in young infants and children. Indeed, previous research has demonstrated that infants forget at a rapid rate (e.g., Fagan, 1977; Rovee-Collier, Sullivan, Enright, Lucas, & Fagen, 1980) and often have smaller memory capacities compared to older children and adults (e.g., Rovee-Collier, Hayne, & Colombo, 2001). Consequently, there may be conditions under which young learners are unable to retrieve prior object-label pairings. Future studies of how infants and children acquire mappings over broad timescales are likely to reveal how young learners overcome memory constraints on the developing ability to retrieve information from the past.

Looking Forward: In-The-Moment Dynamics Across Broader Timescales

Research on cross-situational word learning is in its early stages, largely consisting of demonstrations studies, such as showing that adult learners can acquire cross-situational statistics (e.g., Yu & Smith, 2007), even under conditions of high uncertainty (e.g. Smith et al., 2010), and that this phenomenon can be modeled mathematically (e.g., Fazly et al., 2010; Siskind, 1996). Several researchers have argued that the next step is to engage in more in-the-moment examinations of learning (e.g., Yu & Smith, 2011).

Indeed, the current study is also, in part, a demonstration study. The current experiments show that learners are able to retain cross-situational statistics across extended periods of time, up to one week later. However, this study moves beyond a simple demonstration study and takes that next step—this work provides an examination of the in-the-moment processes underlying learning. This work has revealed how one such set of processes, retrieval dynamics, differ across conditions of cross-situational word learning on a second by second timescale.

We argue for another critical next step—we hope that future research will begin to predict and demonstrate how in-the-moment dynamics may have long-term effects on learning at subsequent moments in time. Why is this step critical? What is observed at one moment in time is not always reflective of another point in time. In the current work, there are many examples of the danger of assuming performance will remain constant across time. For example, if one were to look at the results of the immediate test, they might predict that participants in the 2 x 2 condition would have the highest performance at a delayed test. However, we observed that isn't the case at the one week delayed test.

In conclusion, future research should continue to examine the in-the-moment mechanisms underlying cross-situational word learning over broad timescales. In order to account for real-world learning, research should incorporate learning and testing over longer timescales—over the course of weeks, months, and years. A complete theory of cross-situational word learning not only accounts for learning in the moment and at each time point, but also integrates them in order to understand how they influence each other over time. Taken together, this work will provide a mechanistic account of how we learn new words despite the inherent ambiguity and difficulty of the task.

References for CHAPTER 2

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635-645. doi: 10.1111/j.1467-8624.1996.tb01756.x

Baddeley, A. (1994). The magic number seven: Still magic after all these years? *Psychological Review, 101*, 353-356. doi: 10.1037/0033-295X.101.2.353

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–843. doi: 10.1037/0012-1649.29.5.832

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55-68. doi: 10.1037/0096-3445.127.1.55

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*, 620-642. doi: 10.1111/j.1551-6709.2009.01089.x

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276.

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology. In H.A. Ruger, C.E. Bussenius, E.R. Hilgard (Eds.), Trans. New York: Dover Publications. (Original work published in 1885).

Estes, W. K. (1955a). Statistical theory of distributional phenomenon in learning. *Psychological Review, 62*, 369-377. doi: 10.1037/h0046888

Estes, W. K. (1955b). Statistical theory of spontaneous regression and recovery. *Psychological Review, 62*, 145-154. doi: 10.1037/h0048509

Fagan, J. F. (1977). Infant recognition memory: Studies in forgetting. *Child Development, 48*, 68-78. doi: 10.1111/j.1467-8624.1977.tb04244.x

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science, 6*, 1017-1063. doi: 10.1111/j.1551-6709.2010.01104.x

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*, 367-380. doi: 10.1111/j.1551-6709.2010.01156.x

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*, 578-585. doi: 10.1111/j.1467-9280.2009.02335.x

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 1–55. doi: 10.1207/s15327817la0101_2

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 801-812. doi: 10.1037/a0023219

Karpicke, J. D., & Roediger, H. L. III (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162. doi: 10.1016/j.jml.2006.09.004

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 989-998. doi: 10.1037/a0015729

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease of processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*, 787-794. doi: 10.1177/0956797611407929

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97. doi: 10.1037/h0043158

Pyc, M. A., Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437-447. doi: 10.1016/j.jml.2009.01.004

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243-257. doi: 10.1037/a0016496

Roediger, H .L., & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20-27. doi: 10.1016/j.tics.2010.09.003

Rovee-Collier, C. K., Hayne, H., & Colombo, M. (2001). *The development and implicit and explicit memory*. Amsterdam: John Benjamins.

Rovee-Collier, C. K., Sullivan, M. W., Enright, M., Lucas, D., & Fagen, J. W. (1980). Reactivation of infant memory. *Science, 208*, 1159-1161. doi: 10.1126/science.7375924

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69,* 94–104. doi: 10.2307/1132073

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122*, 163-180. doi: 10.1016/j.cognition.2011.10.010

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*, 179-193. doi: 10.1037/h0027277

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-91. doi: 10.1016/S0010-0277(96)00728-7

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science, 35*, 480-498. doi: 10.1111/j.1551-6709.2010.01158.x

Smith, L. B. (2000). How to learn words: An associative crane. In R. Golinkoff and k. Hirsh-Pasek (Eds.), *Breaking the word learning barrier*, Oxford University Press, Oxford. pp. 51–80.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568. doi: 10.1016/j.cognition.2007.06.010

Templin, M. C. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373. doi: 10.1037/h0020071

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, *30*, 639-650. doi: 10.1037/0012-1649.30.5.639

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 246-254. doi: 10.1037/a0025260

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414-420. doi: 10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165-180. doi: 10.1111/j.1467-7687.2010.00958.x

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21-39. doi: 10.1037/a0026182

Figures for CHAPTER 2

Table 1

Trial Composition for the Three Learning Conditions

Learning Condition	Number of object-label pairings	Number of presentations of each pairing	Number of trials	Time per trial (in secs)	Total learning time (in secs)
2 x 2	18	6	54	6	324
3 x 3	18	6	36	9	324
4 x 4	18	6	27	12	324

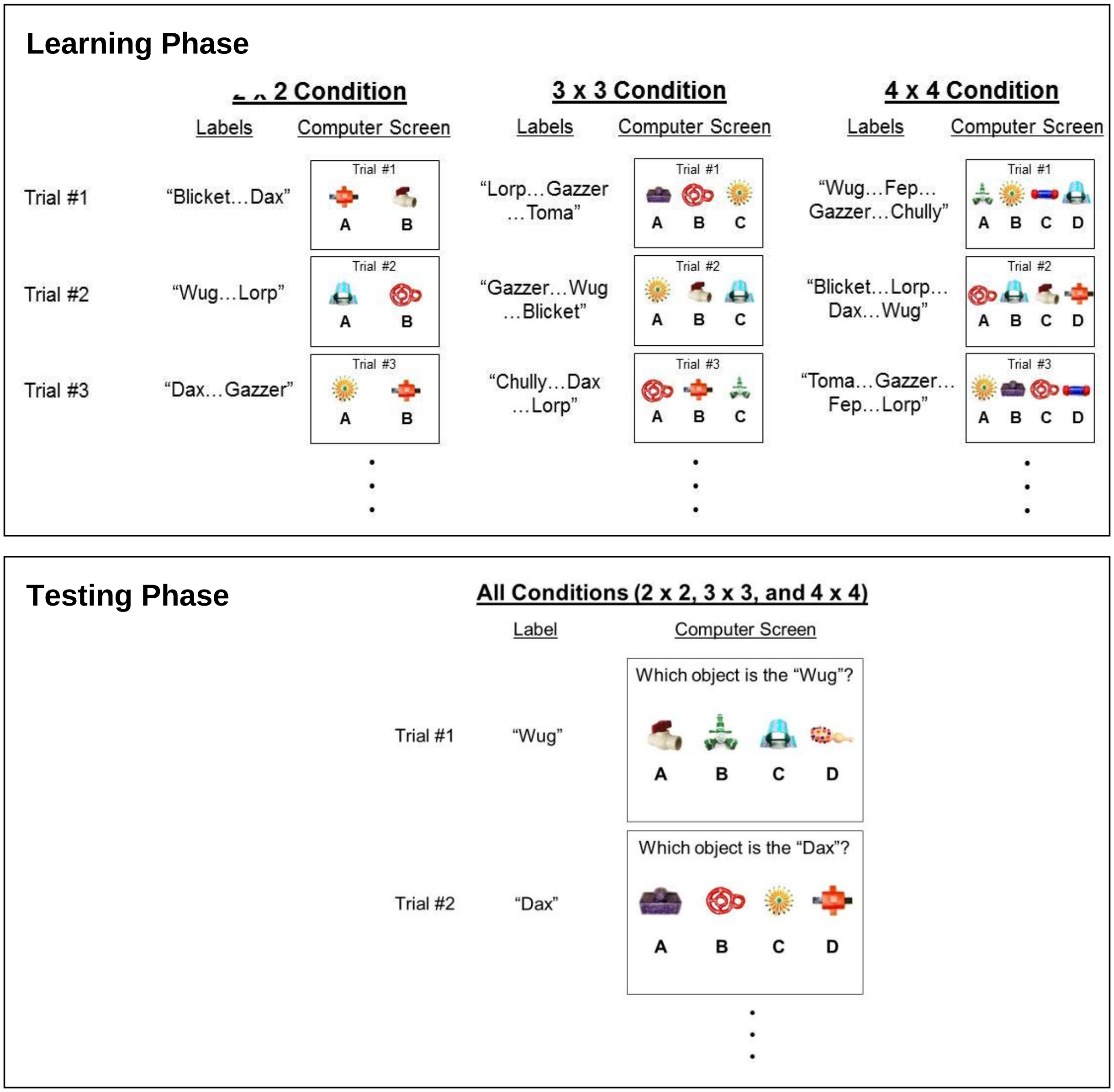

Figure 1. Example trials from the learning and testing phases of Experiments 1 and 2. During the learning phase of both experiments, participants were presented with a series of learning trials according to one of three learning conditions, 2 x 2, 3 x 3, and 4 x 4. See Table 1 for the number and timing of the trials. During the testing phase, participants were presented with four forced-choice testing trials. All labels were presented in the same woman's voice over the computer's speaker system.

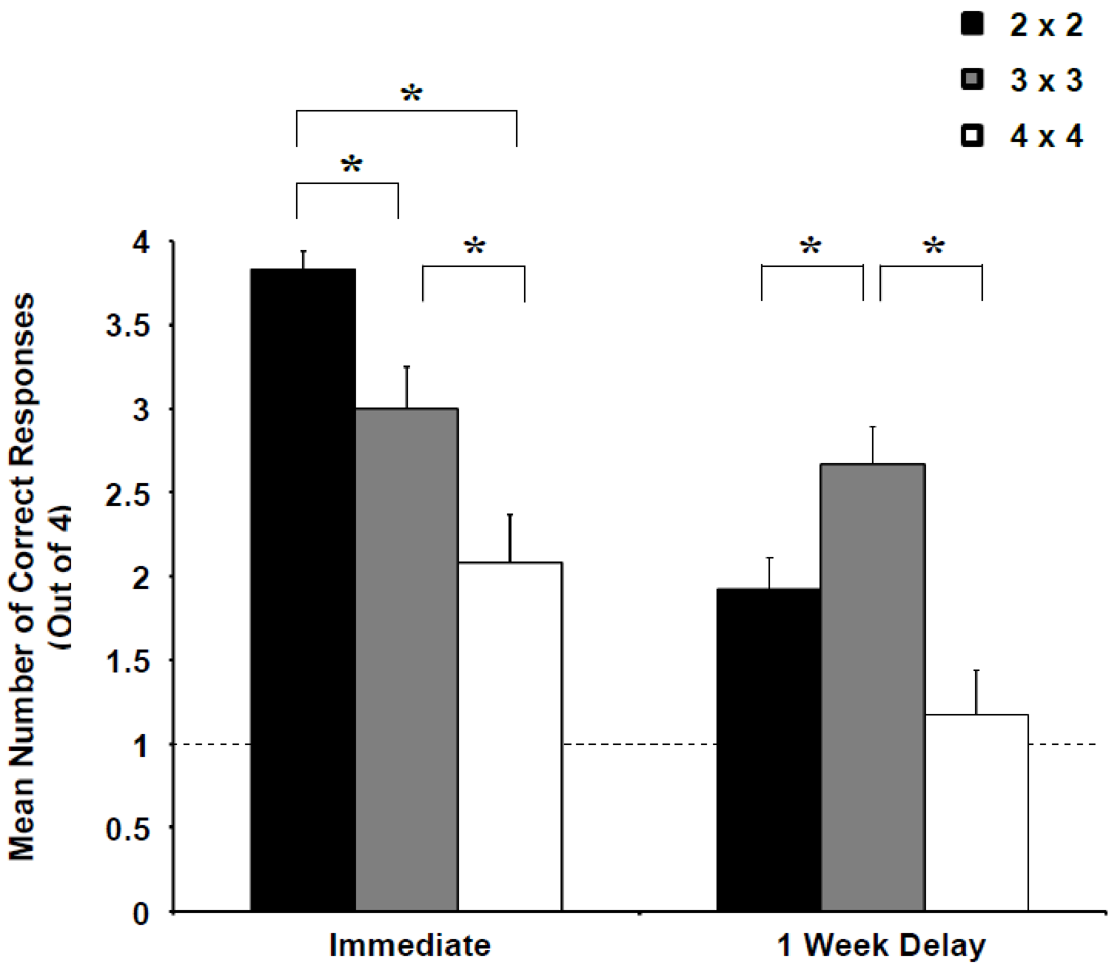

Figure 2. Results of testing phase in Experiment 1. Mean number of correct responses (out of 4) by learning condition (2 x 2, 3 x 3, and 4 x 4) and testing condition (immediate and 1-week delayed). The dashed line represents chance performance and a star indicates a statistically significant difference, $p < .05$.

Training Trial #	Mappings Known		
#1	A	B	None
	C	D	
#2	A	B	None
	C	D	
#3	A	B	None
	C	D	
Learning Trial #	**Mappings Known**		
#1	A	B	None
	C	D	
#2	A	B	None
	C	D	
#3	A	B	None
	C	D	
•	•		
•	•		
•	•		

Figure 3. Example of worksheet used during the self-report retrieval task in Experiment 2. Participants were instructed to record successes of retrieving object-label pairings by circling one letter, multiple letters, or none. Worksheet was used during the training and learning phases of Experiment 2.

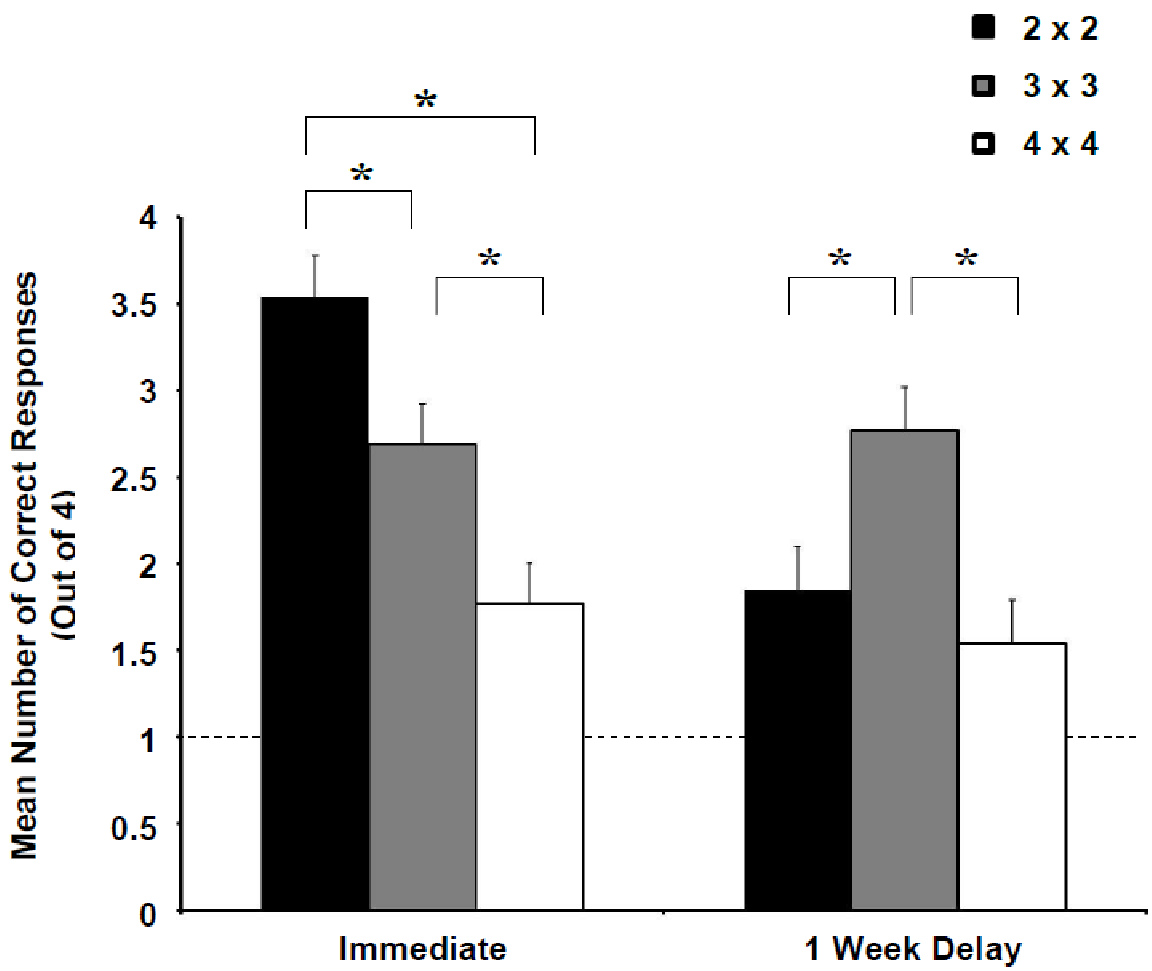

Figure 4. Results of testing phase in Experiment 2. Mean number of correct responses (out of 4) by learning condition (2 x 2, 3 x 3, and 4 x 4) and testing condition (immediate and 1-week delayed). The dashed line represents chance performance and a star indicates a statistically significant difference, $p < .05$.

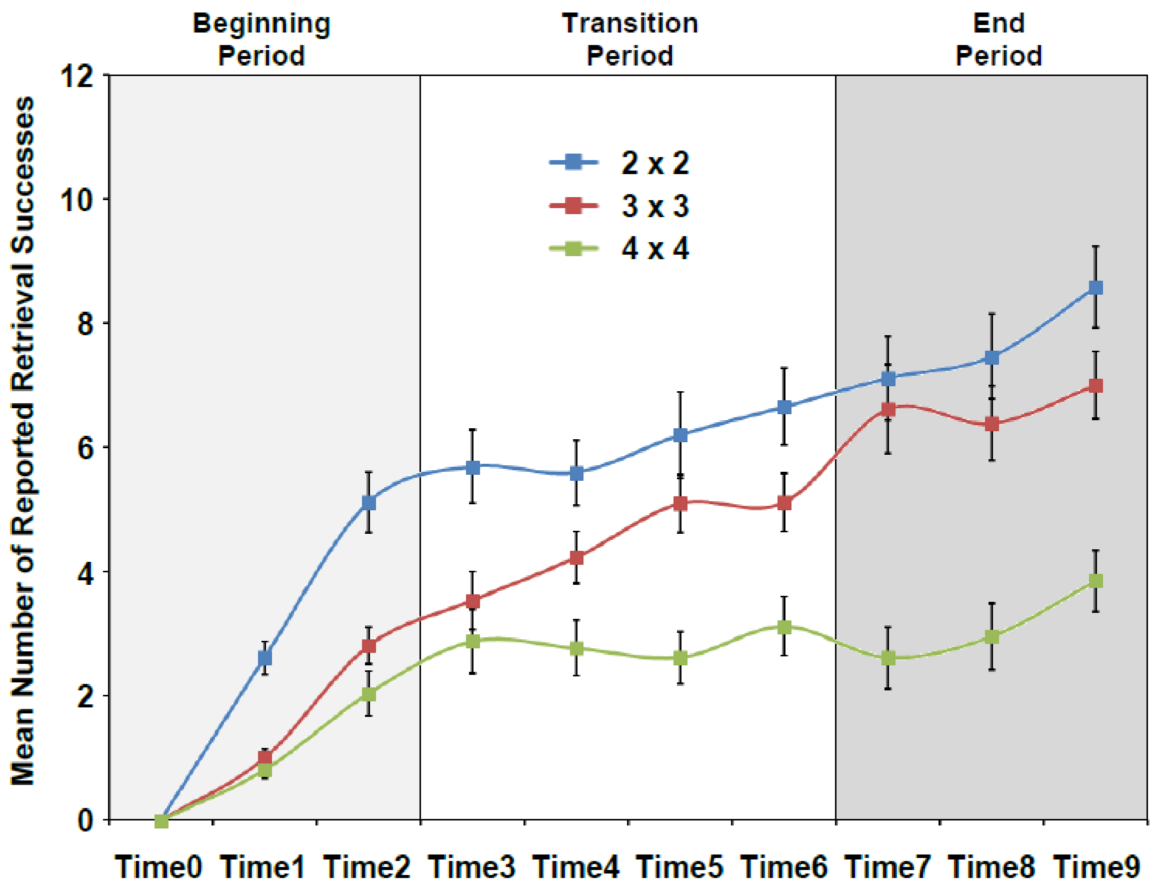

Figure 5. Mean number of reported retrieval successes during the learning phase, for the three learning conditions (2 x 2, 3 x 3, and 4 x 4) by time interval. Error bars represent standard errors. The learning phase was divided into three periods to categorize the nature of the pattern of performance for participants in the 3 x 3 condition. During the Beginning Period, participants in the 3 x 3 condition did not significantly differ in the mean number of retrieval successes from participants in the 4 x 4 condition. However, participants in the 3 x 3 condition did have significantly lower retrieval performance than participants in the 2 x 2 condition. During the Transition Period, participants in the 3 x 3 condition reported an intermediate degree of retrieval success; performance in the 3 x 3 condition was significantly different than performance in both the 2 x 2 and 4 x 4 conditions. Finally, during the End Period, participants in the 3 x 3 condition

reported a significantly higher number of retrieval successes than participants in the 4 x 4 condition, but not the 2 x 2 condition.

CHAPTER 3: An Exploratory Study on Manipulating Retrieval Dynamics

In the previous chapter, Chapter 2, the results of the experiments revealed that retrieval dynamics may be an important mechanism underlying the acquisition and retention of cross-situational statistical learning. In Experiments 1 and 2, the word-referent pairings were presented in random order in each learning condition. A question that naturally follows from this work is: would presenting object-label pairings at certain points in time, in turn altering the ease/difficulty of retrieval, affect the overall performance at the immediate and/or delayed test?

Distributing learning events across time has long been shown to improve retention (e.g., Ebbinghaus, 1885/1964). In this chapter, we have outlined an exploratory study on engendering more favorable retrieval dynamics by distributing object-label pairings events across time. One explanation for this phenomenon is that spacing learning events out in time allows time for forgetting between each event. Consequently, at subsequent learning events, learners must engage in deeper retrieval in order retrieve prior learning events (compared to events that are closer together in time). Thus, in line with this theoretical account (study-phase retrieval theory; see Thios & D'Agostino, 1976; Delaney, Verkoeijen, & Spirgel, 2010), distributing object-label pairings out in time should promote deeper retrieval and stronger retention for pairings.

Method

Participants. There were 68 undergraduate students that participated in this experiment. Participants were randomly assigned to one of the six between-subjects conditions of the experiment. Participants received course credit for their participation and did not participate in Experiment 1 or Experiment 2 (in Chapter 2).

Apparatus & Stimuli. Same as Experiment 2, Chapter 2, with two exceptions. First, the presentation timing of object-label pairings was not randomly assigned. Instead, the object-label pairings were presented according to one of four schedules: (1) massed; there were four object-label pairings that were presented in immediate succession (0 interleaved trials condition), (2) interleaved1; there were four object-label pairings that were presented with one trial in between each pairing (1 interleaved trial condition), (3) interleaved3; there were four object-label pairings that were presented with three trials between each pairing (3 interleaved trials condition), and (4) random; the presentation of the remaining six pairings was randomly assigned. A depiction of the first three conditions can be seen in Figure 1.

Second, in order to capture learning across the different presentation schedules, additional testing trials were included in the testing phase. There were 16 testing trials; four massed testing trials, four interleaved1 testing trials, four interleaved3 testing trials, and 4 random testing trials. The testing trial format was the same as used in Experiments 1 and 2, in Chapter 2.

Design. Same as Experiment 2, Chapter 2. In the current experiment, the timing of the object-label presentations was a within-subjects variable (massed, interleaved1, interleaved3, random). All participants completed the self-report retrieval task, the same task reported in Experiment 2, Chapter 2.

Procedure. Same as Experiment 2, Chapter 2.

Results

The design of this experiment provided a large data set, which revealed several interesting relationships between presentation timing, retrieval dynamics, and retention. Included below are several analyses that highlight many of these relationships.

Overall Test Performance. Did the current experiment replicate the test performance results of Experiment 1 and Experiment 2 in Chapter 2? Because there were different presentation schedules for the object-label pairings, this may have improved and/or deterred immediate and long-term performance. In order to examine this possibility, we conducted a 3 (Learning Condition) x 2 (Testing Delay) ANOVA, with the number of correct responses at the final test as the dependent measure. Results of this test revealed a significant main effect of learning condition, $F(2, 62) = 7.064$, $p = .002$, $\eta_p^2 = .186$, and a significant main effect of testing delay, $F(1, 62) = 43.780$, $p < .001$, $\eta_p^2 = .414$. The results did not suggest an interaction between learning condition and testing delay, $F(2, 62) = 1.533$, $p = .224$.

In order to characterize the nature of the two main effects, we conducted two univariate ANOVAs within each testing condition. We then computed three planned comparisons using t-tests with Bonferroni corrections to determine the nature of the differences between the learning conditions within each testing delay condition. In the immediate testing condition, there was a main effect of learning condition, $F(2, 33) = 5.402$, $p = .009$, $\eta_p^2 = .247$. Participants in the 2 x 2 and 3 x 3 conditions had significantly higher performance than participants in the 4 x 4 condition, $p = .018$ and $p = .030$, respectively. There was not a significant difference between performance in the 2 x 2 and 3 x 3 conditions, $p > .05$. Thus, although there were no differences in immediate test performance between the 2 x 2 and 3 x 3 conditions, participants in both of these conditions outperformed participants in the 4 x 4 condition.

In the one week delayed testing condition, there was a marginal main effect of learning condition, $F(2, 29) = 3.111$, $p = .060$, $\eta_p^2 = .177$. There was only one significant difference in testing performance between the testing conditions; participants in the 3 x 3 condition had marginally higher performance than participants in the 4 x 4 condition, $p = .063$. There was not a significant difference in test performance between the 2 x 2 and 3 x 3 conditions, p = .341. In sum, participants in the 3 x 3 condition outperformed participants in the 4 x 4 condition, but did not perform significantly different than participants in the 2 x 2 condition.

Retrieval Successes Across the Learning Phase. Were there differences in the number and pattern of retrieval successes during the learning phase? We started by examining the overall number of reported retrieval successes by learning condition. We conducted a univariate ANOVA with the overall number of reported retrieval successes as the outcome variable. We found a significant main effect of condition, $F(2, 65) = 4.611$, $p = .013$, $\eta_p^2 = .124$. We then computed three planned comparisons using t-tests with Bonferroni corrections to determine the nature of the differences between the learning conditions. Participants in the 2 x 2 condition (M = 50.52, SD = 24.821) and 3 x 3 condition (M = 47.74, SD = 21.535) reported a marginally and/or significantly higher number of retrieval successes than participants in the 4 x 4 condition (M = 33.09, SD = 13.554), p = .060 and p = .018, respectively. Moreover, there were no significant differences in the overall number of retrieval successes between the 2 x 2 and 3 x 3 conditions. Thus, there were differences in the overall number of retrieval successes across the three learning conditions, except in the case of the 2 x 2 and 3 x 3 learning conditions. This finding is consistent with the fact there were no significant differences in test performance at the immediate test (see Figure 2).

Were there differences in the pattern of retrieval successes across the learning phase? We started by dividing the learning phase into nine timescales, using the same procedure as reported in Experiment 2, Chapter 2. We then conducted a mixed 3 (Learning Condition) x 9 (Time Point) ANOVA, with learning condition as a between-subjects variable and time point as a within-subjects variable. Results of this test revealed a significant main effect of learning condition, $F(2, 65) = 4.611$, $p = .013$, $\eta_p^2 = .124$, a significant main effect of time point, Wilks' Lambda = .456, $F(8, 58) = 8.665$, $p < .001$, $\eta_p^2 = .544$, and a significant interaction of learning condition and time point, Wilks' Lambda = .385, $F(16, 116) = 4.435$, $p < .001$, $\eta_p^2 = .380$.

In order to examine the interaction between learning condition and time point, we conducted a post-hoc analysis using planned comparisons between the three learning conditions, at each time point. The descriptive data can be seen in Figure 3. To correct for all 27 comparisons, we computed a corrected alpha using Bonferroni standards ($\alpha = .05/27$, corrected $\alpha = .00185$). This alpha level was used for all of the planned comparisons.

The results revealed a strikingly different pattern of retrieval successes than seen in Experiment 2, Chapter 2. Only at Time1, Time2, and Time3, did we observe differences in the mean number of self-reported retrieval successes across the three learning conditions. At Time1, participants in the 2 x 2 condition reported significantly more retrieval successes than participants in the 3 x 3 and 4 x 4 conditions, *p*s < .00185. There was not a significant difference in the number of reported retrieval successes between participants in the 3 x 3 and 4 x 4 conditions. At Time2 and Time3, participants in the 2 x 2 and 3 x 3 conditions reported a significantly higher number of retrieval successes than participants in the 4 x 4 condition, *p*s < .00185. Moreover, there were not significant differences in the number of reported retrieval successes between the 2 x 2 and 3 x 3 conditions at both time points. At all other time points (Time3 – Time9), there were not any significant differences in self-reported retrieval successes across the three learning conditions.

In sum, in the current experiment we observed a very different pattern of retrieval successes across the learning phase than reported in Experiment 2, Chapter 2. Overall, there were smaller differences in the overall number of retrieval successes and the trajectory across the learning phase. This was especially the case for the 2 x 2 and 3 x 3 conditions; there was no significant difference in the overall number of retrieval successes during the learning phase and only at Time1 was there a significant difference in the number of retrieval successes across the nine time points. Indeed, these retrieval dynamics were consistent with similar performance at both the immediate and delayed test.

Effects of Presentation Timing of Object-Label Pairings. All object-label pairings were presented on one of four presentation schedules: massed, interleaved1, interleaved3, or random, as described in the Methods section. These presentation schedules allowed for consistency across the three learning conditions in the manner in which the object-label pairings were distributed across learning trials. However, because these presentation schedules were kept consistent in this manner, other learning factors varied.

For example, in the interleaved1 presentation condition, there was one learning trial in between object-label pairing presentations, which was consistent across all three learning conditions. The characteristics of the learning trial in between the object-label presentations varied across the conditions. In the 2 x 2 condition, the trial would have consisted of two object-label pairings and would have lasted six seconds. In the 3 x 3 condition, the trial would have consisted of three object-label pairings and would have lasted nine seconds. In the 4 x 4 condition, the trial would have consisted of four object-label pairings and lasted 12 seconds. Thus, there were differences in the amount of time and intervening object-label pairings across the three learning conditions.

Consequently, in order to examine the effects of presentation timing on immediate and long-term performance, all analyses were conducted within each learning condition. For all learning conditions, we first conducted a multivariate ANOVA with the number of correct responses at test for the massed, interleaved1, interleaved3, and random test items as the outcome measures. If this test revealed a significant difference across the immediate and delayed test, we conducted paired-sample comparisons within each testing condition, using Bonferroni corrections. The descriptive data is presented in Figure 4.

For participants in the 2 x 2 condition, there was a main effect of testing condition for the massed, $F(1, 21) = 17.030$, $p < .001$, $\eta_p^2 = .448$, interleaved1, $F(1, 21) = 18.765$, $p < .001$, $\eta_p^2 = .472$, interleaved3, $F(1, 21) = 4.351$, $p = .049$, $\eta_p^2 = .172$, and random test items, $F(1, 21) = 5.966$, $p = .024$, $\eta_p^2 = .221$. As can be seen in Figure 4, all bars at the one week delayed condition were significantly lower at the one week delayed test. In the immediate test condition, post-hoc analyses revealed that there were no significant differences in test performance between the four presentation conditions, *p*s > .10. However, in the one week delayed testing condition, post-hoc analyses revealed participants' performance on the random test items was significantly higher than that on massed and interleaved1 items, $p = .017$ and $p = .019$, respectively.

For participants in the 3 x 3 condition, there was a main effect of testing condition for the massed, $F(1, 21) = 3.533$, $p = .058$, $\eta_p^2 = .141$, interleaved1, $F(1, 21) = 9.179$, $p = .006$, $\eta_p^2 = .304$, interleaved3, $F(1, 21) = 11.313$, $p = .003$, $\eta_p^2 = .350$, and random test items, $F(1, 21) = 3.796$, $p = .047$, $\eta_p^2 = .117$. As can be seen in Figure 4, all bars at the one week delayed condition were significantly lower at the one week delayed test. In the immediate test condition, post-hoc analyses revealed that there were no significant differences in test performance between the four presentation conditions, *p*s > .10. However, in the one week delayed testing condition, post-hoc analyses revealed that participants' performance on the interleaved3 and random test items was significantly higher than performance on the massed and interleaved1 items, *p*s < .05.

Finally, for participants in the 4 x 4 condition, there was a main effect of testing condition for the massed testing items, $F(1, 20) = 3.772$, $p = .050$, $\eta_p^2 = .155$, but not for the other presentation conditions , *p*s > .10. The only significant difference found in the post-hoc analyses is that, at the immediate test, participants had marginally higher performance on the massed test items compared to the random test items, $p = .061$.

In sum, there were differences in test performance across the four presentation conditions. In the 2 x 2 condition, participants' performance was highest on the random items at the one week delayed test. In the 3 x 3 condition, participants' performance was highest on the interleaved3 and random items the one week delayed test. There was one effect at the immediate test; in the 4 x 4 condition, participants had higher performance on the massed items compared to the random items.

Accuracy of Self-Reported Retrieval Task. In Experiment 2, Chapter 2, we examined whether participants were accurately reporting their retrieval successes during the retrieval task. We conducted the same analyses with the data in this experiment. If participants were accurately reporting their own retrieval dynamics, participants that reported successfully retrieving more object-label pairings should have had higher final test performance at the immediate test. Conversely, participants that reported not being able to retrieve object-label pairings should have had lower final test performance at the immediate test.

We analyzed the relationship between the total number of retrieval successes during learning and immediate test performance using Pearson's r. Results of this test revealed that, for participants in the immediate testing condition, there was a significant relationship between participants' reported number of retrieval successes and their overall test performance, $r(36) =$

.519, $p = .001$. In sum, it appears that participants were able to monitor whether or not they were successfully retrieving correct object-label pairings.

Discussion

The current experiment was designed to be an exploratory study on manipulating retrieval dynamics, by distributing object-label pairings in time, in cross-situational statistical learning. What did the results tell us? The results revealed that manipulating the timing of object-label pairings changed overall test performance, the number reported retrieval successes, and the pattern of retrieval successes across the learning phase, in comparison to experiments reported in Chapter 2.

Why did the changes in presentation timing affect testing performance and retrieval dynamics? There are several possibilities. For example, one possibility is that, by manipulating similar learning schedules across the three learning conditions, this in turn equated many of the retrieval dynamics in learning. By equating retrieval dynamics, this could result in similar test performance across the three learning conditions. Indeed, immediate test performance was not statistically different between the 2 x 2 and 3 x 3 conditions and there was only a marginally significant difference between the 3 x 3 and 4 x 4 conditions at the one week delayed test.

In order to examine this possibility and other explanations for the current results, future research and experiments should be designed to control for a variety of factors of learning that were not controlled in the current experiment. For example, in order to determine if distributing object-label pairings across time promotes learning because time vs. interference of other object-label pairings, controlling the amount of time between learning events and the number of interleaved object-label pairings would be important. Although the current study was not designed to examine these processes, follow-up studies may be able to answer these questions.

In the next chapter, Chapter 4, we extend research in the previous chapters to examining cross-situational learning during development. We see this as a very important step. Memory capacities and language learning undergo dramatic changes over the early part of development – How might retrieval dynamics differ during development and what effect would this have on infants' cross-situational word learning?

References for CHAPTER 3

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H.A. Ruger, C.E. Bussenius, & E.R. Hilgard, Trans.). New York: Dover Publications. (Original work published in 1885).

Delaney, P. F., Verkoeijen, P. J., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *The Psychology of Learning and Motivation: Advances in Research and Theory*, *53*, 63-147. doi: 10.1016/S0079-7421(10)53003-2

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning & Verbal Behavior, 15*, 529-536. doi: 10.1016/0022-5371(76)90047-5

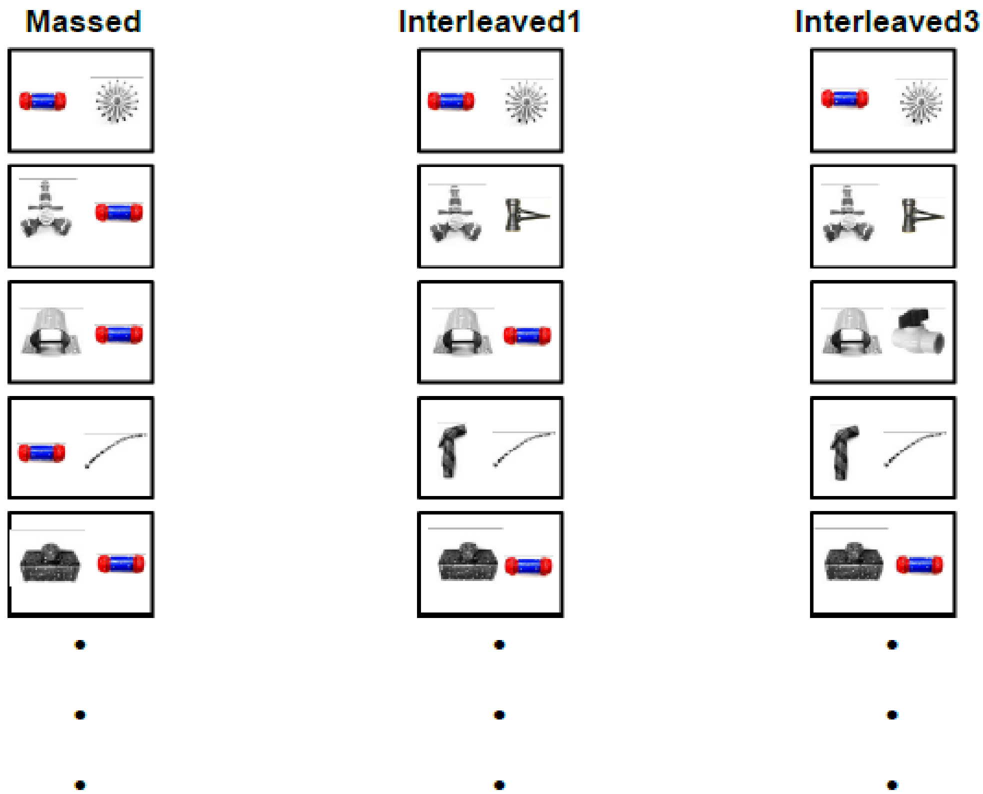

Figure 1. Examples of three of the presentation timing conditions (massed, interleaved1, interleaved3). The random presentation condition is not shown. The examples are shown in the context of the 2 x 2 learning condition. The random presentation trials are not shown. Across all presentation schedules and learning conditions, each object label pairing was presented six times.

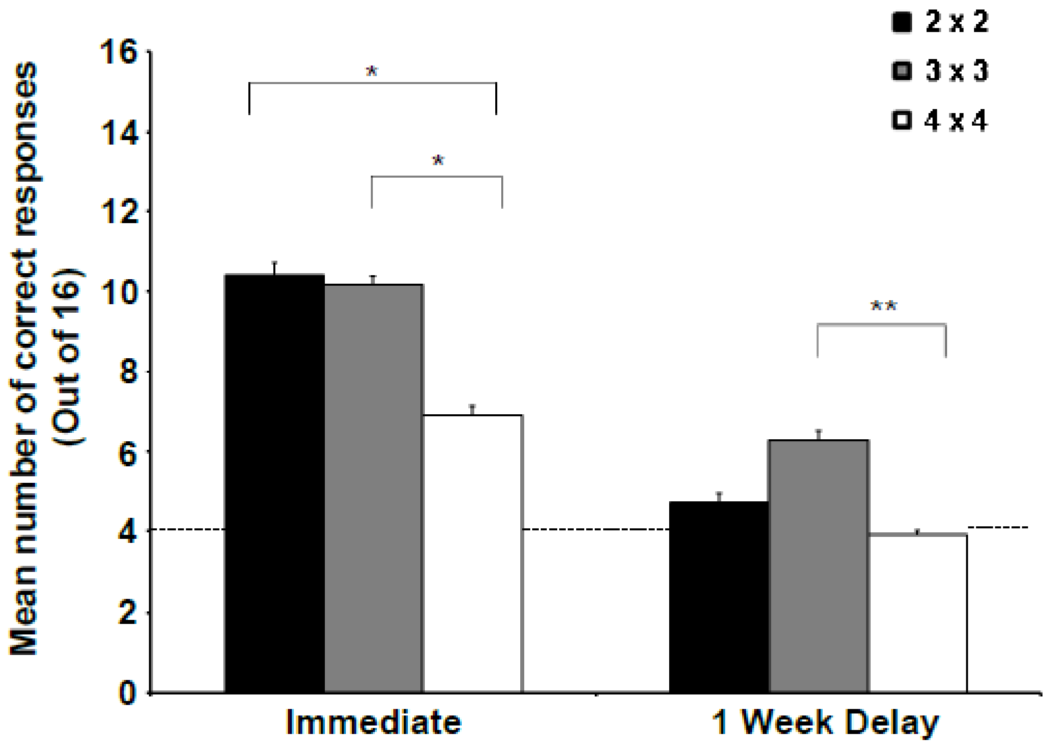

Figure 2. Results of the testing phase. Mean number of correct responses (out of 16) by learning condition (2 x 2, 3 x 3, and 4 x 4) and testing condition (immediate and 1-week delayed). The dashed line represents chance performance and a star indicates a statistically significant difference, * = $p < .05$, ** = $p < .10$.

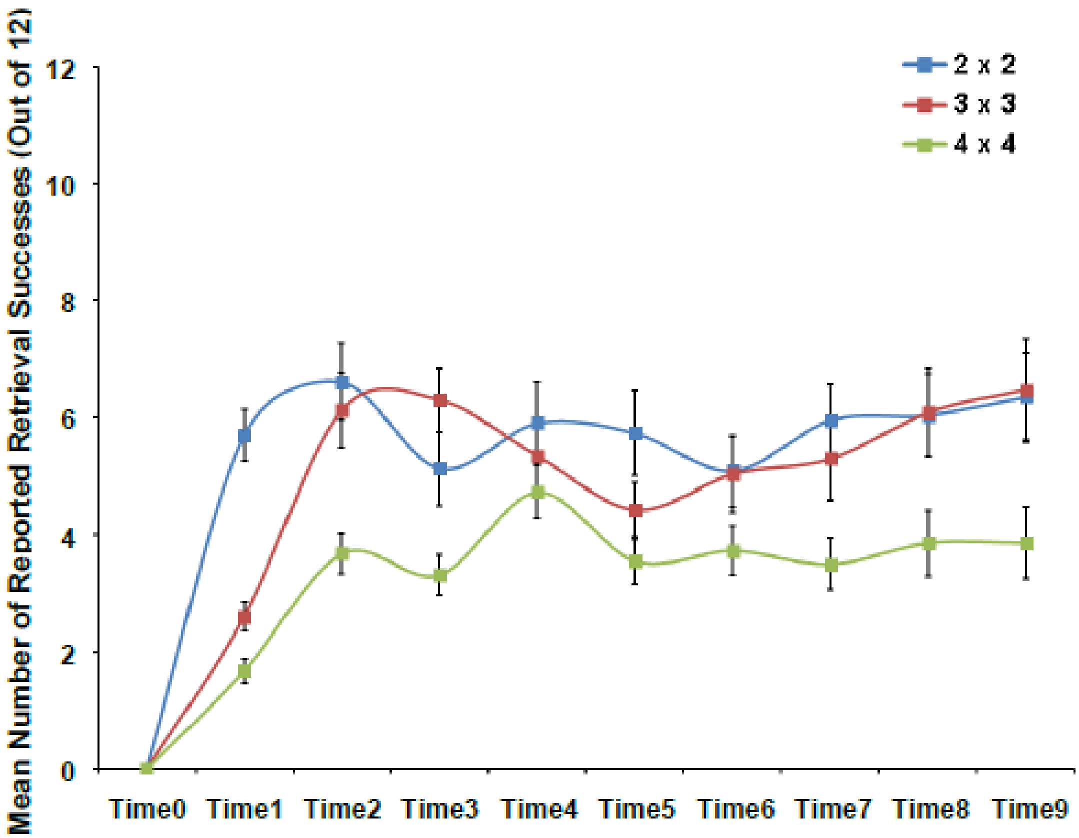

Figure 3. Mean number of reported retrieval successes during the learning phase, for the three learning conditions (2 x 2, 3 x 3, and 4 x 4) by time interval. Error bars represent standard errors. Only at Time1, Time2, and Time3, were there differences in the mean number of self-reported retrieval successes across the three learning conditions. These differences are as follows: At Time1, participants in the 2 x 2 condition reported significantly more retrieval successes than participants in the 3 x 3 and 4 x 4 conditions. At Time2 and Time3, participants in the 2 x 2 and 3 x 3 conditions reported a significantly higher number of retrieval successes than participants in the 4 x 4 condition.

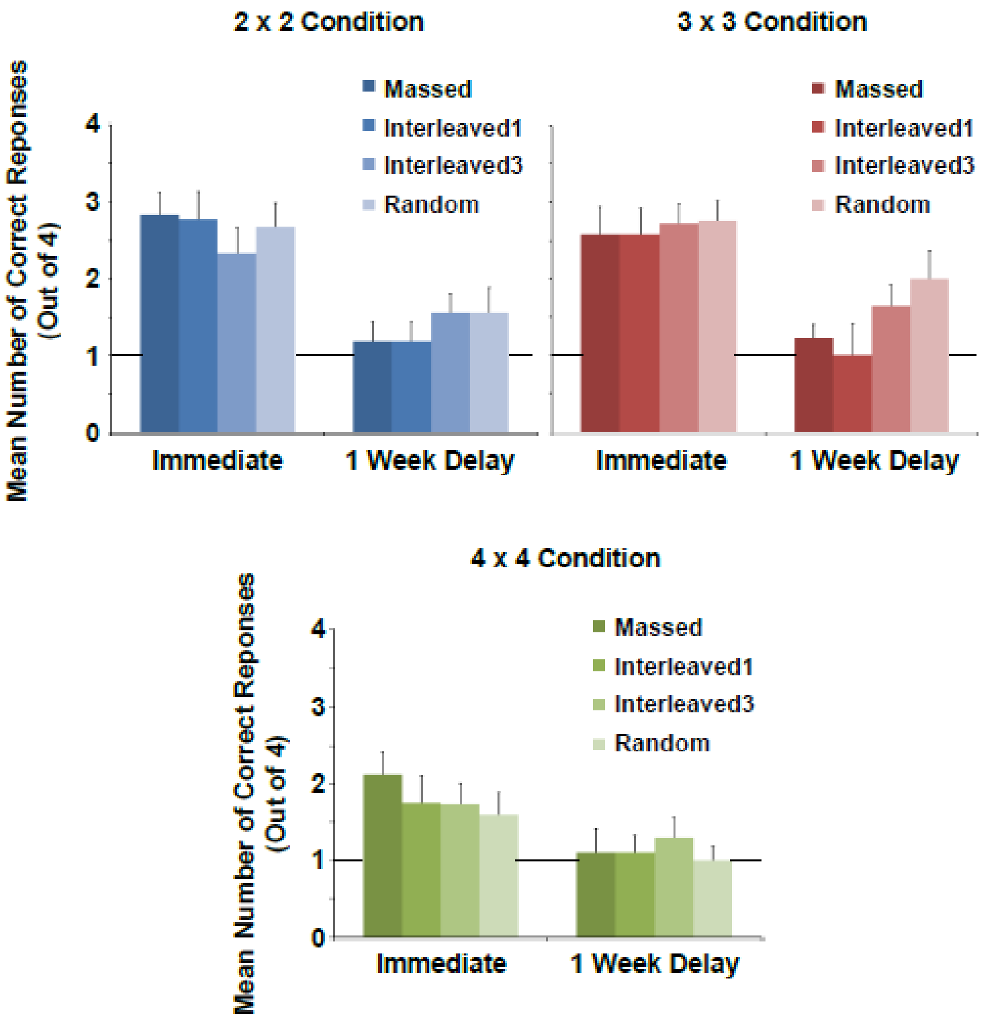

Figure 4. Mean number of correct responses at test, by learning condition (2 x 2, 3 x 3, and 4 x 4) and by presentation timing of object-label pairings during learning (massed, interleaved1, interleaved3, and random). Dashed lines represent chance performance.

CHAPTER 4: Cross-Situational Learning in Development

Word learning is often described as a difficult task because the world offers infants a seemingly infinite number of word-to-world mappings in just one moment in time (Quine, 1960). Historically, researchers have examined the constraints learners use to reduce the possible number of mappings in a single moment. The results of this work have suggested that young learners may use several constraints to reduce ambiguity, such as social/cultural (e.g., Tomasello, 1992), representational (e.g., Merriman & Bowman, 1989), and attentional (e.g., Smith, 2000) constraints.

More recent research on word learning has demonstrated that learners reduce ambiguity by detecting and retaining associations across moments in time in a process often termed *cross-situational statistical learning* (Frank, Goodman, & Tenenbaum, 2009; Siskind, 1996; Smith, Smith, & Blythe, 2010; Smith & Yu, 2008; Vouloumanos, & Werker, 2009; Yu & Smith, 2011, 2012). That is, learners acquire potential word-to-referent associations across learning events and use this information to guide subsequent inference of word meaning. For example, Smith and Yu (2008) used a preferential looking paradigm to examine cross-situational word learning in 12- and 14-month-olds. Infants were presented with two objects and two words in each learning trial such that it was ambiguous as to which word went with which object. However, across the learning trials, the same word co-occurred with one object. Following the learning trials, infants looked significantly longer at the object that co-occurred with a word (compared to a distractor object) as the word was repeatedly presented to the infant. This work suggests that young infants are able to track co-occurrence probabilities in order to map words to referents in the world.

The majority of research on cross-situational statistical learning has focused on adults' learning and/or mathematical models of learning, rather than infants' learning (see Vouloumanos, & Werker, 2009; Yu & Smith, 2011, for recent exceptions). Although some learning processes may operate in a similar manner across the lifespan, it may be the case that the developmental state of the learner mediates many of these learning processes (Bulf, Johnson, & Valenza, 2011). Thus, the current study examines infants' cross-situational statistical learning in order to expand this body of work and examine potential differences across development.

In particular, the current work extends research on infants' cross-situational statistical learning by examining the developing ability to learn and retrieve word-referent pairings over time in order to later infer word mappings. Young learners need to recall the past both during learning and when making subsequent inferences, yet little is known about how they aggregate past and present associations between words and objects, and how these processes of aggregation may develop in infancy. In real-world learning scenarios, there are likely to be frequent temporal gaps between learning events in which word-referent pairings are encoded. Indeed, retrieving prior word-referent pairings over time is a critical process in several theories of cross-situational learning (for a review, see Yu & Smith, 2012). Thus, a complete theory of cross-situational word learning will have to account for how young infants are able to complete this task.

In the current study, we examined 16- and 20-month-olds' ability to learn word mappings via cross-situational statistical learning. We examined word learning in this developmental period for several reasons. First, previous research has indicated that infants learn cross-situational statistics with object-label pairings presented close together in time as early as 12 months (e.g., Smith & Yu, 2008). Thus, infants at 16 and 20 months should be able to learn the pairings presented in immediate succession. Second, this age span is marked by striking differences in language production and development (pre- vs. post-vocabulary boom; Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994). Consequently, we predicted that these age groups might also be marked by developmental changes in the ability to learn cross-situational statistics.

Infants were presented with word-referent pairings on two timescales; half of the pairings were presented in learning trials that occurred in immediate succession (massed) and half of the pairings were presented in learning trials distributed across time (interleaved). These presentation conditions provided a direct examination of the developing ability to learning cross-situational statistics over varying timescales.

Method

Participants. Two groups of infants, 16 16-month-old infants (*M* = 16.1 months; 9 girls) and 16 20-month-old infants (*M* = 20.2 months; 9 girls), participated in the cross-situational word learning task. An additional 7 infants participated but were excluded because of fussiness/inability to complete the experiment (*N* = 4) and technical/experimenter error (*N* = 3). All children were monolingual English speakers and recruited from a university child participant database. In order to determine children's productive vocabulary, parents completed the MacArthur Bates Communicative Development Inventory: Words & Sentences (MCDI) (Fenson et al., 1994). Infants' productive vocabularies were consistent with developmental norms for both the 16-month-old infants (*M* = 81.1 words, percentile range: 15 - 99) and 20-month-old infants (*M* = 210.5 words, percentile range: 15 - 99).

Apparatus and Stimuli. The experiment was presented to infants on a Tobii T60XL eye tracker. The experiment was implemented with Tobii Studio software. The dependent variable was looking at the two objects on the monitor during test trials, operationalized as dwell times (accumulated fixations) in the left or right half of the screen where objects were in view. During all learning and testing trials, the objects were presented on the monitor and the words were presented with the eye tracker's speakers. Infants were seated on a parent's lap ~60 cm from the monitor. Our cross-situational word learning task consisted of a learning phase and a testing phase. During the learning phase, infants were presented with a total 12 novel object-novel label pairings, six of which were presented on a massed schedule and six of which were presented on an interleaved schedule. As seen in Figure 1, each learning trial consisted of two novel objects presented side-by-side on the monitor and two novel words following the phonotactic probabilities of English (e.g., 'blicket', 'toma') played over the speakers. Each of the two words was played once, using one woman's voice, in random order. Over the course of the learning phase, each object-label pairing was presented six times.

During the testing phase, each trial consisted of two novel objects and one novel word, played over the speakers four times, in the same voice used during the learning phase. An attention getter (a small object moving in tandem with a repetitive non-linguistic sound) was shown after every series of three trials.

Design. The study was a 2 (Age Group) x 2 (Presentation Timing of Pairings) mixed design. Age Group (16- and 20-month-olds) was a between-subjects factor and Presentation Timing of Pairings (massed and interleaved) was a within-subjects factor.

Procedure. Prior to the beginning of the experiment, infants were seated on a parent's lap. After the infant and parent were positioned appropriately, the lights in the room were dimmed and the calibration procedure began. A five point calibration procedure was used in order to ensure that the reflections of both eyes were centered in the eye-tracking camera's field of view.

Learning Phase. Following successful calibration, the learning phase of experiment commenced. The learning phase consisted of 36 learning trials (duration: 4s each) and 12 attention getter trials (duration: 3s each), which were grouped into six blocks (see Figure 1). After every three learning trials, an attention getter was presented. The learning phase ended with the last learning trial of Block 6 (see Figure 1).

Testing Phase. The testing phase began after the attention getter that followed the 36 learning trials. There were 12 testing trials (duration: 8s each) and 12 attention getter trials (duration: 3s each). An attention getter was presented before the each test trial to recenter the point of gaze.

Results

We reasoned that infants' inferences about object-label pairings would be revealed by longer looking to the target object than the distractor object on testing trials. The target object was defined as the object that always co-occurred during learning with the one label heard during each test trial. As noted previously, we predicted that infants at both 16 and 20 months would learn object-label pairings during massed presentation, given positive results from younger infants in previous reports (e.g., Smith & Yu, 2008). At issue was the possibility that interleaved presentation challenges the acquisition of object-label pairings, perhaps due to limits in retrieval of pairings from memory, in which case infants would be expected to look at target and distractor objects equally at test.

To examine performance on the testing trials, we calculated the mean looking times (i.e., dwell times) to the target and distractor objects (left vs. right side of the screen) across presentation conditions (massed vs. interleaved) and age groups (16- vs. 20-months). The descriptive data are presented in Figure 2. We also computed the proportion of looking time to the target object (i.e., target looking time/total looking time). We then conducted a multivariate ANOVA with age group as the between subjects factor, and the proportion of time looking to the target object for the massed and spaced pairings as the outcome measures. The results revealed a main effect of age for proportion looking to the target object for spaced items, $F(1, 30) = 5.662$, $p = .024$, $\eta_p^2 = .159$, but not massed items, $F(1, 30) = 0.091$, $p = .765$.

To follow up the differences found across the age groups, post-hoc analyses were conducted using four paired-samples comparisons between the mean looking time to the target and distractor objects for each presentation condition and age group. Bonferroni corrections were used to correct for multiple comparisons. For the 16-month-old infants, results of the post-hoc analyses revealed that infants looked significantly longer toward the target items for the massed pairings, $t(15) = 3.178$, $p = .006$, but not the interleaved pairings, $t(15) = 1.303$, $p = .212$. The 20-month-old infants, in contrast, looked significantly longer toward the target items for both the massed pairings, $t(15) = 2.522$, $p = .023$, and the interleaved pairings, $t(15) = 2.187$, $p = .045$. Only the 20-month-olds, therefore, provided evidence of learning object-label associations under the (presumably) more challenging conditions presented by interleaved presentation.

Why were there differences across the age groups? One possibility is that experience learning words supported infant's developing ability to retrieve object-label pairings. In this case, we would expect to see a relation between test performance and vocabulary level, as measured by the MCDI. Alternatively, if domain general processes supported older infants' ability to retrieve pairings, we should expect to see a correlation between age and test performance. Thus, we examined the relationship between age (in months) and language development (MCDI count) in test performance (proportion looking to the target) on interleaved pairings using Pearson's r. Unsurprisingly, MCDI score was significantly correlated with age,

$r(32) = .393$, $p = .026$. However, test performance was significantly related to age, $r(32) = .404$, $p = .022$, but not MCDI score, $p = .290$. This suggests that development in a domain outside of language—presumably retrieval of associations from memory—supported older infants' ability to learn pairings distributed in time.

Discussion

This study revealed developments during the second year after birth in the ability to learn cross-situational statistics presented over varying timescales. The 16-month-old infants we observed were able to learn the massed, but not interleaved, object-label pairings whereas the 20-month-old infants were able to learn both the massed and interleaved object-label pairings. Test performance was related to age, but not level of vocabulary development, suggesting that domain general developments supported older infants' ability to learn pairings distributed in time. As noted previously, we believe that developing memory and retrieval processes are responsible for supporting infants' ability to aggregate and retrieve pairings over time. Taken together, this work has important theoretical implications for models of cross-situational statistical learning and language development, which are discussed below.

Memory Development & Retrieval Dynamics in Language Learning

In studies of memory development, researchers have sought to characterize developments in retrieval abilities across the first few years after birth (for a review, see Courage & Cowan, 2009). A consistent finding is that, if infants do not recall past events during learning, the current learning event is not related to the previous learning event (e.g., time-window hypothesis, Rovee-Collier, Evancio, & Earley, 1995). That is, the two learning events are not aggregated and/or bound together in memory. This hypothesis provides a ready explanation for 16-month-olds' performance; these infants were likely encoding the information for the interleaved object-label pairings, but not aggregating this information with prior pairings. Thus, at the test, the younger infants did not have a sufficient body of information with which to make inferences about object-label mappings.

Interestingly, the developing ability to retrieve information has also been used to explain rapid periods of language development (e.g., Dapretto & Bjork, 2000; Gershkoff-Stowe, 2002). This work has suggested that infants experience changes in domain general retrieval abilities that later result in an apparent vocabulary burst in language production. The current work is consistent with this theoretical account – there were striking differences in the ability to retrieve and learn object-label pairings across time in the pre-vocabulary burst (16 months) and post-vocabulary burst (20 months) age groups. Future work should continue to examine relations between memory processes in language production, the vocabulary burst, and development.

Theories of Cross-Situational Statistical Learning

To date, there are two primary theories of cross-situational statistical learning, associative learning accounts (e.g., Smith & Yu, 2008; Yu & Smith, 2011) and hypothesis testing accounts (e.g., Frank et al., 2009; Vouloumanos & Werker, 2009; Xu & Tenenbaum, 2007). In both of these theories, retrieving the past is a critical process underlying learning (for a review, see Yu & Smith, 2012). For example, in associative learning accounts, learners track co-occurrence probabilities that eventually develop into a matrix of word-referent associations. Word mapping is subsequently guided by the retrieval of association strengths. Similarly, in hypothesis testing accounts, learners must retrieve specific hypotheses about word–referent pairings and, in the face of current evidence, select among the retrieved hypotheses to infer correct mappings. In sum, according to theories of cross-situational learning, successful word mapping is dependent upon retrieval of past learning.

What happens when infants are unable to retrieve the past? Certainly the current experiment as well as a long history of research on infant memory (e.g., Courage & Cowan, 2009) suggest that there are significant constraints in infants' ability to recall prior learning events. Thus, theories of cross-situational statistical learning need to account for infants' developing ability to retrieve information over time.

It may be that infants go through an extended period of encoding information before it is aggregated, such as in an associative matrix (e.g., Yu & Smith, 2011) or data set by which hypotheses can be formed and tested (e.g., Xu & Tenenbaum, 2007). We suggest that simply encoding co-occurrence information may be the foundation for cross-situational statistical learning and word mapping. Over time and with experience, infants' ability to retrieve the past develops, which supports later aggregation and retrieval of prior learning. Indeed, before we make sense of a seemingly infinite amount of information, we may need to first take it all in – by encoding perceptual, temporal, and contextual information. That is, before fast mapping is possible, there might be *slow* mapping –a simple encoding of potential associations in the world.

References for CHAPTER 4

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121*, 127-132. doi: 10.1016/j.cognition.2011.06.010

Courage, M., & Cowan, N. (2009). *The development of memory in infancy and childhood*. New York, NY: Psychology Press.

Dapretto, M., & Bjork, E. L. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. *Child Development, 71*, 635-648. doi: 10.1111/1467-8624.00172

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59 (5, serial 242). doi: 10.2307/1166093

Frank, M.C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20,* 578-585. doi: 10.1111/j.1467-9280.2009.02335.x

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language, 46*, 665-687. doi: 10.1006/jmla.2001.2830

Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children word learning. *Monographs of the Society for Research in Child Development, 54*, 1-123.

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rovee-Collier, C., Evancio, S., & Earley, L. A. (1995). The time window hypothesis: Spacing effects. *Infant Behavior and Development, 18*, 69-78. doi: 10.1016/0163-6383(95)90008-X

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61,* 39-91. doi: 10.1016/S0010-0277(96)00728-7

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science, 35,* 480-498. doi: 10.1111/j.1551-6709.2010.01158.x

Smith, L. B. (2000). How to learn words: An associative crane. In R. Golinkoff & K. Hirsh-Pasek (Eds.), *Breaking the word learning barrier* (pp. 51–80). Oxford: Oxford University Press.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558-1568. doi: 10.1016/j.cognition.2007.06.010

Tomasello, M. (1992). The social bases of language acquisition. *Social Development, 1*, 67-87. doi: 10.1111/j.1467-9507.1992.tb00135.x

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611-1617. doi: 10.1037/a0016134

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272. doi: 10.1037/0033-295X.114.2.245

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165-180. doi: 10.1111/j.1467-7687.2010.00958.x

Yu, C. & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119,* 21-39. doi: 10.1037/a0026182

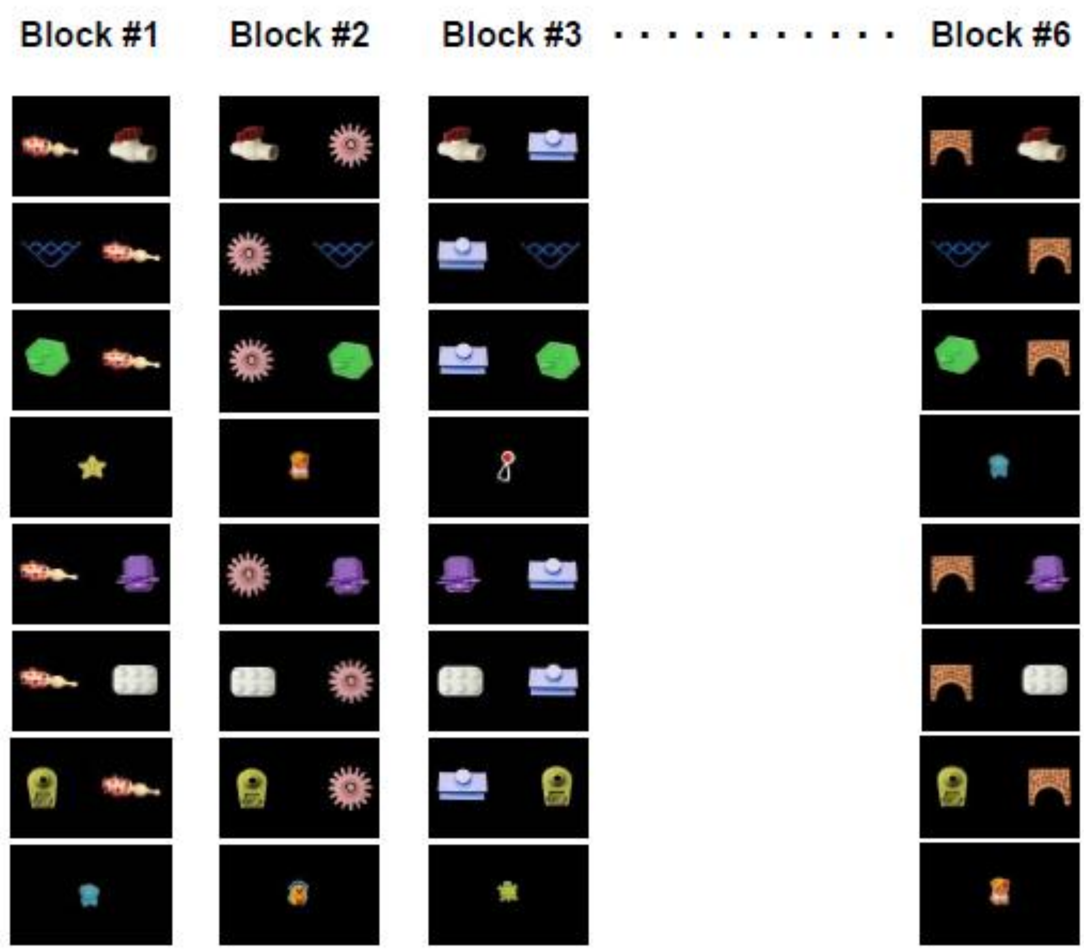

Figure 1. An example of the learning phase of the experiment. Infants were presented with six blocks of learning, which occurred in immediate succession until all six blocks were presented. Half of the object-label pairings were presented on a massed schedule and the other half-were presented on an interleaved schedule. The massed pairings were presented in immediate succession, in the same block. The interleaved pairings were presented at the same position in each block (for example, always presented on the first learning trial of a block). An attention getter trial was presented after every three learning trials.

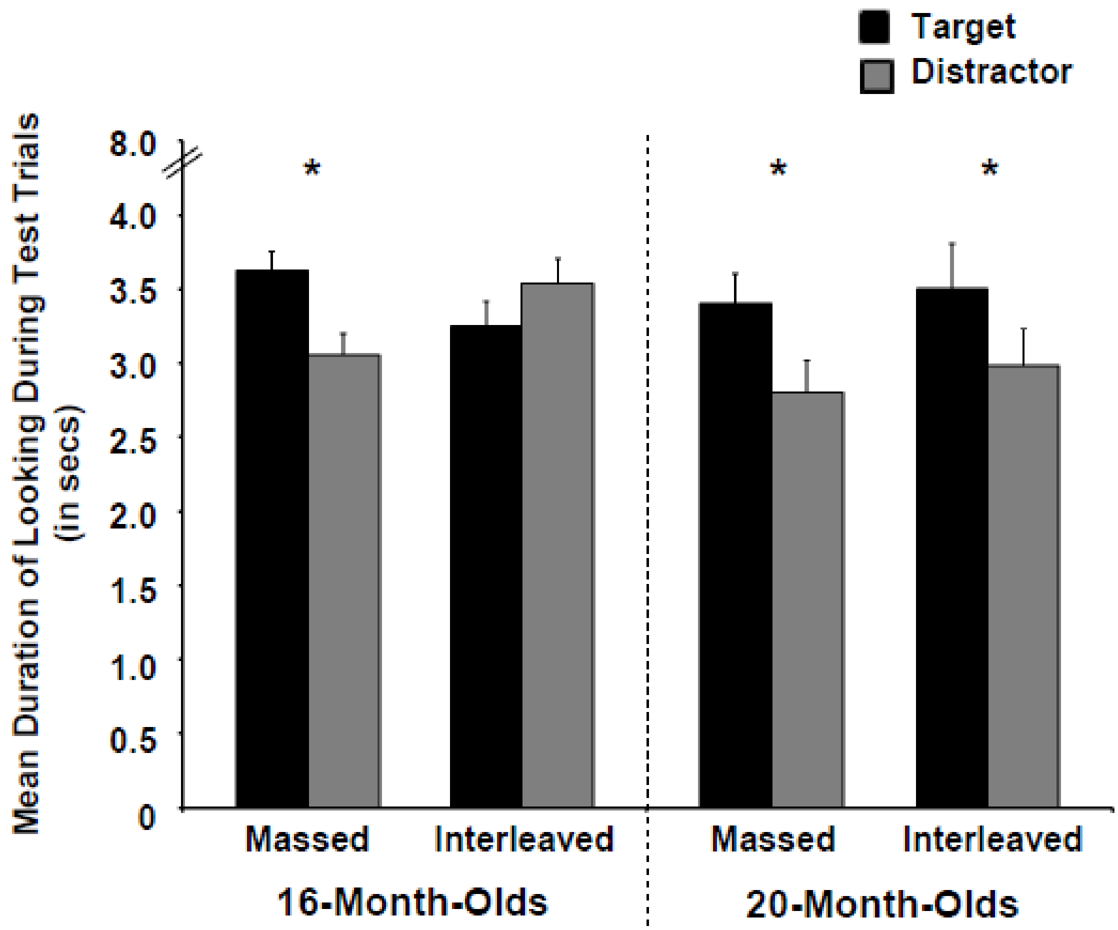

Figure 2. Mean looking time (in secs) to target and distractor objects during test trials, by massed/interleaved items and age group. A star (*) indicates a statistically significant difference between the duration of looking to the target and the duration of looking to the distractor, $p < .05$.

CHAPTER 5: Implications for Theories of Cross-Situational Statistical Learning

The experiments described in Chapters 2 – 4 were designed to expand upon current research on cross-situational learning by identifying mechanisms underlying the acquisition and retention of co-occurrence statistics. The results of this work revealed that, for adult learners, the retrieval dynamics that occurred during learning/acquisition were related to the ability to retain statistics across time. For infant learners, retrieval dynamics during learning were also important for learning outcomes. However, infants demonstrated significant constraints on their ability to aggregate and retrieve information over time. In sum, the current research identifies a mechanism underlying the acquisition and retention of co-occurrence statistics, retrieval dynamics. There are several implications of this work for theories of cross-situational statistical learning and language development, which are discussed below.

Theories & Mathematical Models of Cross-Situational Statistical Learning

Theories of cross-situational statistical learning generally fall into one of two categories: associative accounts (Smith & Yu, 2008; Yu & Smith, 2011) and hypothesis-testing accounts (Frank et al., 2009; Vouloumanos & Werker, 2009; Xu & Tenenbaum, 2007). The associative accounts suggest that learners track co-occurrence probabilities across learning events by simple and unconstrained counting mechanisms. Most associative models propose processes beyond simple counting, such as highly interactive effects of learned associations on each other and/or shifts in attention across different pairings (for a discussion, see Yu & Smith, 2012). Associative accounts are often criticized (e.g., Keil, 1992) with the argument that there are just too many possible associations across situations to keep track of and that learners need heuristics and/or hypotheses in order to reduce the number of associations that are stored. Thus, hypothesis-testing accounts hold a different view of how learners acquire and use cross-situational statistics.

According to hypothesis testing accounts, cross-situational learning is made possible by narrowing the possible number of referents through specific, defined hypotheses (Vouloumanos & Werker, 2009; Xu & Tenenbaum, 2007). In these theories, learners possess hypotheses about word–referent pairings and then, in the face of current evidence, select among those hypotheses based on some principled inference procedure in order to accept, reject, and/or weight current evidence (i.e., potential word-referent pairings). Many have argued that, although hypothesis testing accounts can be used to explain cross-situational statistical learning outcomes, these theories are incomplete because they lack an explanation for how learning transitions from being unconstrained early in infancy to constrained in childhood and adulthood (e.g., Yu & Smith, 2011).

Associative and hypothesis testing accounts of cross-situational statistical learning (and broader theories of cognition and development) often formalize theoretical views by including mathematical models of learning. In all of these theories, retrieving prior learning is both fundamental to successful learning and assumed to operate in an automatic nature. For example, Bayesian models of cognition and cognitive development (e.g., Xu & Tenenbaum, 2007) hold the assumption that prior learning and hypotheses are retrieved from the past in order to evaluate current experience.

However, retrieving prior knowledge may not be automatic and/or successful. The basic process of retrieving the past undergoes rapid development during the first few years after birth (e.g., Courage & Cowan, 2009). Furthermore, retrieval dynamics varied in adulthood, depending on the conditions of the learning environment. Indeed, the results in Chapters 2 - 4 may have been better predicted by mathematical models of memory processes, such as power or power-exponential functions (e.g., Wixted, 2004), rather than mathematical models of cross-situational statistical learning. Future research should examine this possibility.

Developmental psychologists have long critiqued broad theories of cognition for lacking an account of how initial and/or prior knowledge, such as a Bayesian prior, is acquired. The current research presents yet another challenge for such models – these models must account how fundamental processes, on which the theory rests, undergo dramatic changes across the lifespan. We suggest that future models include dynamic variables to represent the inventible dynamics in the development of information processing abilities, such as the ability to retrieve prior information from memory.

Development & Cross-Situational Statistical Learning

To date, the vast majority of research on cross-situational statistical learning has focused on adult learners and/or mathematical models of learning (see Vouloumanos, & Werker, 2009; Yu & Smith, 2011, for recent exceptions). The current work demonstrates that similar mechanisms (i.e., the retrieval dynamics that occur during learning) are involved in cross-situational statistical learning across the lifespan. However, the developmental state of the learner is likely mediating both the ability to retrieve prior learning and the degree to which retrieval difficulty is beneficial for long-term retention. Indeed, the ultimate goal of research on cross-situational statistical learning is to understand how it is the infants and young children acquire word mappings and language. Thus, future research should be grounded in development in order to characterize how learners are able to acquire cross-situational statistics while being constrained by developing information processing abilities.

Examining Learning In-the-Moment & Across Time

Research on cross-situational statistical learning was born out of the goal to characterize how it is that young infants and children learn new words across time (e.g., Siskind, 1996). Rather than concentrating on learning during a single moment in time, as had been the focus of prior research on word learning (see Introduction section titled "How Do We Learn Words?"), researchers reasoned that successful word learning was likely to be a product of several experiences distributed over time.

Although research on cross-situational statistical learning is in its infancy, the vast majority of research has examined in-the-moment processes across brief timescales. That is, experiments have been designed so that learners are presented with a few minutes of learning trials and then tested immediately. In fact, to our knowledge, the experiments reported in Chapter 2 & 3 are the first to examine the retention of co-occurrence statistics learned via cross-situational statistical learning.

In the spirit of the eventual goal of cross-situational statistical learning research, future work should continue to examine the in-the-moment mechanisms underlying cross-situational word learning over broader timescales. In order to account for real-world learning, research should incorporate learning and testing over longer timescales—over the course of weeks, months, and years. A complete theory of cross-situational word learning not only accounts for learning in the moment and at each time point, but also integrates them in order to understand how they influence each other over time. Taken together, this work will provide a mechanistic account of how we learn new words despite the inherent ambiguity and difficulty of the task.

References for CHAPTER 5

Courage, M., & Cowan, N. (2009). *The development of memory in infancy and childhood*. New York, NY: Psychology Press.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20,* 578-585. doi: 10.1111/j.1467-9280.2009.02335.x

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

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39-91. doi: 10.1016/S0010-0277(96)00728-7

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558-1568. doi: 10.1016/j.cognition.2007.06.010

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611-1617. doi: 10.1037/a0016134

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology, 55*, 235-269. doi: 10.1146/annurev.psych.55.090902.141555

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272. doi: 10.1037/0033-295X.114.2.245

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*, 165-180. doi: 10.1111/j.1467-7687.2010.00958.x

Yu, C. & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21-39. doi: 10.1037/a0026182