



# List-strength effects in older adults in recognition and free recall

Lili Sahakyan<sup>1,2</sup>

Published online: 3 January 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

The present investigation provides a novel extension of the retrieving effectively from memory (REM) model to examine free-recall and recognition memory in older adults to inform our understanding of age-related cognitive decline. When some items on a list are strengthened through distributed repetitions, memory for the nonstrengthened items on that list may become impaired depending on how memory is tested—a phenomenon known as the list-strength effect (LSE; e.g., Tulving & Hastie, *Journal of Experimental Psychology*, 92, 297–304, 1972). When the strengthening operation involves distributed repetitions, LSE is robust in free recall, but it is reliably absent in recognition (Malmberg & Shiffrin, *Journal of Experimental Psychology*, 31, 322–336, 2005; Ratcliff, Clark, & Shiffrin, *Journal of Experimental Psychology*, 16, 163–178, 1990). Two experiments investigated LSE in recall and recognition in older adults and young adults who encoded items with full or divided attention. Despite showing impaired recall and recognition, older adults showed patterns of LSEs across both experiments that were similar to young adults with full attention rather than young adults with divided attention. In recognition, there was a null LSE in older and young adults, but a positive LSE was observed under divided attention. In contrast, in free recall, there was a positive LSE in older and young adults, but a null LSE under divided attention. Collectively, the results suggest that older adults do not have impaired encoding of context information (evidenced by intact LSE in recall), and they do not have impaired differentiation of item representations (evidenced by the null LSE in recognition). Age-related impairment in both memory tasks can be accounted for by sparse encoding of item-based information.

**Keywords** Aging · List-strength effect · Strength-based mirror effect · Divided attention · Free recall · Recognition

Fifty years ago, Atkinson and Shiffrin (1968) published their landmark chapter on the modal model of memory, which was revolutionary in guiding the development of new theoretical models and stimulating modern research. Since its initial publication, the model has undergone several major developments, including the search of associative memory (SAM; Shiffrin, Ratcliff, & Clark, 1990) and retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997) models, which have been successful in accounting for a variety of phenomena across different tasks. As successful as these models have been in stimulating new research with young adults, they have not been applied to the issues of aging and memory (with one exception, in which REM was used to

model associative recognition in older adults; Stephens & Overman, 2018). The purpose of the current investigation is to examine age-related changes in episodic memory using a paradigm that was foundational in the development of REM—known as the list-strength effect (LSE), or the mixed/pure list paradigm (Ratcliff, Clark, & Shiffrin, 1990; Shiffrin et al., 1990; Tulving & Hastie, 1972). Examining LSE in aging can inform our understanding of the nature of age-related impairments in memory in part because the LSE phenomenon has been at the center of the modeling literature, and a comprehensive account of LSE in free recall and recognition is provided by REM. The details of the paradigm, its empirical findings, and the underlying mechanisms will be reviewed in greater detail below.

Several factors motivated this investigation. The LSE paradigm produces different effects in free recall and recognition, and REM simultaneously predicts those findings by assuming that item and context features are encoded according to different time courses and are differentially impacted by task demands. Briefly, LSE in recognition is explained through an item differentiation mechanism (i.e., Shiffrin et al., 1990),

✉ Lili Sahakyan  
lsahaky@illinois.edu

<sup>1</sup> Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

<sup>2</sup> Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

whereas LSE in free recall is explained through retrieval competition arising from different amount of context stored in memory traces (Malmberg & Shiffrin, 2005). Thus, examination of LSE in free recall can inform our understanding of older adults' deficits in context processing, which are well established (Burke & Light, 1981; Chalfonte & Johnson, 1996; Spencer & Raz, 1995). What is less clear, however, is whether some of those deficits stem from poor encoding of context (i.e., representational problem), or whether they arise from retrieval deficits. The deficit in context encoding implies reduced magnitude of LSE in free recall (in comparison with young adults), and conversely, intact context encoding in older adults implies that the magnitude of LSE should be unaffected. Whereas examining LSE in free recall can shed light on the encoding of *context* information in older adults, the examination of LSE in recognition can inform about the quality of encoding of *item* information, exploring potential age-related changes in the quality of item representations.

A recent meta-analysis on age-related differences in recognition memory confirms that older adults show impaired discrimination accuracy ( $d'$ ), with multiple variables influencing the magnitude of this effect (Fraundorf, Hourihan, Peter, & Benjamin, 2018). Various conceptualizations of aging have implied that item representations may be less well differentiated in older adults (e.g., Benjamin, 2010; Buchler & Reder, 2007; Li, Naveh-Benjamin, & Lindenberger, 2005). For example, one computational account of age-related changes in recognition suggests that older adults store more sparse representations, encoding fewer and less accurate features (Benjamin, 2010). If sparse representations also contain insufficient diagnostic features to distinguish one trace from another, this would translate into an impaired differentiation problem. However, if older adults compensate for sparsity in encoding by focusing selectively on *the most* diagnostic/distinctive features of a stimulus, sparseness per se may not imply reduced differentiation; the types of features contained in that representation will matter.

Finally, in addition to examining LSE in older adults, this investigation aims to compare older adults with young adults with divided attention at encoding. Several accounts of aging implicate age-related differences in attentional or frontal executive processes by suggesting that older adults have impaired attentional processes ( Craik, 1977; Hasher & Zacks, 1988; West, 1996). The equivalence (or lack thereof) between older adults and young adults with divided attention has been a topic of ongoing investigation. Some studies reported distinct differences between older adults and young adults with divided attention (e.g., Kilb & Naveh-Benjamin, 2007; Smyth & Naveh-Benjamin, 2016), whereas others found that divided attention young adults behaved similar to older adults (e.g., Castel & Craik, 2003; Jacoby, 1999; Kelley & Sahakyan, 2003), and the issue remains unclear (Naveh-Benjamin & Mayr, 2018). Our recent findings with young adults

demonstrate that divided attention at encoding interacts with LSE both in free recall and in recognition (Sahakyan & Malmberg, 2018). Divided attention greatly reduces LSE in free recall, and it also leads to an LSE in recognition, which is a rare outcome (more on this below). It remains an empirical question whether older adults will demonstrate similar types of interactions with LSE. Among young adults, the interaction of divided attention with LSE in recognition indicates impairment of item differentiation that results from difficulty with trace updating, whereas the interaction with LSE in free recall indicates impairment of context encoding. Therefore, observing similar interactions in older adults would inform whether similar processes are implicated in aging. Atkinson and Shiffrin (1968) emphasized control processes as an integral component of their model (i.e., processes that are under the voluntary control of participants, including various encoding and retrieval strategies). Aging (also, divided attention) could disrupt control processes involved in encoding and/or retrieval of memories during encoding, making updating of existing traces more difficult.

## The LSE paradigm

In the LSE paradigm, participants study three lists of items and get tested on each of the lists after encoding. The term *mixed list* refers to a list that contains a mixture of strong and weak items, whereas the term *pure strong* or *pure weak* refer to separate lists, where all items are either strongly or weakly encoded. When the strengthening operation involves distributed repetitions of the items, and memory is tested via free recall, strong items are better recalled in a mixed list compared with a pure strong list, whereas weak items are better recalled on a pure weak list compared with the mixed list (Malmberg & Shiffrin, 2005; Sahakyan, Abushanab, Smith, & Gray, 2014; Wilson & Criss, 2017). In other words, although strong items are better recalled than are weak items when those items are on separate lists (strength effect), this strength effect gets magnified when the items are intermixed on the mixed list. This is known as the positive LSE, and it is robust in free recall only when strengthening manipulations involve distributed repetitions.

In contrast, when memory is tested with cued recall or recognition, a null or even a negative LSE is observed (Ratcliff et al., 1990; Sahakyan & Malmberg, 2018; Wilson & Criss, 2017; Yonelinas, Hockley, & Murdock, 1992). A null LSE refers to the *same* magnitude of the strength effect across the pure lists and the mixed list—that is, strong items are remembered better than weak items, but the strength effect is unaffected by the list composition. A negative LSE refers to the weak items benefiting from being on the mixed list compared with the pure-weak list, whereas strong items suffering

from being on the mixed list compared with the pure-strong list.

The null/negative LSE has posed a major challenge to global matching models of memory because virtually all of those models predicted a positive LSE in recognition (Shiffrin et al., 1990), and yet empirical findings failed to produce that pattern, indicating a fundamental problem with global matching models. It is not an exaggeration to say that the null LSE in recognition served as an impetus for developing a new generation of models, such as REM and SLiM (subjective likelihood model; McClelland & Chappell, 1998), which are known as differentiation models because they implemented differentiation as a fundamental mechanism of memory.

### Assumptions of the REM model and its implications for aging

REM distinguishes between two types of event representations—those that are incomplete and formed during the study, representing *episodic* traces, and those that are relatively complete, accurate, and decontextualized representations of lifelong accumulated knowledge, known as *lexical/semantic* traces. The representations are stored as a vector of feature values, with all episodic traces comprised of item and context features. Context plays a critical role in free recall because context cues are the only cues available to initiate the search process. In contrast, recognition is more dependent on the information presented in the memory cue, and context plays a more limited role in restricting the comparison to the episodic traces from the study list. Thus, item features play a more central role in the recognition decision process as they affect the matching of a test probe to the stored items in memory.

#### Encoding in REM

When an item is presented for study, it is assumed to be an accurate copy of the lexical/semantic representation. Some (but not all) of the features representing that item are stored in an episodic representation, along with the context features. Once a feature is stored, its value does not change throughout the study. Furthermore, even if some features are stored, the values of those features may not necessarily be stored correctly. The feature values are selected from a geometric distribution, in which some features are more common than others. Common features are represented by lower numerical integers and provide less discriminating information about the likelihood that the two traces match compared with higher feature values, which indicate rarer, more discriminating features. Thus, the values of the individual features play a role during the matching process.

Several parameters affect the quality of encoding in episodic representations, and aging may affect any of these parameters, including their combinations. Whether or not a feature gets stored is described by the  $u$  parameter in the model. Variations in the  $u$  parameter reflect the *quantity* of information stored, with lower  $u$  values representing more sparsely encoded representations. Older adults may store fewer features in general (i.e., store reduced amount of information). In REM, reducing the  $u$  parameter lowers hits and increases false alarms in recognition. Importantly, just because a feature is stored, there is no guarantee that the correct value of that feature is stored; this aspect is captured by the  $c$  parameter, which represents the probability that the correct feature value is stored, if a feature is stored. Variations in  $c$  parameter reflect *accuracy* of encoded information, and older adults may store more inaccurate features, contributing to noisier representations. In REM, reducing the  $c$  parameter decreases hits, without affecting false alarms. Finally, the  $g$  parameter determines the likelihood of storing distinctive features, with increases in  $g$  representing a greater likelihood of storing more common features, making representations less distinctive (i.e.,  $g$  parameter is used to model word-frequency effects in recognition by setting higher values of  $g$  for high-frequency words; thus, increasing  $g$  leads to lower hits and higher false alarms). Older adults may store less diagnostic features during encoding, making their representations more generic, or they may focus their limited encoding bandwidth on features that are highly diagnostic; these effects would be captured by varying the system  $g$  parameter between older and younger adults.<sup>1</sup> Importantly, it is reasonable to assume that lexical/semantic representations may differ between older adults and young adults (i.e., in terms of the number of features or types of features). It is well known that older adults have a superior semantic knowledge relative to young adults, despite declines in many other domains (e.g., Park et al., 2002; Verhaeghen, 2003). How differences in knowledge interact with event memory is a complex problem in itself, and efforts have been made to model those processes in young adults (e.g., Cox & Shiffrin, 2012; Nelson & Shiffrin, 2013), but not yet in older adults (see, however, Benjamin, 2010).

#### Retrieval in REM

During the recognition test, the lexical/semantic vector representing the test probe is compared in parallel with each of the noisy episodic memory traces stored during learning. Some of the presented test items are from the study list (targets), whereas others are new (lures). The match between the

<sup>1</sup> Reduction of quantity of features stored ( $u$ ) does not interact with the type of features stored ( $g$ ), whereas reduction in accuracy ( $c$ ) interacts with feature distinctiveness such that the hit rate advantage for distinctive representations (i.e., low-frequency words) reverses at lowest levels of  $c$  (Malmberg, Zeelenberg, & Shiffrin, 2004).

test item and the contents of memory grows as a function of their overlap. Importantly, during this comparison process, the diagnosticity of features plays an important role. Matching a rare feature value provides more diagnostic evidence that the test item is old than does matching a common feature value. At the same time, all mismatching features (regardless of feature values) provide equivalent amount of evidence that the test item is new. The comparison process for each stored episodic trace yields the likelihood value that the trace resulted from the presented test item. The decision to call the test item “old” is based on the mean of the likelihood values for all memory traces.

## Comprehensive account of LSEs

### LSE in recognition

An important assumption made by REM is that repetitions of an item during encoding accumulate additional features (about the item and the context) in the *original* episodic trace that was laid during the first presentation of the item, as opposed to being stored as a new trace. Repetitions thus update the memory trace, making it more complete, and informative (modeled by incrementing  $u$  parameter). The result of this updating is that the trace becomes more differentiated from other traces in memory (Criss, 2006; Shiffrin et al., 1990). Hence, strengthening manipulations increase the match between the test item and its episodic trace (increasing the hits), while simultaneously decreasing the match between the target trace and other traces in memory (decreasing the false alarms). In other words, the more we learn about one thing, the less like other things it becomes. This mechanism, known as *differentiation*, allows REM to account for the null LSE in recognition. By strengthening some items in memory, the similarity of those traces to other items decreases, which counteracts the increase in interference that might otherwise be caused by increases in the variability of encoding. Thus, adding strong representations to memory reduces the noise associated with the global-matching process as long as the repetitions accumulate information in the original trace, enabling them to be updated. In the absence of trace accumulation/updating, a positive LSE is predicted in recognition. The differentiation mechanism also accounts for the strength-based mirror effect (SBME), which refers to the increase in hit rates and decrease in false alarm-rates when strength is manipulated between the lists (Wixted & Stretch, 2004). For a computational example demonstrating how differentiation explains both LSE and SBME, see Criss (2006).

### LSE in free recall

The same encoding assumption that explains the accumulation of additional features in the original memory trace also

predicts a positive LSE in free recall (Malmberg & Shiffrin, 2005). Free recall is assumed to be initiated with context cues, and the match between the context stored in the traces and context used to probe memory determines which traces are sampled. The traces that contain more context features are more likely to be sampled compared with traces containing fewer context features.

According to the one-shot hypothesis for context encoding (Malmberg & Shiffrin, 2005), a fixed amount of item and context information is stored at the start of the study trial. Additional study time, as well as deeper encoding, allow storage of additional item features, whereas the context features remain relatively unchanged as context drifts relatively slowly during the experiment. With distributed repetitions, however, there is an opportunity to update both item features and context features. Therefore, when a pure list is studied, all traces have about the same amount of context stored in them, and thus, all other things being equal, have the same chance to be sampled. In contrast, in a mixed list, items strengthened via distributed repetitions have more context stored in their traces than weakly encoded items do. Hence, the traces of items studied more than once in a distributed fashion have advantage over the traces of items studied only once (or presented in a massed fashion) in the retrieval competition. This produces the positive LSE in free recall when distributed repetitions are used as a strengthening operation. In contrast, extra study time and depth of processing manipulations produce a null LSE in free recall, consistent with the predictions of the model (Malmberg & Shiffrin, 2005).

## Experiment 1: Recognition

The purpose of Experiment 1 is to assess LSE in recognition among older adults and compare its magnitude with young adults who encode the items with full or divided attention. As mentioned, trace differentiation is the mechanism that explains the null LSE in recognition, which is the typical outcome in recognition. If older adults do not update the traces during repetitions, then trace differentiation will be hampered, producing a positive LSE (and impairing the SBME compared with younger adults). Failure of trace updating could take place because older adults may be less likely to be reminded of the previous occurrence of an item during its repetition and instead store a new memory trace. Research suggests that repetition itself is not the underlying factor for determining trace updating. Rather, when an item is remembered as having occurred in a specific context, then only the trace gets updated (Criss, Malmberg, & Shiffrin, 2011). Wahlheim (2014) showed that older adults were less likely to detect changed pairings in stimuli across the lists, which suggests that older adults might be less likely to be reminded of the previous occurrence of the item during its repetition.

If older adults have less differentiated traces, then a positive LSE should be observed in recognition and the SBME should be disrupted—akin to the effects of divided attention in young adults, which produce a positive LSE in recognition and disrupt the SBME (Sahakyan & Malmberg, 2018). However, there are reasons to be skeptical of such outcomes in older adults, based upon Criss, Aue, and Kilic (2014), who found that older adults showed intact SBME, similar in magnitude to that of younger adults. Because the SBME and LSE are explained through the same mechanism of differentiation, Criss et al.'s (2014) study tentatively suggests that older adults may not have impaired differentiation, and therefore in the current study they may show a null or slightly negative LSE, which is the typical finding in the literature with young adults who encode with full attention.

## Method

**Participants.** Participants were 100 University of Illinois at Urbana–Champaign (UIUC) undergraduates (ages 18–25 years) who participated for course credit, and 50 community dwelling older adults ages 62–75 years, who were paid \$10 for participating. Older adults were community members in the Greensboro, North Carolina, area who were recruited from the Adult Cognition Lab at the University of North Carolina at Greensboro (UNCG). No participant reported a history of dementia, stroke, or seizure, and none took medications that made them feel drowsy. In addition to demographics/health questionnaire, participants completed a vocabulary test, and a processing speed test (participant characteristics are reported in Table 1). Young adults were randomly assigned to full attention (YA) or divided attention (YA-DIV) conditions, with

50 participants in each group, whereas the older adult (OA) group performed the task under full attention. Sample sizes were based on the previous studies on LSE and divided attention from our lab.

**Materials** The stimuli were the same as in Experiment 3 of Sahakyan and Malmberg (2018). These involved 144 unrelated English words, divided into 72 targets and 72 foils, equated on word length and frequency (the details about the word characteristics are reported in Sahakyan & Malmberg, 2018). A *pure-strong* list contained 24 items presented twice in a spaced fashion, with seven items in between the two presentations. A *pure-weak* list contained 24 items presented twice, consecutively. A *mixed-list* contained 12 strong and 12 weak items, each presented twice either in spaced or massed fashion. The presentation order of the three lists was fully counterbalanced during encoding, and so was the assignment of items to the strong/weak conditions.

**Procedures** During encoding, items were presented at a rate of 1 s per word, with a 250-ms interstimulus interval, and there was no orienting task during encoding. Participants were told to study the items for an upcoming memory test. After each list, participants completed a 30-s distractor task, during which random letters appeared one at a time on the computer screen, along with a blank space next to the letter, and participants were to input the next sequential letter of the alphabet into the blank space. Afterwards, they completed a yes/no recognition test on each study list. Testing was self-paced, and the targets and foils were randomly intermixed for each participant. Participants who encoded the words with divided attention were asked to monitor random single digits (from 1 to 9) presented auditorily at a rate of 1 s per digit, and to press a button whenever they detected three odd digits in a row. Digit monitoring accuracy and memory accuracy were emphasized as being equally important.

**Design** The design was a List (pure vs. mixed) × Strength (strong vs. weak) × Group (OA vs. YA vs. YA-DIV) mixed factorial design, with group as the only between-subjects factor.

## Results

**Analytic plan** The analyses first report the magnitude of LSE, by analyzing recognition accuracy ( $d'$ ) across the type of list, item strength, and group. Whenever a null LSE is predicted/obtained, Bayes factors are included to assess the evidence for or against the null hypothesis of no List × Strength interaction (using default priors in JASP; JASP Team, 2016). The evidence for a model that includes only the main effects of strength and list is evaluated against an alternative model that additionally includes the List × Strength interaction. The

**Table 1** Participant characteristics in Experiment 1 and Experiment 2

	Young adults		Old adults	
	Mean	SD	Mean	SD
<b>EXPERIMENT 1</b>				
Age, years	19.12	0.84	68.40	4.01
Education, years	13.02	0.69	15.47	2.46*
Vocabulary	30.21	3.97	34.57	3.73*
Processing speed	37.27	7.38	29.14	6.32*
<b>EXPERIMENT 2</b>				
Age, years	18.97	0.78	70.25	4.55
Education, years	13.12	0.87	16.22	1.89*
Vocabulary	15.78	2.89	24.26	6.65*
Processing speed	36.18	7.12	27.84	6.41*

*Note.* Vocabulary was measured by Shipley Institute of Living Scale (Zachary, 1991) in Experiment 1, and by the Advanced Vocabulary Test (Ekstrom, French, & Harman, 1976) in Experiment 2. Processing speed was measured by the Pattern Comparison Test (Salthouse, 1993)

\*Denotes significant age differences at  $p < .001$

Bayes factor  $BF_{10}$  may be interpreted as the ratio of evidence in favor of a model that includes the interaction (implying the presence of LSE) in contrast to a “null” model that includes only the main effects of list and strength (implying a null LSE). In addition, hits and false alarms across pure lists are analyzed to gauge the magnitude of the SBME across three groups. Finally, the relationship between LSE and SBME is evaluated at the level of individual participants, and regression analyses are carried out to assess potential differences in the slopes between the groups.

**Analyses of LSE** The raw hits and false alarms are reported in Table 2. Recognition accuracy ( $d'$ ) was computed after hits and false alarms were transformed using a loglinear correction (Hautus, 1995; Stanislaw & Todorov, 1999). A mixed factorial ANOVA was conducted on  $d'$  scores, using list (pure vs. mixed) and strength (strong vs. weak) as the within-subjects variables, and group (OA vs. YA vs. YA-DIV) as the between-subjects variable. The results are summarized in Fig. 1.

There was a significant main effect of strength,  $F(1, 147) = 59.73, MSE = .164, p < .001$ , indicating that strong items were better recognized than weak items. There was also a significant main effect of group,  $F(2, 147) = 44.82, MSE = 1.46, p < .001$ , indicating that accuracy was significantly worse in the YA-DIV group compared to either the OA group,  $t(98) = 7.35, p < .01, Cohen's d = 1.48$ , or the YA group,  $t(98) = 10.18, p < .01, Cohen's d = 2.04$ , with a nonsignificant difference between the YA group and OA group,  $t(98) = 1.74, p = .08, Cohen's d = .35$ . Importantly, there was a significant three-way interaction, indicating that there were differences in the magnitude of LSE,  $F(2, 147) = 8.67, MSE = .164, p < .001$ . To follow up the interaction, List  $\times$  Strength repeated-measures ANOVAs were used to assess the magnitude of LSE within each group.

In the YA group, there was a significant main effect of strength,  $F(1, 49) = 19.14, MSE = .214, p < .001$ , and a significant interaction with list,  $F(1, 49) = 4.67, p = .04$ . The weak interaction observed between the list and strength was more indicative of a *negative* LSE, as opposed to a positive LSE in the YA group, such that strong items were numerically

but not statistically better recognized on the pure list than on the mixed list,  $t(49) = 1.71, p = .09, Cohen's d = .25$ , whereas weak items were numerically better recognized on the mixed list than on the pure list,  $t < 1, Cohen's d = .04$ . This pattern is consistent with previous reports in the literature (e.g., Ratcliff et al., 1990). Comparing the evidence for a model that includes only the main effects of strength and list against an alternative model that additionally includes the interaction term using Bayesian analyses, yielded  $BF_{10} = 0.41$ , which suggests rather weak evidence for an interaction. It suggests that the data were 2.44 times more likely to be the outcome of a model that does *not* include an interaction (i.e., a null LSE).

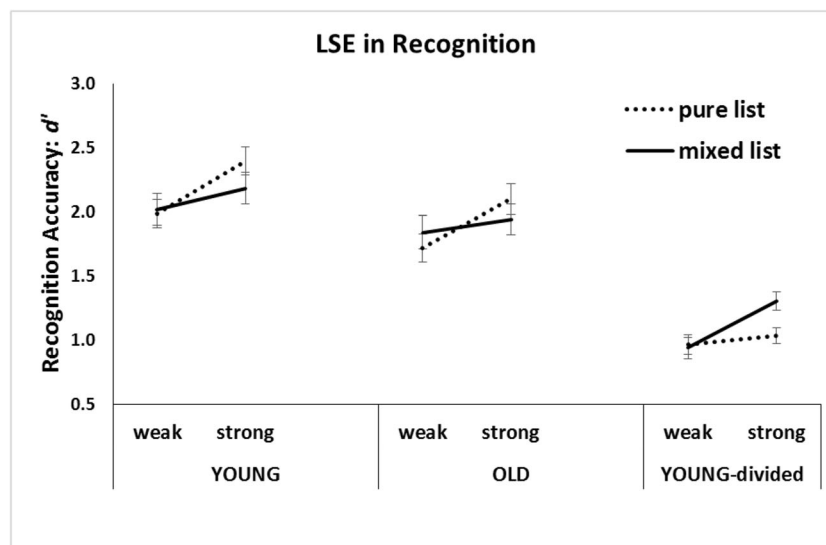
In the OA group, there was also a significant main effect of strength,  $F(1, 49) = 20.78, MSE = .167, p < .001$ , and a significant interaction with list,  $F(1, 49) = 5.54, MSE = .247, p = .008, BF_{10} = 1.1$ . The interaction was again indicative of a *negative* LSE, with strong items being numerically better recognized on the pure list than on the mixed list,  $t(49) = 1.75, p = .09, Cohen's d = .24$ , whereas weak items were numerically better recognized on the mixed list than the pure list,  $t(49) = 1.09, p = .28, Cohen's d = .09$ .

In the YA-DIV group, there was a significant main effect of strength,  $F(1, 49) = 21.16, MSE = .112, p < .001$ , and a significant interaction with list,  $F(1, 49) = 10.71, MSE = .104, p = .002$ . Unlike in the previous two groups, the interaction in the YA-DIV group was indicative of the *positive* LSE. Strong items were better recognized on the mixed list than on the pure list,  $t(49) = 3.46, p = .001, Cohen's d = .58$ , whereas recognition of weak items did not differ between mixed and pure lists,  $t < 1, Cohen's d = .06$ . In the YA-DIV group,  $BF_{10} = 9.79$ , which suggests that the data were 9.79 times more likely to be the outcome of a model that includes an interaction (suggesting the presence of LSE) rather than one without an interaction. The results replicate the previous report in the literature, where divided attention at encoding led to a small positive LSE in young adults (Sahakyan & Malmberg, 2018).

**Analyses of the SBME** As explained earlier, according to the REM model, trace differentiation explains the null LSE in recognition and the SBME. The empirical investigations of

**Table 2** Untransformed hits and false-alarm rates across list, strength, and group in Experiment 1. Numbers in brackets represent *SD*

	Young adults		Old adults		Young divided	
<b>HIT RATES</b>						
	Weak	Strong	Weak	Strong	Weak	Strong
Pure list	.77 (.15)	.83 (.13)	.72 (.15)	.78 (.15)	.56 (.15)	.64 (.12)
Mixed list	.77 (.16)	.82 (.15)	.73 (.15)	.75 (.17)	.54 (.17)	.67 (.14)
<b>FALSE-ALARM RATES</b>						
Pure strong	.11 (.07)		.14 (.14)		.27 (.12)	
Pure weak	.15 (.10)		.17 (.12)		.24 (.15)	
Mixed list	.16 (.16)		.16 (.14)		.24 (.15)	



**Fig. 1** Recognition accuracy as a function of list, strength, and group in Experiment 1. Error bars reflect *SE* of the mean

these two phenomena have been typically carried out in separate studies, but the design of the current experiment allowed examining both effects simultaneously. The SBME was assessed by analyzing the hits and false alarms in pure-strong and pure-weak lists. Therefore, in the YA group and in the OA group, where the typical null/negative LSE was observed, one would expect to observe the SBME, whereas in the YA-DIV group, where a positive LSE was observed, the SBME should be reduced/disrupted. The hits and false alarms across the pure lists are displayed in Fig. 2.

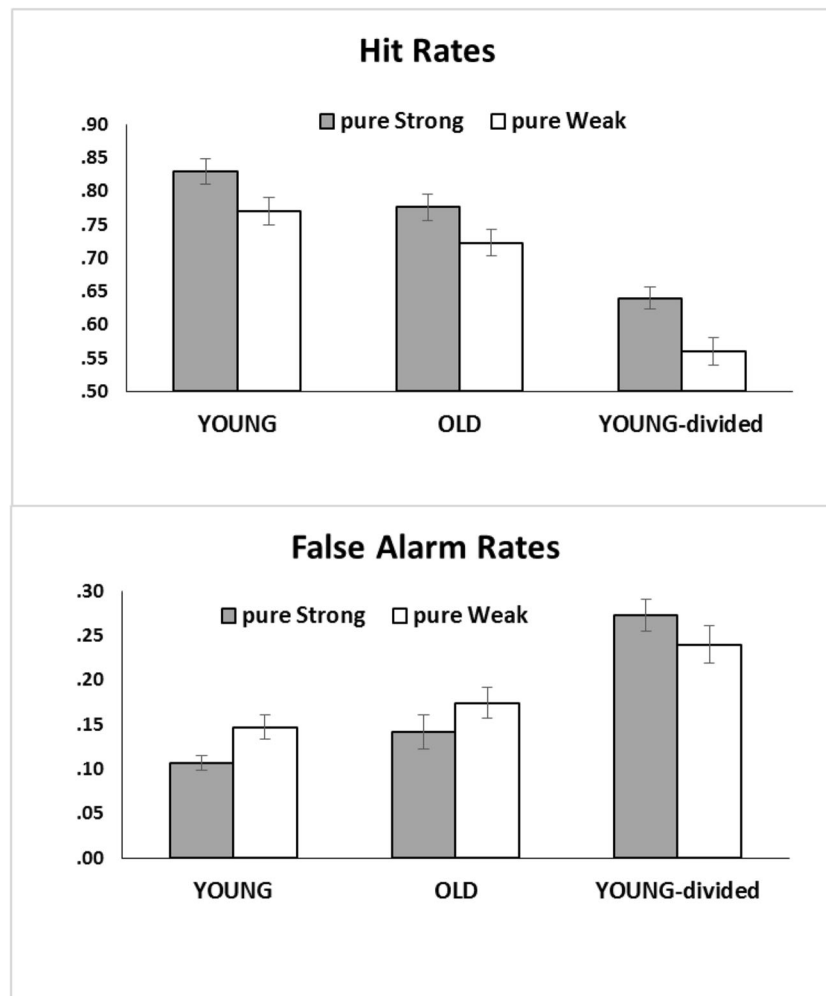
Hits were analyzed using mixed ANOVA, with strength (pure-weak vs. pure-strong) as the within-subjects variable, and group (OA vs. YA vs. YA-DIV) as the between-subjects variable. There was a significant main effect of strength, indicating a higher hit rate for strong items than weak items,  $F(1, 147) = 36.45$ ,  $MSE = .009$ ,  $p < .001$ . There was also a significant main effect of group,  $F(2, 147) = 35.06$ ,  $MSE = .031$ ,  $p < .001$ , indicating that YA group had significantly higher hit rates than OA group,  $t(98) = 2.04$ ,  $p = .04$ , Cohen's  $d = .40$ ,  $BF_{10} = 1.32$ , who in turn had significantly higher hit rates than YA-DIV group,  $t(98) = 5.76$ ,  $p < .001$ , Cohen's  $d = 1.15$ . There was no interaction between these variables,  $F < 1$ .

False alarms were analyzed using the same factors as the hits. There was a significant main effect of group,  $F(2, 147) = 17.91$ ,  $MSE = .025$ ,  $p < .001$ . Overall, false alarms were significantly higher in the YA-DIV group than either the YA group,  $t(98) = 5.97$ ,  $p < .001$ , Cohen's  $d = 1.18$ , or the OA group,  $t(98) = 4.02$ ,  $p < .001$ , Cohen's  $d = .82$ , with OA making overall more false alarms than YA, but the difference was not statistically significant,  $t(98) = 1.27$ ,  $p = .21$ , Cohen's  $d = .25$ . There was also a significant group by strength interaction,  $F(1, 147) = 7.33$ ,  $MSE = .005$ ,  $p = .001$ . It indicated that in the YA group, false alarms were significantly lower on the pure-strong list than pure-weak list,  $t(49) = 3.37$ ,  $p = .001$ ,

Cohen's  $d = .45$ , and the same pattern was found also in the OA group,  $t(49) = 2.20$ ,  $p = .03$ , Cohen's  $d = .26$ . Together with the pattern of hits, this indicates the presence of SBME both in the YA group and in the OA group. In contrast, in the YA-DIV group, false alarms were *higher* on the pure-strong list than on the pure-weak list in the YA-DIV group, although the difference was short of significance,  $t(49) = 1.91$ ,  $p = .055$ , Cohen's  $d = .22$ ,  $BF_{10} = 1.27$ . Thus, in the YA-DIV group, divided attention disrupted the SBME, consistent with the notion of impaired differentiation (Sahakyan & Malmberg, 2018).

**Relationship between LSE and SBME** The purpose of this analysis was to evaluate the relationship between LSE and SBME. The model predicts that the same mechanism gives rise to these empirical outcomes, and therefore one would expect the LSE and SBME to be correlated at the level of individual participants. To evaluate the relationship between the two effects, the empirical pattern of LSE and SBME was first converted into a single value. In the LSE literature, it is common to compute the mix/pure “ratio of ratios” (e.g., Ratcliff et al. 1990; Shiffrin et al., 1990).<sup>2</sup> It is the ratio of the performance ( $d'$ ) of strong/weak items on the mixed list divided by the ratio of strong/weak items on the pure lists. Values greater than 1 signify a positive LSE, a value of 1 denotes a null LSE, and values lower than 1 refer to a negative LSE. The SBME can be captured through a similar “ratio of ratios.” Computing the ratio of HT/FA in the pure-strong list in SBME is analogous to computing the ratio of strong/weak on the mixed list LSE, whereas the ratio of HT/FA in the pure-weak list in SBME is an analog of strong/weak ratio on the pure list in LSE. Thus,

<sup>2</sup> For assumptions underlying the ratio-of-ratio measure, see Ratcliff, Sheu, and Gronlund (1992), Fig. 8.



**Fig. 2** Hits (top panel) and false alarms (bottom panel) as a function of strength and group in Experiment 1. Error bars reflect *SE* of the mean

dividing the ratio in the pure-strong list by the ratio in the pure-weak list in SBME converts the empirical pattern of SBME into a single value, akin to the measure used for LSE. The transformed hits and false alarms were used in computation of the ratio measures. When the LSE-ratio and the SBME-ratio were computed, the resulting distributions were skewed, and therefore the data were log-transformed (the LSE ratio had a skew of 7.07,  $SE = .20$ , and a kurtosis of 59.80,  $SE = .40$ , whereas the SBME ratio had a skew of 2.61,  $SE = .21$ , and a kurtosis of 8.82,  $SE = .41$ ). The positive values on the log-transformed variables indicate a positive LSE and SBME, with zero denoting null effects in both.

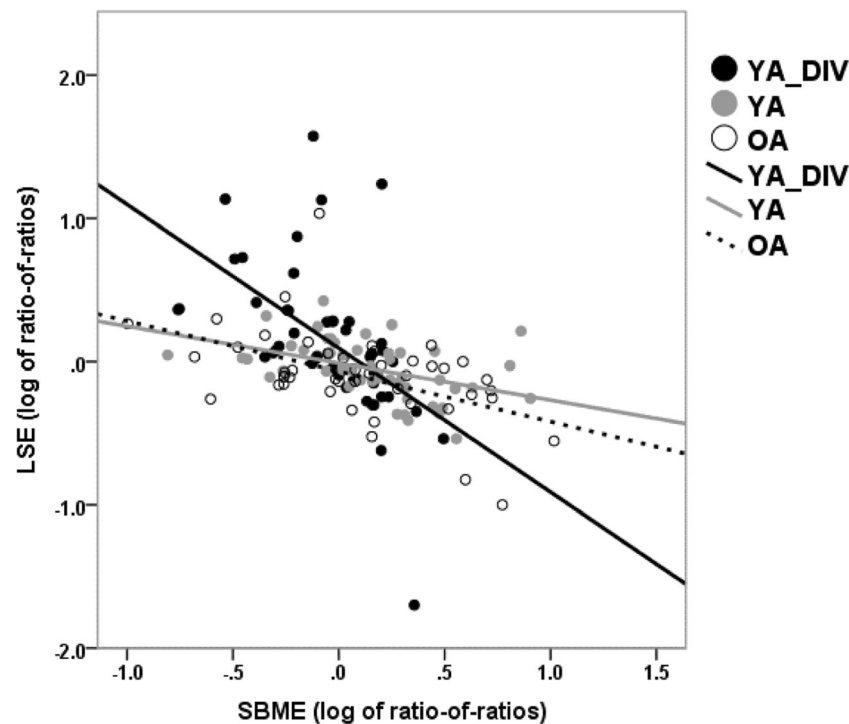
The relationship between LSE and SBME is shown in Fig. 3. As the figure demonstrates, the SBME and LSE were significantly associated in the overall data set,  $r(143) = -.46$ ,  $p < .001$  (a medium-to-large effect size). Thus, participants who showed the positive LSE in recognition tended to be the same participants who showed reduced SBME, and vice versa, enhanced SBME was associated with a null or negative LSE. This relationship was

present in all three groups: in the YA group,  $r(47) = -.44$ ,  $p = .002$ ; in the OA group,  $r(45) = -.49$ ,  $p = .001$ ; in the YA-DIV group,  $r(47) = -.53$ ,  $p < .001$ . To assess whether this effect differed between the groups, linear regressions were computed, with the groups dummy coded. The results showed that the slopes differed significantly between the YA-DIV and OA groups ( $beta = .42$ ,  $p = .001$ ), and between the YA-DIV and YA groups ( $beta = .40$ ,  $p = .001$ ); however, the slopes between the YA and OA did not differ significantly ( $beta = .05$ ,  $p = .59$ ).

## Discussion

The main purpose of the investigation was to compare the magnitude of LSE in OA and YA, and compare it with YA-DIV group. The results in YA and YA-DIV replicate the previous findings of Sahakyan and Malmberg (2018), in which divided attention at encoding produced a small positive LSE in recognition and disrupted the SBME. Both of these effects indicate that differentiation was impaired in young adults under divided





**Fig. 3** Relationship between LSE and SBME across three groups, with linear best fit lines

attention. Against this backdrop, the results in the OA group were more similar to the YA group, and did not mimic the YA-DIV group. Older adults had worse recognition accuracy overall ( $d'$ ) compared with YA with full attention, consistent with the reported findings in the recent meta-analysis (Fraundorf et al., 2018). However, both OA and YA showed similar magnitude of LSE and SBME, suggesting that there was no impaired differentiation in OA in the current study. The different pattern of results between the OA and YA-DIV groups are in line with other research that finds nonequivalence between aging and young adults with divided attention (e.g., Kilb & Naveh-Benjamin, 2007; Naveh-Benjamin et al., 2003; Smyth & Naveh-Benjamin, 2016). Although there were overall recognition differences between young adults and older adults, the results suggest that age-related differences in recognition cannot be solely accounted for by the reduced attentional resources of older adults. The current findings are also consistent with Criss et al. (2014), who found intact SBME between older and younger adults.

Consistent with the notion that the same underlying mechanism explains LSE and SBME, there was a negative association between LSE and SBME at the individual level. Participants who showed the SBME showed a null/negative LSE, whereas the positive LSE in recognition was associated with reduced SBME. Although this relationship was observed in all three groups, the divided attention manipulation made this association even stronger, and it further highlighted the distinction from the OA group, which was similar to the YA group in all reported analyses.

## Experiment 2: Free recall

The purpose of the current experiment is to evaluate LSE in free recall between older adults, younger adults, and younger adults with divided attention. As explained previously, REM explains LSE in free recall through the strength of encoded contextual information, which produces interference at the time of retrieval. Thus, the relative magnitude of LSE in free recall between older adults and younger adults can be informative for whether older adults have deficits in *encoding* of contextual information (i.e., storing fewer context features in their episodic representations). Reduced LSE in older adults suggests impaired encoding of contextual information.

A recent computational account of aging that focused primarily (but not exclusively) on modeling free recall dynamics in aging concluded that older adults' deficits are localized in the retrieval of context, rather than encoding of context (Healey & Kahana, 2016). Therefore, this model would predict that LSE in older adults should not be particularly affected. Another computational account of aging suggests that older adults have a global deficit, combined with representational sparsity of context features compared with young adults (Benjamin, 2010). This model was not designed to explain free recall and speculating about its predictions runs the risk of overinterpretation. Nevertheless, the same encoding assumptions that are built into the model and explain a host of recognition findings should have implications for free recall. Sparse encoding of context should translate into a reduced

LSE in older adults in free recall as long as context is used to probe the retrieval process.

## Method

**Participants** Participants were 80 UIUC undergraduates who participated for course credit, and 36 community dwelling older adults (OA), who were paid \$10 for participating (demographics are reported in Table 1). Older adults were recruited from the Adult Cognition Lab at UNCG. Half of the young adults were randomly assigned to full attention or divided attention conditions, with 40 participants in each group, whereas old adults performed the task under full attention.

**Materials and design** The stimuli were a subset of 72 words, selected from Experiment 1 materials. The presentation format and the counterbalancing conditions were the same as in Experiment 1. The design was List (pure vs. mixed)  $\times$  Strength (strong vs. weak)  $\times$  Group (YA vs. OA vs. YA-DIV) mixed factorial design.

**Procedure** With a few notable changes, the encoding procedures were similar to Experiment 1. The presentation rate of items was 5 s per word. Participants were given an orienting task. Participants were asked to make a yes/no pleasantness judgment during the first presentation of the word, and to make a more nuanced pleasantness judgment (using a 1–5 scale) during the second presentation of the word. Each list was followed by 30 s of the same filler task as in Experiment 1. Afterwards, participants had 90 s to recall as many words as they could remember in any order they wished. During encoding, participants in the YA-DIV group performed the same digit monitoring task as in Experiment 1, except that digits were presented at a rate of 2 s per digit. A slightly slower presentation rate of digits was chosen in this experiment because this rate roughly equated performance between OA and YA-DIV in a cued-recall task in our previous research (Kelley & Sahakyan, 2003).

## Results

LSE in free recall across the three groups is shown in Fig. 4. In the YA group, there was a significant main effect of strength,  $F(1, 39) = 30.53$ ,  $MSE = .020$ ,  $p < .001$ , and a significant interaction with the list,  $F(1, 39) = 13.65$ ,  $MSE = .014$ ,  $p = .001$ ,  $BF_{10} = 30.94$ . Strong items were significantly better recalled in the mixed list than on the pure list,  $t(39) = 3.20$ ,  $p = .003$ , Cohen's  $d = .50$ , whereas weak items were significantly better recalled on the pure list than on the mixed list,  $t(39) = 2.27$ ,  $p = .03$ , Cohen's  $d = .40$ . This pattern confirms the positive LSE in free recall in the YA group.

In the OA group, there was also a significant main effect of strength,  $F(1, 35) = 35.09$ ,  $MSE = .012$ ,  $p < .001$ , and a

significant interaction with the list,  $F(1, 35) = 5.05$ ,  $MSE = .018$ ,  $p = .03$ ,  $BF_{10} = 2.34$ , confirming the presence of the positive LSE in the OA group. Strong items were significantly better recalled on the mixed list than on the pure list,  $t(35) = 2.05$ ,  $p = .048$ , Cohen's  $d = .38$ , whereas weak items were numerically better recalled on the pure list than on the mixed list, but the difference was not significant,  $t(35) = 1.34$ ,  $p = .19$ , Cohen's  $d = .30$ .

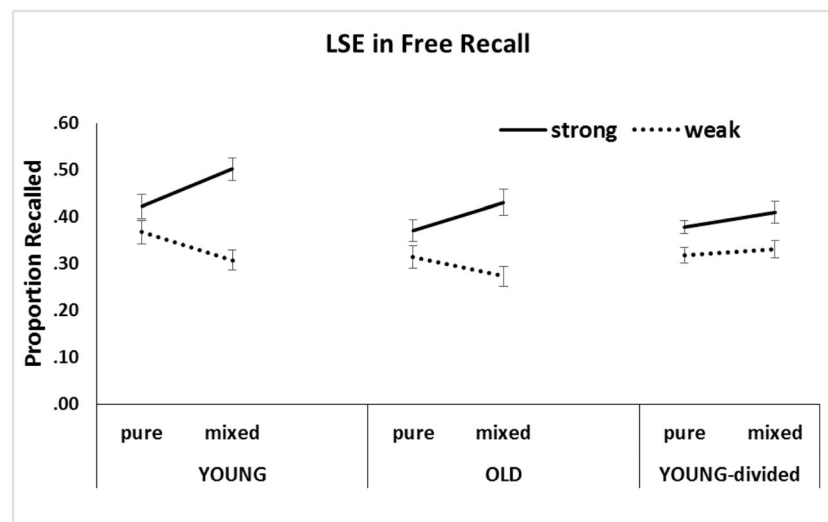
In the YA-DIV group, there was only a main effect of strength,  $F(1, 39) = 23.66$ ,  $MSE = .008$ ,  $p < .001$ , but no interaction with list,  $F < 1$ ,  $BF_{10} = 0.39$  (the data are 2.56 times more likely to be an outcome of a model that does *not* include an interaction). Strong items did not benefit from being on a mixed list compared with a pure list,  $t(39) = 1.35$ ,  $p = .19$ , Cohen's  $d = .26$ , and weak items did not suffer from being on a mixed list compared to a pure list,  $t < 1$ , Cohen's  $d = .12$ . Thus, neither the strengthening component, nor the weakening component were observed.

When YA and YA-DIV groups are included in the List  $\times$  Strength  $\times$  Group analyses, in addition to the main effect of strength, and List  $\times$  Strength interaction, there is a three-way interaction with the group,  $F(1, 78) = 5.70$ ,  $MSE = .013$ ,  $p = .02$ ,  $BF_{10} = 3.63$ , indicating that the size of LSE differed between the two conditions. There was also a significant Strength  $\times$  Group interaction,  $F(1, 78) = 4.26$ ,  $MSE = .014$ ,  $p = .04$ , indicating that the strength effect was more robust in the YA group (12%),  $t(39) = 5.53$ ,  $p < .001$ , than in the YA-DIV group (7%),  $t(39) = 4.87$ ,  $p < .001$ . This could happen if participants occasionally miss noticing repetitions of items due to divided attention task. In contrast, when YA group and OA group are included in the same analyses, only the main effect of strength, and List  $\times$  Strength interaction are significant, but the three-way interaction is not,  $F < 1$ ,  $BF_{10} = 0.09$ , suggesting that the size of LSE did not differ between the OA and YA groups. There is also no Strength  $\times$  Group interaction,  $F < 1$ , indicating the strength effect was of the same magnitude between the YA group and OA group (11%),  $t(365) = 5.92$ ,  $p < .001$ .

When all three groups are included in the same analyses, the three-way interaction was not statistically significant,  $F(2, 113) = 2.63$ ,  $p = .07$ . However, there is a main effect of group,  $F(2, 113) = 3.01$ ,  $MSE = .039$ ,  $p = .05$ . Overall, recall was significantly lower in the OA group than in the YA group,  $t(74) = 2.14$ ,  $p = .04$ , and it did not differ between OA and YA-DIV groups,  $t < 1$ . The overall recall in the YA-DIV group was lower than in the YA group, but it was short of significance,  $t(78) = 1.87$ ,  $p = .06$ .

## Discussion

Experiment 2 results provided yet another dissociation between the OA group and YA-DIV group. Whereas divided attention impaired LSE in free recall, aging did not. The



**Fig. 4** Proportion recalled as a function of list, strength, and group in Experiment 2. Error bars reflect *SE* of the mean

findings in the YA group and YA-DIV group fully replicate Sahakyan and Malmberg (2018). Although repetitions of the items were stored, as indicated by the main effect of strength in both groups, the divided attention manipulation drastically reduced LSE in free recall. These findings support the model that divided attention results in separate traces to be stored for repeated events as opposed to accumulating the information in the original trace (an assumption made by REM for full attention). Divided attention likely disrupted the control processes involved in the retrieval of memories during encoding (cf. Atkinson & Shiffrin, 1968), making the trace updating difficult. There was no evidence that trace accumulation fails to take place in older adults because LSE was approximately of the same magnitude in the OA group and the YA group.

According to SAM/REM models, successful recall performance is contingent on two processes—sampling of the trace, followed by the recovery of information from that trace (Gillund & Shiffrin, 1981; Raaijmakers & Shiffrin, 1980). Sampling is governed by the *relative* strength rule (i.e., the strength of context in the traces of remaining items in memory affect the probability of sampling), whereas recovery is determined by the *absolute* strength rule (i.e., the remaining items in memory play no role, and only the item information contained in the trace of the item matters). The intact LSE in the OA group in the current study suggests that context encoding was not impaired in older adults. What accounts for the main effect of the age group is most likely driven by impaired encoding of item information in the memory trace. Additional evidence for this claim comes from the recognition results in Experiment 1, showing impaired overall recognition in aging compared with young adults with full attention. In the current experiment, it appears that sampling of traces was unaffected in older adults (i.e., intact LSE), whereas the recovery of information was impacted by aging due to encoding fewer item features, explaining the overall reduced recall in old adults.

In contrast, divided attention impaired encoding of both item information and context information. If repetitions do not accumulate in the same trace, and instead a new trace is stored for repetitions under divided attention, then on the mixed list, strong items will be more likely to be sampled due to multiple traces being stored for those items; however, this advantage will be diminished because those traces will be rather weak and their contents will be less recoverable (for additional discussion of the mechanisms of divided attention in free recall, see Sahakyan and Malmberg, 2018).

## General discussion

The present special issue offers an opportunity to celebrate the contribution of Atkinson and Shiffrin's (1968) seminal chapter on the information processing model of memory. The present article sought to highlight the relevance of their contribution by demonstrating how it (and the subsequent models it gave rise to) continue to influence the study of memory processes, including their implications for understanding age-related memory impairments.

The purpose of this investigation was to examine the magnitude of LSE in recognition and free recall among older adults and compare it with young adults who encoded the materials with full or divided attention. The LSE phenomenon has received considerable attention in the modeling literature, and it was the impetus for the development of the REM model. The distinct patterns of LSE across free recall and recognition that are typically observed with young adults with full attention are fully explained by the REM model, and therefore the magnitude of LSE can inform about the underlying processes that might be implicated in aging. Experiment 1 assessed age-related changes in encoding of item information by examining LSE in recognition, whereas Experiment 2

assessed age-related changes in encoding of context information by examining LSE in free recall. In addition, older adults' performance was compared with a group of young adults who encoded with divided attention across both experiments.

Experiment 1 findings showed that OA group demonstrated a null LSE and an intact SBME in single-item recognition, replicating the typical findings observed with young adults with full attention, including the findings obtained in the YA group in Experiment 1. Overall, the pattern of the null LSE and intact SBME suggests that OA in the current study did not have impaired differentiation of item representations. They appeared to have successfully updated memory traces during distributed repetitions, as evidenced by the markers of successful trace updating and differentiation (i.e., null LSE and intact SBME). During repetitions, older adults likely updated/incremented both item features and context features. However, despite storing equivalent amount of context features as YA, OA most likely stored fewer item features overall compared with YA. Several findings support these claims. In Experiment 2, older adults showed intact LSE in free recall, which is a sign of successful context updating/incrementing. If older adults had stored fewer context features (or failed to update them during repetitions), one would expect a reduced LSE in free recall in OA, but there was no evidence of such pattern in Experiment 2. Furthermore, despite demonstrating LSE patterns that resembled those of YA with full attention, across both experiments, OA had overall worse performance both in free recall and in recognition.

A common mechanism might explain the overall impaired memory of OA across both memory tasks, which has to do with sparse encoding of item features (but not context features, as discussed above). Sparse encoding of item features would produce impaired recognition in OA by reducing the hits and increasing the false alarms, and it would also lower free recall by reducing the probability of recovery of information from episodic trace (note that intact encoding of context features in OA would help with successful sampling of the traces, but sparse encoding of item features would impair recovery of information from successfully sampled traces, impairing overall free recall in OA). Furthermore, intact encoding of context information in OA would produce similar magnitude of LSE in free recall, whereas sparse encoding of item features would explain why LSE was obtained at lower levels of overall free-recall performance. In addition, sparse encoding of item features could also explain overall impaired recognition in the OA group.

In Experiment 1, OA had overall worse recognition accuracy compared to YA, and this was driven by OA having significantly lower hit rates and a nonsignificant increase in the false-alarm rates. Sparse encoding of item features would typically increase the false-alarm rates (Malmberg, Zeelenberg, & Shiffrin, 2004). Nevertheless, in the current study, the false alarms were not significantly inflated in the

OA group. The nonsignificant increase in the false alarms may be a chance finding in the current study, but it may also be a real finding that is noteworthy because it has implications for the nature of the memory impairment underlying aging. Although OA are known for their high error rates, both in terms of intrusions in free recall and increased false alarms in recognition (e.g., Healey & Kahana, 2016), there are also reports of relatively stable false alarms across increasing age (e.g., Balota, Burgess, Cortese, & Adams, 2002). If false alarms remain relatively constant in light of decreasing hit rates, this would indicate a problem with storage of *accurate* item features. In contrast, as described earlier, an increase in false alarms coupled with a decrease in hit rates would be consistent with sparse encoding of item features, such as storing reduced quantity of item features. Disentangling these two interpretations (i.e., accuracy vs. quantity of features stored) in the current data set is difficult because accuracy tends to interact with feature distinctiveness, such as word-frequency manipulation (Malmberg et al., 2004), and there was no manipulation of word frequency in the current study. A useful avenue for future research might be crossing the manipulation of episodic strength (i.e., repetitions) with manipulation of normative word frequency (i.e., feature distinctiveness). The distinction between how storing accurate features versus amount of stored features might affect recognition highlights the importance of assessing the hits and false alarms separately, as opposed to combining them into a corrected recognition measure, which is a common practice in the literature. The latter not only conflates discrimination accuracy with response bias, but it could also disguise the underlying processes that contribute to recognition differences.

How does intact encoding of context information fit with other research in aging documenting context processing deficits in older adults (Burke & Light, 1981; Chalfonte & Johnson, 1996; Spencer & Raz, 1995)? As mentioned earlier, Healey and Kahana's (2016) computational account of aging pins older adults' deficits in processing of context to their retrieval deficits rather than to their encoding deficits. The results of Experiment 2 are consistent with their claim. An established line of work shows that older adults are impaired in their ability to accurately remember the associations between components of the episode, known broadly as the associative binding deficits (e.g., Naveh-Benjamin, 2000). Although prior research assumed that aging impairs formation of associations between all stimuli equally, recent research suggests that older adult's deficits are larger for item-to-item associations compared with item-to-context associations (Overman, McCormich-Huhn, Dennis, Salerno, & Giglio, 2018). Thus, Experiment 2 results are consistent with the relatively spared encoding of item-to-context information (at least among "young-old").

Across both experiments, OA did not resemble YA-DIV participants, consistent with some reports in the literature in

which divided attention at encoding did not mimic performance of OA in associative recognition studies ( Craik, Luo, & Sakuta, 2010; Naveh-Benjamin, Guez, & Marom, 2003; Naveh-Benjamin, Guez, & Shulman, 2004). In Experiment 1, divided attention in the YA-DIV group affected the pattern of false alarms across pure strong and weak lists differently from OA. Interestingly, an opposite dissociation between the YA-DIV and OA groups was reported in a study of associative recognition (Castel & Craik, 2003), where YA-DIV and OA groups both showed reduced hits, but there was a substantial increase in false alarms in the OA group, whereas in YA-DIV, the false alarms remained unchanged. Assessing how a manipulation affects hits and false alarms separately can improve our understanding of the mechanisms producing recognition differences in older adults.

Not only did OA not resemble YA-DIV participants in the current experiments but they also did not mimic the performance of young adults with low working memory capacity that were found in other reports. Our previous lab studies showed that low working memory capacity participants (measured using the span tasks) showed reduced LSE in free recall compared with high working memory participants (Sahakyan et al., 2014). This was not the case in Experiment 2, where OA did not differ from YA in terms of the magnitude of LSE. Given OA's established deficits in working memory (e.g., Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Engle, Kane, & Tuholski, 1999), it may seem unusual that they did not show reduced LSE in free recall. However, there are reports in the literature in which older adults' performance in other cognitive tasks does not necessarily mimic performance of low-span participants, and might even resemble performance of high-span participants (e.g., Naveh-Benjamin et al., 2014).

In modeling age-related changes in memory, Healey and Kahana (2016) demonstrated that many prominent theories of aging can explain singular effects quite well, but they all failed when attempting to account for multiple effects simultaneously. Assessing the explanatory power of any theory requires evaluating multiple effects simultaneously and across different tasks, and this includes accounting for LSE in recognition and in free recall. Differentiation is not the only explanation of LSE in recognition. Another class of models, known as context-noise models, assume that recognition is affected by the similarity between the test context and the previous contexts in which the test item appeared (Dennis & Humphreys, 2001). Because context is the only factor that affects recognition decision in these models, and the strength of other items on the list is never compared to the test item, context-noise models make the general prediction that list-composition manipulations, like those used in the LSE paradigm, should produce null effects.

Therefore, they easily account for the typical null LSE in recognition that is observed with full attention. However, divided attention produces a positive LSE, which is a challenge for context-noise models. Importantly, context-noise models are silent about the mechanisms producing LSE in free recall, whereas REM provides a comprehensive account of LSE in free recall and recognition.

In conclusion, the current studies were inspired by the developments in the modeling literature, and the implications of their predictions for memory performance in aging. Computational models of memory have been successful in explaining young adult data, and extending and implementing them on aging can advance our understanding of the nature of their cognitive decline. All models are approximations, and even though the present studies focused exclusively on the REM model, it is important to examine the extent to which it and other models can account for the current findings.

**Acknowledgements** The author thanks UNCG's Adult Cognition Lab directed by Dr. Dayan Touron for the assistance with participant recruitment, and Dr. David Frank and Dr. Megan Jordano for their assistance with the data collection of older participants.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (pp. 89–195). New York: Academic Press.
- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). Memory for the infrequent in young, old, and early stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*, 199–226.
- Benjamin, A. (2010). Representational explanations of “process” dissociations in recognition: The dryad theory of aging and memory judgments. *Psychological Review*, *117*, 1055–1079.
- Buchler, N. E. G., & Reder, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology and Aging*, *22*, 104–121.
- Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of the retrieval processes. *Psychological Bulletin*, *90*, 513–546.
- Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging*, *18*, 873–885.
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*, 403–416.
- Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Lifespan development of visual working memory: When is feature binding difficult? *Developmental Psychology*, *42*, 1089–1102.
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*, 135–150.

- Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 384–420). New York: Van Nostrand Reinhold.
- Craik, F. I. M., Luo, L., & Sakuta, Y. (2010). Effects of aging and divided attention on memory for items and their contexts. *Psychology and Aging, 25*, 968–979.
- Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55*, 461–478.
- Criss, A. H., Aue, W. R., & Kilic, A. (2014). Age and response bias: Evidence from the strength based mirror effect. *Quarterly Journal of Experimental Psychology, 67*, 1910–1924.
- Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory testing. *Journal of Memory and Language, 64*, 316–326.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Kit of factor-referenced cognitive tests*. Princeton: Education Testing Service.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.
- Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2018). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*.
- Gillund, G., & Shiffrin, R. M. (1981). Free recall of complex pictures and abstract words. *Journal of Verbal Learning and Verbal Behavior, 20*, 575–592.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). San Diego, CA: Academic Press.
- Hautus, M. J. (1995). Corrections for extreme proportion and their biasing effects on estimated values of  $d'$ . *Behavior Research Methods, Instruments, & Computers, 27*, 46–51.
- Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123*, 23–69.
- JASP Team (2016). JASP [Computer software] <https://jasp-stats.org>. Accessed 26 Nov 2016.
- Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 3–22.
- Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring and control in attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–721.
- Kilb, A., & Naveh-Benjamin, M. (2007). Paying attention to binding: Further studies assessing the role of reduced attention resources in the associative deficit of older adults. *Memory & Cognition, 35*, 1162–1174.
- Li, S.-C., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging neuromodulation impairs associative binding: A neuro-computational account. *Psychological Science, 16*, 445–450.
- 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.
- Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 540–549.
- McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.
- 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.
- Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition, 31*, 1021–1035.
- Naveh-Benjamin, M., Guez, J., & Shulman, S. (2004). Older adults’ associative deficit in episodic memory: Assessing the role of decline in attentional resources. *Psychonomic Bulletin & Review, 11*, 1067–1073.
- Naveh-Benjamin, M., Kilb, A., Maddox, G., Thomas, J., Fine, H., Chen, T., & Cowan, N. (2014). Older adults don't notice their names: A new twist to a classic attention task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1540–1550.
- Naveh-Benjamin, M., & Mayr, U. (2018). Age-related differences in associative memory: Empirical evidence and theoretical perspectives. *Psychology and Aging, 33*(1), 1–6.
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review, 120*, 356–394.
- Overman, A. A., McCormick-Huhn, J. M., Dennis, N. A., Salerno, J. M., & Giglio, A. P. (2018). Older adults’ associative memory is modified by manner of presentation at encoding and retrieval. *Psychology and Aging, 33*, 82–92.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*, 299–320.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic Press.
- 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.
- Ratcliff, R., Sheu, C.-f., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518–535.
- Sahakyan, L., Abushanab, B., Smith, J. R., & Gray, K. J. (2014). Individual differences in contextual storage: Evidence from the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 873–881.
- Sahakyan, L., & Malmberg, K. (2018). Divided attention during encoding causes separate memory traces to be encoded for repeated events. *Journal of Memory and Language, 101*, 153–161.
- Salthouse, T. A. (1993). Speed mediation of adult age differences in cognition. *Developmental Psychology, 29*, 722–738.
- 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.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145–166.
- Smyth, A. C., & Naveh-Benjamin, M. (2016). Can DRYAD explain age-related associative memory deficits? *Psychology and Aging, 31*, 1–13.
- Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*, 527–539.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31*, 137–149.
- Stephens, J. D. W., & Overman, A. A. (2018). Modeling age differences in effects of pair repetition and proactive interference using a single parameter. *Psychology and Aging, 33*, 182–194.
- Tulving, E., & Hastie, R. (1972). Inhibition effects in intralist repetitions in free recall. *Journal of Experimental Psychology, 92*, 297–304.

- Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging, 18*, 332–339.
- Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition, 42*, 950–964.
- West, R. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*, 272–292.
- Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language, 95*, 78–88.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.
- Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 345–355.
- Zachary, R. A. (1991). *The manual of the Shipley Institute of Living Scale*. Los Angeles: Western Psychological Services.