Item noise in the Sternberg paradigm

Allison Chapman (allison.chapman.m@gmail.com)
225 Psychology Building, 1835 Neil Avenue
Columbus, OH 43210 USA

Simon Dennis (simon.dennis@gmail.com)
225 Psychology Building, 1835 Neil Avenue
Columbus, OH 43210 USA

Abstract

The list length effect in recognition memory is the finding that performance improves as the number of items studied decreases. Models have attempted to operationalize interference in recognition memory as noise that accumulates over items in the study list, other contexts in which studied items have appeared, or a combination of both sources. Item noise models predict a list length effect. The list length effect has been eliminated in long-term recognition tasks. We demonstrate that the length effect may also be eliminated in short-term recognition (the Sternberg paradigm) if a filled delay is introduced. List length and recency effects were eliminated following an engaging 15 second distractor task. Articulatory suppression invoked the same results at a 2 second delay.

Keywords: Recognition memory; Sternberg paradigm; list length; recency; rehearsal; Bind Cue Decide Model of Episodic Memory

Introduction

Global-matching models of recognition memory propose that a test probe is compared to every item on the study list, and familiarity is summed over all items. As such, each new item introduces additional noise. Item-noise accounts predict a list length effect in which recognition performance decreases as a function of the number of items studied. Context-noise models suggest that the list length effect in a long-term setting derives from confounding factors that are not inherent to the studied items themselves. Null length effects have been found by introducing a delay between study and test. The delay encourages reinstatement of the study list context as a unit rather than referencing the end-of-list context to assess familiarity. The delay also equates the overall time before test to eliminate the retention interval confound (Dennis & Humphreys, 2001). Because shorter lists may be rehearsed more completely during any delay, the delay must be filled with a distractor task (Dennis Lee & Kinnell, 2008).

If set size is the driving source of interference, the latency to respond to items in a short-term recognition task should increase with list length. Saul Sternberg (1966) established a paradigm for observing list length effects in short-term memory. Study lists were comprised of 1 to 6 digits presented for 1.2 seconds (s) each. There was a 2 s, unfilled delay between study and test. At test, one number was presented and participants pulled a lever to indicate that the number had been seen in the preceding list. Sternberg found that latency for both distractors and targets increases linearly with set size, and concluded that retrieval occurs via an exhaustive serial scan. Exhaustive serial scanning, however, cannot account for null list length effects (Townsend & Fifić, 2003).

Recency has been purported to induce list length effects in a short-term paradigm (Monsell, 1978; McElree & Dosher, 1989). The recency hypothesis is concurrent with the Sternberg paradigm only when the delay between study and test is relatively short. Monsell, for example, tested lengths 1 through 4 after a 0.1 s delay (1978). When the delay is greater than 2 seconds, the effect of recency is diminished (Forrin & Cunningham, 1973; Jensen & Lisman, 1998); yet, Sternberg (1966) found list length effects after a 2 s delay. Moreover, Sternberg (1969) identified primacy but not recency following 2.5 s of rehearsal. Participants exhibited a range of serial position results from flat functions to strong primacy effects. Sternberg hypothesized that individual differences may reflect starting strategies in scanning. In particular, rehearsal may induce a circular scan that begins with the last item rehearsed (Sternberg, 1969).

The original Sternberg paradigm included a serial recall task following each study-test block (Sternberg, 1966). Duncan & Murdock (2000) found an interaction between intentionality and serial position for a mixed recognition-recall paradigm with a 1 s study-test delay. Each study list was randomly assigned to either a recall or recognition test. When participants were pre-cued, there was a clear recency effect in the recognition task trials. However, when post-cued, the recognition serial position function was flat. A questionnaire confirmed that participants prepared for recall in the absence of cue. The results suggest that the recall portion of the Sternberg paradigm may have confounded the recency results by inducing task-specific encoding or rehearsal strategies.

The current research aims to disambiguate factors underlying list length findings in the Sternberg paradigm and to determine if, as in long-term paradigms, it is possible to eliminate the length effect. Such a result would undermine item-noise accounts and would suggest that short- and long-term memory are not subsumed by fundamentally different systems.

Experiment 1

The main objectives of Experiment 1 were to both replicate and eliminate the list length effect in the Sternberg paradigm (1966). The study-test delay was increased from 2 s to 15 s to encourage contextual reinstatement and reduce recency effects. To counter the confounding effects of rehearsal, the 15 s delay was filled with a distractor task and serial recall

was not tested. Studied stimuli were words rather than digits to reduce proactive interference from guessing strategies for a closed stimuli set.

Method

Participants Participants were 76 undergraduate students enrolled in an introductory psychology course at The Ohio State University. Participants selected this experiment to fulfill a partial-credit course requirement and all provided informed consent. The 2 s delay condition included 19 women and 20 men averaging 19.63 years in age. The 15 s delay condition included 20 women and 17 men with a mean age of 19.39 years.

Materials Stimuli were lists of length 1 through 6, composed of 5-, 6-, and 7-letter low-frequency (1-4 Google counts per million) and high-frequency (100-200 counts) words. Each of 6 stimuli groups was composed of 68 words. An additional 26 words drawn from mid-frequency ranges (42-62 counts) were presented during the practice portion of the experiment. Distractor words were drawn from the same stimuli pool as test words, but did not appear in any iterations of the experiment. Math questions were randomly generated addition/subtraction problems using a combination of two digits. Half the problems were incorrect by one numeric interval. An example of a math question would be: 9 - 7 = 3 (Q = Yes / P = No). The same stimuli were used between-subjects across conditions.

Design

The within-subjects factors of Experiment 1 were the length of the study list and the serial position of the test item. The between-subjects factor was the study-test delay. The dependent measure was response time (RT). There were 6 levels of length and 2 levels of delay (2 s unfilled and 15 s filled).

Procedure Each serial position of each list was probed an equal number of times for a total of 21 position by length combinations. Each iteration of Experiment 1 included 21 trials with target probes and also 21 distractor trials. The experiment was divided into two sessions, each containing two-iterations. In sum there were 4 observations of each length by position combination per participant. Sessions ranged from 2-7 days apart. The same stimuli were used in both sessions, but the order of individual stimuli, list lengths, and probe positions were randomized between days. Words were combined and shuffled within frequency-type for each iteration.

Participants were seated at individual desks separated by partitions. Up to 8 participants were tested at once. Participants were instructed to remember a list of study words for a test where they must indicate whether a word was studied in the prior list or was completely novel (Q = Old / P = New). The 15 s math task appeared in both conditions (either between or after each study-test cycle) to equate overall testing time. Participants were informed of when the math test would appear, and were instructed to respond both accurately and quickly on all tasks. Response times were logged utilizing

the Python-Experiment Programming Library (PyEPL).

Words were presented to the center of the screen for 1.2 s each. At the end of study, there was a 2 s delay followed by 0.5 s of fixation preceding the test item. Participants were allotted 2.7 s to respond. When a response was registered, the math portion began. Each of the 8 problems was presented at fixed intervals of 1.875 seconds for a total filler period of 15 s. A 1 s delay and 0.5 s fixation interval signaled the start of the next study block. In the delayed condition, the first of 8 math problems appeared on the screen immediately following study. Then there was a 2 s delay followed by a 0.5 s fixation period and lastly the test probe. Then followed 1 s delay and 0.5 s fixation interval before the next cycle.

Results

The list length results for Experiment 1 are presented in Figure 1[1]. Participants with $d' < 0.9$ were eliminated. An ANOVA revealed a non-significant interaction of length by delay, $F(5, 370) = 0.928$, $p = 0.462$. A linear regression was run on each condition. The regression assessed the average of the median RTs across all positions for each length per participant. In the 2 s delayed condition, the slope was statistically significantly greater than zero: $t = 4.458$, $p < 0.001$. The trend in the 15 s delay condition was also statistically significant: $t = 2.852$, $p < 0.01$.
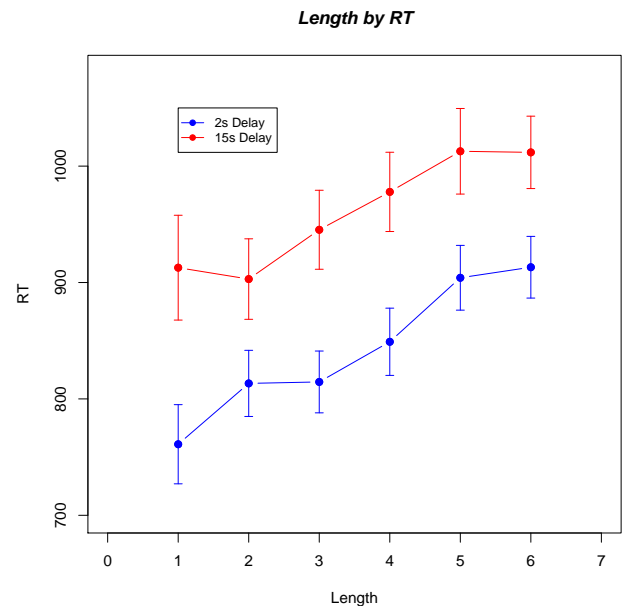

Figure 1: The averaged median latencies to correctly respond "yes" in Experiment 1 as a function of list length

A linear regression analysis was conducted on the median RTs across positions for each given length to examine the recency effects. The results are as follows for each position in

---

[1]Error bars depict standard errors.

the 2 s delay condition: $t(2) = 0.516$, $p = 0.608$; $t(3) = 0.132$, $p = 0.896$; $t(4) = 1.327$, $p = 0.187$; $t(5) = 0.176$, $p = 0.86$; $t(6) = -0.479$, $p = 0.633$. The results for the 15 s delay condition are: $t(2) = 2.251$, $p < 0.05$; $t(3) = 0.001$, $p = 1.00$; $t(4) = 0.755$, $p = 0.451$; $t(5) = 1.612$, $p = 0.109$; $t(6) = 0.929$, $p = 0.354$. The overwhelming pattern indicates that recency is not evident in either data set. Recency results for the 2 second unfilled delay and the 15 second delay with distractor are presented in Figures 2 and 3[1], respectively.
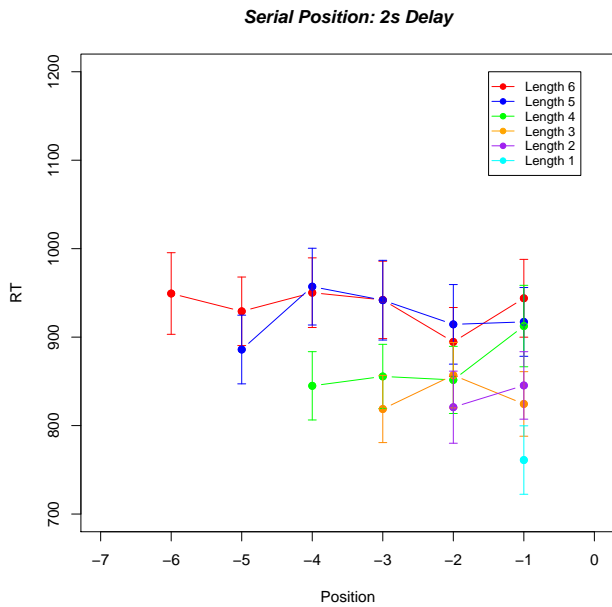

Figure 2: The averaged median latencies to correctly respond "yes" in Experiment 1 as a function of position for each list length in the 2 s delay condition.

Discussion

The Sternberg findings were successfully replicated; however, instantiating a 15 s delay is not sufficient to induce a null length effect. Interestingly, however, recency effects were not found in the 2 s delay condition despite increases in latency as a function of length. In this instance, recency alone does not capture the short-term list length effect.

Experiment 2

Experiment 2 was designed to determine if the 15 s distractor task was sufficiently engaging to reduce rehearsal. The distractor task was modified to be self-paced with feedback.

Method

Participants   Participants were 36 undergraduate students enrolled in an introductory psychology course at The Ohio State University. Participants selected this experiment to fulfill a partial-credit course requirement and all provided informed consent. The 2 s delay condition included 11 women
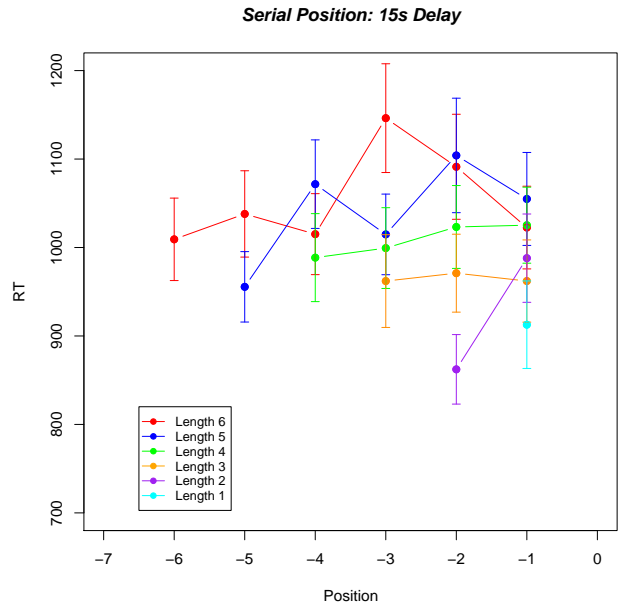

Figure 3: The averaged median latencies to correctly respond "yes" in Experiment 1 as a function of position for each list length in the 15 s delay condition.

and 7 men averaging 21.44 years in age. There were 6 women and 12 men with a mean age of 20.11 years in the $\sim$15 s delay condition.

Materials   Materials were lists of lengths 2, 4, and 6 composed of 5-, 6-, and 7-letter low-frequency (1-4 Google counts per million) and high-frequency (100-200 counts) words. There were a total of 190 words (including 22, mid-frequency practice words).

Procedure   The stimuli were randomized and repeated at one-week intervals. List lengths were reduced from 6 to 3 but observations per position were increased from 4 to 6 to reduce error in the recency analyses. There were 12 position by length list combinations. Each of the 12 types was tested in pure high-frequency and pure low-frequency cycles such that one iteration contained 24 target trials and 24 distractor trials.

The experiment was completed in 3 (single iteration) sessions for a total of 6 observations per positive probe position. The study-test lag was filled with math problems identical in form to the previous experiment. However, the task was self-paced and include feedback (Correct / Wrong). 16 math questions were presented for up to 1.875 seconds each. The total possible time was 30 seconds; but, on average, the delay was approximately 15 seconds.

Results

The list length results for Experiment 2 are presented in Figure 4[1]. Participants with $d' < 0.9$ were eliminated.

An ANOVA revealed a statistically significant interaction of length by delay, $F(2, 68) = 5.2076$, $p < .01$. A linear regression analysis of length was run on the average of the median RTs across all positions for each length per participant. In the 2 s delayed condition, the slope was statistically significantly greater than zero: $t = 3.272$, $p < 0.01$. Critically, the slope in the ~15 s delay condition was not statistically significantly different from zero: $t = 0.63$, $p = 0.531$.
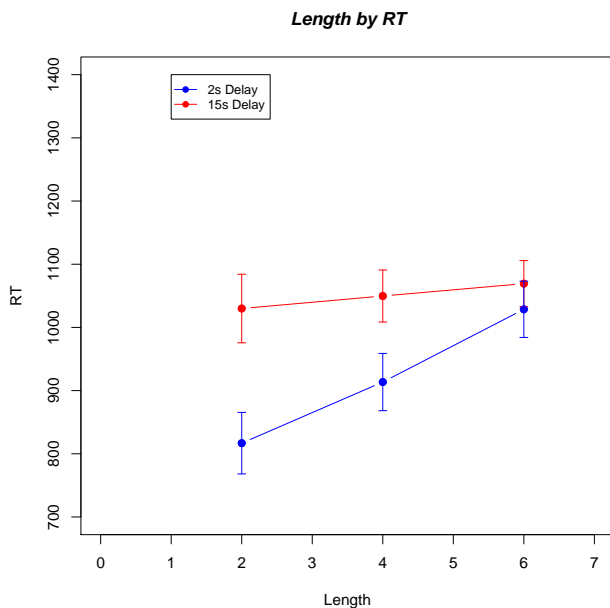

Figure 4: The averaged median latencies to correctly respond "yes" in Experiment 2 as a function of list length.

A linear regression analysis was conducted to examine the recency effects. RTs were combined such that the medians for positions 1-2, 3-4, and 5-6 were averaged for each participant. The results for combined positions across each length in the 2 s delay condition are as follows: $t(4) = 0.163$, $p = 0.871$; $t(6) = 0.549$, $p = 0.586$. The results for the ~15 s delay condition are: $t(4) = 0.305$, $p = 0.762$; $t(6) = 0.248$, $p = 0.805$. Serial position graphs for the 2 s unfilled delay and the ~15 s filled delay are presented in Figures 5 and 6[1].

Discussion

It appears that a null list length effect is possible in a Sternberg paradigm when the design includes both a longer study-test lag and a sufficiently engaging distractor task. As with Experiment 1, recency effects do not appear to underlie the list length effects.

Experiment 3

A potential caveat of Experiment 2 was that increasing the study-test delay actually shifted the paradigm from short-term memory to a long-term task. The purpose of Experiment 3 is to examine the extent to which null length effects hold at a
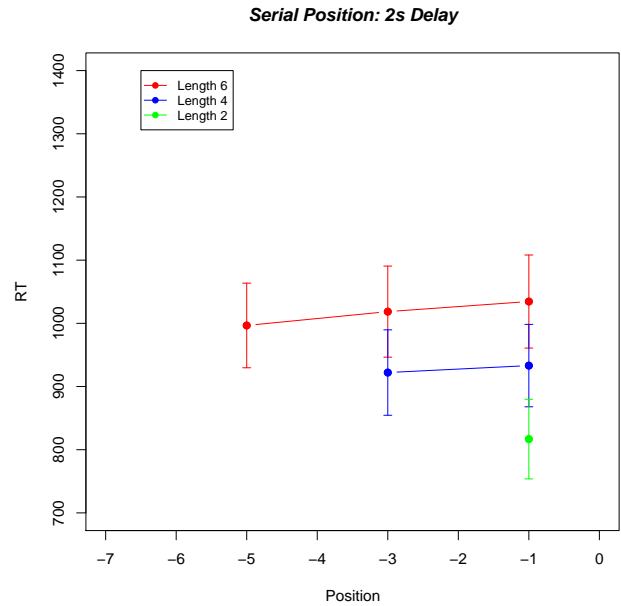

Figure 5: The averaged median latencies to correctly respond "yes" in Experiment 2 as a function of position for each list length in the 2 s delay condition.

2 s delay consisting of articulatory suppression. Conclusions will be more readily applicable to short-term memory.

Method

Participants Participants were 44 undergraduate students enrolled in an introductory psychology course at The Ohio State University. Participants selected this experiment to fulfill a partial-credit course requirement and all provided informed consent. The 2 s delay included 11 women and 11 men averaging 18.59 years in age. Participants in the 2 s delay with suppression were 10 women and 12 men with a mean age of 19.27 years.

Materials Word stimuli were identical to those used in Experiment 2. The distractor task was modified such that a single number between 10,000 and 90,000 appeared on the screen for 2 s. The five digits were random integers presented without commas.

Procedure The experiment was completed in one session that contained 3 iterations. Each iteration consisted of 24 target trials and 24 distractor trials. There were 6 observations of each probe position per participant. The study-test delay was 2 s in both conditions. The delay in the manipulation condition was filled with an articulatory suppression task to prevent subvocal rehearsal.

In addition to the standard instruction (to remember a list of study words and respond yes/no to a single test word), participants were informed that digits would appear and they should quickly and accurately say the digits aloud as a number. Par-
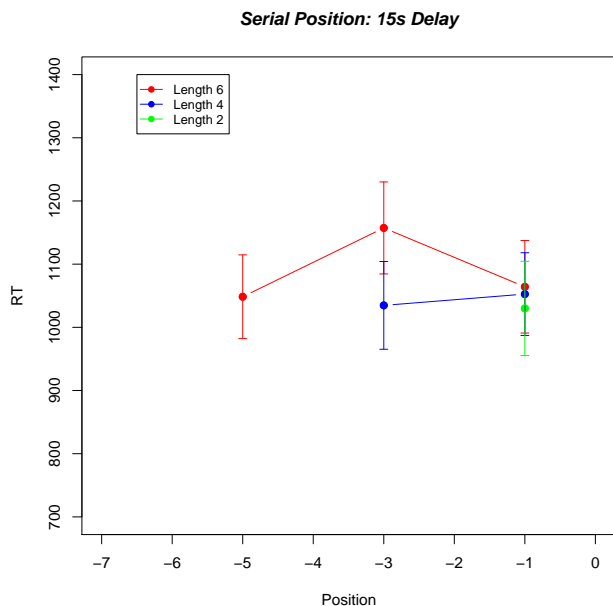
Figure 6: The averaged median latencies to correctly respond "yes" in Experiment 2 as a function of position for each list length in the ∼15 s delay condition.

Figure 7: The averaged median latencies to correctly respond "yes" in Experiment 3 as a function of list length.

ticipants in the suppression condition were tested individually. The experimenter noted any abnormal responses.

Results

The list length results for Experiment 3 are presented in Figure 7[1]. Participants with $d' < 0.9$ were eliminated. An ANOVA revealed a statistically significant interaction of length by delay, $F(2, 84) = 8.0441$, $p < 0.001$. A linear regression assessed the average of the median RTs across combined positions (1-2, 3-4, and 5-6) for each length per participant. In the 2 s delayed condition, the slope was statistically significantly greater than zero: $t = 2.656$, $p < 0.001$. In the 2 s articulatory suppression condition, the slope was not statistically significantly different from zero: $t = 0.457$, $p = 0.649$.

Recency trends were conducted on the average of median RTs for combined positions for each given length. The results are as follows for each position in the 2 s delay condition: $t(4) = -1.267$, $p = 0.212$; $t(6) = -0.726$, $p = 0.471$. The results for the suppression condition are: $t(4) = -0.774$, $p = 0.443$; $t(6) = -1.789$, $p = 0.0783$. The articulatory suppression linear trends approach significance statistically, which may indicate that recency is a factor to consider. However, recency effects would be surprising given the null list length effect. The recency results for the 2 s unfilled delay and the 2 s delay with articulatory suppression are presented in Figures 8 and 9[1].

Discussion

Results demonstrate that a null list length effect is observed at a 2 s delay in which rehearsal is prevented. When rehearsal
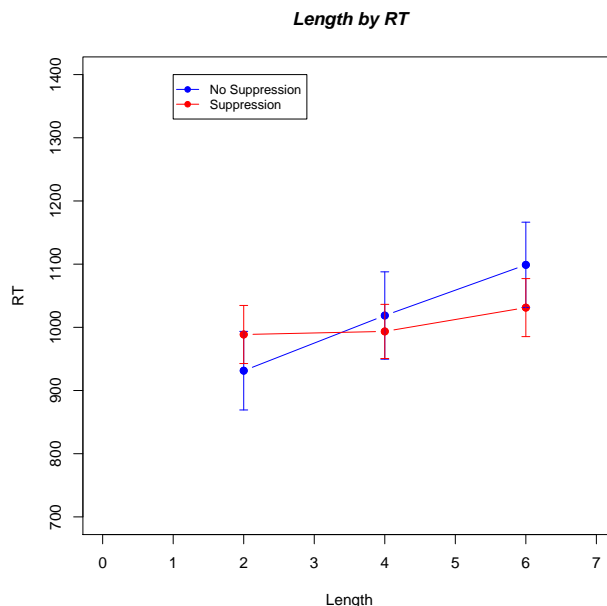
is possible, the list length effects are maintained. However, in either case recency does not appear to affect the list length results.

Discussion

Sternberg's concept of forgetting as a function of set size informed exemplar-based recognition memory models. Here, we elucidate features underlying forgetting in a short-term memory task. Evidence for the null list length effect in both short- and long-term memory underscores continuity between the two classifications of memory. The results suggest that study items themselves are not predictive of length and recency trends across recognition memory. In no case does the list length effect appear to be a clear consequence of item interference. When there are list length effects, the causes are likely due to various confounds: (1) recency in immediate tests; (2) serial rehearsal in recall-oriented tasks; and (3) rehearsal when either a 2 s or 15 s delay does not include an engaging distractor task.

The impact of short-term rehearsal on recency is a topic that merits further exploration. It is possible that rehearsal eliminates recency at short delays (Kahana & Loftus, 1999), and that at longer delays recency is eliminated even when rehearsal is reduced. If the recency effects in the articulatory suppression condition were statistically significant, it would suggest that both study-test lag and relative temporal distinctiveness contribute to the short-term memory trace (Morin Brown & Lewandowsky, 2010).

Alternatively, the nature of contextual representation could shift over the course of study-test delay. Rehearsal may re-
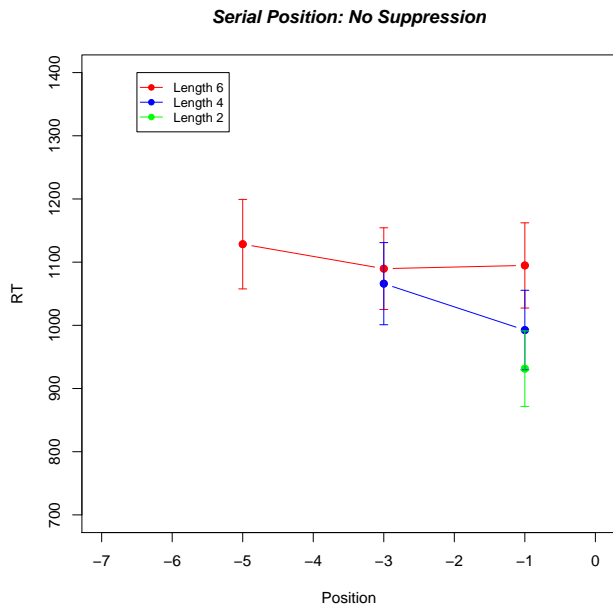
Serial Position: No Suppression

Figure 8: The averaged median latencies to correctly respond "yes" in Experiment 3 as a function of position for each list length in the unfilled 2 s delay.

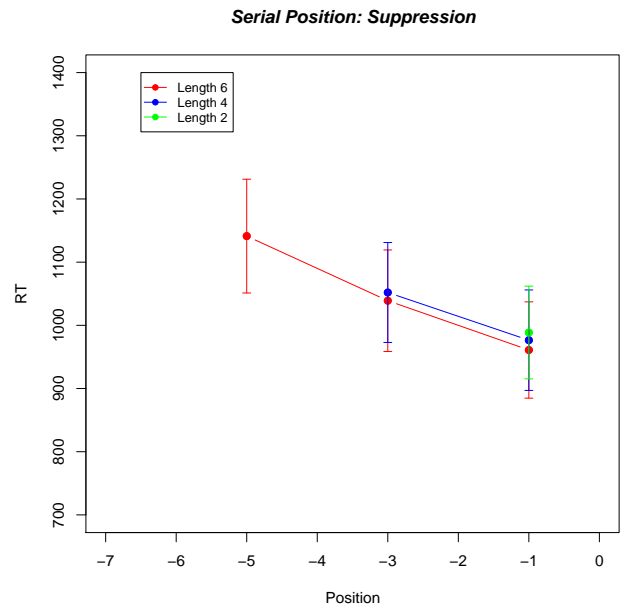**Serial Position: Suppression**

Figure 9: The averaged median latencies to correctly respond "yes" in Experiment 3 as a function of position for each list length in the 2 s articulatory suppression delay.

duce the extent to which the studied context, as a unit, may be reinstated. Similar to the concept of circular scanning (Sternberg, 1969), rehearsal may perturb the positional markers of the study-list context in short delays. Over longer time frames, the distinctiveness of serial position information may be offset by accumulating contextual noise that arises during the study-test lag. These results suggest that both delay and rehearsal function in consort to predict serial position and list length trends. It seems conclusive that there is continuity in the relative contribution of item-interference across short and long-term memory – specifically that item interference alone is not explicative of forgetting in recognition memory.

References

Dennis, S., & Humphreys, M. S. (2001). The role of context in episodic recognition: The bind cue decide model of episodic memory. *Psychological Review*, *108*, 452–478.

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*, 361–376.

Duncan, M., & Murdock, B. (2000). Recognition and recall with precuing and postcuing. *Journal of Memory and Language*, *42*, 311–301.

Forrin, B., & Cunningham, K. (1973). Recognition time and serial position of probed item in short-term memory. *Journal of Experimental Psychology*, *99*, 272–279.

Jensen, O., & Lisman, J. E. (1998). An oscillatory short-term memory buffer model can account for data on the Sternberg task. *The Journal of Neuroscience*, *18*, 10688–10699.

Kahana, M., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. Sternberg (Ed.), *The nature of cognition.* Boston: MIT.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*, 346–373.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465–501.

Morin, C., Brown, G. D. A., & Lewandowsky, S. (2010). Temporal isolation effects in recognition and serial recall. *Memory and Cognition*, *38*, 849–859.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652-654.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*, 421–457.

Townsend, J. T., & Fifić, M. (2004). Parallel versus serial processing and individual differences in human memory. *Perception and Psychophysics*, *66*, 953–962.