UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Desirable Difficulties in Cross-Situational Word Learning

Permalink
https://escholarship.org/uc/item/2151507t

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Vlach, Haley
Sandhofer, Catherine

Publication Date
2010

Peer reviewed

Desirable Difficulties in Cross-Situational Word Learning

Haley A. Vlach (haleyvlach@ucla.edu)
Catherine M. Sandhofer (sandhof@psych.ucla.edu)
Department of Psychology, University of California, Los Angeles
1285 Franz Hall, Los Angeles, CA 90095 USA

Abstract

The world offers learners a seemingly infinite number of word-to-world mappings (Quine, 1960). In order to account for how learners manage to accomplish such a difficult task, theories of word learning have proposed different tools that make the task of learning words easier. However, we propose that reducing difficulty may be detrimental—difficulty may promote long-term word learning. We tested this hypothesis in a cross-situational paradigm in which object-label mappings were ambiguous during each learning event. The three conditions of learning (2 x 2, 3 x 3, and 4 x 4) varied in the degree of difficulty. Results revealed that, although difficulty deterred immediate performance, difficulty promoted long-term performance. We suggest that theory and research should shift from focusing on in-the-moment learning to examining both immediate *and* long-term learning. A complete theory of word learning not only accounts for word learning in the moment and on each time scale, but also integrates them in order to understand how they influence each other over time.

Keywords: word and category learning; statistical learning; cross-situational learning; long-term memory

Introduction

The world offers learners a seemingly infinite number of word-to-world mappings (Quine, 1960). Thus, an essential question for research on word learning is: How do learners manage to accomplish the difficult task of mapping words to objects, actions, and events in the world?

Theories of word learning typically focus on tools that learners use to make word learning easier. In this study, we examine word learning from the radical perspective that reducing difficulty may be detrimental. This study explores the idea that some difficulty may promote word learning, even in difficult tasks in which learners must track mappings across events, such as cross-situational word learning.

Theories of Word Learning. Three main classes of theories have sought to explain word learning: the Constraints/Principles theories, the Social-Pragmatic theories, and the Domain-General theories. All three of these theories propose tools that make word learning easier but differ in the nature of the task simplification tools.

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 an unfamiliar label to an unfamiliar object rather than an object that has already been named.

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 learners are already focused on and consequently make it easier for learners to make word-to-world mappings (Bloom, 1993).

A final 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., 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, suggesting that saliency cues alone guided children's word learning.

Word Learning and Memory. Although the three classes of theories make different predictions about many aspects of word learning, in this study we investigated a cognitive mechanism that is inevitably a critical part of each theory: memory. For example; the Constraint/Principles theories argue that constraints promote memory for words—if everything had a multiple unique labels it would be impossible to store and recall all of these words from memory. Social-Pragmatic theories rely on processes of memory and attention for establishing joint attention among two people—learners must attend and remember what others are focusing on in order to adequately label words and actions (e.g., Bloom, 1993). Domain-General theories assert that word learning is guided by global principles of attention, association, and frequency, which are basic cognitive mechanisms associated with memory (e.g., Smith, 1995). In sum, memory is a critical component to word learning theories because it supports every part of the word learning process—learning words requires attending to

words, encoding the properties of the word, binding words to objects in the world, and recalling words when needed in order to communicate with others.

Although it is clear that memory matters for learning words, relatively little work has investigated the role of memory and retention in word learning. In fact, only a handful of studies have imposed a delay between learning and testing (see Horst & Samuelson, 2008, for a discussion of this issue). Consequently, the vast majority of word learning theories are based upon immediate performance rather than performance over time.

Examining word learning both immediately and over time is essential for two reasons. First, a complete theory of word learning accounts for developmental changes in word learning and retention abilities. Moreover, such a theory not only accounts for word learning and retention on each time scale, but also integrates them in order to understand how they influence each other over time.

Second, immediate performance may not be a reflection of long-term performance. The few studies that have examined word learning and retention have yielded mixed results as to whether performance remains constant over time (e.g., Horst & Samuelson, 2008; Markson & Bloom, 1997). Alternatively, the memory literature has provided countless examples of how the factors that promote immediate learning do not necessarily promote long-term learning (e.g., Bjork, 1994). Immediate performance may not predict long-term performance because the degree of difficulty in learning influences long-term performance.

Desirable Difficulties in Learning. There has been a long history of research that has investigated the conditions under which long-term memory is enhanced. The principle goals of this research have been to discover factors that promote adults' ability to (1) produce and store a representation of knowledge and (2) create a representation that remains accessible and recallable over extended periods of time. Research has revealed that several manipulations of learning events can enhance long-term memory, such as varying the conditions of practice (e.g., Smith & Rothkopf, 1984), providing contextual interference (e.g., Mannes & Kintsch, 1987), distributing practice and the spacing effect (e.g., Cepeda et al., 2006), and reducing feedback to the learner over time (e.g., Schmidt, 1991).

These manipulations promote long-term memory because they introduce difficulty for learners while knowledge is being acquired (e.g., Bjork, 1994). Although learning tasks that are designed to make learning easy initially show greater learning, retention tests reveal that more difficult learning tasks promote more long-term memory and learning (and hence the term 'desirable' difficulty is commonly used). Thus, the memory literature suggests that, instead of making tasks easy for learners, the best way to promote long-term memory is to create difficulty for learners during learning.

An example of a desirable difficulty in learning is the spacing effect (e.g., Cepeda et al., 2006). The spacing effect describes the robust phenomenon whereby memory is enhanced when learning events are distributed across time (i.e., spaced), instead of being presented in immediate succession (i.e., massed). Spaced learning is more difficult than massed learning because the time between learning events creates greater opportunities for forgetting (e.g., Bjork & Allen, 1970). Massed presentations prevent forgetting because presentations are in immediate succession. In fact, upon immediate testing, massed presentations lead to a greater amount of learning than spaced presentations. However, if a test is administered following a delay, a spaced presentation schedule will yield more learning than the massed presentation schedule.

Several researchers have long suggested that, although introducing difficulty during memory tasks is beneficial, these difficulties may be detrimental in more difficult cognitive tasks (e.g., Gagne, 1950). For example, spaced learning was thought to be particularly detrimental in generalization tasks. In fact, spaced learning was coined the "enemy of induction" (e.g., Gagne, 1950; see Kornell & Bjork, 2008, for a discussion).

Desirable Difficulties in Word Learning. Despite speculations that desirable difficulties may be the "enemy of induction", recent research suggests that imposing difficulty during learning promotes long-term word learning and generalization (e.g., Kornell & Bjork, 2008; Vlach et al., 2008). For example, one study (Vlach et al., 2008) presented children with novel objects and labels in an object category learning paradigm. Category exemplars were presented on two schedules, massed and spaced. Children were tested after a three minute delay and were required to generalize a word to a novel instance of a category. The results revealed that spaced presentations promoted more learning than massed presentations. Thus, a spaced learning schedule, a more difficult learning schedule, promoted word learning and generalization.

One limitation of research on desirable difficulties in word learning is that all of the studies have been artificially simplistic—a linguistic label could only be mapped onto one object. In real word learning contexts, mapping words to objects is generally not this straightforward. Word learners must figure out what words map onto in the world (Quine, 1960). Thus, because learners must track possible mappings across learning events, real world word learning is much more difficult than tested in recent research on desirable difficulties in word learning.

Research on cross-situational learning has indicated that the more objects and labels in each learning event, the more difficult it is for learners to determine mappings (e.g., Smith & Yu, 2008; Yu & Smith, 2007; Yurovsky & Yu, 2008). For example, when adult learners are presented with two words and two objects in learning events, they demonstrate relatively high performance, ~90% correct mappings on an immediate test. However, when learners are presented with four objects and four labels in each word learning event, learners perform significantly lower, ~55% correct

mappings on an immediate test (Yu & Smith, 2007). Thus, it appears that the more objects and labels in each learning event, the more difficult it is to track mappings across learning situations.

The memory literature would suggest that increasing the number of objects and labels in each word learning event presents several forms of desirable difficulty. First, increasing the number of object and labels in each learning event creates more spaced learning because each object-label pairing is interleaved between more possible pairings (e.g., Cepeda et al, 2006; Vlach et al., 2008). Second, an increase in the number of objects and labels in each learning event creates more contextual variation and interference between word learning events (e.g., Mannes & Kintsch, 1987). Both of these factors have been shown to promote long-term retention (e.g., Bjork, 1994).

Although recent research suggests that difficulty may promote word learning, this hypothesis has only been tested in artificially simple tasks where object-label mappings are straightforward. It may be the case that adding more difficulty to an already difficult task of mapping words to objects is not beneficial. Consequently, too much difficulty may deter both in-the-moment and long-term word learning. The current study investigates this possibility.

Current Study. The current study investigated the role of difficulty during word learning in a cross-situational word learning paradigm. Participants were presented with word learning events in which determining the object-label mappings were increasingly difficult. In the 2 x 2 condition, each trial presented two words and two objects. In the 3 x 3 condition, each trial presented three words and three objects. Finally, in the 4 x 4 condition, each trial presented four words and four objects. There were also three testing delay conditions: immediate, 30 minute delay, and one week delay. These conditions allowed for a direct comparison of the effects of varying degrees of difficulty in both in-the-moment and long-term word learning.

Method

Participants
Participants were 95 undergraduates at University of California, Los Angeles. Participants received course credit for their participation.

Design
This study used a 3 x 3 design. Learning Condition (2 x 2, 3 x 3, and 4 x 4) and Testing Delay (immediate, 30 minutes, and one week) were both between-subjects factors. Participants were randomly assigned to one of the nine conditions of the study.

Stimuli
Pictures of objects were presented on a 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. The labels were novel words following the phonotactic probabilities of English (e.g., 'blicket', 'dax'). There were a total of 18 labels. Objects and labels were randomly paired together, for a total of 18 object-label pairs. In all conditions, there were a total of 6 presentations of each of the 18 object-label pairs. There were also an additional four objects and four labels presented during the training trial. These objects and labels were not used during the learning phase of the experiment.

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.
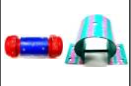
	Labels	Computer Screen
Trial #1	"Blicket"…"Dax"	
Trial #2	"Wug"…"Lorp"	
Trial #3	"Blicket"…"Spog"	
Trial #4	"Gazzer"…"Wug"	

Figure 1. Example stimuli from the 2 x 2 condition.

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).

Table 1. Three Learning Conditions

Condition	Number of Trials	Time per Trial (in secs)	Total Time (in secs)
2 x 2	54	6	324
3 x 3	36	9	324
4 x 4	27	12	324

Procedure
Participants were told that they would be shown children's toys and it was their job to figure out which word went with which toy. They were also instructed that it would be

ambiguous as to which words went with which objects on each trial. Participants were then given a brief training exercise to demonstrate what the experiment would be like. The training consisted of three learning trials, each with two objects and two labels, immediately followed by a forced-choice test. Objects and labels used during training were not included during the rest of the experiment.

After the training trial, participants were informed that they would now be beginning the learning phase of the experiment. Participants were presented with learning trials according to the condition in which they were assigned (2 x 2, 3 x 3, or 4 x 4). The number and length of trials was also set according to the condition (see Table 1).

After viewing all of the trials, participants were given a forced-choice test, depending upon the testing condition in which they were assigned. In the immediate condition, participants were given a test immediately following learning. In the 30 minute delay condition, participants were asked to play tetris for 30 minutes, and then were given a test. In the one week delay condition, participants were asked to come back exactly 7 days after the learning session and complete a test.

The test consisted of four force-choice questions. Each question presented one label and asked participants to identify the corresponding object among four objects. The three foil objects were other objects used in the experiment. No one object was repeated in the tests. Thus, 16 of the 18 objects were used in the test. The labels and objects used during the test were randomly assigned.

Results

We asked whether difficulty would promote learners' long-term word learning in a cross-situational learning paradigm. If difficulty promoted word learning, we would expect to see lower performance immediately, but stronger performance long-term. However, if difficulty did not promote word learning, we would expect to see lower performance regardless of testing delay.

We first conducted a 3 (Learning Condition) x 3 (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, $F(2, 86) = 20.582$, $p < .001$, a significant main effect of testing delay, $F(2, 86) = 17.294$, $p < .001$, and a significant interaction of learning and testing delay, $F(4, 86) = 2.542$, $p = .045$.

First, three univariate ANOVAs were conducted within each testing condition. We then computed three planned comparisons using t-tests with Bonferroni corrections ($p < .05$) to determine the nature of the differences between learning conditions within each testing delay condition. If difficulty promoted word learning, we expected there to be differences in performance between learning conditions across the testing conditions.

When the immediate testing condition, there was a main effect of learning condition, $F(2, 32) = 10.997$, $p < .001$. Participants in the 2 x 2 condition ($M = 3.85$ correct mappings out of 4, $SD = .376$) had significantly higher performance than in the 4 x 4 condition ($M = 2.00$ correct mappings out of 4, $SD = 1.195$), $p < .001$. Performance was also marginally higher in the 2 x 2 condition than the 3 x 3 condition ($M = 3.07$ correct mappings out of 4, $SD = .997$),
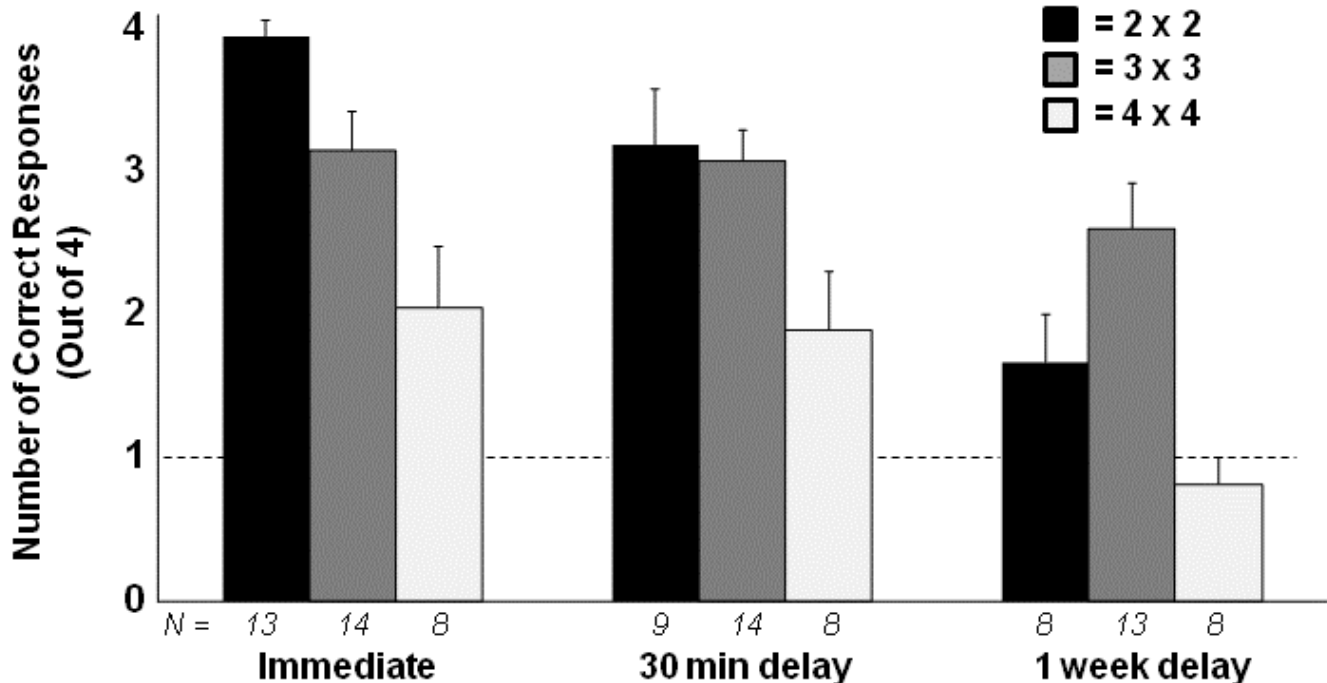

Figure 2. Average number of correct responses (out of 4) by learning condition (2 x 2, 3 x 3, 4 x 4) and testing condition (immediate, 30 minute delay, one week delay). The dashed line represents chance performance.

$p = .086$. Finally, performance in the 3 x 3 condition was significantly higher than the 4 x 4 condition, $p = .029$. Thus, greater the number of object-label pairings in each learning trial, the lower the performance.

However, there was a different pattern of results in the 30 minute delay condition. There was a main effect of learning condition, $F_{(2, 28)} = 5.304$, $p = .011$. Participants in the 2 x 2 condition ($M = 3.11$ correct mappings out of 4, $SD = 1.167$) had similar performance to participants in the 3 x 3 condition ($M = 3.00$ correct mappings out of 4, $SD = .784$), $p > .05$. Participants in the 2 x 2 and 3 x 3 conditions both had significantly higher performance than participants in the 4 x 4 condition ($M = 1.75$ correct mappings out of 4, $SD = 1.035$), $p = .022$ and $p = .021$.

In the one week testing delay condition there was a particularly interesting pattern of results. There was a main effect of learning condition, $F_{(2, 26)} = 11.286$, $p < .001$. Participants in the 3 x 3 condition ($M = 2.54$ correct mappings out of 4, $SD = 1.127$) had higher performance than both the 2 x 2 condition ($M = 1.62$ correct mappings out of 4, $SD = .518$), $p = .071$, and 4 x 4 condition ($M = .75$ correct mappings out of 4, $SD = .463$), $p < .001$. Participants in the 4 x 4 condition performed similarly to participants in the 2 x 2 condition, $p > .05$. Thus, although initially participants in the 3 x 3 condition had lower performance than the 2 x 2 condition, one week later participants in the 3 x 3 condition had higher performance than participants in the 2 x 2 condition.

In addition to examining the differences within each testing condition, we also examined differences in each learning condition across the testing conditions using ANOVAs and three planned comparisons with Bonferroni corrections ($p < .05$). In the 2 x 2 condition, there was a main effect of testing delay, $F_{(2, 27)} = 12.255$, $p < .001$. Immediate performance was marginally higher than performance in the 30 minute delay condition, $p = .085$, and the performance in the 30 minute delay condition was significantly higher than performance in the 1 week condition, $p = .001$. Thus, there was significant decrease in retention across each of the testing delay conditions.

There was also a main effect of testing delay in the 4 x 4 condition, $F_{(2, 21)} = 3.868$, $p = .037$. There was not a significant difference in performance between immediate and 30 minute delay conditions, $p > .05$, or the 30 minute delay and one week delay conditions, $p > .05$. However, there was a significant difference in performance between the immediate and one week delay condition, $p = .047$. Thus, there was a significant decrease in retention between the immediate test and one week delayed test. Finally, in the 3 x 3 condition, there was not a main effect of testing delay, $F_{(2, 38)} = 1.172$, $p > .05$. Thus, there was not a significant decrease in retention between the immediate and delayed tests.

Discussion

The results of this study support the idea that difficulty imposed during learning can promote long-term word learning (e.g., Vlach et al., 2008). Moreover, difficulty promoted word learning in the already difficult task of cross-situational word learning, in which learners must track mappings across events. We found that, when tested immediately, learners had the highest performance in the 2 x 2 condition and the lowest performance in the 4 x 4 condition. Performance in the 3 x 3 condition was somewhere in between. These findings replicate that of Yu & Smith (2007). However, when tested 30 minutes later, there were no differences in the performance between the 2 x 2 and 3 x 3 conditions. Finally, when tested a week later, performance in the 3 x 3 condition was higher than performance than in both the 2 x 2 and 4 x 4 conditions. Thus, although difficulty yielded lower immediate performance (i.e., the 2 x 2 condition had higher performance than the 3 x 3 condition), there was a benefit of difficulty for long-term performance (i.e., one week later the 3 x 3 condition had higher performance than the 2 x 2 condition). This study demonstrates that, even in the seemingly difficult task of mapping words to objects (Quine, 1960), adding difficulty promoted long-term word learning.

The findings from this study also have implications for research on cross-situational word learning and, more generally, statistical word learning. Recent research on statistical word learning has focused on the factors that promote immediate performance in order to discover the mechanisms by which words are acquired over time (e.g., Kachergis, Yu, & Shiffrin, 2009; Lany & Saffran, 2010). However, this study suggests that this may not be the best approach for describing long-term trajectories of word learning. This study clearly demonstrates that immediate performance does not always reflect long-term performance. Thus, in order to assert that a mechanism promotes word learning over time, evidence should be provided from not just an immediate test, but an immediate *and* delayed test.

The broader impact of this study is that it highlights the intimate relationship between word learning and memory. Learning new words and categories requires perceiving an object, attending to relevant features, mapping a label to the object, binding this mapping to other instances of the label and object, abstracting across instances, and, finally, generalizing to novel objects. Memory is a critical factor in this process, both during category formation (e.g., remembering relevant features and binding instances and labels) and recall (e.g., retrieving stored instances and categories).

Despite the clear relationship between word learning, memory, and retention, we have failed as word learning researchers and developmentalists to explore the mechanisms underlying this relationship. Fundamental questions have remained unexamined. For example, the few studies that have asked whether children retain words over time have provided conflicting evidence. While one study finds children retain words for a month (e.g., Markson & Bloom, 1997), other studies have found that children forget words in a matter of minutes (e.g., Horst & Samuelson,

2008; Vlach et al., 2008). What are the implications of our research if participants do not remember words after a few minutes? Why are we speculating about long-term word learning from immediate performance, rather than empirically investigating word learning over time? Are we really uncovering the mechanisms of word learning?

In sum, future research should investigate both in-the-moment and long-term word learning. Exploring in-the-moment word learning is essential for understanding how words and categories are initially encoded. However, in theories of word learning, the common assumption is that performance will remain constant over time. This study clearly demonstrates that this is not always the case.

In order to account for real-world word learning, research should incorporate testing over longer time-scales—over the course of weeks, months, and years. A complete theory of word learning not only accounts for word learning in the moment and on each time scale, but also integrates them in order to understand how they influence each other over time.

Acknowledgments

References

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalf and A. P. Shimura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567-572.

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

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA. MIT Press.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

Gagne, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired-associates. *Journal of Experimental Psychology, 40*, 61-73.

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.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*, 128-157.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society* . Austin, TX: Cognitive Science Society.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592.

Lany, J., & Saffran, J. R. (2010). From statistics to meaning. *Psychological Science*, published online January 2010.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91-115.

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

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 285*, 813-815.

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

Samuelson, L. R., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94-104.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor neuroscience (pp. 59-75). Dordrecht: Kluwer.

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, 333-338.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*, 341-358.

Tomasello, M., & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology, 30*, 639-650.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*, 163-167.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414-420.

Yurovsky, D. & Yu, C. (2008). Mutual Exclusivity in Cross-Situational Statistical Learning. In B. C. Love, K. McRae, & V.M. Sloutsky (Eds.). *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.