Psychology and Aging

Adult Age Differences in the Use of Temporal and Semantic Context in Dual-List Free Recall

Christopher N. Wahlheim and Sydney M. Garlitch

Online First Publication, November 14, 2019. http://dx.doi.org/10.1037/pag0000425

CITATION
Wahlheim, C. N., & Garlitch, S. M. (2019, November 14). Adult Age Differences in the Use of Temporal and Semantic Context in Dual-List Free Recall. *Psychology and Aging*. Advance online publication. http://dx.doi.org/10.1037/pag0000425

American Psychological Association

Psychology and Aging

ISSN: 0882-7974

http://dx.doi.org/10.1037/pag0000425

Adult Age Differences in the Use of Temporal and Semantic Context in Dual-List Free Recall

Christopher N. Wahlheim and Sydney M. Garlitch
University of North Carolina at Greensboro

Healthy older adults experience episodic memory deficits when temporal context reinstatement is required, but they also have preserved semantic memory. Semantic associations can therefore support or impair older adults' retrieval from a specific temporal context. The present experiment characterized the roles of pre- and postretrieval processing in age-related memory differences when semantic and temporal contexts worked together or in opposition. Participants studied 2 lists of exemplars from either the same category or different categories and recalled from one list. During recall, participants reported all words that came to mind and made source monitoring judgments. Both groups initiated first retrievals similarly from primacy positions on delayed tests, but older adults initiated first retrievals from later recency positions on immediate tests. Older adults took longer on average to initiate subsequent retrievals, especially when recalling from List 1 and when exemplars from the same category appeared in both lists. Further, trial-level analyses showed that retrieval latencies were longer when fewer responses were produced, and older adults produced fewer responses. When response production was equated, retrieval latencies were more comparable for both age groups. Finally, when lists included exemplars from the same category, older adults produced intrusions earlier and monitored them less effectively on immediate tests, but both age groups showed near-perfect intrusion monitoring when lists included exemplars from different categories. Collectively, these findings show that both pre- and postretrieval processing contributed to age-related recall differences when semantic associations facilitated or opposed reinstatement and monitoring of temporal context.

Keywords: context, free recall, preretrieval processing, semantic associations, source monitoring

Supplemental materials: http://dx.doi.org/10.1037/pag0000425.supp

Healthy older adults experience episodic memory deficits (for reviews, see Balota, Dolan, & Duchek, 2000; Zacks, Hasher, & Li, 2000). These deficits occur primarily in tasks that require self-initiated retrieval (e.g., Craik, 1986), memory for temporal context (for reviews, see Burke & Light, 1981; Spencer & Raz, 1995) and the exclusion of information from irrelevant sources (Hasher & Zacks, 1988; Jacoby, 1999; Wahlheim, Ball, & Richmond, 2017; Wahlheim & Huff, 2015; Wahlheim, Richmond, Huff, & Dobbins, 2016). These studies extended work on age differences in interference effects (e.g., Smith, 1979) by characterizing retrieval dynamics in older and younger adults under conditions of proactive and retroactive interference. The main advantage of this approach is that it enabled inferences about age differences in temporal context processing from analyses of recall initiation (Healey & Kahana, 2016; Unsworth & Engle, 2007), response production, and source monitoring (Kahana, Dolan, Sauder, & Wingfield, 2005). We extend this approach here to examine age differences in both temporal and semantic context processing in free recall.

Research has shown that older adults' impaired temporal context processing contributes to their episodic memory deficits (e.g., Burke & Light, 1981). Older adults' intact semantic memory can alleviate those deficits (e.g., Wingfield, Lindfield, & Kahana, 1998), but it can also lead to associative memory errors when the temporal context associated with those errors cannot be recalled (e.g., Norman & Schacter, 1997). These findings suggest that older adults' memory performance should be facilitated when semantic associations are unique to a temporal context and impaired when semantic associations overlap across temporal contexts. To foreshadow, we demonstrate this pattern in summary scores here, and decompose those scores into measures of recall initiation, response production, and output monitoring. Our primary goal was to use those measures to assess the roles of pre- and postretrieval processing in age differences in temporal and semantic context use. Our general perspective was inspired by the suggestion that those stages of processing are mediated by different cognitive control mechanisms (e.g., Goldsmith, 2016; Jacoby, Shimizu, Velanova,

---

Christopher N. Wahlheim and Sydney M. Garlitch, Department of Psychology, University of North Carolina at Greensboro.

Some of the results and ideas from the present research were presented as at the 13th Biennial Conference of the Society for Applied Research in Memory and Cognition. The materials, data, and analysis scripts are available on the Open Science Framework (https://osf.io/qrn69/).

For their assistance with data collection, we thank Tim Alexander, Carson Peske, Sydney Smith, Crystal Thinzar, and Steven Windsor.

Correspondence concerning this article should be addressed to Christopher N. Wahlheim, Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart Building, P.O. Box 26170, Greensboro, NC 27402. E-mail: cnwahlhe@uncg.edu

& Rhodes, 2005). We expected the present experiment to further illuminate age differences in the use of such mechanisms. Prior to describing the present experiment and hypotheses in more detail, we review relevant literatures on age and individual differences in recall initiation, response production, and source monitoring in free recall.

Recall Initiation

Older adults are presumed to have impaired preretrieval processing that diminishes their ability to elaborate cues to constrain retrieval to a source (e.g., Jacoby et al., 2005). Most studies showing these differences have used behavioral and brain imaging measures in recognition memory paradigms that involve manipulations of preretrieval cueing (for a review, see Morcom, 2016). However, these differences can also be examined by assessing retrieval dynamics in free-recall tasks. One way to examine preretrieval processing in free recall is to compute first-recall probabilities (FRPs), which are serial position curves for first retrievals (Hogan, 1975; Howard & Kahana, 1999; Laming, 1999). When recalling from single lists of unrelated items in standard free recall, FRPs show age-invariant recency-oriented functions on immediate tests and primacy-oriented functions on delayed tests (Golomb, Peelle, Addis, Kahana, & Wingfield, 2008; Healey & Kahana, 2016; Kahana, Howard, Zaromb, & Wingfield, 2002). These findings suggest that preretrieval processing deployed prior to a first retrieval attempt may operate similarly for both age groups when recalling from single lists.

However, one limitation of using standard free recall to examine preretrieval processing is that participants are instructed to withhold incorrect responses, which can mask age differences in FRPs. In contrast to standard free recall, an externalized free recall procedure (EFR; Bousfield & Rosner, 1970; Roediger & Payne, 1985) may be ideal for assessing age differences in recall initiation because it encourages participants to output all responses that come to mind while recalling from a target list. Consistent with this suggestion, Wahlheim et al. (2017) found age differences in FRPs when participants recalled from single lists of unrelated words in a dual-list EFR procedure. Their results showed that on a delayed test (List 1 recall), younger adults initiated recall primarily from List 1 primacy positions, whereas older adults initiated recall primarily from List 2 recency positions (i.e., intrusions). Further, on an immediate test (List 2 recall), younger adults initiated recall from earlier List 2 recency positions than older adults. These findings suggested that older adults' preretrieval processing deficit impaired their ability to reinstate temporal contexts when local sources of interference were present.

The conflicting outcomes from standard and externalized free recall procedures regarding age differences in FRPs indicate that further investigation is needed to illuminate the role of preretrieval processing in first recall initiation. Toward this goal, we extended the approach of Wahlheim et al. by varying semantic and temporal contexts between lists. Based on their findings, we expected the EFR procedure to be sensitive to age differences in FRPs, if they exist. However, the provision of semantic associations within lists could bolster recall initiation for older adults, leading to smaller age differences than observed by Wahlheim et al. We also tested predictions from extant models of free recall to evaluate evidence for a variety of proposed mechanisms. The goal of this nascent approach to examining age-related episodic memory differences (also see, Healey & Kahana, 2016) was to leverage theoretical perspectives from memory models built on younger adult data to characterize processing deficits in older adults.

The two classes of models that we tested were retrieved context models and active control models. Retrieved context models predict age invariant FRPs in standard free recall by making two key assumptions. The first is that temporal context changes at the same rate during study for both age groups (Kahana et al., 2002), which explains age invariant FRPs on immediate tests. The second is that both groups use temporal context similarly to cue their first retrieval attempt (Healey & Kahana, 2016), which explains age invariant FRPs on immediate tests. In contrast, active control models (e.g., Lehman & Malmberg, 2013; Unsworth & Engle, 2007) assume that context reinstatement partly depends on the efficacy of control processes. This assumption leads to the general prediction that recall initiation should differ between older and younger adults if older adults experience deficits in the control processes involved in recall initiation (cf. Braver & West, 2008). However, active control models differ in their predictions about the association between control processes and recall initiation. The buffer model (Lehman & Malmberg, 2013) proposes that people with poorer control should strategically initiate recall across more input positions to improve performance. In contrast, the working memory model (Unsworth & Engle, 2007) proposes that people with poorer control should initiate recall from later recency positions on immediate tests because they hold fewer items in working memory than people with superior control abilities. In the present experiment, we examined the extent to which each of these models could account FRP functions in older and younger adults.

Collectively, the findings and theoretical perspectives described above suggest that FRPs could serve to index age differences in the preretrieval processing used for first-retrieval initiation. However, inferences from FRPs are limited because that measure does not assay the repeated use of preretrieval processing across retrieval attempts. One way to examine the repeated use of preretrieval processing is to measure the latencies between retrievals. To explain potential age differences in preretrieval processing using retrieval latencies, we adopted a perspective from search models of memory (e.g., Raaijmakers & Shiffrin, 1980, 1981; Rohrer, 1996; Shiffrin, 1970). Search models generally assume that retrieval cues can be self-generated to elicit select memory representations referred to as a search set. The models further assume that search set contents are determined by context reinstatement that is governed by preretrieval processing. Better matches between features of generated cues and target memories produce smaller sets that include mostly target representations. Set size can be inferred from latencies between retrievals called interresponse times (IRTs; Murdock & Okada, 1970; Wixted & Rohrer, 1994). Slower IRTs are assumed to indicate larger search sets that are the consequence of noisier context reinstatement, which produces more competition among memory representations at retrieval (e.g., Unsworth, Brewer, & Spillers, 2013; Unsworth & Engle, 2007).

If these assumptions are correct, then IRTs can be used to make inferences about age differences in the use of preretrieval processing to reinstate context and avoid interference after the first retrieval. Based on the assumption that older adults have impaired temporal context processing and intact processing of semantic context (e.g., Wingfield, Lindfield, & Kahana, 1998), IRTs should be disproportionately slower for older adults when they retrieve from a temporal context with semantic associations that are not unique to that context. According to search models, this would suggest that older adults elicit larger search sets with more irrelevant memory representations. In addition, this predicted age difference in IRTs should partly reflect that older adults output the fewest items under such conditions, as IRTs slow more rapidly across retrievals when fewer responses are output (e.g., Murdock & Okada, 1970). These predicted differences would implicate a role for preretrieval processing in age-related deficits in temporal context reinstatement that leads older adults to be more susceptible to semantic interference.

Response Production and Source Monitoring

In addition to affecting the measures of preretrieval processing described above, age-related deficits in preretrieval processing should affect overall response production frequencies, especially when there are local sources of interference. According to inhibition deficit theory (Hasher & Zacks, 1988; Lustig, Hasher, & Zacks, 2007) and dual-process models of memory (Jacoby, 1999; Jennings & Jacoby, 1993), older adults should produce more intrusions than younger adults (albeit through different mechanisms according to each account). Such age-related deficits have been found in several studies (for a review, see Zacks et al., 2000), and these could be a consequence of older adults' noisier context reinstatement prior to retrievals (cf. Unsworth et al., 2013). After intrusions are produced, older adults often monitor the source of those intrusions less effectively than younger adults, which also contributes to older adults' interference susceptibility (for a review, see Dodson, 2017). In fact, a recent context-based model proposes that a postretrieval monitoring deficit is a primary determinant of age-related deficits in free recall (Healey & Kahana, 2016). This proposed combination of processing differences leads to the prediction that not only will older adults produce more intrusions, but they will also reject proportionally fewer of those intrusions than younger adults.

Consistent with this prediction, older adults have been shown to produce more intrusions than younger adults in standard free recall (e.g., Healey & Kahana, 2016; Kahana et al., 2002; Wahlheim & Huff, 2015; Wahlheim et al., 2016). However, the fact that standard free recall encourages covert withholding indicates that this difference could reflect age-related deficits in production, monitoring, or a combination of the two. One way to tease apart these processes is to use the EFR procedure discussed above. The first EFR study to examine age differences included single-list trials with delayed tests (Kahana et al., 2005). Inconsistent with inhibition deficit theory and dual process models, younger adults produced more intrusions than older adults. However, older adults attributed proportionally more intrusions to target lists, indicating a monitoring deficit. Wahlheim et al. (2017) replicated these differences in a dual-list EFR paradigm, showing that older adults produced and rejected proportionally fewer intrusions than younger adults on both immediate and delayed tests.

Together, the results from these EFR studies are inconsistent with the well-established view that preretrieval processing deficits should lead to greater intrusion production for older adults. Given this discrepancy, we sought to determine if the pattern showing greater intrusion production for younger adults and poorer intrusion monitoring for older adults would replicate under conditions similar to Wahlheim et al. We also sought to determine whether including distinct semantic associations in separate temporal contexts would alleviate age-related monitoring deficits by exploiting older adults' intact semantic memory.

The Present Experiment

The overarching goal of the present experiment was to examine the roles of pre- and postretrieval processing in age differences in temporal and semantic context use in free recall. To accomplish this, we used a variant of the dual-list EFR procedure that included exemplars from either the same or different categories in each list. We first assessed age differences in response production and monitoring in summary scores. Next, we examined FRPs to assess age differences in preretrieval processing prior to the first retrieval attempt. We then examined IRTs to assess age differences in preretrieval processing across retrieval attempts. Finally, we examined age differences in intrusion monitoring across the recall period.

Given the similarity to the Wahlheim et al. (2017) paradigm, we expected that younger adults would produce more correct recalls and intrusions than older adults, and that older adults would monitor responses less effectively than younger adults, especially for intrusions. However, we expected that unique semantic associations in each list would bolster memory accuracy and monitoring, especially for older adults because they could exploit their intact semantic memory. Regarding retrieval dynamics, we expected to replicate the general patterns of Wahlheim et al., but we also expected effects of semantic associations. For FRPs, we expected both groups to show functions that are more primacy-oriented for delayed tests and recency-oriented for immediate tests. However, older adults' intact semantic processing may support their preretrieval processing, leading to patterns more similar to younger adults than observed by Wahlheim et al. For IRTs, we expected the most slowing for older adults when semantic associations overlapped between lists. We also expected this difference in aggregate IRTs to be associated with older adults' lower response production (cf. Murdock & Okada, 1970). For intrusion monitoring, we expected to observe more source memory errors for older adults early in the recall period when semantic associations overlapped between lists, extending Wahlheim et al. (2017). It was unclear whether this predicted age difference would vary between immediate and delayed tests, which could occur if older adults are affected differently by proactive and retroactive interference.

Method

Below we report how we determined our sample size, all data exclusions, all manipulations, and all measures (Simmons, Nelson, & Simonsohn, 2012). The materials, data, and analysis scripts can be found on the Open Science Framework (OSF) website: https://osf.io/qrn69/. The research reported here was approved by the Institutional Review Board at The University of North Carolina at Greensboro (UNCG).

Participants

We originally planned to test at least 36 participants per age group, which is equivalent to the largest sample of younger adults that the first author recruited in an earlier study examining adult age differences in dual-list free recall (Wahlheim et al., 2016; Experiment 2). However, given that more than 36 younger adults are often available during a semester, we decided to test as many as possible during that time and to stop at a number that was divisible by 12 (i.e., the number of experimental formats). We tested 60 younger adults in one semester and then matched that sample size for older adults. We conducted a power analysis to determine our ability to detect three-way interactions, which represented the largest number of factors that we included in our analyses. According to G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), our final sample size of 120 people was sufficient to detect a three-way interaction with a medium effect size.

The participants were 60 younger adults (39 women, 21 men), ages 18–29 ($M = 19.42$, $SD = 2.13$) from UNCG, and 60 older adults (42 women, 18 men), ages 65–81 ($M = 70.10$, $SD = 3.80$), from Greensboro and the surrounding areas. For their participation, younger adults received partial course credit, and older adults were paid $10/hr. The experiment lasted approximately 1 hr. The average years of education were significantly greater for older adults ($M = 16.68$, $SD = 2.10$, *range* = 12–19) than younger adults ($M = 12.92$, $SD = 1.37$; *range* 12–18), $t(101) = 11.63$, $p < .001$. The average score on the Shipley Vocabulary Test (Shipley, 1986) was also greater for older adults ($M = 34.89$, $SD = 2.97$, *range* = 28–40) than younger adults ($M = 27.52$, $SD = 4.13$, *range* = 19–38), $t(107) = 11.22$, $p < .001$.

We assessed older adults' cognitive health by administering the Short Blessed Test (SBT; Katzman et al., 1983) over the phone during recruitment, and the Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) in the lab after the experiment. Our criteria for including older adults in the final sample were: an error weighted score $\leq$4 on the SBT, and a score $\geq$25 on the MMSE. Two participants were replaced because they met the SBT criterion but scored <25 on the MMSE (i.e., we tested a total of 62 older adults and excluded two from analyses). Older adults also completed the Digit Symbol Substitution Task (DSST) taken the from the WAIS–R (Wechsler, 1981). Table 1 displays the older adults' cognitive ability scores. All older adults had a normal or corrected normal visual acuity score of 20/50 or better on the Snellen Eye Chart Test (Hetherington, 1954).

Table 1
Cognitive Performance Measures for Older Adults

Task	*M*	*SD*	Range
SBT (weighted errors)	0.47	0.92	0–4
MMSE (out of 30)	28.18	1.23	25–30
DSST (in 90 s)	52.55	10.50	29–71
DSST (out of 9)	6.23	2.29	1–9

Note. SBT = Short Blessed Test (Katzman et al., 1983); MMSE = Mini Mental State Exam (Folstein, Folstein, & McHugh, 1975); DSST = Digit Symbol Substitution (Wechsler, 1981).

Design

The experiment used a 2 (Age: Younger vs. Older) × 2 (Recall: List 1 vs. List 2) × 2 (Category: Same vs. Different) mixed design. Age was treated as a between-subjects variable, and the Recall and Category variables were manipulated within subjects.

Materials

The materials were taken from the VanOverschelde, Rawson, and Dunlosky (2004) category norms. The stimulus set consisted of 20 words that were exemplars from 24 different categories (480 exemplars total). Words were 2–12 letters in length ($M = 6.00$, $SD = 2.00$) and had Hyperspace Analog to Language (HAL) log frequency (Lund & Burgess, 1996) counts of 3.61–12.05 ($M = 8.39$, $SD = 1.63$). Note that HAL log frequency ratings were obtained from the English Lexicon Project database (Balota et al., 2007), which included most (93%) but not all of the words in the material set. The typicality ratings for exemplars within their respective categories according to VanOverschelde et al. (2004) ranged from 1–34 ($M = 11.45$, $SD = 6.71$). The complete material set is available on OSF (https://osf.io/qrn69/).

The Category conditions were created by first dividing each set of 20 exemplars per category into two groups of 10 exemplars. Exemplars were then assigned to groups such that the average HAL log frequency and typicality rating for groups within each category were matched as closely as possible given the constraints of the material set. The counterbalancing of the assignment of categories to conditions was achieved in the following way. For the Same condition, each group of 10 exemplars within a category (e.g., Animals) was assigned to one of the study lists with a trial. For the Different condition, one group from one category (e.g., Fabrics) and one group from another category (e.g., Birds) were each assigned to a study list within a trial. Four sets of six categories (two in the Same condition; four in the Different condition) were randomly assigned to each of four trial blocks (one set per block). The assignment of category sets to blocks remained constant across experimental formats. In contrast, the assignment of categories to lists within Recall conditions was rotated such that each group of 10 exemplars within a category appeared equally often in each list in each Recall condition across participants. This arrangement produced 12 experimental formats.

Procedure

All participants were tested individually with an experimenter present. The experiment consisted of a series of study-test trials that each included two study lists and a recall test for one of those lists (for a schematic of the procedure, see Figure 1). Before the experiment began, participants were told that they would complete several trials of this sort. Participants first completed a short practice trial that included two 5-word study lists and a 30 s recall period where they were asked to recall from List 1. They were instructed to read words in the study lists aloud to study them for an upcoming test. They were also told that they would not know which of the two lists would be the target until after they completed the study phase, so they should study both lists equally well to maximize their recall performance. In addition, they were told that during recall they should attempt to recall words from the target

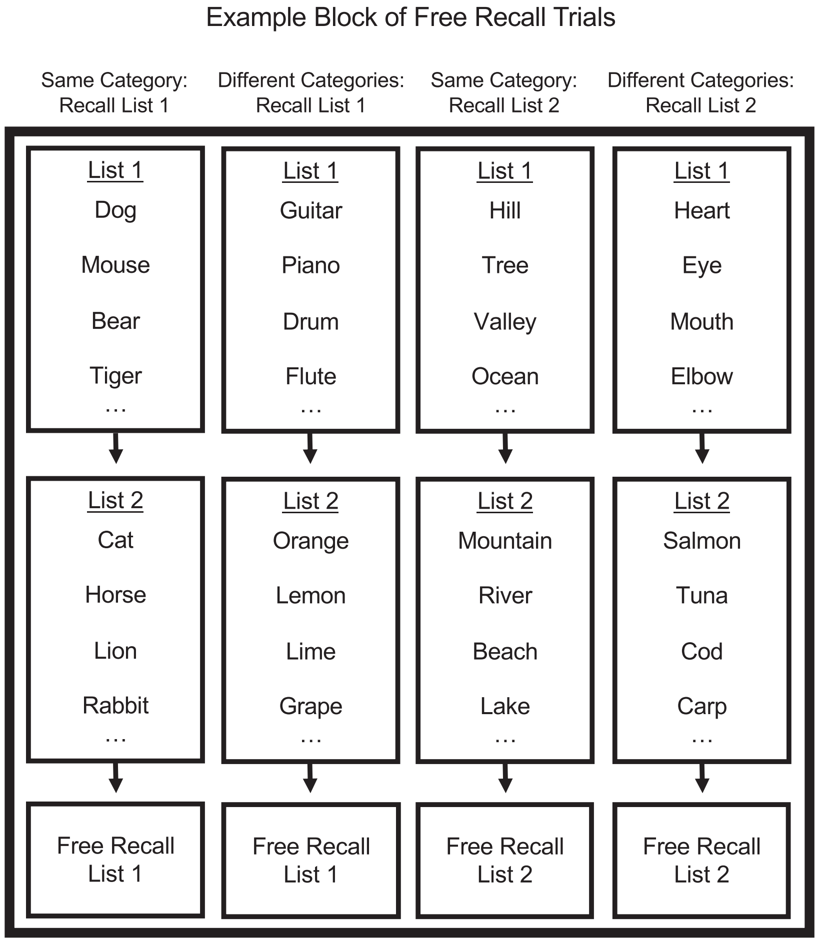

Figure 1. Schematic of the procedure for the free-recall task. Participants completed four blocks of four trials (16 total) that each included one trial from each combination of the category and recall conditions. Note that although the schematic above only includes four exemplars per list, the actual experiment included 10 exemplars per list.

list in any order and that they should type in any other words that came to mind while they did this. Finally, they were told that they would be able to indicate whether each response was from the target list (correct) or from another source (incorrect).

After completing the practice phase, participants started the actual experiment, which consisted of 16 study-test trials that each included two 10-word study lists and a 60 s recall period. The 16 trials comprised four trial blocks that each included one trial from each within-subjects condition. The presentation order for trials within each block was randomized in order to minimize the extent to which participants could predict the associations among exemplars in each list and which list would be the target on the recall test.

In every study phase, the prompts “List 1” and “List 2” appeared individually on the screen for 3 s before each respective study list appeared. Each study item appeared individually for 3 s followed by a 500 ms ISI. After both study lists had appeared, participants were prompted to recall from one of the lists. Either the prompt “Recall from List 1” or “Recall from List 2” appeared for 1.5 s before the recall period began. During the recall period, participants typed their responses onto the screen. After participants pressed “Enter” to submit a response, a prompt appeared indicating which keys to press to classify the response. Participants indicated target list responses by pressing the “1” key, and responses from other sources by pressing the “2” key. After each recall period ended, participants pressed the spacebar to begin the subsequent trial.

After completing the experiment, all participants completed the Shipley vocabulary test (Shipley, 1986). Older adults then completed both the DSST and the MMSE in that order.

Results

All statistical tests were performed using R software (R Development Core Team, 2008). The data were fitted with linear and logistic mixed effects models from the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). The models included experimental factors as fixed effects and subjects and trials as random effects. Hypothesis tests were performed with the Anova function from the car package (Fox & Weisberg, 2011), and post hoc comparisons were performed using Tukey method from the emmeans package (Lenth, 2018). The level for significance was set at α = .05.

Overview of Analysis Plan

The analyses below are organized in the following way. We first compute summary scores for response frequencies and accuracy classifications to assess overall age differences in response production and monitoring. We then decompose summary scores to examine FRPs, IRTs, and output monitoring profiles to examine the contributions of pre- and postretrieval processing to summary scores differences. For IRTs, we first compute functions aggregated across participants. To foreshadow, age differences in IRTs were largest when participants recalled from List 1 and semantic associations overlapped between lists (i.e., List 1-Same condition). We further examine age differences in this most sensitive cell by computing IRT functions for select response frequencies. For output monitoring profiles, we focused primarily on intratrial intrusions because intrusion monitoring differences have direct implications for computational models of age differences in free recall (e.g., Healey & Kahana, 2016). These profiles characterize monitoring dynamics throughout recall by including both response production and accuracy classification frequencies across output positions. Because both age groups correctly rejected nearly every intrusion when lists included exemplars from separate categories (the Different conditions), we compute output monitoring profiles only for intrusions in the Same conditions. We provide more detailed rationales below.

Overall Recall and Accuracy Classifications

We first examined age differences in production and monitoring separately for correct recalls and intratrial intrusions. To do this, we fitted three separate Age × Recall × Category models to each response type. First, we assessed the accessibility of responses by fitting a model to the total number of responses output (Figure 2, total height of bars). Next, we assessed the number of responses that participants would presumably output on a standard free recall test by fitting a model to the total number of responses classified as correct (Figure 2, darker colored bars). Finally, we assessed monitoring accuracy by fitting a model to the proportion of accurately classified responses, calculated by dividing the number of correct recalls classified as correct by the number produced, and

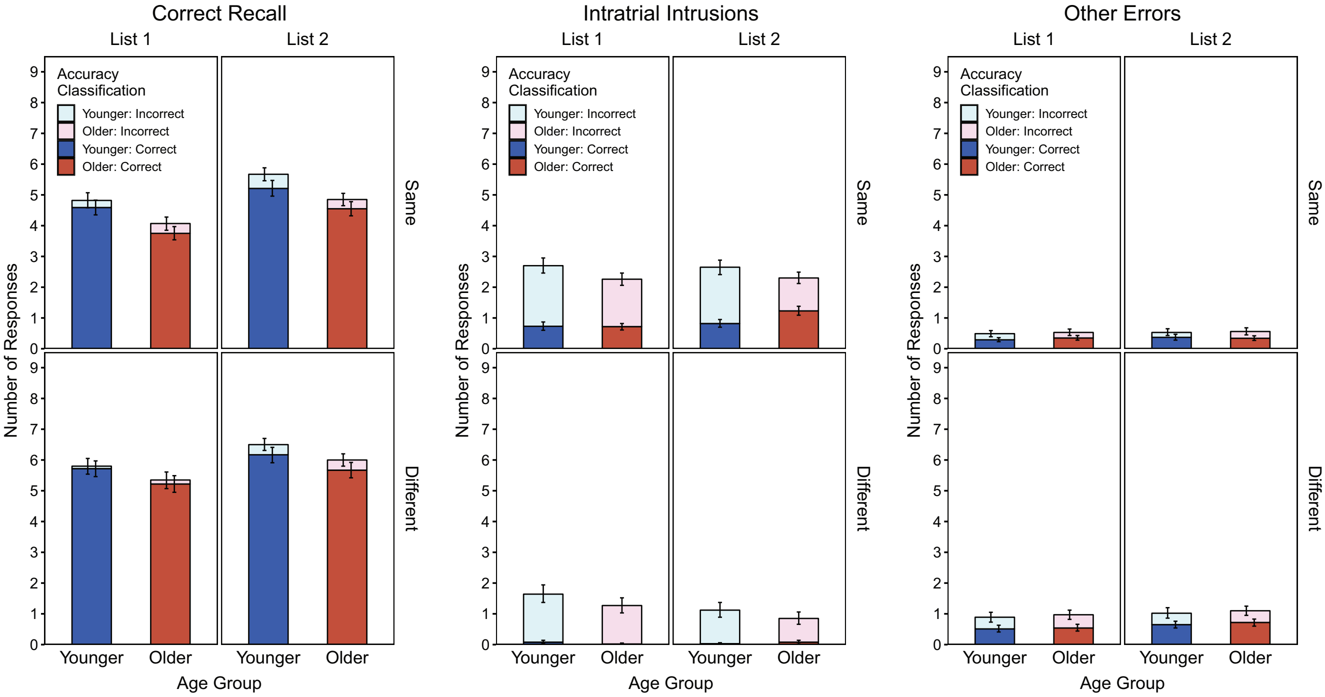

Figure 2. Mean frequencies of correct recall (left panels), intratrial intrusions (middle panels), and other errors (right panels) per trial in each combination of the category and recall conditions for younger and older adults. The total possible number of responses for correct recall and intratrial intrusions is 10. Overall response production is displayed as the total height of each bar (including both the dark-colored and light-colored bars). Responses classified as "correct" are displayed as dark-colored bars, and responses classified as "incorrect" are displayed as light-colored bars. The legend labels for accuracy classifications indicated whether participants responded "correct" or "incorrect." "Correct" accuracy classifications were accurate when participants produced correct recalls and inaccurate when they produced intratrial intrusions or other errors. In contrast, "incorrect" accuracy classifications were accurate when participants produced intratrial intrusions or other errors and inaccurate when participants produced correct recalls. Error bars showing bootstrap 95% confidence intervals are displayed for each type of accuracy classification. See the online article for the color version of this figure.

dividing the number of intrusions classified as incorrect by the number produced. We assumed that higher estimates of accurate classifications indicated better monitoring accuracy. Note that "Correct" accuracy classifications were accurate when participants produced correct recalls and inaccurate when they produced intratrial intrusions or other errors. In contrast, "Incorrect" accuracy classifications were accurate when participants produced intratrial intrusions or other errors and inaccurate when participants produced correct recalls. Note also that we did not model the other error responses collapsed across prior-trial intrusions, extraexperiment intrusions, and repetitions of responses output earlier in the same recall period (Figure 2, right panel) because we were primarily interested in production and monitoring processes for within-trial events.

Figure 2 (left panel) displays the mean number of correct recalls produced per trial. We examined correct recall production by fitting a model to all correct recalls produced. The model indicated significant effects of Age, $\chi^2(1) = 12.19$, $p < .001$, Recall, $\chi^2(1) = 118.79$, $p < .001$, and Category, $\chi^2(1) = 237.37$, $p < .001$, and a significant Age × Category interaction, $\chi^2(1) = 4.95$, $p = .03$. No other effects were significant, *largest* $\chi^2(1) = 1.04$, $p = .31$. The main effects showed that correct recall production was greater in: younger than older adults, the List 2 than List 1 condition, and the Different than Same condition. Further, the interaction showed that correct recall production was significantly greater for younger than older adults in the Same condition, $t(155) = 4.06$, $p < .001$, but was not significantly different between age groups in the Different condition, $t(155) = 2.47$, $p = .07$. This interaction shows that older adults experienced more interference when both lists shared semantic associations.

Next, we examined the number of correct recalls classified as correct by fitting a model to those responses (Figure 2, left panel, darker bars). The model indicated significant effects of Age, $\chi^2(1) = 8.70$, $p = .003$, Recall, $\chi^2(1) = 63.32$, $p < .001$, and Category, $\chi^2(1) = 254.87$, $p < .001$. No other effects were significant, *largest* $\chi^2(1) = 3.39$, $p = .07$. These effects were mostly consistent with the results above in showing more correct recalls classified as such by younger than older adults, in the List 2 than List 1 condition, and in the Different than Same condition. However, the lack of a significant Age × Category interaction, which contrasted with that observed in overall response production, indicated an age difference in monitoring.

To characterize the age difference in monitoring, we modeled the proportion of correct recalls classified as correct. The model indicated significant effects of Recall, $\chi^2(1) = 16.42$, $p < .001$, Category, $\chi^2(1) = 28.39$, $p < .001$, and the following significant interactions: Age × Recall, $\chi^2(1) = 4.18$, $p = .04$, and Recall × Category, $\chi^2(1) = 5.05$, $p = .02$. No other effects were significant, *largest* $\chi^2(1) = 3.39$, $p = .07$. These effects showed that monitoring accuracy was significantly greater in the List 1 than List 2 condition for younger adults, $t(1763) = 4.32$, $p < .001$, but not older adults, $t(1744) = 1.44$, $p = .48$. Monitoring accuracy was also significantly greater when recalling from List 1 compared to List 2 in the Different condition, $t(1756) = 4.44$, $p < .001$, but did not differ in the Same condition, $t(1738) = 1.29$, $p = .57$. These result show that younger adults remembered the source of remote memories better than more recent memories, whereas older adults showed comparable remembering of the source of both remote and recent memories, and that source memory was better when lists included distinct semantic associations.

Figure 2 (middle panel) displays the mean number of intratrial intrusions produced per trial. We examined intratrial intrusion production by fitting a model to all intrusions produced. The model indicated significant effects of Recall, $\chi^2(1) = 15.76$, $p < .001$, and Category, $\chi^2(1) = 418.89$, $p < .001$, that were qualified by a significant Recall × Category interaction, $\chi^2(1) = 12.82$, $p < .001$. No other effects were significant, *largest* $\chi^2(1) = 2.25$, $p = .13$. These effects showed that significantly more intrusions were produced in the List 1 than List 2 condition in the Different condition, $t(1785) = 5.33$, $p < .001$, but did not differ between Recall conditions in the Same condition, $t(1787) = 0.27$, $p = .99$. These results suggest that distinct semantic associations led to fewer intrusions from a remote source.

Next, we examined the number of intratrial intrusions classified as correct by fitting an Age × Recall model to those responses (Figure 2, middle panel, darker bars). Here, we only analyzed data from the Same condition because participants rarely classified intratrial intrusions as correct in the Different condition. The model indicated significant effects of Age, $\chi^2(1) = 4.52$, $p = .03$, and Recall, $\chi^2(1) = 26.08$, $p < .001$, that were qualified by a significant Age × Recall interaction, $\chi^2(1) = 12.04$, $p < .001$. The interaction showed that significantly more intratrial intrusions were inaccurately classified as being from the correct list by older than younger adults in the List 2 condition, $t(230) = 3.66$, $p = .001$, but not in the List 1 condition, $t(230) = .09$, $p = 1.00$, indicating poorer intrusion monitoring for older than younger adults under conditions of proactive interference.

Finally, we examined intratrial intrusion monitoring accuracy in the Same condition by fitting an Age × Recall model to the proportion of responses classified as incorrect. The model indicated significant effects of Age, $\chi^2(1) = 4.88$, $p = .03$, and Recall, $\chi^2(1) = 16.40$, $p < .001$, that were qualified by a significant Age × Recall interaction, $\chi^2(1) = 9.10$, $p = .003$. These results showed that younger adults correctly rejected a significantly greater proportion of intrusions than older adults in the List 2 condition, $t(164) = 3.22$, $p = .009$, but not in the List 1 condition, $t(172) = 0.72$, $p = .89$. Together with the previous analyses, these results confirm that older adults exhibited a monitoring deficit that rendered them more susceptible to proactive interference from semantic associations than younger adults.

The summary scores for correct recall and intratrial intrusions were generally consistent with our predictions. Replicating Wahlheim et al. (2017), younger adults produced more correct recalls and monitored those responses more effectively than older adults. In contrast to Wahlheim et al., intrusion production did not differ between age groups, which may have resulted from the inclusion of semantic associations in each list. However, older adults still did not produce more intrusions than younger adults, which contrasts with predictions from extant accounts of age-related memory deficits (e.g., Hasher & Zacks, 1988; Jacoby, 1999). In addition, when unique semantic associations appeared in each list, both age groups showed higher memory accuracy and better monitoring. This indicated that both groups could leverage intact semantic processing to improve recall accuracy. However, when semantic associations overlapped between lists, older adults were more susceptible to semantic interference when its effects were proactive, as revealed by their increased intrusion monitoring errors when recalling from List 2. In what follows, we examine the contributions of pre- and postretrieval processing to these effects by decomposing summary scores as described in the Introduction.

First Recall Probabilities

We first assessed age differences in preretrieval processing prior to first retrievals by examining FRPs in each Recall condition (see Figure 3). FRPs were computed as the mean number of correct recalls produced on the first retrieval attempt per trial at each input position. We fitted separate Age × Category × Position models to the aggregate FRP data in each Recall condition to independently examine functions for immediate and delayed tests. The model for the List 1 condition (delayed test) indicated a significant effect of Position, $\chi^2(9) = 517.65$, $p < .001$, showing that recall was initiated primarily from the first input position. No other effects were significant, *largest* $\chi^2(9) = 9.91$, $p = .36$. The model for the List 2 condition (immediate test) indicated a significant effect of Position, $\chi^2(9) = 1128.37$, $p < .001$, showing that List 2 functions were characterized by larger recency than primacy effects. In addition, a significant Age × Position interaction, $\chi^2(9) = 34.22$, $p < .001$, showed that primacy effects were larger for younger than older adults, whereas recency effects were larger for older than younger adults. Finally, a significant Category × Position interaction, $\chi^2(9) = 39.83$, $p < .001$, showed that primacy effects were larger in the Different than Same condition, whereas recency effects were larger in the Same than Different condition. No other effects were significant, *largest* $\chi^2(9) = 9.05$, $p = .43$.

Consistent with our predictions, immediate and delayed tests were characterized by recency- and primacy-oriented functions, respectively. One theoretically important finding was that both age groups showed comparable primacy-oriented functions in the List 1 condition. This did not replicate Wahlheim et al. (2017), as older adults were more likely to report recency items from List 2 on first retrieval attempts than younger adults in that study. The present results suggest that the provision of semantic associations within lists supported older adults' preretrieval processing when reinstating a remote temporal context. Moreover, these results are somewhat consistent with context-based models (e.g., Healey & Kahana, 2016; Kahana et al., 2002), which propose that both age groups can reinstate context similarly at the outset of delayed recall tests. Another theoretically important finding was that older adults showed more pronounced recency effects and less pronounced primacy effects than younger adults in the List 2 condition, which replicated Wahlheim et al. These results are consistent with the working memory model (Unsworth & Engle, 2007), which proposes that people with lower working memory capacity, presumably older adults, should initiate retrieval from later recency positions. Finally, we did not predict that both age groups would show larger List 2 primacy effects for Different than Same categories, but this could indicate an effect of

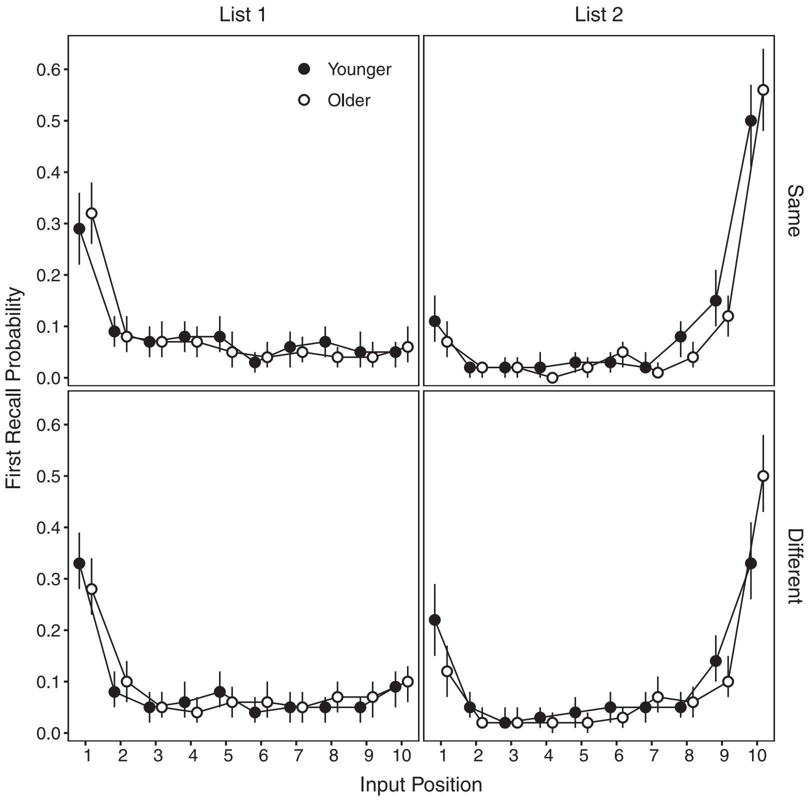

Figure 3. First-recall probability functions displaying the mean frequencies of correct recall in the first output position as a function of input position in each combination of category and recall conditions for younger and older adults. The maximum recall probability is 1.0. Error bars are bootstrap 95% confidence intervals.

conceptual change between lists on attention to List 2 primacy items.

Interresponse Times

We examined age differences in preretrieval processing on retrievals subsequent to the first attempt by computing IRTs. We measured IRTs as the time between the offset of the button press to indicate whether a response was correct or incorrect and the onset of the first key press when entering a subsequent response. We chose this interval to minimize the contribution of the decision time involved in making source monitoring judgments to response latencies. We first examined aggregate IRTs for older and younger adults in each experimental condition. To simplify these comparisons, we only included the first four IRTs in the analyses (see Figure 4). This approach of truncating IRT functions was inspired by earlier work comparing IRTs for higher- and lower-working memory capacity younger adults (e.g., Unsworth & Engle, 2007).

We examined the effects of age and semantic associations on aggregate IRTs by fitting separate Age × Category × Interval models to each recall condition. The model for the List 1 condition indicated significant effects of Age, $\chi^2(1) = 16.92$, $p < .001$, Category, $\chi^2(1) = 80.29$, $p < .001$, and Interval, $\chi^2(3) = 237.46$, $p < .001$, and a significant Age × Category interaction, $\chi^2(1) = 7.64$, $p = .006$. No other effects were significant, largest $\chi^2(3) = 5.30$, $p = .15$. The main effects showed that on average IRTs were slower for older than younger adults, were slower in the same than different condition, and increased across output position intervals. The significant interaction showed that the difference in IRTs between older and younger adults was greater in the Same than Different condition. The model for the List 2 condition indicated significant effects of Age, $\chi^2(1) = 28.76$, $p < .001$, Category, $\chi^2(1) = 18.55$, $p < .001$, and Interval, $\chi^2(3) = 296.37$, $p < .001$. No other effects were significant, *largest* $\chi^2(3) = 3.76$, $p = .29$. Consistent with the List 1 condition results, average IRTs were slower for older than younger adults, were slower in the Same than Different condition, and increased across output position intervals.

One potential explanation for these overall age differences in aggregate IRTs is that an age-related deficit in processing speed (e.g., Salthouse, 1996) led older adults to transition between responses more slowly on all trials. An account of this sort may also assume that older adults' greater susceptibility to semantic interference further slowed their search among competing representations for those trial types. However, we know that older and younger adults show comparable IRT functions when recall transitions occur within categories (e.g., Wingfield et al., 1998), and that IRT functions are steeper when fewer responses are output (Murdock & Okada, 1970). Consequently, an alternative explanation is that when controlling for the number of responses output, the patterns underlying aggregate IRTs may be similar for both age groups, but because older adults produce fewer responses on more trials, they also produce steeper IRT functions on more trials than younger adults. We evaluated these potential explanations by

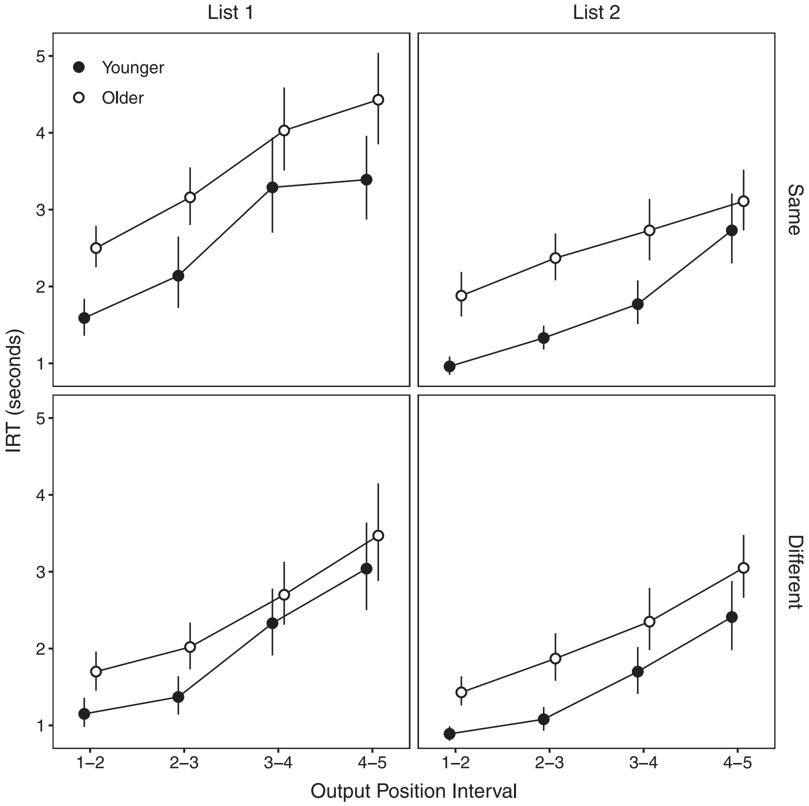

Figure 4. Mean interresponse times (IRTs) in each category and recall condition aggregated over response frequency for younger and older adults. Error bars are bootstrap 95% confidence intervals.

decomposing aggregate IRTs into functions binned by response frequency.

To simplify the presentation of results, we only report trial-level IRT functions in the condition that produced the most robust age difference in aggregate IRTs for a range of response frequencies that each provided a reasonable number of observations. Specifically, we computed IRT functions for both age groups in the List 1-Same condition for response frequencies (bins) 5–10. The trial and participant frequency data appearing in the upper portion in each panel of Figure 5 shows clearly that older adults contributed more observations than younger adults to smaller bins, and this pattern eventually reversed for larger bins. Figure 5 also shows that functions became shallower as bin size increased, replicating earlier findings (e.g., Murdock & Okada, 1970). We compared the shapes of IRT functions for older and younger adults in each bin by fitting separate models including Age and Interval as factors. All models indicated a significant effect Interval, *smallest* $\chi^2(3) = 43.76$, $p < .001$, showing that IRTs slowed across subsequent retrieval attempts. All models also indicated no significant effect of Age, *largest* $\chi^2(1) = 1.57$, $p = .21$. The models for bins 5, 9, and 10 indicated no significant Age × Interval interaction, *largest* $\chi^2(3) = 10.37$, $p = .24$; whereas the models for bins 6–8 indicated significant Age × Interval interactions, *smallest* $\chi^2(4) = 17.73$, $p = .001$, indicating slightly steeper functions for younger than older adults with the age differences occurring at the last interval.

The results from the aggregate IRT analyses show that older adults were slower to transition between responses during the recall period. Consistent with our predictions, this age-related difference was greatest when semantic associations overlapped between lists (the Same condition) and participants were tasked with reinstating a remote temporal context (the List 1 condition). Taken with the findings from the trial-level IRTs, these results indicate that older adults' disproportionately slower aggregate IRTs in the List 1-Same condition resulted from the combination of their lower response production than younger adults and IRTs being steeper when fewer responses were produced. According to search models, this disproportionate slowing suggests that semantic interference led to larger search sets on more trials for older adults, resulting from their noisier context reinstatement (e.g., Unsworth et al., 2013; Unsworth & Engle, 2007). We believe that this finding, coupled with the earlier finding that older adults also produced fewer correct recalls in the List 1-Same condition, suggests that older adults were the most impaired in preretrieval processing in this condition, which was associated with rapidly accelerated slowing across subsequent IRTs.

Output Profiles

In the final analyses, we examined age differences in the production and perceived accuracy of intratrial intrusion errors across the recall period. This approach follows the Wahlheim et al. (2017) study, which showed that older adults' production of intratrial intrusions peaked at earlier output positions, and that older adults rejected proportionally fewer of those intrusions than younger adults. This combination of results was likely the consequence of age-related deficits in preretrieval processing to reinstate context prior to early retrieval attempts and postretrieval processing to evaluate the retrieved context associated with intrusions. Here, we examined whether this pattern would replicate when semantic associations appeared in both lists by computing intrusion output profiles in the Same condition. Note that we did not compute profiles for the Different condition because both age groups showed near perfect monitoring.

Figure 6 displays the frequencies for intratrial intrusions production (top panels) and monitoring errors (bottom panels). We examined age differences on these measures separately for each Recall condition using 2 (Age) × 12 (Position) models. The models for intrusion production indicated no significant effects of Age, *largest* $\chi^2(1) = 1.35$, $p = .25$, significant effects of Position, *smallest* $\chi^2(11) = 344.29$, $p < .001$, and significant Age × Position interactions, *smallest* $\chi^2(11) = 59.82$, $p < .001$, showing that intrusion production peaked earlier in the recall period for older than younger adults. The models for intrusion monitoring errors both indicated significant effects of Position, *smallest* $\chi^2(11) = 101.72$, $p < .001$, and significant Age × Position interactions, *smallest* $\chi^2(11) = 27.84$, $p = .003$, showing that intrusion monitoring errors also peaked earlier for older than younger adults. In addition, the model for List 1 indicated no significant effect of Age, $\chi^2(1) = 0.15$, $p = .70$, whereas the model for List 2 indicated a significant effect of Age, $\chi^2(1) = 11.61$, $p < .001$. These results showed that older adults' monitoring deficit in early positions was greater in the List 2 condition.

These results extend on Wahlheim et al. by showing that age-related deficits in pre- and postretrieval processing led to greater

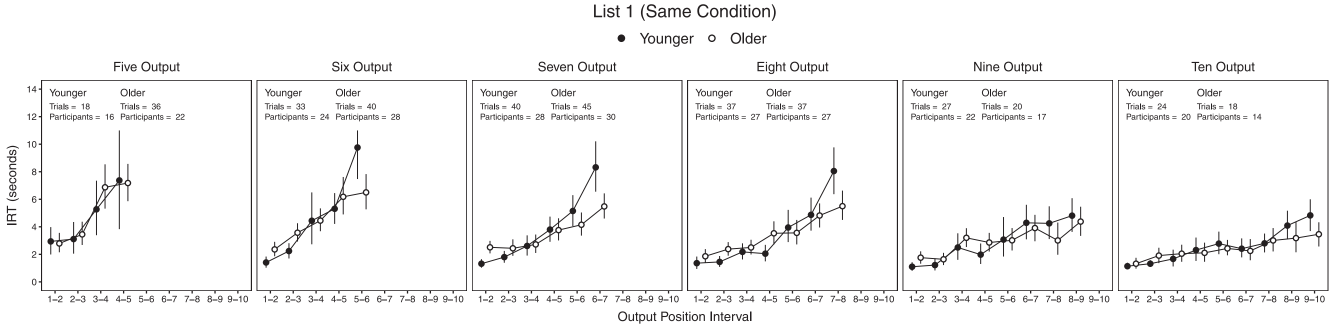

Figure 5. Mean interresponse times (IRTs) in the List 1–Same condition for younger and older adults within response frequency bins. The total number of trials (out of 240) and the number of unique participants contributing to each bin are displayed in the top of each panel. Error bars are bootstrap 95% confidence intervals.

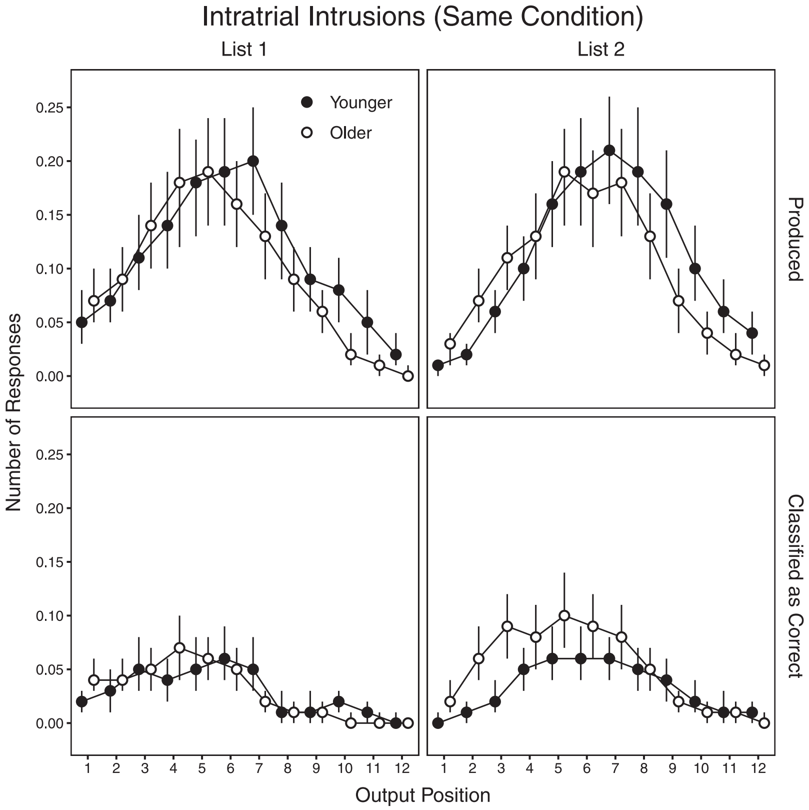

Figure 6. Mean output frequencies of intratrial intrusions produced (top panels) and classified as correct (bottom panels) in each Recall condition for younger and older adults in the Same condition as a function of output position (up to position 12). The maximum output probability is 1.0. Error bars are bootstrap 95% confidence intervals.

semantic interference that was especially prevalent when its effects were proactive. We interpreted older adults' earlier intrusion output as suggesting that their impaired preretrieval processing led to an earlier accrual of poorer-quality context representations that more rapidly impaired their reinstatement of temporal context across subsequent retrievals. This idea is based on the assumption of retrieved-context models that the quality of retrieved context is autocorrelated across retrievals, such that the quality of context retrieved on one attempt determines the efficacy of context reinstatement on the next retrieval attempt and so on (e.g., Healey & Kahana, 2016). Further, older adults' monitoring deficit for early intrusions in the List 1-Same condition may have reflected their reduced reliance on recollected context as a basis for intrusion rejection (e.g., Dodson, Bawa, & Krueger, 2007). We based this on the possibility that early intrusions of semantic associates have characteristics, such as high retrieval fluency, that lead older adults to more often mistake them for correct recalls.

Discussion

The present experiment examined the roles of pre- and postretrieval processing in adult age differences in semantic and temporal context use in dual-list free recall. Younger adults produced more correct recalls and comparable intratrial intrusions compared to older adults, but older adults showed poorer intrusion monitoring when semantic associations created proactive interference. For recall initiation, both groups showed primacy-oriented functions on a delayed test and recency-oriented functions on an immediate test. Both groups showed comparable functions on delayed tests, whereas older adults showed larger recency and smaller primacy effects than younger adults on immediate tests. Aggregate latencies between retrieval attempts (IRTs) were longer for older adults, especially for delayed tests when study lists included overlapping semantic associations. Trial-level of analyses of those IRTs revealed comparable functions for both age groups when response frequencies were equated. IRT functions were steeper when fewer responses were produced, and older adults more frequently produced fewer responses. Intrusion production peaked earlier during recall for older adults, and older adults rejected proportionally fewer intrusions when semantic associations overlapped across lists. Collectively, these results suggest that older adults' susceptibility to proactive interference from semantic associations reflected deficits in pre- and postretrieval processes. In what follows, we consider how the retrieval dynamics reported here relate to prior findings and inform theory.

First Recall Probabilities

We examined age differences in the use of preretrieval processing to initiate first retrieval attempts by comparing FRPs between groups. The results from the List 1 condition replicated results showing age-invariant primacy effects in delayed free recall (Golomb et al., 2008; Kahana et al., 2002), and suggest that both age groups were able to reinstate List 1 context similarly on the first retrieval attempt. This finding is inconsistent with the prediction that age differences in cognitive control should lead to differences in recall initiation from a remote temporal context (cf. Lehman & Malmberg, 2013). This finding is also inconsistent with the results from Wahlheim et al. (2017) showing that older adults were more likely to initiate recall from List 2 recency positions when recalling from List 1 in an EFR procedure. Finally, this finding is inconsistent with the view that older adults have impaired preretrieval processing (for a review, see Morcom, 2016). The comparable FRPs between age groups here could indicate that older adults were able to use semantic associations to bolster their recall initiation.

In contrast to FRPs in the List 1 condition, the patterns of FRPs in the List 2 condition replicated Wahlheim et al. (2017). Older adults produced larger recency effects than younger adults on an immediate test. This finding is inconsistent with studies showing age invariance in retrieval initiation patterns in immediate free recall (e.g., Healey & Kahana, 2016; Kahana et al., 2002), and is more consistent with the view that people with better control abilities, in this case younger adults, should distribute their retrieval initiation more broadly across input positions when retrieving from an immediately preceding temporal context (Unsworth & Engle, 2007). Finally, the finding showing greater primacy for both age groups when lists included different categories suggests that conceptual changes increased the accessibility of the first item in List 2. This effect is reminiscent of studies showing that category changes alleviate proactive interference (e.g., Wickens, 1970) and may reflect an increase in attention to the first items that followed such changes.

These comparisons of recall initiation also have implications for context-based models of age differences in free recall. According to the temporal context framework (Howard & Kahana, 1999; Kahana et al., 2002; Murdock, 1997), FRPs should be comparable for younger and older adults because context changes at the same rate during study. Consistent with this assumption, the age invariance in FRPs for List 1 recall suggests that temporal context changed at the same rate during List 1 study for both age groups. However, the smaller recency effects in FRPs for younger than older adults in the List 2 condition suggest that context changed more slowly during encoding for younger adults. This finding is inconsistent with the predictions of early context-based models that predict slower context change for older than younger adults (see Balota, Duchek, & Paullin, 1989). To further understand the role of context change in these differences, one could fit a recent context-based computational model (context maintenance and retrieval 2 [CMR2]; Healey & Kahana, 2016) to these data to see if the model performs well when the context drift parameter varies freely during List 2 study.

Interresponse Times

We examined age differences in the use of preretrieval processing to initiate subsequent retrievals after the first attempt by computing IRTs. Aggregate IRT functions showed that, on average, older adults transitioned more slowly between retrievals than younger adults. This slowing was also disproportionately greater on delayed tests when lists included overlapping semantic associations. According to a working memory model (Unsworth & Engle, 2007), these results suggest that older adults experienced more semantic interference during memory search. Presumably, this age difference reflected impaired preretrieval processing that was less effective at constraining search sets to target lists. This view is consistent with other perspectives on age-related episodic memory deficits that posit a critical role for preretrieval processing (for a review, see Morcom, 2016). However, as we described above, the interpretation of aggregate IRTs requires further refinement because IRT functions vary systematically with the number of responses output on any given trial (Murdock & Okada, 1970).

To more precisely characterize the observed age differences in aggregate IRTs, we examined IRTs separately within select response bins for the List 1-Same condition. Replicating earlier findings, IRTs slowed more rapidly when fewer responses were output (Murdock & Okada, 1970). Importantly, when response frequency was equated, both age groups showed comparable IRT functions in most bins with the only consistent difference being that younger adults showed more slowing in the last output position interval than older adults. This finding of comparable IRT functions for younger and older adults when recalling exemplars from a category is consistent with earlier work showing age-invariant in IRTs when recalling within category clusters (Wingfield et al., 1998). Collectively, these findings show that older adults' greater slowing when semantic associations overlapped between lists primarily reflected that they produced the fewest responses for those trials. These results suggest that future studies should examine factors that lead to older adults' diminished response production to further understand the role of preretrieval processing in age-related memory deficits. Context-based models may provide guidance for such an endeavor by accounting for factors that lead to age differences in IRTs that are associated with differences in response production frequencies at the time of recall termination (for a similar suggestion, see Miller, Weidemann, & Kahana, 2012).

Output Production and Monitoring

Finally, we used an EFR procedure to characterize the downstream consequences of age differences in preretrieval processing for response production and to examine age differences in the use of postretrieval processing to monitor those responses. The present findings generally replicated earlier EFR studies showing that younger adults produced more correct recalls than older adults (Kahana et al., 2005; Wahlheim et al., 2017). However, intratrial intrusion production did not differ between age groups, even though the trends in the Same condition indicated numerically more intrusions for younger adults. This finding of comparable intrusion production suggests that older adults leveraged their intact processing of semantic associations to produce exemplars from nontarget lists. In addition, the present monitoring results replicated findings showing that older adults endorsed proportionally more intrusions as correct (Kahana et al., 2005; Wahlheim et al., 2017). However, this age difference was only significant in the List 2 condition, suggesting that older adults were more susceptible to monitoring errors when semantic interference had proactive effects. This may have reflected older adults' being more likely to endorse recalls on the basis of nonrecollective information (e.g., Dodson et al., 2007). Finally, both age groups produced fewer intratrial intrusions in the Different than Same condition, presumably due to weaker activation of nontarget categories following successful retrieval from target categories in the Different condition. Both groups also showed near-perfect intrusion monitoring in the Different condition, which presumably reflected older adults' preserved use of semantic associations as a basis for rejecting intrusions.

One potential concern with our interpretation of the age differences in monitoring observed in the List 2 condition is that older adults may have misunderstood the response mapping for those judgments for the first several retrieval attempts. Specifically, older adults may have initially thought that by pressing the "1" key, they were accurately indicating the source of their intrusions (List 1). By this account, this mis-mapping should lead to symmetrical monitoring errors when older adults produced correct responses in the List 2 condition, as they would press the "2" key to indicate the source of correct recalls (List 2), which actually indicated incorrect responses. Contrary to this prediction, both age groups showed highly accurate correct recall monitoring across all output positions in all conditions (see Figure S1). Also contrary to this account, both age groups correctly rejected nearly every intratrial intrusion in the List 2 condition when study lists included different categories. Finally, evidence against this account was shown by Wahlheim et al. (2017) as older adults also rejected proportionally fewer List 2 intrusions when recalling from List 1 than younger adults. This collection of evidence against the mismapping account suggests that although mapping errors could have added noise to the current point estimates, such errors could not fully explain the observed monitoring differences.

More generally, the patterns of production and monitoring observed here have implications for memory models. Search models assume that participants self-generate retrieval cues and monitor retrieved context to evaluate the source of those memories (e.g., Raaijmakers & Shiffrin, 1980, 1981; Shiffrin, 1970). To the extent that cues include features from a target source, they should elicit more memories from target than nontarget contexts, and the retrieved context associated with those memories should better facilitate monitoring. From the perspective of search models, the present results suggest that when only temporal context was diagnostic of list membership (i.e., the Same condition), older adults activated fewer representations from target lists. These results also suggest that the context retrieved by older adults in the Same condition was less diagnostic of source information when intrusions were from a remote source. Further, when semantic associations appeared in distinct temporal contexts (i.e., the Different condition), younger adults activated more target-list memories than older adults, but semantic associations provided comparably diagnostic bases for monitoring decisions for both age groups.

Although search models can reasonably account for the age differences observed here, they cannot explain why both age groups produced comparable intratrial intrusions (cf. Kahana et al., 2005; Wahlheim et al., 2017). This pattern is also inconsistent with prominent theories of cognitive aging. For example, inhibitory deficit theory (Hasher & Zacks, 1988; Lustig et al., 2007) posits that older adults' episodic memory deficits reflect a diminished ability to suppress irrelevant information when retrieving from a specific source. Similarly, dual process theory (e.g., Jennings & Jacoby, 1993; Jacoby, 1999) posits that such deficits partly reflect impaired used of controlled retrieval to constrain access to a source. Both of these views predict that older adults should produce more intratrial intrusions than younger adults, which is inconsistent with the present results. This may suggest that older adults' episodic memory deficit uniformly affects response accessibility in dual-list free recall. However, given that production peaks later for intrusions than correct recalls, EFR procedures may not have provided sufficient time for older adults to finish production. We will eventually investigate this using a self-terminating EFR procedure (cf. Dougherty & Harbison, 2007).

The present findings also test assumptions of CMR2 (Healey & Kahana, 2016). CMR2 assumes that older adults have impaired intrusion monitoring because they set a lower editing threshold than younger adults (also see, Kahana et al., 2005). This could explain why older adults showed impaired intrusion monitoring when semantic context was not diagnostic of list origin. However, this proposed difference in editing thresholds cannot explain the lack of age differences in intrusion monitoring observed in several other conditions. CMR2 also assumes that older adults have a noisier evidence accumulation process that leads them to produce more intrusions than younger adults. However, this prediction is inconsistent with the finding that both age groups produced comparable intratrial intrusions. Based on this latter result, it would be worthwhile to model older and younger adult data from an experiment using an EFR procedure that allows self-termination of recall. Moreover, these discrepancies between model predictions and empirical findings suggest that one challenge for CMR2 descendants will be to explain the interaction of semantic and temporal associations in the production and monitoring of intrusions.

Concluding Remarks

In the present study, we demonstrated that pre- and postretrieval processing contributed to age-related differences in the use of temporal and semantic associations in dual-list free recall. The findings reported here suggest that semantic associations can support such processes and improve recall for older adults when associations appear in distinct temporal contexts. However, semantic associations can also create less favorable retrieval conditions for older adults when associations overlap across temporal contexts. The present study provides a substantive contribution to the cognitive aging literature by characterizing age differences in the retrieval dynamics associated with the interaction of semantic and temporal context use in free recall. These findings have implications for theories of cognitive aging and could serve to inspire the evolution of a recently proposed computational model of age differences in free recall.

References

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F. Craik (Eds.), *Oxford handbook of memory* (pp. 395–410). Oxford, UK: Oxford University Press.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4,* 3–9. http://dx.doi.org/10.1037/0882-7974.4.1.3

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445–459. http://dx.doi.org/10.3758/BF03193014

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48.

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20,* 75–76. http://dx.doi.org/10.3758/BF03335608

Braver, T. S., & West, R. (2008). Working memory, executive control, and aging. In F. I. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed., pp. 311–372). New York, NY: Psychology Press.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90,* 513–514. http://dx.doi.org/10.1037/0033-2909.90.3.513

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities, mechanisms and performance* (pp. 409–422). Amsterdam, the Netherlands: North-Holland.

Dodson, C. S. (2017). Aging and memory. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 403–421). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/B978-0-12-809324-5.21053-5

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory,and high-confidence errors: A misrecollection account. *Psychology and Aging, 22,* 122–133. http://dx.doi.org/10.1037/0882-7974.22.1.122

Dougherty, M., & Harbison, J. (2007). Motivated to retrieve: How often are you willing to go back to the well when the well is dry? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1108–1117. http://dx.doi.org/10.1037/0278-7393.33.6.1108

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE. Available at http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Goldsmith, M. (2016). Metacognitive quality-control processes in memory retrieval and reporting. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 357–385). New York, NY: Oxford University Press.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition, 36,* 947–956. http://dx.doi.org/10.3758/MC.36.5.947

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* (Vol. 22, pp. 193–225). New York, NY: Academic Press.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123,* 23–69. http://dx.doi.org/10.1037/rev0000015

Hetherington, R. (1954). The Snellen chart as a test of visual acuity. *Psychologische Forschung, 24,* 349–357. http://dx.doi.org/10.1007/BF00422033

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition, 3,* 197–209. http://dx.doi.org/10.3758/BF03212898

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 923–941. http://dx.doi.org/10.1037/0278-7393.25.4.923

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. http://dx.doi.org/10.1037/0278-7393.25.1.3

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language, 52,* 493–504. http://dx.doi.org/10.1016/j.jml.2005.01.007

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8,* 283–293. http://dx.doi.org/10.1037/0882-7974.8.2.283

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60,* 92–97. http://dx.doi.org/10.1093/geronb/60.2.P92

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. http://dx.doi.org/10.1037/0278-7393.28.3.530

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *The American Journal of Psychiatry, 140,* 734–739. http://dx.doi.org/10.1176/ajp.140.6.734

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology, 34,* 419–426. http://dx.doi.org/10.1080/002075999399774

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120,* 155–189. http://dx.doi.org/10.1037/a0030851

Lenth, R. (2018). emmeans: Estimated marginal means, aka leastsquares means. R package (Version 1.2.3) [Computer software]. https://CRAN.R-project.org/package=emmeans

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208. http://dx.doi.org/10.3758/BF03204766

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view." *Inhibition in Cognition, 17,* 145–162. http://dx.doi.org/10.1037/11587-008

Miller, J. F., Weidemann, C. T., & Kahana, M. J. (2012). Recall termination in free recall. *Memory & Cognition, 40,* 540–550. http://dx.doi.org/10.3758/s13421-011-0178-9

Morcom, A. M. (2016). Mind over memory: Cuing the aging brain. *Current Directions in Psychological Science, 25,* 143–150. http://dx.doi.org/10.1177/0963721416645536

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839–862. http://dx.doi.org/10.1037/0033-295X.104.4.839

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology, 86,* 263–267. http://dx.doi.org/10.1037/h0029993

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25,* 838–848. http://dx.doi.org/10.3758/BF03211328

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York, NY: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13,* 1–7. http://dx.doi.org/10.3758/BF03198437

Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition, 24,* 188–201. http://dx.doi.org/10.3758/BF03200880

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428. http://dx.doi.org/10.1037/0033-295X.103.3.403

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-521350-9.50017-6

Shipley, W. C. (1986). *Shipley Institute of Living Scale*. Los Angeles, CA: Western Psychological Services.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21-word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26,* 4–7.

Smith, A. D. (1979). The interaction between age and list length in free recall. *Journal of Gerontology, 34,* 381–387. http://dx.doi.org/10.1093/geronj/34.3.381

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. http://dx.doi.org/10.1037/0882-7974.10.4.527

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Focusing the search: Proactive and retroactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1742–1756. http://dx.doi.org/10.1037/a0033743

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. http://dx.doi.org/10.1037/0033-295X.114.1.104

VanOverschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289–335. http://dx.doi.org/10.1016/j.jml.2003.10.003

Wahlheim, C. N., Ball, B. H., & Richmond, L. L. (2017). Adult age differences in production and monitoring in dual-list free recall. *Psychology and Aging, 32,* 338–353. http://dx.doi.org/10.1037/pag0000165

Wahlheim, C. N., & Huff, M. J. (2015). Age differences in the focus of retrieval: Evidence from dual-list free recall. *Psychology and Aging, 30,* 768–780. http://dx.doi.org/10.1037/pag0000049

Wahlheim, C. N., Richmond, L. L., Huff, M. J., & Dobbins, I. G. (2016). Characterizing adult age differences in the initiation and organization of retrieval: A further investigation of retrieval dynamics in dual-list free recall. *Psychology and Aging, 31,* 786–797. http://dx.doi.org/10.1037/pag0000128

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised*. New York, NY: Psychological Corporation.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15. http://dx.doi.org/10.1037/h0028569

Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology and Aging, 13,* 256–266. http://dx.doi.org/10.1037/0882-7974.13.2.256

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. http://dx.doi.org/10.3758/BF03200763

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In T. Salthouse & F. Craik (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Lawrence Erlbaum.

Received June 14, 2019
Revision received October 5, 2019
Accepted October 11, 2019 ■