Spacing Meets Cross-Situational Word Learning: How the Temporal Structure of Labeling Events Affects Word Learning

Yi Tong (tong33@wisc.edu)
Department of Educational Psychology, 1025 W. Johnson Street
Madison, WI 53706 USA

Melina L. Knabe (melina.knabe@utexas.edu)
Department of Psychology, 108 E Dean Keeton Street
Austin, TX 78712 USA

Haley A. Vlach (hvlach@wisc.edu)
Department of Educational Psychology, 1025 W. Johnson Street
Madison, WI 53706 USA

Abstract

Limited work has considered how the temporal distribution of labeling events affects word learning in ambiguous contexts, such as the cross-situational word learning paradigm, over real-world timescales. The temporal distribution of learning events can impact how well information is retained: spacing out information promotes retention more than presenting information in close succession. In the current study, adults were presented with novel object-word pairings across six different temporal schedules over four consecutive training days. Word learning was assessed either immediately after the final training session ($N = 50$) or after a week ($N = 54$). Results revealed that adults successfully disambiguated word-object mappings across all learning schedules at both test times, except for the massed and most spaced schedules at the 1-week delay. These findings suggest that temporal distribution effects emerge across extended timescales, but there might be constraints on the amount of spacing that is optimal for word learning.

Keywords: cross-situational word learning; temporal structure; spacing effect; retrieval; retention

Introduction

Word learning often takes place in ambiguous contexts, where a word can theoretically refer to an infinite number of possible referents (Quine, 1960). Imagine walking through a grocery store in a foreign country and hearing someone say the word "wug" while pointing toward a shelf full of products. You might wonder whether "wug" refers to a specific item, the entire shelf, or something else completely. This type of referential ambiguity poses a critical challenge for learners: to successfully acquire the meaning of a word, they must first identify the correct referent and retain this mapping across time. While it may seem impossible to resolve such ambiguity in a single labeling event, a growing body of research has demonstrated that learners can overcome this challenge by tracking statistical regularities across multiple labeling events—a process known as cross-situational word learning (CSWL; see Roembke et al., 2023, for a review).

In CSWL studies, learners are typically presented with a series of trials in which two novel objects are shown concurrently and labeled in random order. With no explicit feedback, it is difficult to identify the correct word-object mapping in a single trial. However, as objects are paired with different foils across trials, learners can track which object consistently co-occurs with a specific label to disambiguate mappings over time. Studies have shown that both children (e.g., Smith & Yu, 2008; Suanda et al., 2014; Vlach & DeBrock, 2019) and adults (e.g., Poepsel & Weiss, 2016; Smith et al., 2011; Yu & Smith, 2007; Vlach & Johnson, 2013; Vlach & Sandhofer, 2014) can successfully acquire words through CSWL, mapping objects to their corresponding labels significantly above chance at test.

While much of the CSWL literature has focused on how these initial mappings are formed (e.g., Smith & Yu, 2008; Yu & Smith, 2007), less attention has been given to how these mappings are retained over time. In real-world settings, learners often need to retrieve word mappings days, weeks, or even years later. Examining retention in CSWL is particularly interesting because the ability to track, store, and integrate co-occurrence statistics across labeling events is crucial to resolving referential ambiguity (Vlach & Johnson, 2013; Vlach & DeBrock, 2017, 2019; Vlach, & Sandhofer, 2014). This reliance on domain-general memory processes suggests that CSWL may be subject to broader learning mechanisms, such as the effects of temporal distribution of learning events. That is, how labeling events are spaced over time may influence the learning and retention of object-word mappings.

Research on memory and learning has consistently shown that the temporal structure of learning events can affect retention. Distributing information across time (i.e., spacing) enhances long-term retention compared to presenting information in close succession (i.e., massing) (Janiszewski et al., 2003; Vlach et al., 2008). This spacing effect is arguably one of the most robust findings in cognitive science, which has been replicated across a wide range of tasks and contexts (see Cepeda et al., 2006, for a review). However, spaced learning is not always advantageous. When

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

performance is measured immediately after training, massed learning can sometimes lead to equivalent or even better outcomes than spaced learning. The benefits of spacing often become more apparent at delayed tests (Smith & Scarf, 2017; Vlach et al., 2012). Beyond a simple massed-versus-spaced distinction, learning schedules can also vary in the degree of spacing, with some studies examining intermediate schedules that include a combination of massed and spaced learning events (e.g., clumping; Gluckman et al., 2014; Vlach & Sandhofer, 2012).

Despite extensive research on spacing effects in other domains, relatively few studies have examined the effects of temporal structure in the context of CSWL (Benitez et al., 2020; Smith et al., 2011; Yurovsky & Frank, 2015). These studies primarily focused on the acquisition of word-object mappings across short timescales. That is, participants were presented with ambiguous word learning trials and tested immediately following the learning phase. Findings from these studies revealed that adults performed better when target objects were presented in consecutive trials (massing) compared to interleaving trials between repetitions. While these studies provide valuable initial insights into how temporal structure impacts word learning in ambiguous contexts, they leave open an important question: How does temporal structure influence the retention of word-object mappings in CSWL over time?

In line with the spacing literature, we hypothesized that, compared to massing, spaced learning would enhance the retention of word-object mappings at a delayed test. Prior research suggests that while massed learning may boost immediate performance, the benefits of spacing typically emerge when a delay is introduced between learning and test (e.g., Vlach et al., 2012). Alternatively, massed learning may be particularly helpful for resolving ambiguity in CSWL; minimizing time between learning events may aid in aggregating word-object associations across learning trials.

To address this question, the current study manipulated the temporal structure of labeling events within-subjects in a CSWL paradigm and assessed learning either immediately after training or following a 1-week delay. This design allowed for a direct comparison of the effects of temporal distribution across real-world timescales. Prior studies have manipulated temporal structure in various ways, such as comparing massed and interleaved schedules (Smith et al., 2011), adding an unstructured schedule (Benitez et al., 2020), and varying the number of interleaved trials between repetitions (Yurovsky & Frank, 2015). Notably, all of these manipulations occurred within a single training session. To better reflect real-world scenarios where learning unfolds over time, the current study implemented six distinct learning schedules across four consecutive training days.

Methods

Participants

A total of 104 adults were recruited through Amazon Mechanical Turk (MTurk) and Prolific. Participants were recruited separately for each of the two testing conditions: an immediate test (N = 50, M_{age} = 42.48 years, SD_{age} = 10.58 years) or a 1-week delayed test (N = 54, M_{age} = 34.28 years, SD_{age} = 7.73 years).

Design

To investigate the effects of the temporal distribution of labeling events on word learning, this study used a 6 × 2 mixed design with the item presentation schedule (Massed, Clumped 1, Clumped 2, Spaced 1, Spaced 2, Spaced 3) as a within-subjects variable and the timing of post-tests (immediate, one-week delay) as a between-subjects variable. The primary dependent variable was the proportion of objects correctly identified at test.

Material

The stimuli consisted of 48 novel objects and 48 English-sounding pseudowords, sourced from the Novel Objects and Unusual Names (NOUN) database (Horst & Hout, 2016). Each pseudoword was consistently paired with a corresponding novel object to create 24 target items and 24 distractors. During training, both target and distractor items co-occurred with their assigned pseudowords. However, only target items were assigned to specific learning schedules, whereas distractors were not assigned to any schedule and served solely as filler items to ensure a consistent number of trials across sessions. Because the amount and timing of distractor presentations were not experimentally manipulated, learning outcomes for these items were not included in the analysis. The number of distractors used during each session varied depending on the number of target words scheduled for that day. The pseudowords were recorded as single tokens by a female native English speaker. These recordings served as auditory cues during both the training and testing phases.

The 24 target items were distributed across six presentation schedules, with four items assigned to each schedule. Objects in the Massed schedule were presented six times within a single training session. Participants were randomly assigned to learn the massed objects either at the beginning (Time 1) or the end (Time 4) of the four-day training period. Objects in the clumped schedules (Clumped 1, Clumped 2) were presented six times across two consecutive training days, with three repetitions per session. Similarly, objects in the spaced schedules (Spaced 1, Spaced 2, Spaced 3) were presented six times across two training sessions, but these sessions were separated by longer intervals to manipulate the amount of temporal spacing. See Figure 1 for a visual depiction of these schedules.

Each training session included 36 trials, each with two objects displayed side-by-side. Trials could include two target items, one target and one distractor, or two distractors. The naming of the two objects on each trial was randomized and did not always follow a left-to-right order. Within each session, target items scheduled for presentation that day were presented in a randomized order, regardless of condition. That is, while the temporal spacing manipulation was implemented across four days of training, the exact spacing of target presentations within each session was randomized and did not follow a specific spaced schedule. Trials were presented in the form of pre-recorded videos, and each trial lasted approximately five seconds before automatically proceeding to the next. The orders of learning trials for Time 2 and Time 3 were fixed across all participants, while Time 1 and Time 4 had two possible trial orders depending on whether the massed items were shown. No pair of objects appeared together more than once across the four training days.

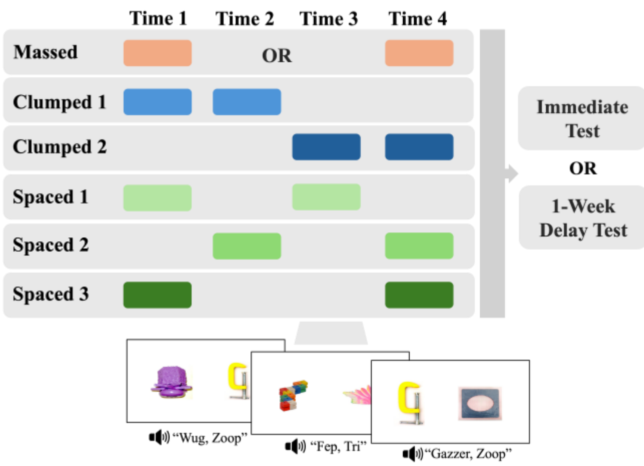

Figure 1: Experimental procedure.

The post-test consisted of 33 trials, each presenting a correct object and a foil side-by-side on the screen. The location of the correct object (left or right) was randomized across trials. A total of 33 objects, 24 target items and 9 distractors, were included in the test. Each object appeared twice across all testing trials, once as the correct object and once as a foil. Object pairings for each test trial were selected using a predetermined rule: they were either objects that had co-occurred during training or objects introduced during the same training session but never directly paired. No pair of objects appeared together more than once during the test. Because distractor items were included only as fillers in training, their accuracy was not analyzed. Two versions of the test were created depending on whether participants had viewed the massed objects at Time 1 or Time 4.

Procedure

Participants were recruited online via MTurk and Prolific, and all training and testing sessions were conducted through Qualtrics. All participants provided informed consent and were compensated for their participation. At the start of the study (Time 1), participants were informed that they would participate in a four-day longitudinal study in which they would watch short videos to learn new words.

Each training session included one pre-recorded video lasting approximately two minutes. Before watching the video, participants completed a sound check by listening to a short audio clip and indicating what sound they heard. Participants were then instructed that during the video, they would see pictures of objects appear on the screen while hearing their corresponding labels and that the objects would not necessarily be labeled from left to right. Participants viewed the videos passively, as no responses were required during training. After each video, participants completed an attention check question. Only participants who passed both the sound check and the attention check were invited to continue with subsequent training sessions. Participants who failed either check were permanently excluded from the study.

Participants assigned to the immediate test condition completed all testing trials immediately following the final training video. During each test trial, participants viewed a pre-recorded 5-second clip showing two objects displayed side by side, accompanied by an auditory cue (e.g., "Which one is ___?"). They were asked to select the object that corresponded to the label. Participants assigned to the delayed test condition completed the same testing trials one week after the last training session.

Results

Learning Performance at Immediate Test

We first conducted one-sample t-tests to examine whether participants' performance was above chance immediately after the last training session (Figure 2). One-sample t-tests indicated that participants correctly identified target objects above chance for all learning schedules: Massed, $t(49) = 4.72$, $p < .001$; Clumped 1, $t(49) = 3.02$, $p = .004$; Clumped 2, $t(49) = 7.51$, $p < .001$; Spaced 1, $t(49) = 4.54$, $p < .001$; Spaced 2, $t(49) = 4.38$, $p < .001$; and Spaced 3, $t(49) = 2.51$, $p = .016$. Further analyses revealed that participants learned massed objects above chance regardless of whether they were presented at Time 1, $t(26) = 3.55$, $p = .002$, or Time 4, $t(22) = 3.05$, $p = .006$. A two-sample t-test confirmed no significant difference between these two groups for massed objects, $t(46.80) = 0.18$, $p = .861$.

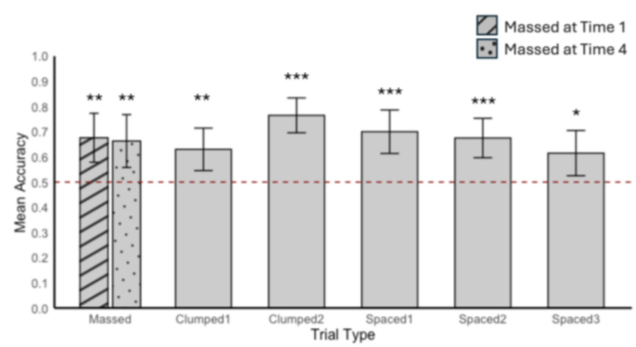

Figure 2: Proportion of accurate responses for each learning schedule at the immediate test. The dashed red line indicates chance-level performance (0.5). Asterisks denote schedules where performance was significantly above chance.
*** $p < .001$, ** $p < .01$, * $p < .05$

To assess the overall effect of learning schedules on performance at the immediate test, we fitted a linear mixed model predicting accuracy with schedule as a fixed effect (dummy coded with Massed as the reference group) and a random intercept by participants to account for individual variability. A likelihood ratio test showed that including schedule significantly improved the model fit compared to a reduced model with random intercepts only, $\chi^2(5) = 12.194$, $p = .032$. We then conducted Bonferroni-corrected pairwise comparisons of all levels of schedule. Results revealed that participants were significantly more accurate for objects in the Clumped 2 schedule ($M = .765$, $SD = .250$) compared to the Spaced 3 schedule ($M = .615$, $SD = .324$), $b = 0.15$, $SE = 0.048$, $t = 3.10$, $p = .033$. No other comparisons were significant (see Table 1).

Table 1: Comparisons of accuracy across schedules at immediate test.

Comparison	Estimate	SE	t-value	p-value
M vs. C 1	0.04	0.048	0.826	1
M vs. C 2	-0.095	0.048	-1.961	0.765
M vs. S 1	-0.03	0.048	-0.619	1
M vs. S 2	-0.005	0.048	-0.103	1
M vs. S 3	0.055	0.048	1.135	1
C 1 vs. C 2	-0.135	0.048	-2.787	0.086
C 1 vs. S 1	-0.07	0.048	-1.445	1
C 1 vs. S 2	-0.045	0.048	-0.929	1
C 1 vs. S 3	0.015	0.048	0.31	1
C 2 vs. S 1	0.065	0.048	1.342	1
C 2 vs. S 2	0.09	0.048	1.858	0.965
C 2 vs. S 3	0.15	0.048	3.096	0.033*
S 1 vs. S 2	0.025	0.048	0.516	1
S 1 vs. S 3	0.085	0.048	1.755	1
S 2 vs. S 3	0.06	0.048	1.239	1

Learning Performance at Delayed Test

Next, we examined participants' performance one week after the final training session (Figure 3). One-sample t-tests revealed above-chance performance for objects presented in Clumped 1, $t(53) = 2.712$, $p = .009$, Clumped 2, $t(53) = 4.864$, $p < .001$, Spaced 1, $t(53) = 4.198$, $p < .001$, and Spaced 2 schedules, $t(53) = 5.164$, $p < .001$. Performance for Spaced 3, $t(53) = 0.910$, $p = .367$, and Massed objects, $t(53) = 0.785$, $p = .436$, did not differ from chance overall. However, for Massed objects, participants who learned them at Time 1 performed significantly above chance, $t(26) = 2.267$, $p = .032$, while those who learned them at Time 4 did not, $t(26) = -1.192$, $p = .244$. A two-sample t-test confirmed that Massed accuracy was significantly higher when learned at Time 1 compared to Time 4, $t(51.87) = 2.464$, $p = .017$.

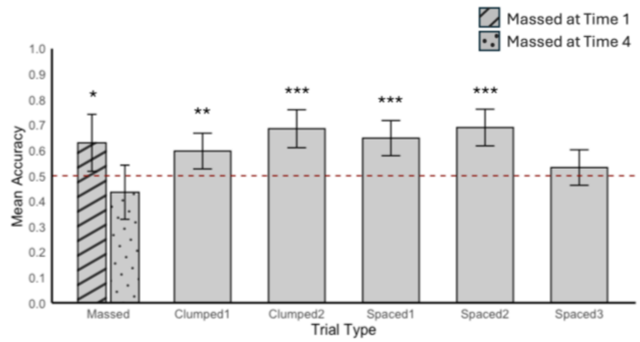

Figure 3: Proportion of accurate responses for each learning schedule at the 1-week delayed test. The dashed red line indicates chance-level performance (0.5). Asterisks denote schedules where performance was significantly above chance. *** p <.001, ** p <.01, *p < .05

Additionally, we fitted another linear mixed model predicting delayed performance with schedule (dummy coded with Massed as the reference group) as the fixed effect and a random intercept by participants. A likelihood ratio test showed that the inclusion of schedule significantly improved model fit compared to a reduced model, $\chi^2(5) = 25.523$, $p < .001$. Bonferroni-corrected pairwise comparisons revealed that accuracy on the delayed test for Massed objects ($M = 0.532$, $SD = 0.303$) was significantly lower than Clumped 2 ($M = 0.685$, $SD = 0.280$), $b = -0.152$, $SE = 0.044$, $t = -3.462$, $p = .009$, and Spaced 2 ($M = 0.690$, $SD = 0.270$), $b = -0.157$, $SE = 0.044$, $t = -3.567$, $p = .006$. Additionally, Spaced 3 accuracy ($M = 0.532$, $SD = 0.262$) was significantly lower than Clumped 2, $b = 0.153$, $SE = 0.044$, $t = 3.462$, $p = .009$, and Spaced 2, $b = 0.157$, $SE = 0.044$, $t = 3.567$, $p = .006$. Other comparisons were not significant (Table 2).

Comparison of Immediate and Delayed Performance

To assess how different learning schedules influenced retention over time, we examined differences in performance between the immediate and delayed tests. Specifically, we fitted another linear mixed model with test (dummy coded with the immediate test as the reference group), schedule (dummy coded with Massed as the reference group), and their interaction as fixed effects, with a random intercept for participants. Results revealed a significant interaction between test and the Spaced 2 schedule, indicating that the decrease in performance from the immediate to the delayed test was smaller for Spaced 2 compared to Massed items, $b = 0.152$, $SE = 0.065$, $t = 2.331$, $p = .020$. Interaction terms comparing the change in accuracy from immediate to delayed test for other schedules (Clumped 1, Clumped 2, Spaced 1, Spaced 3) relative to Massed were not significant.

Table 2: Comparisons of accuracy across schedules at delayed test.

Comparison	Estimate	SE	t-value	p-value
M vs. C 1	-0.065	0.044	-1.469	1
M vs. C 2	-0.153	0.044	-3.462	0.009**
M vs. S 1	-0.116	0.044	-2.623	0.138
M vs. S 2	-0.157	0.044	-3.567	0.006**
M vs. S 3	0	0.044	0	1
C 1 vs. C 2	-0.088	0.044	-1.993	0.708
C 1 vs. S 1	-0.051	0.044	-1.154	1
C 1 vs. S 2	-0.093	0.044	-2.098	0.552
C 1 vs. S 3	0.065	0.044	1.469	1
C 2 vs. S 1	0.037	0.044	0.839	1
C 2 vs. S 2	-0.005	0.044	-0.105	1
C 2 vs. S 3	0.153	0.044	3.462	0.009**
S 1 vs. S 2	-0.042	0.044	-0.944	1
S 1 vs. S 3	0.116	0.044	2.623	0.138
S 2 vs. S 3	0.157	0.044	3.567	0.006**

Follow-up Bonferroni-corrected pairwise comparisons between immediate and delayed performance for each schedule revealed a significant decline in accuracy for Massed items at the delayed test compared to the immediate test, $b = 0.138$, $SE = 0.055$, $t = 2.495$, $p = .013$. None of the other schedules showed significant differences in performance across tests (Table 3).

Table 3: Immediate vs. delayed test comparisons by schedule.

Schedule	Estimate	SE	t-ratio	p-value
M	0.138	0.055	2.495	0.013*
C 1	0.033	0.055	0.594	0.553
C 2	0.08	0.055	1.447	0.149
S 1	0.052	0.055	0.94	0.348
S 2	-0.015	0.055	-0.269	0.788
S 3	0.083	0.055	1.497	0.135

Discussion

Much of the CSWL literature has focused on how learners acquire word-object mappings in ambiguous contexts, yet limited research has explored how these mappings are retained over time. Retention is critical for understanding how CSWL supports real-world learning, where associations must be stored across extended timescales to resolve ambiguity and learn word-object mappings. The current study manipulated the temporal structure of labeling events in CSWL over four consecutive training days to examine how different learning schedules affected short-term retrieval and long-term retention. Despite task difficulty, adults successfully disambiguated 24 word-object pairings across learning schedules at the immediate test. These findings contribute to the existing body of evidence highlighting the robustness of CSWL as a mechanism for resolving referential ambiguity, even under demanding conditions (e.g., Smith et al., 2010; Vlach & Sandhofer, 2014).

The effects of learning schedules were minimal at the immediate test, with the only significant difference observed between Clumped 2 and Spaced 3 schedules. However, these effects became more pronounced at the delayed test, showing an overall inverted U-shaped pattern in retention performance. That is, mappings were retained after a week, but memory performance was significantly poorer for items presented in the Massed schedule at Time 4 and the most spaced schedule (Spaced 3). Schedules with intermediate time intervals between labeling events resulted in better long-term retention. This inverted U-shaped trend is consistent with previous research on spacing and long-term retention (Cepeda et al., 2009; Rovee-Collier et al., 1995).

Why would the effects of how labeling events are temporally structured emerge over time delays? The desirable difficulty account of long-term memory (e.g., Bjork & Bjork, 2011; Bjork, 1994; Kornell et al., 2009; Vlach et al., 2012; Vlach & Sandhofer, 2014) provides a useful framework for interpreting these results. This theory posits that learning conditions involving more effortful—but successful—retrieval enhance long-term retention. In the context of CSWL, learners must remember previously seen word-object pairs across trials to resolve referential ambiguity. In the Massed schedule, where the same object (e.g., a red round object) consistently co-occurs with the word "wug" in close succession, retrieval requires minimal effort as memories of previous trials are still recent. While this facilitates rapid mapping, it may lead to weaker encoding and poorer retention over time. Consistent with this theory, our results showed that Massed items showed more forgetting after a week. On the other hand, the excessive spacing in Spaced 3 may also hinder retention because the intervals between repetitions exceeded learners' ability to reliably retrieve the association between the red round object and the word "wug". Although retrieval was still possible immediately after the last training day, the weak memory trace of this association might be susceptible to rapid forgetting. In contrast, learning schedules with intermediate intervals between repetitions (e.g., Clumped 2, Spaced 1) appear to optimize retrieval difficulty: they introduce enough delay to challenge learners while ensuring retrieval success, resulting in more robust long-term retention.

Interestingly, while the Massed schedule as a group followed the overall inverted U-shaped retention curve, memory performance differed significantly depending on when these items were shown. Results showed that participants who learned Massed items at Time 1 retained these mappings better than those who learned at Time 4. This finding is counterintuitive as a longer delay would typically result in more forgetting. Why might earlier learning lead to better long-term retention? One possibility is that items learned earlier in the training sequence might face less interference from subsequently introduced associations (e.g., Postman & Underwood, 1973). Alternatively, the longitudinal design of this study may have contributed to a decline in participants' engagement or attention over multiple training days, resulting in weaker encoding of items shown at Time 4. Future studies should explore this pattern further to better understand the interaction between learning order and temporal structure in CSWL. Moreover, using a purer Massed schedule, in which items are always tested directly after learning, may yield different results.

Nevertheless, the current study provides initial insights into the impact of temporal structure on CSWL, highlighting the importance of considering retrieval dynamics in existing theories. Two major accounts have been proposed to explain the underlying mechanisms of CSWL. The associative account (e.g., Kachergis et al., 2012; Yu & Smith, 2007) argues that learners establish associations between each novel word and its potential referents. Across trials, the correct mapping becomes more robust through repeated reinforcement. In contrast, the hypothesis-testing account (e.g., Trueswell et al., 2013) posits that learners only form and test one hypothesis at a time, revising it based on subsequent evidence. More recently, some have suggested that both mechanisms might operate simultaneously during CSWL (e.g., McMurray et al., 2012; Roembke et al., 2023; Yurovsky & Frank, 2015). Importantly, all of these accounts assume successful retrieval of previous associations. To fully account for CSWL across extended timescales, theories must consider the temporal distribution of learning events and how forgetting and retrieval processes shape retention of word-object associations across time. Future research should examine the role of retrieval dynamics in CSWL by manipulating the timing of the delayed tests or the levels of retrieval difficulty during training.

In sum, this study was the first to examine how different presentation schedules of labeling events in ambiguous word learning contexts influence short-term retrieval and long-term retention. By manipulating temporal structure, this work extends the currently limited body of studies examining CSWL over time (e.g., Walker et al., 2020; Vlach & Sandhofer, 2014). The results showed that the effects of temporal distribution on word learning become more pronounced over time, but there are constraints on the optimal amount of spacing, as both minimal and excessive spacing may hinder long-term retention.

References

Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of experimental child psychology*, *200*, 104961. https://doi.org/10.1016/j.jecp.2020.104961

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). The MIT Press.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: theoretical analysis and practical implications. *Experimental psychology*, *56*(4), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

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*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology, 28*(2), 266–273. https://doi.org/10.1002/acp.2997

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of consumer research*, *30*(1), 138-149. https://doi.org/10.1086/374692

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. https://doi.org/10.3758/s13423-011-0194-6

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of experimental psychology. Learning, memory, and cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, *152*, 9–19. https://doi.org/10.1016/j.cognition.2016.03.001

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & cognition*, *1*(1), 19–40. https://doi.org/10.3758/BF03198064

Quine, W. V. O. (1960). *Word and Object: An Inquiry Into the Linguistic Mechanisms of Objective Reference.* Cambridge, MA, USA: The MIT Press.

Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. *Frontiers in psychology*, *14*, 1175272. https://doi.org/10.3389/fpsyg.2023.1175272

Rovee-Collier, C., Evancio, S., & Earley, L. A. (1995). The time window hypothesis: Spacing effects. *Infant Behavior & Development, 18*(1), 69–78. https://doi.org/10.1016/0163-6383(95)90008-X

Smith, C. D., & Scarf, D. (2017). Spacing Repetitions Over Long Timescales: A Review and a Reconsolidation Explanation. *Frontiers in psychology*, *8*, 962. https://doi.org/10.3389/fpsyg.2017.00962

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science, 35*(3), 480–498. https://doi.org/10.1111/j.1551-6709.2010.01158.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of experimental child psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

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*(1), 246–254. https://doi.org/10.1037/a0025260

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of memory and language*, *93*, 217–230. https://doi.org/10.1016/j.jml.2016.10.001

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 700–711. https://doi.org/10.1037/xlm0000611

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382. https://doi.org/10.1016/j.cognition.2013.02.015

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*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cognitive science*, *38*(4), 757–774. https://doi.org/10.1111/cogs.12092

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, *70*(S2), 221-254. https://doi.org/10.1111/lang.12395

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013