Journal of Experimental Psychology: Learning, Memory, and Cognition

Individual Differences in Contextual Storage: Evidence From the List-Strength Effect

Lili Sahakyan, Branden Abushanab, James R. Smith, and Kendra J. Gray

Online First Publication, December 2, 2013. doi: 10.1037/a0035222

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/13/$12.00 DOI: 10.1037/a0035222

RESEARCH REPORT

Individual Differences in Contextual Storage: Evidence From the List-Strength Effect

Lili Sahakyan, Branden Abushanab, James R. Smith, and Kendra J. Gray
University of North Carolina at Greensboro

Strengthening some items in a list of words impairs free recall of the remaining items in the list—a phenomenon known as the list-strength effect (LSE; e.g., Tulving & Hastie, 1972). Research indicates that whether the LSE is observed depends on the nature of the strengthening manipulation, and the effect is attributed to the enhancement of the contextual information in the memory trace of the items (e.g., Malmberg & Shiffrin, 2005). We investigated the magnitude of the LSE as a function of individual differences in working memory capacity (WMC). The findings indicate that low-WMC participants do not show the LSE, suggesting that they do not accumulate as much contextual information in the memory trace as the remaining participants do. These results suggest that the low-spans' deficits in utilizing contextual cues during retrieval (e.g., Spillers & Unsworth, 2011) could be partly linked to their deficits in encoding and storing contextual information. Implications for global theories of memory are discussed.

Keywords: list-strength effect, context strength, working memory capacity, individual differences

Research on the list strength effect (LSE) shows that memory is influenced by the relative strength of items in a study list. Studies have shown that strengthening some items on a list has a negative effect on the recall of the remaining items in the list (e.g., Malmberg & Shiffrin, 2005; Ratcliff, Clark, & Shiffrin, 1990; Tulvig & Hastie, 1972). The LSE is typically demonstrated by having participants study a pure-strong list, a pure-weak list, and a mixed list containing both strong and weak items. An LSE is said to emerge when strong items on the mixed lists are recalled better than strong items on the pure lists, whereas weak items on the pure lists are recalled better than weak items on the mixed lists. The LSE is most robust in free recall than in cued recall or recognition, and we therefore limit the discussion to free recall findings. The LSE is attributed to contextual competition at retrieval that arises from some items being more strongly associated with their episodic context (Malmberg & Shiffrin, 2005; Shiffrin, Ratcliff, & Clark, 1990). Specifically, it is assumed items containing strong context in their memory trace are activated and sampled preferentially over the items that contain weak amounts of context in their trace.

Whether the LSE is observed depends on the manner in which the items are strengthened (Malmberg & Shiffrin, 2005). Specifically, when items are strengthened via increased study time, massed repetitions, or depth of processing, a null LSE is observed—that is, strong items are better recalled than weak items (i.e., main effect of strength), but the effect is invariant across the pure-lists and mixed-lists. Conversely, when items are strengthened via spaced repetitions, then a significant LSE is observed (i.e., the effect of strength is magnified on the mixed lists compared to the pure lists).

It is typical to think of memory strength as a unitary construct defined usually in terms of item recallability. However, Malmberg and Shiffrin (2005) suggested a more complex view of strength, which they decompose into item strength and context strength and demonstrate that these two forms of strength have different effects on memory. Their view was motivated in part by trying to explain the presence or absence of the LSE with different strengthening manipulations. Because learning always takes place in a certain context, Malmberg and Shiffrin suggested that when an item is initially presented, a fixed amount of episodic context is stored in the memory trace of that item (along with the item information). Item information is typically described in terms of the item's meaning (e.g., lexical/semantic representation), whereas context refers to the global environment that is present at the time of encoding the item (e.g., spatial–temporal environment, as well as the emotional and internal environment of the participant). Thus, contextual information is stored only at the beginning of the presentation, presumably because context is more novel at the start of the event. When the exposure duration of an item is increased, or when encoding requires more elaborate processing, then additional item features are stored in the memory trace, thereby enhancing item strength. However such manipulations do not enhance context strength, which explains why extra study time and depth of processing produce a null LSE. In contrast, when the item

Lili Sahakyan, Branden Abushanab, James R. Smith, and Kendra J. Gray, Department of Psychology, University of North Carolina at Greensboro.

The authors thank Paige Henline and Jason May for their help in data collection and scoring.

Correspondence concerning this article should be addressed to Lili Sahakyan, University of North Carolina at Greensboro, Department of Psychology, PO Box 27610, Greensboro, NC 27402. E-mail: l_sahaky@uncg.edu

is repeated again in a spaced fashion, there is an additional opportunity to store more contextual information, thereby incrementing context strength in addition to item strength. Malmberg and Shiffrin termed this the one “shot” hypothesis of context storage (for a full description and the formal model, refer to Malmberg & Shiffrin, 2005).

The goal of this investigation was to examine whether individual differences in working memory capacity (WMC) might predict the magnitude of LSE in different participants. We had theoretical reasons to suspect that low-spans (as indexed by WMC measures) might show a smaller LSE. Research documents many cognitive task deficits in low-spans (for a review, see, e.g., Unsworth & Engle, 2007). We were particularly interested in their episodic memory deficits that are attributed to difficulty in utilizing contextual cues at retrieval (e.g., Delaney & Sahakyan, 2007; Spillers & Unsworth, 2011). For example, it is known that low-spans' deficits are much larger in tests of free recall, which requires self-initiated search processes, as opposed to recognition (e.g., Unsworth, 2009). Not only do low-spans recall fewer items overall, but they also recall them at a slower rate, and they make many more intrusions than high-spans (e.g., Unsworth & Spillers, 2010). These findings are interpreted in terms of a deficient ability to guide search processes during the free recall task. Specifically, low-spans tend to include more irrelevant items in their search set (e.g., intrusions), suggesting that they have deficits in self-generating appropriate retrieval cues. The more intrusions that are included in the search set, the lower the chances of sampling correct items (e.g., Unsworth, 2009; Wixted & Rohrer, 1994). Finally, as the size of the search set increases, the time to sample any item from that set also increases, which explains longer recall latencies in low-spans. Thus, low-spans have lower recall accuracy, longer recall latency, and make more intrusion errors compared to high-spans—effects that indicate deficient use of contextual cues to delimit the search processes (e.g., Unsworth & Spillers, 2010).

Spillers and Unsworth (2011) also examined retrieval dynamics in delayed free recall as a function of WMC, including serial position functions, first response probability functions (which are an indicator of how participants initiate recall), and conditional response probability (CRP) functions, which are an indicator of how people transition between their responses during recall (e.g., Howard & Kahana, 2002). Although Spillers and Unsworth (2011) did not find any differences in the serial position functions or the first response functions between high-spans and low-spans, suggesting that low-spans initiate recall in a similar fashion as high-spans, they found significant differences in CRP functions, indicating more erratic retrieval dynamics in low-spans. In general, the shape of the CRP function could be used to diagnose how participants relied on context cues in free recall (explained further below).

Typically, CRP functions reveal a tendency to successively recall items from the nearby list positions (known as the contiguity effect), and they tend to be characterized by a marked asymmetry, favoring forward transitions over backward transitions. For example, if a participant recalls an item from serial position 7, their next recalled item is more likely to come from adjacent serial positions (e.g., position 8 or 6) than remote positions (e.g., position 15 or 2), demonstrating a contiguity effect. Furthermore, an item from position 8 will be more likely to be recalled than an item from position 6, demonstrating forward bias in recall. Spillers and Unsworth (2011) showed that low-spans were less likely to recall items presented contiguously in a list, demonstrating a diminished contiguity effect. These findings resemble those found with older adults, who also show diminished contiguity (Kahana, Howard, Zaromb, & Wingfield, 2002). Furthermore, low-spans' CRP functions also showed reduced asymmetry compared to high-spans. This means that low-spans were no more likely to transition in a forward fashion than in a backward fashion, showing that they were sampling more indiscriminately from the list after the initial recall.

The contiguity effects are explained in terms of retrieved context (e.g., Howard & Kahana, 2002; Lehman & Malmberg, 2013; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008). Specifically, according to the temporal context model (TCM) of Howard and Kahana (2002), the presentation of each new item during encoding leads to the retrieval of the preexperimental context associated with that item, which is then used to update the experimental context of the list. A similar process presumably takes place during retrieval. The recall of any item reinstates the experimental context associated with that item, which is then used as a retrieval cue for other items on the list. The retrieval of the experimental context leads to the contiguity effect because items with similar context are more likely to be items from the nearby serial positions, and hence they are more likely to be recalled in response to the context cue. Kahana and colleagues described this process as a mental time travel, whereby by recalling an item, participants “jump back” into the list and are then more likely to recall the items from the nearby serial positions. The retrieval of preexperimental context is largely responsible for the asymmetry component of the contiguity effect in TCM. Because the preexperimental context associated with an item updates the experimental context during encoding, that part of the retrieved context is similar to the context associated with subsequent list items but not prior list items. Therefore, a combination of the retrieved preexperimental context and the experimental context lead to a forward-bias in recall in TCM.

The diminished contiguity effect among low-spans observed by Spillers and Unsworth (2011) could indicate that low-spans are less able to formulate and/or utilize retrieved context to guide retrieval of other items from the list. They might be unable to reinstate the context of the retrieved item (i.e., have difficulty jumping back in time), and/or the retrieved context may fail to serve as a good cue for other items on the list. The evidence thus far suggests that low-spans have specific deficits in being able to use contextual cues during retrieval to delimit the search processes, and/or to use the retrieved context as a cue in the recall of subsequent items. Based on findings presented thus far we had hypothesized that low-spans might be deficient in encoding and storing contextual information in the first place, which could explain why they do not benefit from contextual cues at the time of retrieval.

According to the newly proposed buffer model of memory (Lehman & Malmberg, 2013), which is an extension of Atkinson and Shiffrin's (1968) model, the rehearsal buffer is not only limited by the capacity of items it may process but also by its capacity to associate items to the study context. WMC might place constraints on the number of items that are rehearsed together during study, and according to the model, the number of items residing in the buffer is negatively related to the strength with which they are associated to the study context. Another novel addition to the buffer model is what Lehman and Malmberg (2013) refer to as compartmentalization, which is an active controlled process that is used to intentionally drop items from the buffer in order to process other items. If low-spans have deficits in the operations of rehearsal and/or compartmentalization processes, this would mean that some list items would be very weakly encoded or perhaps not at all. Not only might low-spans maintain fewer items in the buffer, they might also create weaker item-to-context associations overall. Indeed, Unsworth and Spillers (2010) documented inefficient use of repetition strategies among low-spans. On these assumptions, the buffer model would predict a smaller LSE for low-spans, along with lower level of accuracy for low-spans.

If low-spans do not store sufficient contextual information in the memory trace as high-spans, then even if they retrieve the context along with the recalled item, that context may be rather impoverished and may not be an effective cue for retrieving other items. This could explain why low-spans are deficient in utilizing context cues. Ironically, reduced storage of context in low-spans will protect them when context changes between the study and test as a consequence of directed forgetting instructions or mental context-change instructions (e.g., Aslan, Zellner, & Bäuml, 2010; Delaney & Sahakyan, 2007; Soriano & Bajo, 2007) or from the mismatch of encoding and retrieval conditions (Unsworth, Brewer, & Spillers, 2011).

Experiment

Method

Participants. Participants were 112 University of North Carolina at Greensboro undergraduates who participated for course credit. They were tested individually.

Design. The experiment involved a *List* (pure vs. mixed) × *Strength* (spaced vs. massed) repeated-measures design.

Materials. The stimuli involved 72 unrelated English nouns of medium frequency, which were randomly divided to create three different study lists for each participant. A *pure-strong* list contained 24 items presented twice, with seven intervening items in between. A *pure-weak* list contained 24 items presented twice, back-to-back. A *mixed* list contained 12 strong and 12 weak items each presented twice either in a spaced or massed fashion. The lag between the presentations of spaced items was always seven words, and items were repeated in the same order in which they appeared during the first presentation. Each tercile of the mixed list contained an equal number of spaced and massed items. We counterbalanced whether a mixed list began with a spaced item or a massed item. We also counterbalanced assignment of words to spaced or massed presentation conditions. Finally, the presentation order of the three types of lists was fully counterbalanced across participants.

After the study and recall of all three lists, participants completed the standard operation span (OSPAN) task measuring working memory capacity (e.g., Kane, Bleckley, Conway, & Engle, 2001; Turner & Engle, 1989). In this task, participants have to verify arithmetic equations while trying to remember words. For example, participants might see "(9/3) +2 = 5 *DRILL*." They read aloud the equation, verified whether it was correct by saying Yes or No, then read aloud the word and tried to remember it for a later test. The next equation/word then immediately appeared. When three question marks (???) appeared on the screen, participants were to write all the words they could remember in the order of presentation in that trial. A trial involved between two to five equations/words, presented one at a time. Three sets of each length (from two to five equation/word presentations) were presented in an unpredictable order (i.e., participants did not know how many words would appear until they received the cue to recall). An item was scored if it was correct and in the correct position. The span score was the proportion of items that were correct and in the correct position (e.g., A. R. Conway et al., 2005).

Procedure. Prior to the experiment, all participants were told that they were going to study and recall three different lists of words. They were also told that they would be performing specific rating tasks on each word in order to help them remember the words for a later test. Items were presented at a rate of 4 s per word. Participants were told that the words were going to be repeated within the list. During the first presentation of the word, they were told to make a yes/no pleasantness judgment by circling either a "+" or "−" sign on a sheet of paper, for pleasant and unpleasant, respectively. During the second presentation, they were asked to make a more graded pleasantness judgment, using a scale from 1 to 5, where higher numbers indicated stronger liking. We used two versions of the same orienting task (as opposed to two different orienting tasks) based on a pilot study in which participants found the constant switching between the orienting tasks distracting from studying the words, especially on the mixed lists. We also avoided using the same orienting task during both presentations out of concern that during the second presentation, participants might simply retrieve their first judgment instead of using the second presentation as an additional study trial. After encoding each list, participants performed a brief distractor task for 30 s. A series of letters were printed on a sheet of paper followed by a blank, and the task was to fill in the blank with the next sequential letter of the alphabet. Afterward, participants were given 90 s to write as many words as they could remember from the study list. The procedure was repeated until all three lists had been studied and recalled. Finally, participants completed the OSPAN task.

Results

The overall LSE. First we examined whether we obtained any LSE by analyzing proportion recalled with repeated-measures analysis of variance (ANOVA), using List (pure vs. mixed) and Strength (strong vs. weak) as factors. The results are summarized in Figure 1. The presence of the LSE should be signified by a significant List × Strength interaction. There was a main effect of Strength, $F(1, 111) = 54.89$, $MSE = 0.016$, $p < .001$, $\eta_p^2 = .33$, qualified by a significant List × Strength interaction, $F(1, 111) = 21.51$, $MSE = 0.014$, $p < .001$, $\eta_p^2 = .16$. Overall, the strong items were recalled better than weak items on both types of lists, but the advantage was more pronounced on mixed lists, $t(111) = 7.71$, $p < .001$, than on pure lists, $t(111) = 2.56$, $p = .012$. This pattern emerged because compared to pure lists, strong items improved on mixed lists, $t(111) = 2.78$, $p = .006$, whereas weak suffered on

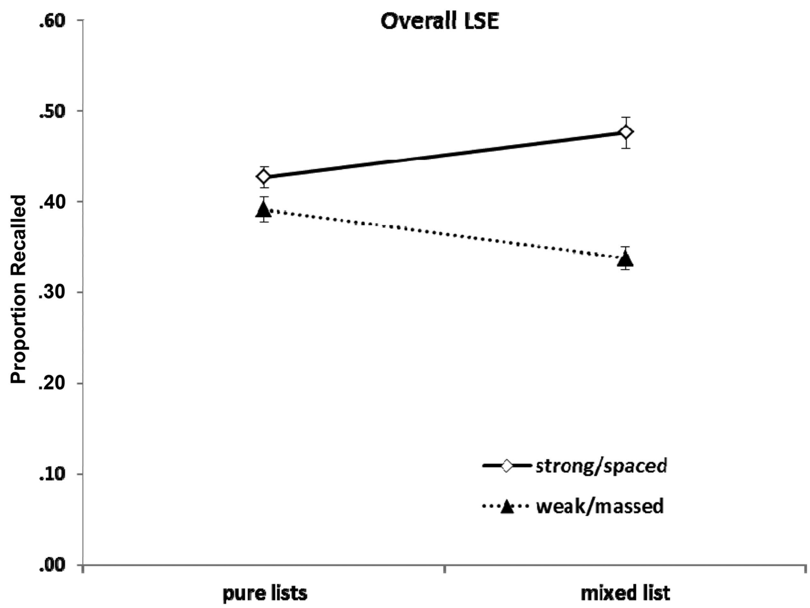

Figure 1. Mean proportion recalled by list type, group, and strength. Error bars represent ± *SE* of the mean. LSE = list-strength effect.

mixed lists, $t(111) = 3.21$, $p = .002$. Thus, the data indicated a significant LSE.

Individual differences in the LSE. Next, we evaluated whether the magnitude of LSE varied as a function of individual differences in WMC. Finding individual differences in the magnitude of the LSE would have implications for the role of attentional control in the storage of contextual information and could provide insight about the nature of the episodic memory deficits experienced by low-spans.

OSPAN scores ranged from .14–.93, with an average score of .65 (*SD* = .14). We reran the analyses described above by entering span scores as a continuous covariate. There was a three-way interaction of List × Strength × WMC, $F(1, 110) = 5.86$, $MSE = 0.013$, $p = .017$, $\eta_p^2 = .05$. Figure 2 shows the strengthening

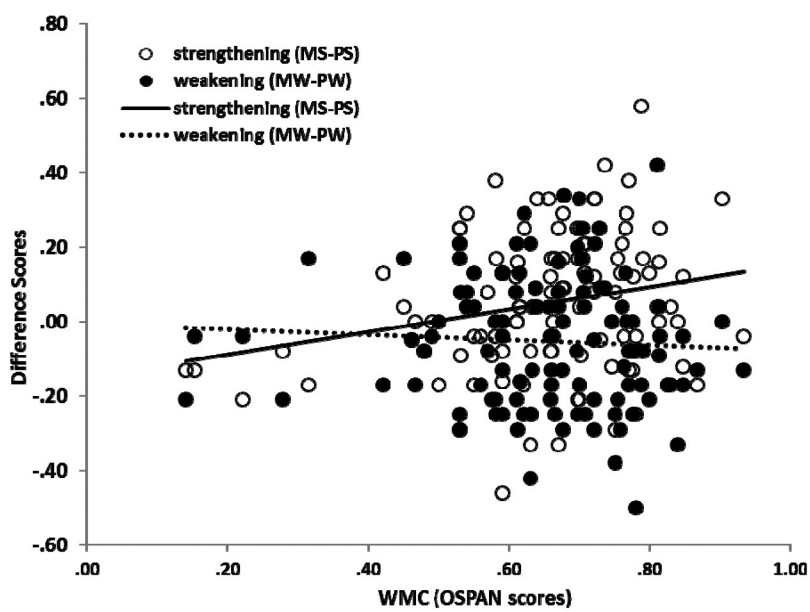

Figure 2. Scatterplot of the strengthening and the weakening components of the list-strength effect calculated for each participant as a function of working memory capacity (WMC; as indexed by operation span [OSPAN] scores). MS = mixed-strong; MW = mixed-weak; PS = pure-strong; PW = pure-weak.

component of the LSE (the difference between mixed-strong and pure-strong recall) and the weakening component of the LSE (the difference between mixed-weak and pure-weak recall) as a function of WMC. There was a significant correlation between WMC and the strengthening component, $r(112) = .23$, $p = .015$, but not between WMC and the weakening component ($r = -.06$, $p = .560$). Given that these components are based on the difference scores, which have reliability issues, and the fact that the weakening component of the LSE is less robust than the strengthening component (e.g., Malmberg & Shiffrin, 2005), the relationship with WMC may be more pronounced in the strengthening than weakening component.

To examine the interaction of LSE with WMC, we divided the participants into terciles based on their span scores and contrasted LSE in the lower tercile against the upper tercile. If low-spans have deficient context storage, then they should show smaller or null LSE. Indeed, among low-spans (e.g., lower tercile, where span scores $< .61$), there was a significant main effect of strength, $F(1, 36) = 13.95$, $MSE = 0.014$, $p = .001$, $\eta_p^2 = .28$, but no List × Strength interaction, $F(1, 36) = 1.28$, $p = .265$, implying that low-spans did not show the LSE despite benefiting from repeated presentations (i.e., showing 7% overall strengthening effect). In contrast, the high-spans (e.g., upper tercile, where span scores $>.72$) showed both the main effect of strength, $F(1, 37) = 20.40$, $MSE = 0.018$, $p < .001$, $\eta_p^2 = .36$, and the List × Strength interaction, $F(1, 33) = 22.42$, $MSE = 0.013$, $p < .001$, $\eta_p^2 = .38$, implying a significant LSE. When Group (first tercile vs. third tercile) was entered as a between-subjects variable in the List × Strength analyses, the three-way interaction was significant, $F(1, 73) = 5.05$, $MSE = 0.015$, $p = .028$, $\eta_p^2 = .07$ (see Figure 3). The main effect of group was also significant, $F(1, 73) = 6.05$, $MSE = 0.033$, $p = .016$, $\eta_p^2 = .08$, indicating that low-spans had overall worse memory than high-spans.

Another interesting observation is the presence of spacing effect on the pure lists among low-spans, $t(36) = 2.08$, $p = .045$, but complete absence of such pattern among high-spans ($t < 1$). Prior research showed that with instructions to rehearse each item alone, there is a spacing effect on pure lists, but with instructions to rehearse items together (i.e., the usual way of rehearsal), the spacing effect is reduced on pure lists (e.g., Delaney & Knowles, 2005; Delaney & Verkoeijen, 2009; Verkoeijen & Delaney, 2008). The reason rehearsal attenuates the spacing effect on pure lists is because it raises the recall of massed items by converting them into functionally spaced items. We suspect that even though we controlled the encoding strategy in our experiment, high-spans managed to nonetheless rehearse after completing the orienting task, whereas low-spans did not engage in rehearsal and merely followed the orienting task instructions, thereby focusing only on a given item. This could explain why the spacing effect was eliminated on pure lists among high-spans but not low-spans.

Analyses of first responses. Research from our lab indicates that when participants study a mixed list of strong and weak items, context strength influences the output dynamics during free recall, whereas item strength does not. Specifically, when strength was varied by spaced presentations, participants output spaced items first compared to massed items, but they did not show the same bias when items were strengthened via depth of processing or extra study time (Sahakyan, Delaney, & Waldum, 2008). Given our previous findings, and the assertion that the LSE is driven by items with stronger context coming to mind more easily during recall, we expected that the first recalled item in the mixed list should be a spaced item. However, this should be found only among those participants who showed a significant LSE, whereas low-spans should not show this advantage. Indeed, among low-spans, 41% of the participants began their recall in mixed list with spaced items,

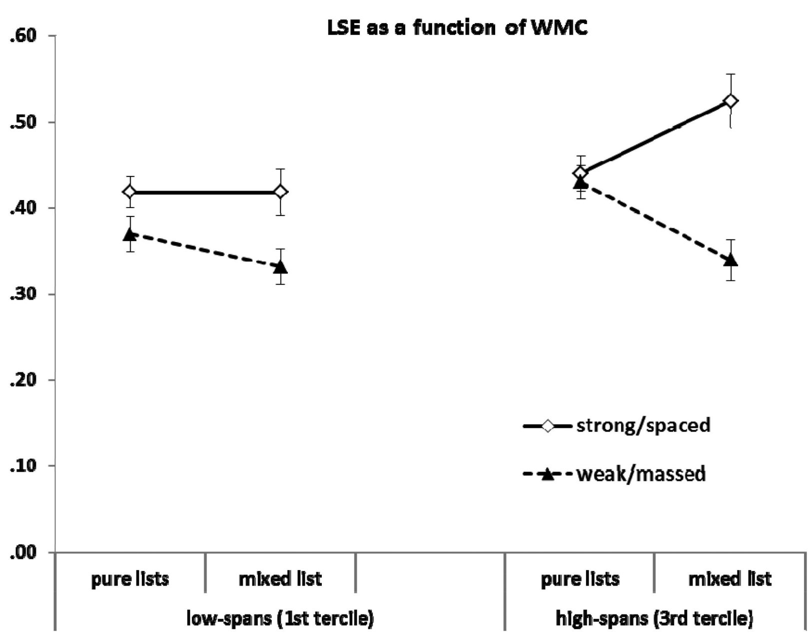

Figure 3. Mean proportion recalled by list type, strength, and working memory capacity (WMC; as indexed by operation span scores) split by higher and lower tercile. Error bars represent ± *SE* of the mean. LSE = list-strength effect.

whereas among high-spans, the majority of participants (68%) initiated recall with a spaced item, $\chi^2(1, N = 75) = 5.88, p = .015$.

Finally, we examined first response function among low-spans and high-spans, because it is informative for understanding how participants initiate recall. First response probability (FRP) refers to the number of times the first word recalled comes from a given serial position, divided by the number of times the first recalled word could have come from that serial position. It indicates those responses that theoretically have the strongest overlap with the context cues being used at retrieval. For example, if participants initiate recall from the end of the list, it would suggest that they are relying on the end of the list context to probe memory, whereas initiating retrieval from the beginning of the list implies that participants are actively reinstating the beginning of list context rather than relying on the recent context.

Given that all of the 24 items were repeated twice, analyzing first responses as a function of nominal serial position was feasible only for the pure-weak list, where items were repeated back-to-back. In contrast, on pure-spaced list or on mixed list, the serial position of the recalled item would be impossible to infer because the recalled item could refer either to the first or second repetition of that item within the list. Figure 4 demonstrates FRPs on pure-weak list as a function of serial position and WMC. The figure demonstrates that both high-spans and low-spans showed a tendency to initiate recall from the beginning of the list rather than the end of the list, but the advantage was numerically higher for high-spans. This is consistent with prior research, where a numerical advantage to initiate from the beginning of the list was found among high-spans (e.g., Spillers & Unsworth, 2011; Unsworth & Spillers, 2010). Note, that on the pure-weak list high-spans never initiated recall from the end of the list, whereas some low-spans did. In immediate recall, retrieving from the buffer (i.e., end of the list) makes sense, and that is what high performers do (e.g., Lehman & Malmberg, 2013). However, in delayed free recall, retrieval from the end of the list will make it much more difficult to reinstate the beginning of list context. Hence, some low-spans are utilizing inefficient retrieval strategy when they begin delayed free recall with the end of the list.

According to Lehman and Malmberg's (2013) model, the ability to initiate recall at the beginning of the list requires two things—(a) to reinstate the beginning of the list context on the assumption that context changes over the course of study, and (b) that context is most strongly associated with the first item on the list. The latter falls out of the critical assumption of the model that there is a limited capacity to encode item-to-context associations and the first item on the list therefore has an advantage. The tendency for high-spans to initiate recall from the beginning of pure-weak list implies that they encoded context most strongly for the beginning of the list and reinstated that beginning of list context more successfully than low-spans. If low-spans create weaker item-to-context associations as evidenced for example by the null LSE, it would make it more difficult for them to reinstate the beginning of list context.

Although we could not plot FRPs for pure-spaced and mixed-lists as a function of serial position, we nevertheless assessed those functions by splitting the lists into three terciles (T1, T2, T3), with each tercile containing eight items. Our lists were constructed such that repetitions of items occurred only within the terciles, but never across the terciles, and hence the origin of the first recalled item could be traced to the beginning, middle, or end of the list. Figure 5 shows FRPs as a function of list type, serial position terciles, and WMC. Interestingly, both low-spans and high-spans showed similar FRPs on pure-spaced and mixed-lists, but they differed on pure-massed list. We suspect that the reason we did not observe the same pattern for high-spans on pure-spaced and on mixed-list as we did on pure-massed list is because other items on those lists had multiple "shots" of context and therefore the advantage for reinstating the beginning of the list was reduced in light of other items with strong context associations. On a pure-massed list, however, context is stored relatively weakly because massed presentations

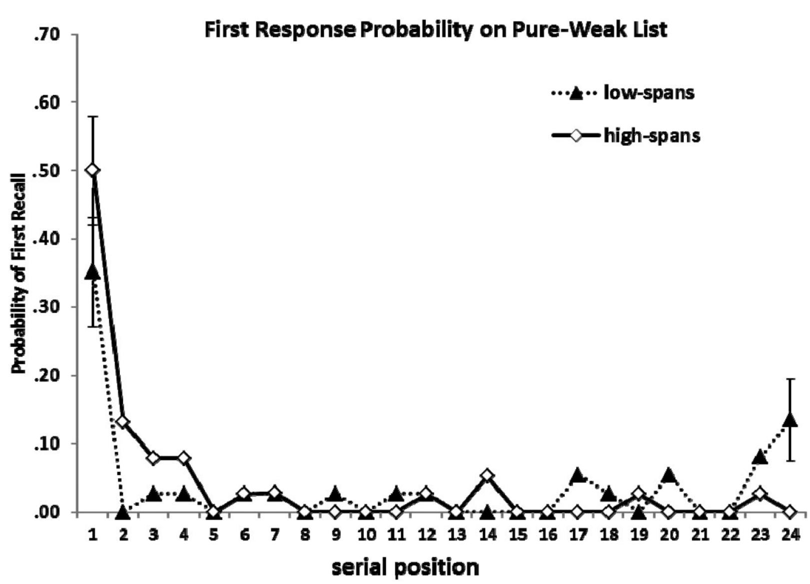

Figure 4. First response probability as a function of serial position on pure-weak list for low-spans and high-spans. Error bars represent ± *SE* of the mean.

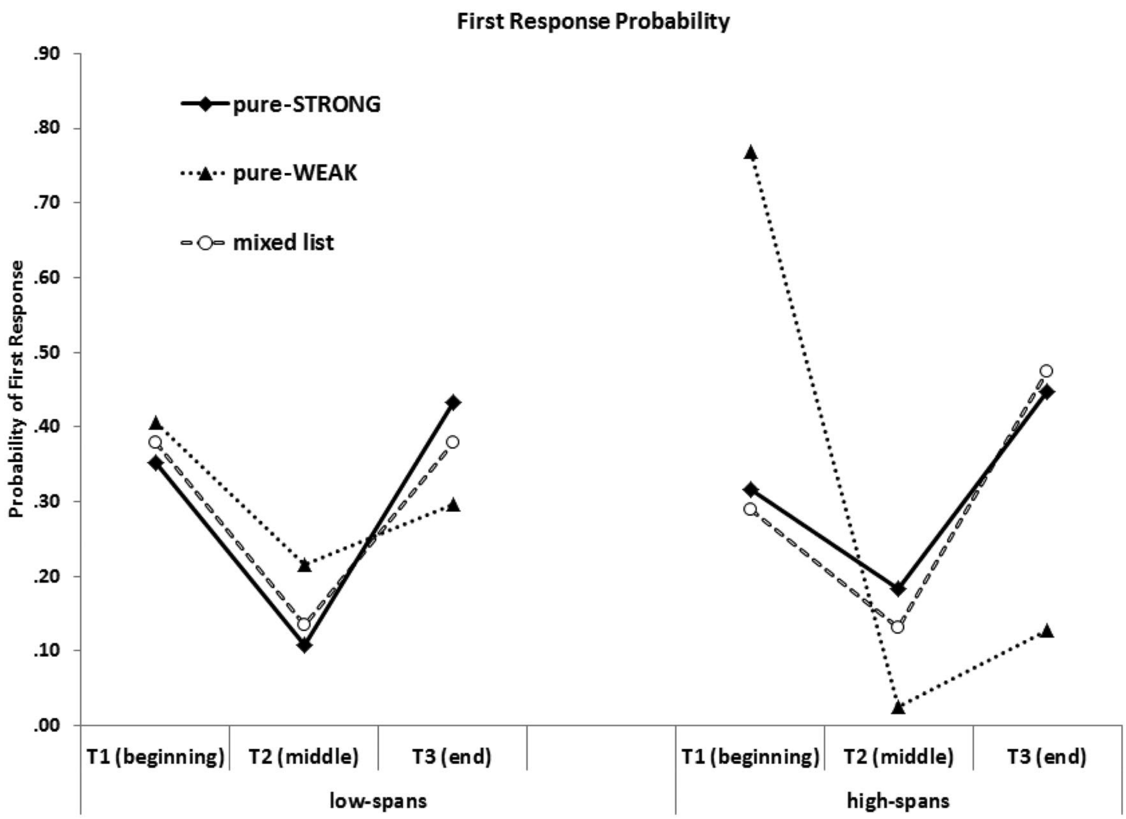

Figure 5. First response probability on pure-weak, pure-strong, and mixed list as a function of serial position tercile for low-spans and high-spans.

do not increment context strength (Malmberg & Shiffrin, 2005), and hence there is no competition from other list items, making it easier to reinstate the beginning of list context.

Overall, our analyses of FRPs further confirm that context is encoded and reinstated better by high-spans than low-spans and that such advantage is evident on pure-weak lists, where there is no competition from other items with stronger context associations.

General Discussion

The goal of the study was to examine if individual differences in WMC might predict who shows an LSE and who does not. Consistent with our predictions, the results showed that low-spans showed a null LSE despite showing an overall strengthening effect. In fact, the pattern of recall in low-spans was reminiscent of the pattern that is obtained when items are strengthened via additional study time, massed repetitions, or depth of processing in prior research (e.g., Malmberg & Shiffrin, 2005). To explain why spaced manipulations produced LSE, whereas other manipulations did not, Malmberg and Shiffrin (2005) proposed that a fixed amount of context is stored with each presentation. When items are repeated spaced apart, there is an additional opportunity to store more contextual information. When the presentation duration is increased, or when the encoding requires more elaborate processing, additional item features are being encoded, but the contextual information is no longer strengthened because it is stored only at the beginning of presentation.

The null LSE results in low-spans suggest that these participants do not store additional contextual information with repeated presentations (at least, when two repetitions are concerned). This might explain their deficits in capitalizing on contextual information that is retrieved with each item (e.g., Spillers & Unsworth, 2011). Our findings suggest that one reason low-spans may have deficits in utilizing the contextual information is that they do not necessarily store as much contextual information during encoding as high-spans do. They may store contextual information along with the item, but they may store a smaller "shot" of context than high-spans. If this is the case, low-spans might require many more spaced presentations to eventually benefit from stored context at the time of memory search. Low-spans may need more time to store a shot of context, or more time to retrieve it because they experience more competition at recall (e.g., Unsworth & Spillers, 2010). The null LSE findings with low-spans suggest that populations with low WMC (e.g., older adults), may also be deficient in storage of contextual information, which could explain why older adults experience particular difficulty in tasks that rely heavily on formation and retrieval of episodic associations (e.g., Hoyer & Verhaeghen, 2006; Kahana et al., 2002; Naveh-Benjamin, 2000; Wingfield & Kahana, 2002).

Given that low-spans have deficits in attentional control (e.g., Engle & Kane, 2004), and given that low-spans did not show LSE, the results imply that contextual encoding relies on attentional control. Additional evidence supporting this idea comes from list-method directed forgetting studies. For example, M. A. Conway, Harries, Noyes, Racsma'ny, and Frankish (2000) showed that when participants performed a divided attention task during the encoding of the second list, they failed to show directed forgetting. Although the authors have interpreted those findings as reflecting attentional demands of inhibition, we have argued that they could also reflect attentional demands of contextual encoding and maintenance (e.g., Sahakyan & Kelley, 2002). The null LSE findings among low-spans support the role of attention in contextual encoding and maintenance.

Collectively, the results suggest that low-spans' difficulty in utilizing contextual cues during retrieval may be linked to their failure to store contextual information in the first place. It could also explain why they do not suffer as much from directed forgetting, from changes in mental context between the study and test (e.g., Delaney & Sahakyan, 2007), or from mismatch of encoding and retrieval conditions (Unsworth et al., 2011). It could also explain why they show a smaller spacing effect than the remaining participants. Since low-spans do not store as much contextual information in the memory trace, then even if they partly reinstate the context along with the recalled item, that context cannot be an effective cue for retrieving other items, explaining why low-spans show diminished contiguity effects. The diminished contiguity effects among low-spans could also be interpreted in terms of the attentional demands of context retrieval. Since contiguity effects are attributed to the effects of retrieved context, and since low-spans show diminished contiguity effects, it could also be the case that reinstatement of episodic context requires attentional control. In other words, mental time travel is attentionally demanding, and participants who have deficits in this capacity experience difficulty in contextual reinstatement. That low-spans have deficits in contextual reinstatement is further confirmed by the analyses of how high-spans and low-spans initiate retrieval in delayed free recall. Our FRP analyses suggest that high-spans are more adaptable, and they use a strategy that is appropriate for what they have stored in memory. They appear better at using control processes as their FRP patterns look more varied across condition than the low-spans. Overall, our findings along with those reported by Unsworth et al. (2011) provide converging evidence for the deficits in storage and utilization of contextual information in memory among low-spans.

Finally, our findings constrain global theories of memory, and they pose a challenge for single-store models of memory like TCM and their variants because according to TCM, WM plays no role in producing temporal contingencies during retrieval or encoding. Our findings, however, suggest that in delayed free recall participants initiate retrieval from the beginning of the list, which is not predicted by TCM, which assumes that participants rely on the recent context to probe memory on immediate test and that such tendency is diminished on delayed test because the overlap between the test context and study context is reduced. TCM has no built-in mechanism by which it could explain the tendency to start retrieval from the beginning of the list (see also Lehman & Malmberg, 2009, 2013; Sahakyan & Hendricks, 2012; Spillers & Unsworth, 2011). What distinguishes TCM from the Atkinson-Shiffrin model is the assumption that control processes implemented in STM/WM/Buffer do not play a role in primacy, recency, or contiguity effects. However, the newly proposed model includes a buffer or working memory component (cf. Lehman & Malmberg, 2013), and it explains a wide range of findings, including individual differences in the storage and utilization of contextual information.

References

Aslan, A., Zellner, M., & Bäuml, H. (2010). Working memory capacity predicts listwise directed forgetting in adults and children. *Memory, 18,* 442–450. doi:10.1080/09658211003742698

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposal system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). London, England: Academic Press.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786. doi:10.3758/BF03196772

Conway, M. A., Harries, K., Noyes, J., Racsma'ny, M., & Frankish, C. R. (2000). The disruption and dissolution of directed forgetting: Inhibitory control of memory. *Journal of Memory and Language, 43,* 409–430. doi:10.1006/jmla.2000.2706

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing in free recall. *Journal of Memory and Language, 52,* 120–130. doi:10.1016/j.jml.2004.09.002

Delaney, P. F., & Sahakyan, L. (2007). Effects of working memory capacity on context maintenance: Adding individual differences to the contextual-change account of directed forgetting. *Memory & Cognition, 35,* 1074–1082. doi:10.3758/BF03193479

Delaney, P. F., & Verkoeijen, P. P. J. L. (2009). Rehearsal strategies can enlarge or diminish the spacing effect: Pure versus mixed lists and encoding strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1148–1161. doi:10.1037/a0016380

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York, NY: Elsevier.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. doi:10.1006/jmps.2001.1388

Hoyer, W. J., & Verhaeghen, P. (2006). Memory aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (6th ed., pp. 209–232). San Diego, CA: Elsevier. doi:10.1016/B978-012101264-9/50013-6

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 530–540.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 970–988. doi:10.1037/a0015728

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*. Advance online publication. doi:10.1037/a0030851

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336. doi:10.1037/0278-7393.31.2.322

Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170–1187. doi:10.1037/0278-7393.26.5.1170

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). Episodic and semantic organization during free recall: The control of memory search. *Psychological Review, 116,* 129–156. doi:10.1037/a0014420

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178. doi:10.1037/0278-7393.16.2.163

Sahakyan, L., Delaney, P. F., & Waldum, E. R. (2008). Intentional forgetting is easier after two "shots" than one. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 408–414. doi:10.1037/0278-7393.34.2.408

Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Memory & Cognition, 40,* 844–860. doi:10.3758/s13421-012-0198-0

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072. doi:10.1037/0278-7393.28.6.1064

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912. doi:10.1037/a0013396

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195. doi:10.1037/0278-7393.16.2.179

Soriano, M. F., & Bajo, M. T. (2007). Working memory resources and interference in directed forgetting. *Psicologia, 28,* 63–85.

Spillers, G., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1532–1539. doi:10.1037/a0024852

Tulving, E., & Hastie, R. (1972). Inhibition effects in intralist repetitions in free recall. *Journal of Experimental Psychology, 92,* 297–304. doi:10.1037/h0032367

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N. (2009). Individual differences in self-initiated processing at encoding and retrieval: A latent variable analysis. *The Quarterly Journal of Experimental Psychology, 62,* 257–266. doi:10.1080/17470210802373092

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Variation in working memory capacity and episodic memory: Examining the importance of encoding specificity. *Psychonomic Bulletin & Review, 18,* 1113–1118. doi:10.3758/s13423-011-0165-y

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. doi:10.1037/0033-295X.114.1.104

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review, 17,* 200–205. doi:10.3758/PBR.17.2.200

Verkoeijen, P. P. J. L., & Delaney, P. F. (2008). Rote rehearsal and spacing effects in the free recall of pure and mixed lists. *Journal of Memory and Language, 58,* 35–47. doi:10.1016/j.jml.2007.07.006

Wingfield, A., & Kahana, M. J. (2002). The dynamics of memory retrieval in older adults. *Canadian Journal of Experimental Psychology, 56,* 187–199. doi:10.1037/h0087396

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1,* 89–106. doi:10.3758/BF03200763

Received July 22, 2013
Revision received October 11, 2013
Accepted October 14, 2013 ■