COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 38 (2014) 757–774

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12092

Retrieval Dynamics and Retention in Cross-Situational Statistical Word Learning

Haley A. Vlach,[a] Catherine M. Sandhofer[b]

[a]*Department of Educational Psychology, University of Wisconsin*
[b]*Department of Psychology, University of California, Los Angeles*

Received 2 May 2012; received in revised form 17 November 2012; accepted 1 April 2013

Abstract

Previous research on cross-situational word learning has demonstrated that learners are able to reduce ambiguity in mapping words to referents by tracking co-occurrence probabilities across learning events. In the current experiments, we examined whether learners are able to retain mappings over time. The results revealed that learners are able to retain mappings for up to 1 week later. However, there were interactions between the amount of retention and the different learning conditions. Interestingly, the strongest retention was associated with a learning condition that engendered retrieval dynamics that initially challenged the learner but eventually led to more successful retrieval toward the end of learning. The ease/difficulty of retrieval is a critical process underlying cross-situational word learning and is a powerful example of how learning dynamics affect long-term learning outcomes.

Keywords: Cross-situational word learning; Statistical learning; Word learning; Language development; Memory and learning; Retrieval dynamics; Retrieval effort hypothesis

1. Introduction

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?

Correspondence should be sent to Haley A. Vlach, Department of Educational Psychology, University of Wisconsin, 859 Educational Sciences, 1025 W. Johnson Street, Madison, WI, 53706. E-mail: hvlach@wisc.edu

Historically, word learning research has focused on identifying the processes involved in resolving ambiguity in one moment in time. This body of 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; Kachergis, Yu, & Shiffrin, 2012; Scott & Fisher, 2012; Siskind, 1996; Smith, Smith, & Blythe, 2010; Smith & Yu, 2008; Yu & Smith, 2007, 2011, 2012; Yurovsky, Yu, & Smith, 2013). This body of work has revealed that adult learners can track co-occurrence of word-referent pairings with varying degrees of within-trial ambiguity (e.g., numbers of words and referents; see 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 mappings.

Do learners retain cross-situational mappings over 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 study takes an important first step in examining whether learners can retain cross-situational mappings over time and, if they are able to retain mappings, how low-level memory processes support the ability to do so.

In this article, we report two experiments that were designed to examine learners' long-term retention of cross-situational mappings. In both Experiment 1 and 2, learners' acquisition and retention of word-referent (i.e., object–label) pairings was tested at an immediate or 1-week delayed forced-choice test. The pairings were presented in three learning conditions, which varied the amount of within-trial ambiguity, to capture an array of conditions under which learners are typically presented with cross-situational statistics (e.g., Yu & Smith, 2007). Because these learning conditions typically present learners with varying numbers of objects and labels, we predicted that there may be different memory demands and processes operating in each of the learning conditions.

Experiment 2 was also designed to reveal how memory processes may be supporting and/or deterring the ability to retain cross-situational mappings. Specifically, we examined

the retrieval dynamics occurring during learning. We predicted that the ease and/or difficulty in retrieving information during learning may affect learners' ability to retrieve information at a later point in time. Indeed, previous research has indicated that difficult but eventually successful retrieval (e.g., Carpenter & DeLosh, 2006; Halamish & Bjork, 2011; Kornell, Hays, & Bjork, 2009; Pyc & Rawson, 2009; Richland, Kornell, & Kao, 2009; Vlach, Ankowski, & Sandhofer, 2012) and retrieval practice (e.g., Karpicke & Roediger, 2007; Roediger & Butler, 2011) can support the long-term retention of information. We examined whether these dynamics occur during cross-situational word learning and, if so, how they may be related to retention. We predicted that learning conditions that engender the most optimal retrieval dynamics would result in higher levels of retention than other learning conditions. In sum, these experiments took the important first steps in elucidating the mechanisms that support the long-term ability to retain cross-situational mappings.

2. Experiment 1

In this experiment, we started by examining whether learners would be able to retain cross-situational mappings over a real-world period of time: 1 week. Learners were presented with a cross-situational word learning task, across three learning conditions which varied the number of objects and labels, and tested immediately or 1 week later. If learners are able to retain cross-situational mappings, we predicted that performance would be above chance at the 1-week delayed test. If participants are not able to retain these mappings over the 1-week period, we predicted that performance would be at chance at the 1-week delayed test.

2.1. Method

2.1.1. 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.

2.1.2. 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 Fig. 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 (2×2,

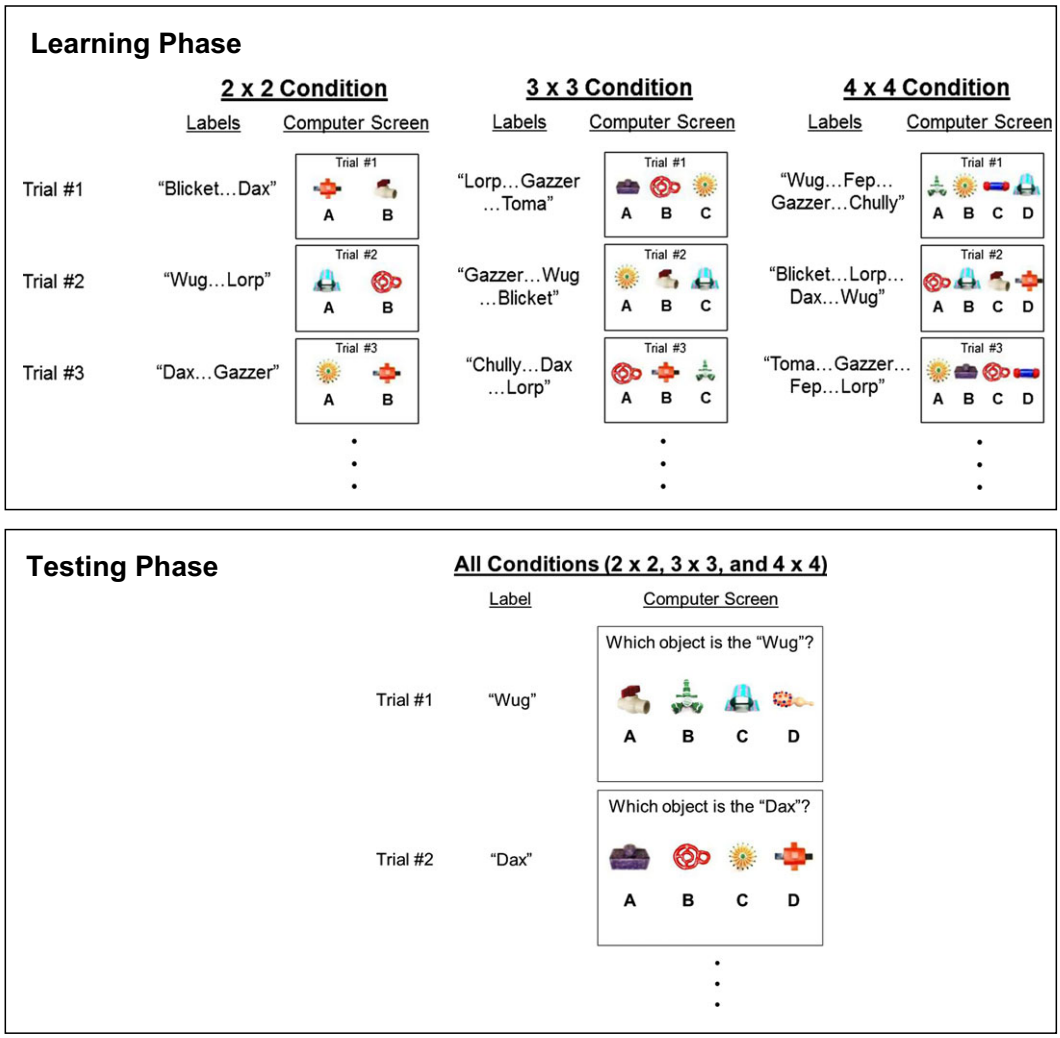

Fig. 1. Example trials from the learning and testing phases of Experiment 1 and 2. During the learning phase of both experiments, participants were presented with a series of learning trials according to one of three learning conditions, 2 × 2, 3 × 3, and 4 × 4. See Table 1 for the number and timing 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.

3 × 3, and 4 × 4), there were a total of six presentations of each of the 18 object–label pairs during the learning phase.

2.1.3. Design

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

Table 1
Trial composition for the three learning conditions

Learning Condition	No. of Object–Label Pairings	No. of Presentations of Each Pairing	No. of Trials	Time per Trial (in seconds)	Total Learning Time (in seconds)
2 × 2	18	6	54	6	324
3 × 3	18	6	36	9	324
4 × 4	18	6	27	12	324

2.1.4. 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 three training trials, each with two objects and two labels (similar to the 2 × 2 condition, see Fig. 1), immediately followed by one forced-choice testing trial (similar to the testing phase, see Fig. 1). 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 × 2, 3 × 3, and 4 × 4. In the 2 × 2 condition, two objects and two words were presented in each learning trial (see Fig. 1). In the 3 × 3 condition, three objects and three labels were presented. In the 4 × 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 (six presentations each), other presentation factors varied across conditions 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); the duration of the individual trials varied across the three conditions, but the time allocated to each object (3 s) and overall amount of exposure time (324 s) was constant across the three conditions.

Participants were presented with learning trials according to the condition in which they had been randomly assigned (2 × 2, 3 × 3, or 4 × 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 1 week later and complete the testing phase.

Testing Phase. The testing phase consisted of four forced-choice trials (see Fig. 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 16 labels and objects used during the testing phase were randomly assigned for each participant.

2.2. Results and discussion

We were interested in whether learners would be able to learn and retain cross-situational mappings over a real-world period of time. To examine this question, we first conducted a 3 (Learning Condition) × 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, $F(2, 66) = 6.070$, $p = .004$, $\eta_p^2 = .155$.

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 learning processes that affected the long-term ability to retain cross-situational mappings, we would expect there to be differences in performance between the learning conditions and across the testing conditions. As can be seen in Fig. 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 Fig. 2). Participants in the 2 × 2 condition had significantly higher performance than participants in the 3 × 3, $p = .043$, and 4 × 4 conditions, $p < .001$. Moreover, performance in the 3 × 3 condition was significantly higher than the 4 × 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 and Smith (2007).

However, in the 1-week delayed testing condition, there was a strikingly different pattern of performance (see Fig. 2). There was a main effect of learning condition, $F(2, 33) = 10.482$, $p < .001$, $\eta_p^2 = .388$. Participants in the 3 × 3 condition had significantly higher performance than participants in the 2 × 2, $p = .048$, and 4 × 4 conditions, $p < .001$. Moreover, performance in the 2 × 2 condition was marginally significantly higher than performance in the 4 × 4 condition, $p = .086$. Hence, although initially participants in the 3 × 3 condition had lower performance than participants in the 2 × 2 condition, 1 week later participants in the 3 × 3 condition had higher performance than participants in both the 2 × 2 and 4 × 4 conditions.

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

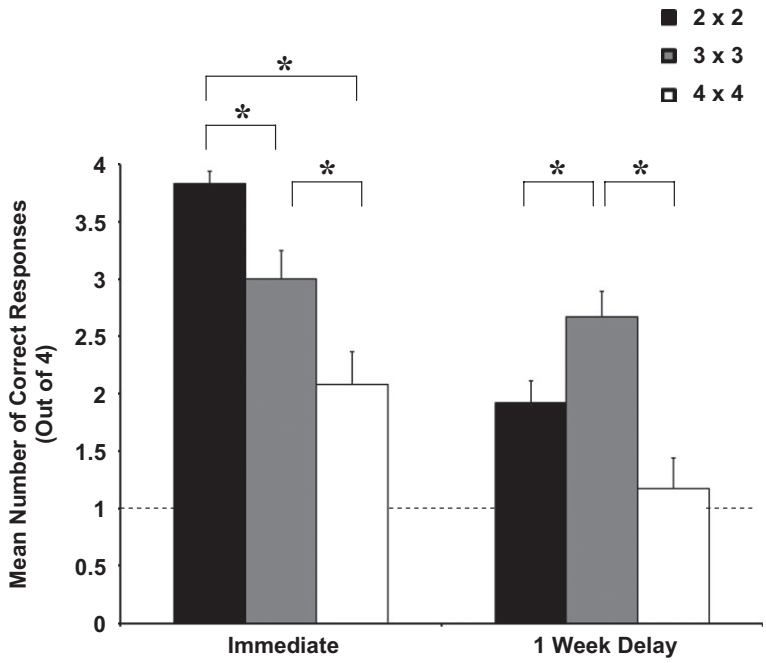

Fig. 2. Results of testing phase in Experiment 1. Mean number of correct responses (out of 4) by learning condition (2 × 2, 3 × 3, and 4 × 4) and testing condition (immediate and 1-week delayed). The dashed line represents chance performance, and a star indicates a statistically significant difference,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

How did learning dynamics differ across the three 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,b; 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 who 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 in the 2 × 2 and 3 × 3 conditions in Experiment 1. In the case of the 4 × 4 condition, the overall lower performance may be an indication that learners struggled and were not successful at retrieving information by the end of the learning phase.

Were there different retrieval dynamics occurring across the three learning conditions? We predicted that participants in the 2 × 2 condition were experiencing the greatest ease of retrieving object–label pairings during learning, compared to participants in the 3 × 3

and 4 × 4 conditions. For participants in the 3 × 3 condition, we predicted that they were experiencing an intermediate degree of difficulty retrieving object–label pairings. Because participants in the 3 × 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 × 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.

3. Experiment 2

In this experiment, we examined whether there were retrieval dynamics that differed across the three learning conditions. We hypothesized that participants in the 3 × 3 condition were experiencing more optimal retrieval processes than participants in the 2 × 2 and 4 × 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 ease and/or difficulty of retrieving information occurring during learning.

3.1. Method

3.1.1. 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.

3.1.2. Apparatus and stimuli

The computer and stimuli used in Experiment 1 were also used in Experiment 2. In addition, 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 Fig. 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.

3.1.3. Design

Same as Experiment 1.

3.1.4. 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 Fig. 3). During the training phase, the experimenter provided instructions on 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 × 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, he or she was 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).

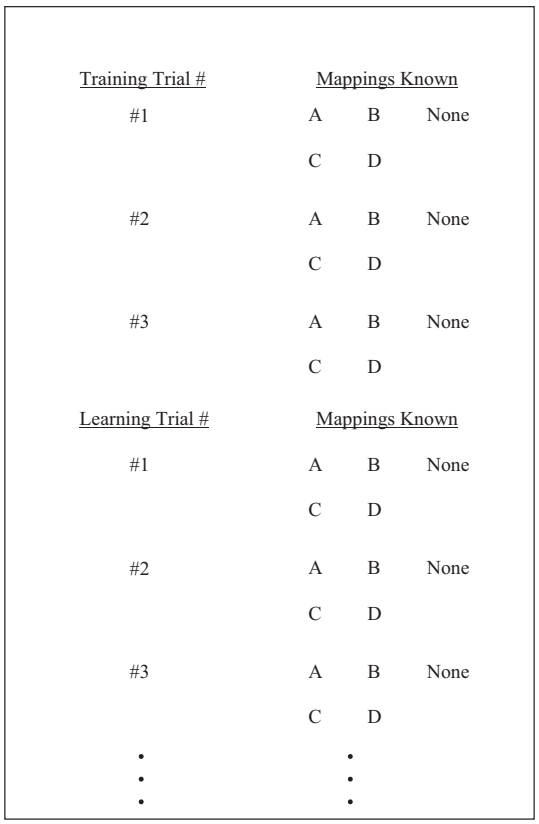

Fig. 3. Example of worksheet used during the self-report retrieval task in Experiment 2. Participants were instructed to record successes of retrieving object–label pairings by circling one letter, multiple letters, or none. Worksheet was used during the training and learning phases of Experiment 2.

3.2. Results and discussion

3.2.1. Final test performance

We first examined the final test performance to see whether 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) × 2 (Testing Delay) ANOVA, with the number of correct responses at the final test as the dependent measure (see Fig. 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, $F(2, 72) = 5.808$, $p = .001$, $\eta_p^2 = .173$.

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 × 2 condition had significantly

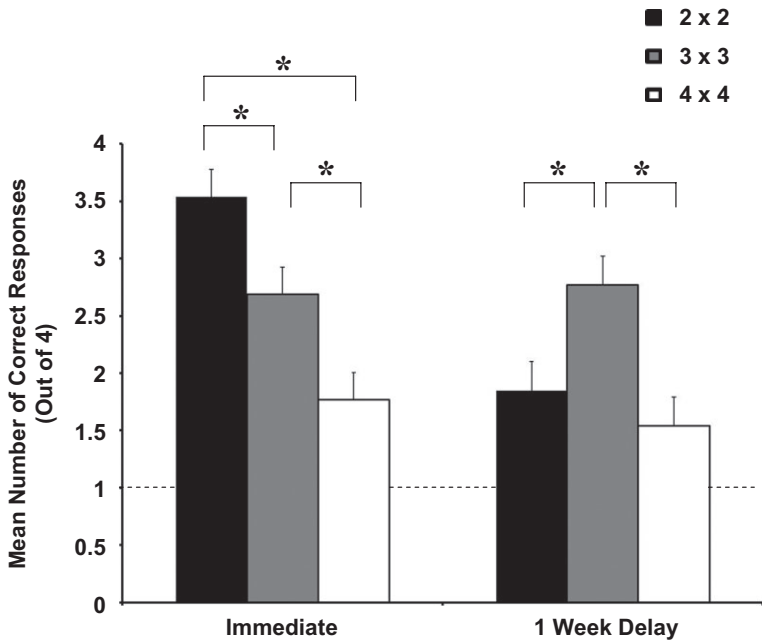

Fig. 4. Results of testing phase in Experiment 2. Mean number of correct responses (out of 4) by learning condition (2 × 2, 3 × 3, and 4 × 4) and testing condition (immediate and 1-week delayed). The dashed line represents chance performance, and a star indicates a statistically significant difference,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

higher performance than in the 4 × 4 condition, $p < .001$. Participants' performance was also higher in the 2 × 2 condition than the 3 × 3 condition, $p = .049$. Finally, participants' performance in the 3 × 3 condition was significantly higher than the 4 × 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 1 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 × 3 condition had higher performance than both the 2 × 2 condition, $p = .039$, and 4 × 4 condition, $p = .004$. Participants in the 4 × 4 condition did not have significantly different performance than participants in the 2 × 2 condition, $p > .05$. In sum, the pattern of final test performance seen in Experiment 1 was replicated in Experiment 2 (compare Figs. 2 and 4).

3.2.2. 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 in the ease and/or difficulty in retrieving information that could be contributing to differences in long-term performance. Specifically, we predicted that retrieval dynamics during learning could be a mechanism underlying immediate 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 × 2 condition reported a significantly higher number of retrieval successes ($M = 53.81$, $SD = 21.83$) than participants in the 3 × 3 condition ($M = 42.23$, $SD = 14.02$), $p = .058$, and participants in the 4 × 4 condition ($M = 23.65$, $SD = 15.48$), $p < .001$. Moreover, participants in the 3 × 3 condition reported a significantly higher number of retrieval successes than participants in the 4 × 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. To examine the ability to successfully retrieve object–label pairings during learning, we started by dividing the learning phase into nine blocks of time, 36 s each. We chose this time scale because, over 36 s, participants in all of the conditions were exposed to the same number of object–label pairings. For example, in the 2 × 2 condition, there were six trials with

two object–label pairings, for a total of 12 object–label pairings. In the 3 × 3 condition, there were four trials with three object–label pairings, for a total of 12 object–label pairings. Finally, in the 4 × 4 condition, there were three trials with four object–label pairings each, for a total of 12 object–label pairings.

After dividing the learning phase into nine time scales, we then computed the mean number of reported retrieval successes during each time scale, for each learning condition. Each time point (Time1–Time9) represents the mean number of reported 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, Time2 represents the mean number of retrieval successes between Time1 and Time2, and so forth. The descriptive results can be seen in Fig. 5.

We then conducted a mixed 3 (Learning Condition) × 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,

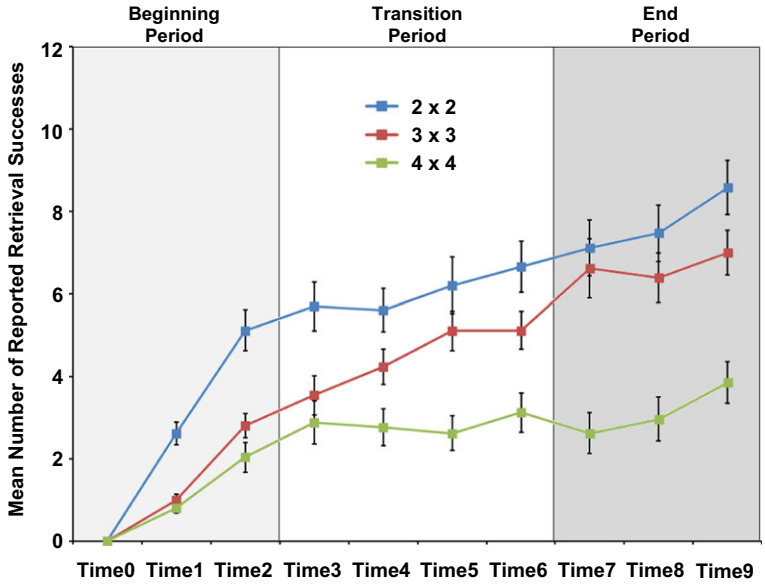

Fig. 5. Mean number of reported retrieval successes during the learning phase, for the three learning conditions (2 × 2, 3 × 3, and 4 × 4) by time interval. Error bars represent standard errors. The learning phase was divided into three periods to categorize the nature of the pattern of performance for participants in the 3 × 3 condition. During the Beginning Period, participants in the 3 × 3 condition did not significantly differ in the mean number of retrieval successes from participants in the 4 × 4 condition. However, participants in the 3 × 3 condition did have significantly lower retrieval performance than participants in the 2 × 2 condition. During the Transition Period, participants in the 3 × 3 condition reported an intermediate degree of retrieval success; performance in the 3 × 3 condition was significantly different from performance in both the 2 × 2 and 4 × 4 conditions. Finally, during the End Period, participants in the 3 × 3 condition reported a significantly higher number of retrieval successes than participants in the 4 × 4 condition, but not the 2 × 2 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, $F(16, 136) = 3.412$, $p < .001$, $\eta_p^2 = .286$.

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 Fig. 5). During the first period, the Beginning Period (Time0–Time2), participants in the 2 × 2 condition reported significantly more retrieval successes than participants in the 3 × 3 and 4 × 4 conditions, at Time1 and Time2, ps <. 00185. There were no significant differences between the number of reported retrieval successes in the 3 × 3 and 4 × 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 × 2 condition were experiencing a greater ease in retrieval compared to participants in the 3 × 3 and 4 × 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 × 2 condition reported significantly more retrieval successes than participants in the 3 × 3 and 4 × 4 conditions. Moreover, participants in the 3 × 3 condition reported significantly more retrieval successes than participants in the 4 × 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 × 3 condition experiencing an intermediate degree of difficulty compared to participants in the 2 × 2 and 4 × 4 conditions.

Finally, during 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 × 2 and 3 × 3 conditions at all time points, Time7–Time9, ps > .00185. Moreover, participants in the 2 × 2 and 3 × 3 conditions reported significantly more retrieval successes than participants in the 4 × 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 × 2 and 3 × 3 conditions.

In sum, there were significant changes in the pattern of reported retrieval successes across the learning phase of the 3 × 3 condition. Initially (during the Beginning Period), participants in the 3 × 3 condition appeared to struggle to retrieve object–label pairings; retrieval performance was significantly lower than retrieval performance in the 2 × 2 condition and not significantly different than that of retrieval performance in the 4 × 4 condition. In the middle part of the learning phase (during the Transition Period),

participants in the 3 × 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 × 2 and 3 × 3 conditions, suggesting that by the end of the learning phase, participants in the 3 × 3 condition were experiencing a greater degree of retrieval success.

These findings confirm our hypothesis that the retrieval dynamics during learning differed across the three conditions. There were differences in the overall number and pattern of retrieval successes across the three learning conditions. Participants in the 3 × 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 1-week delayed test. Indeed, the pattern of retrieval successes in the 3 × 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.

3.2.3. 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 who reported successfully retrieving more object–label pairings should have had higher final test performance at the immediate test. Conversely, participants who 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 1-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 they were successfully retrieving correct object–label pairings.

4. General discussion

The experiments in this study were designed to examine whether cross-situational mappings are retained over a real-world period of time. To our knowledge, this is the

first study to demonstrate that learners are able to retain cross-situational mappings up to 1 week later. Moreover, the ability to retain word-referent pairings was related to the retrieval dynamics that occurred during learning. Interestingly, participants who 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 × 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.

4.1. Memory processes and cross-situational statistical word learning

This study demonstrates that different conditions of cross-situational word learning engender varying retrieval dynamics during learning and over real-world periods of time, such as at a 1-week delayed test. In this study, there were striking differences in the number and pattern of retrieval successes across the 2 × 2, 3 × 3, and 4 × 4 learning conditions. Why did the 3 × 3 condition engender more favorable retrieval dynamics for long-term retention?

We do not hypothesize that there is something unique about the 3 × 3 learning condition versus the other learning conditions. What is more likely is that working memory and/or short-term memory moderated the ability to hold enough information in memory 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 × 2 condition, there were two words and two objects presented in each presentation, for a total of four items. In the 3 × 3 condition there were three words and three objects, for a total of six items. Finally, in the 4 × 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 × 3 condition may have been more taxing of working memory than the 2 × 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 × 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 mappings to explore this possibility.

The current research also has important implications for broad theories of cross-situational statistical learning. Theories of cross-situational statistical learning generally fall into one of two categories: associative accounts (e.g., Smith & Yu, 2008; Yu & Smith, 2011) and hypothesis-testing accounts (e.g., Frank et al., 2009). In both categories of theories, including computational/mathematical models of these theories, retrieving prior learning is both fundamental to successful learning and assumed to operate in an automatic nature. However, retrieving prior knowledge may not be automatic and/or successful. Indeed, the learners in this study demonstrated a wide array of retrieval dynamics during learning and at test. Thus, theories and models of cross-situational learning should be revised to account for varying retrieval dynamics. For example, mathematical models of cross-situational learning should begin to incorporate variables of forgetting and retrieval variability (such as power or power-exponential functions of forgetting; see Wixted, 2004, for a review).

Finally, it is also important to note that 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 time scales are likely to reveal how young learners overcome memory constraints on the developing ability to retrieve information from the past.

4.2. Looking forward: Learning across broader time scales

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

Acknowledgments

We thank the research assistants of the UCLA Language and Cognitive Development Lab for their assistance in collecting the data for this project. Research discussed in this article was supported by NICHD grant R03 HD064909-01.

References

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.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word-referent mappings. *Psychonomic Bulletin & Review*, *19*(2), 317–324. doi: 10.3758/s13423-011-0194-6.

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, and 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, 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, England: Oxford University Press.

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., & 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.

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

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

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, and Cognition*, *38*, 246–254. doi: 10.1037/a0025260.

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.

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.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*, 891–921. doi: 10.1111/cogs.12035