Aging and false memories: Comparing effects of item-relatedness and list position

Gurguryan, L.[1], Kensinger, E.A. [2], Reuter-Lorenz, P.A. [3], & Dimsdale-Zucker, H.R.[1]

Author details:
[1]Department of Psychology, University of California, Riverside, Riverside, USA
[2]Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA, USA
[3]Department of Psychology, University of Michigan, Ann Arbor, MI, USA

Author emails:
Lauri Gurguryan: laurig@ucr.edu
Elizabeth Kensinger: elizabeth.kensinger@bc.edu
Patricia Reuter-Lorenz: parl@umich.edu
Halle Dimsdale-Zucker: halledz@ucr.edu

Corresponding author:
Lauri Gurguryan
Department of Psychology, University of California Riverside, Riverside, USA
Tel: 323-949-8043
Email: laurig@ucr.edu

Funding details: This work was supported by funding from the Searle Scholars Program and Boston College awarded to EAK.

Keywords: false memory, semantic false memory, phonological false memory, study list position, memory store

Abstract

Semantic false memories are traditionally more frequent from early list positions and thought to arise from presumed long-term memory stores whereas phonological false memories traditionally are more frequent from late list positions and thought to arise from presumed short-term memory stores. Yet, recent research with younger adults challenges this distinction, finding semantic and phonological false memories across memory stores and list positions (Atkins & Reuter-Lorenz, 2008; Dimsdale-Zucker et al., 2018; Flegal et al., 2010). Since aging is understood to differently influence integrity of short- and long-term memory, we compared performance between younger (YAs) and older (OAs) adults using a modified Deese-Roediger-McDermott paradigm to understand whether semantic and phonological memories depend on distinct or shared underlying memory stores. OAs recalled fewer items than YAs, but both groups generated more phonological than semantic false memories. Notably, list position did not influence false memories in either YAs or OAs, suggesting false memory generation is driven by list composition rather than putative memory store. These findings challenge traditional views of memory organization by revealing shared cognitive processes, which transcend traditional memory system distinctions, underlie semantic and phonological false memories.

Introduction

Although colloquially we think of memory as a factual recapitulation of our experiences, extensive laboratory work has shown that, instead, memory is prone to distortions and confabulations (Loftus, 1997). These inaccuracies, which can be measured in the lab as false memories, underscore the dynamic and reconstructive nature of memory (Bartlett, 1993; Schacter, 2012). This fallibility of memory has been extensively studied using the Deese-Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995). The DRM paradigm is a verbal list-learning task designed to elicit false memories (for a review, see Gallo, 2010). In this task, participants study a list of words related to a central theme and later recall those words. A striking finding is that a significant number of participants falsely remember non-presented words related to the theme in meaning or the theme word itself (e.g., if participants studied '*bed*', '*pillow*', '*dream*', '*nap*', they often falsely endorse seeing the word '*sleep*'). While false memories have traditionally been studied as failures of long-term memory (Gallo, 2013), growing evidence demonstrates they can arise after relatively short delays and short lists of memoranda (e.g., Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008; Coane et al., 2007; Dimsdale-Zucker et al., 2018; Festini & Reuter-Lorenz, 2013; Flegal et al., 2010). This raises important questions about the memory systems involved—putative long- and short-term memory stores—in generating false memories. Here, we address this by examining how age-related changes, particularly those that traditionally are theorized to differentially affect long- and short-term memory systems (Brickman & Stern, 2009), can offer new insights into our understanding of false memory as well as the structure of memory itself.

In the canonical DRM paradigm, participants study 12-to-15 words semantically related to a non-presented theme (e.g., words like '*bed*,' '*rest*,' and '*pillow*' are associated with the theme

'*sleep*'), which exceeds the capacity of short-term memory (Miller, 1956). After encoding, participants freely recall the studied items from memory or complete a recognition memory task. Research shows that false memories manifest in both tasks, with participants erroneously recalling or endorsing non-presented words that are semantically related to the theme word (e.g., '*sleep*' in the example above; Deese, 1959; McDermott, 1996; McDermott & Roediger, 1998; Robinson & Roediger, 1997; Roediger & McDermott, 1995; Seamon et al., 1998; for a comprehensive review, refer to Gallo, 2013). These memory errors, or false memories, are thought to arise from the automatic activation of related semantic lexical networks during encoding, whereby studied words activate related but unstudied concepts and words from long-term memory stores. In turn, these activated items interfere during remembering, producing false memories (Roediger et al., 2001). Such memory errors reflect that memory is not a verbatim recording of past events, but, instead, a constructive process influenced by prior knowledge and associations (Roediger & McDermott, 1995; Schacter et al., 2007). Taken together, these findings reinforce the view that false memory is a long-term memory phenomenon.

However, evidence also shows that false memories can occur even with shorter 3-item lists (Festini & Reuter-Lorenz, 2013) or 4-item lists and immediate testing with no retention delay, which fall within the capacity of short-term memory (e.g., Abadie & Camos, 2019; Atkins & Reuter-Lorenz, 2008; Coane et al., 2007; Dimsdale-Zucker et al., 2018; Flegal et al., 2010). For example, Atkins and Reuter-Lorenz (2008) found that participants falsely recognized semantically-related lures after studying 4-item lists and being tested within seconds of encoding. Flegal and colleagues (2010) extended this work by testing false memories under both short- and long-term retrieval conditions. Participants studied 4-item lists and were tested either immediately after encoding each list (i.e., retrieval from short-term memory stores) or after a

delay, once all the lists had been encoded, in a surprise memory test (i.e., retrieval from long-term memory stores). False memories occurred at similar rates in both conditions and participants rated true and false memories as subjectively similar, regardless of whether their memory was tested in the short-term or long-term version of the task. Together, these results suggest that common mechanisms may be at play during the generation of false memories from putative short- and long-term memory stores.

Much of this research, however, often involves different experimental designs to elicit false memories from putative short- versus long-term memory system. To address this, Dimsdale-Zucker and colleagues (2018) used a modified serial position curve paradigm to investigate false memories across short- and long-term memory within a single task and recall window (see also Flegal et al. 2010 for use of a within-subjects design with identical study conditions across immediate and delayed tests). In this paradigm, participants studied 12-item lists with three distinct nested 4-item "sublists" (Glanzer & Cunitz, 1966; Murdock, 1962). Each sublist consisted of four words that were semantically related to a unique theme (e.g., *'hive', 'bumble,' 'sting,' and 'buzz'* are all semantically related to the theme *'bee'*). Critically, the early sublist approximates long-term memory retrieval and most recent sublist approximates short-term memory, allowing false memories from both memory systems to be assessed simultaneously within the same task and individual. False memories occurred across all sublists, although they were less frequent for the most recent sublist, especially when the retention interval was unfilled. These findings support the idea that false memories can emerge from both short- and long-term memory stores.

Additionally, research on false memory has largely focused on semantic false memories, leaving open questions about the role of phonological similarity in generating false memories.

When items are recalled from the beginning of the list (i.e., from putative long-term memory stores), this is thought to be preferentially supported by gist- or meaning-based traces that are more resilient than fragile sensory coding that fades more rapidly (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). A consequence of this is an increased likelihood of semantic, as compared to sound-based or phonological, false memories. In contrast, when items are retrieved from the end of the list (i.e., putative short-term memory stores), phonological or acoustic information is more likely to be accessible (Baddeley & Hitch, 1974), which increases the likelihood of phonological errors (although this is contrasted by a greater likelihood of verbatim traces; Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). Consistent with this distinction, studies demonstrate that items in the primacy position are predominantly susceptible to semantic errors (Craik & Levy, 1970), while items in the recency position are selectively susceptible to acoustic or phonological errors (Craik, 1968). Phonological false memories are also more common when there is a short rather than a long delay between encoding and retrieval (Lim & Goh, 2019; McBride et al., 2019 but see Atkins & Reuter-Lorenz, 2008), likely reflecting retrieval from short-term memory stores (Coane et al., 2024).

However, while this theory would predict that phonological false memories should primarily arise from short-term memory (particularly affecting the most recently studied items), Dimsdale-Zucker and colleagues (2018) found that phonological false memories occurred across all list positions, albeit their occurrence was rare given that study lists were constructed on the basis of semantic similarity and were therefore intentionally biased to elicit semantic false memories. There is extant evidence that, when lists are constructed based on phonological, as compared to semantic similarity, false memories can occur in the phonological domain at similar rates as seen in semantically related lists (e.g., recalling '*mat*' instead of '*cat*' or '*bat*'; (McBride

et al., 2019; Sommers & Lewis, 1999; Watson et al., 2003). These findings suggest the intriguing possibility that a strict interpretation of phonological representations as falling within the purview of short-term memory stores (Baddeley, 1966, 1968) may need to be reconsidered.

Healthy aging offers a valuable lens through which to examine the differential access to verbatim and gist traces, which may underlie phonological and semantic false memories, respectively. The semantic memory system, which is responsible for storing and recollecting general facts and knowledge about the self and the world (Tulving, 1972) remains relatively stable throughout the lifespan (Grady, 2012; Nyberg & Pudas, 2019; Salthouse, 2004). In contrast, the episodic memory system, which is responsible for recalling personal experiences and specific events (Tulving, 1972), becomes increasingly impaired with age. This decline in episodic memory often leads to difficulties in the retrieval of specific contextual information (e.g., retrieving the time, place, or source of information; for a review of memory and aging, see Craik, 2023). As a result, older adults may preferentially rely on gist-based memory traces, drawing on general knowledge and abstracted information from their preserved semantic memory system, rather than on precise, verbatim details from their impaired episodic memory system. Accordingly, research consistently shows that older adults generally recall less than their younger counterparts (Brickman & Stern, 2009; Craik & Salthouse, 2011; Rönnlund et al., 2005) and their recollections tend to be more reliant on gist-based traces rather than perceptual, verbatim traces (Abadie & Camos, 2019; Brainerd et al., 2008; Grilli & Sheldon, 2022; Schacter et al., 1997). Furthermore, this shift toward a greater reliance on gist-based memory processing may also contribute to an increased susceptibility to false memories in older adults, particularly semantic false memories. Given that gist-based memory traces are more abstract and less

detailed, they may create opportunities for older adults to incorporate erroneous information that aligns with their existing general knowledge, leading to semantic distortions.

Consistent with this, research shows that younger adults generally outperform older adults on DRM tasks. Specifically, older adults correctly report fewer studied items and also report a greater number of semantic false memories as compared to younger adults (Balota et al., 1999; Budson et al., 2003; Dehon, 2020; Dennis et al., 2007; Norman & Schacter, 1997; Tun et al., 1998). This increase in false memories in older individuals appears to be resistant to additional control processes—even when cues are provided that inform individuals about the nature of the DRM paradigm, these cues either fail to reduce false memories or only partially mitigate them in older as compared to young adults (McCabe & Smith, 2002).

Additionally, aging is known to be associated with declines in sensory processing, including reductions in vision and hearing, with research showing that the degree of these sensory declines is linked to rates of cognitive decline (Roberts & Allen, 2016; for review, see Monge & Madden, 2016). One well-supported hypothesis is the information degradation hypothesis, which posits that sensory degradation leads to errors during perceptual processing, ultimately resulting in poor performance on higher-order cognitive tasks that may not necessarily be perceptual in nature (Schneider & Pichora-Fuller, 2000). Following this, one possibility is that declines in sensory processing contribute to the higher rates of false memories observed in older adults. Consistent with this, research indicates that older adults experience declines in memory precision, particularly the ability to discriminate between perceptually similar stimuli (Ferguson et al., 1992; Korkki et al., 2020; Stark et al., 2013; Toner et al., 2009; Yassa et al., 2011). This decline in memory discrimination may in turn contribute to difficulties in accurately recalling specific details and correctly distinguishing between similar items.

Taken together, aging provides a unique opportunity to examine how putative short- and long-term memory stores contribute to the generation of semantic and phonological false memories. Given that older adults exhibit preserved semantic but reduced episodic memory, they may rely more heavily on gist-based processing and thus be especially prone to semantic false memories, particularly for items presented early in a list. Alternatively, this shift toward gist-based processing may not reflect preserved semantic function per se but rather may be a consequence of reduced access to item-specific or sensory-based representations, which support verbatim memory and can be used to reject false information (Dennis et al., 2008). Such reductions may stem from weaker initial encoding (Glisky et al., 2001) or difficulties in binding contextual details (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). In contrast, phonological false memories tend to be associated with short-term memory processes and are more prevalent for items presented toward the end of a list where verbatim traces remain the strongest. As such, aging may influence not only the type of false memories (e.g., semantic vs. phonological) but also how these memory errors are distributed across list positions, suggesting that certain types of false memories may be more likely to occur at specific list positions as a function of age. Understanding these patterns can clarify the mechanisms underlying the differential roles of short- and long-term memory in the generation of false memories.

To provide a rigorous test of whether semantic and phonological false memories occur across list positions putatively associated with short- and long-term memory stores, we adapted the paradigm used by Dimsdale-Zucker and colleagues (2018) and constructed lists where words either shared semantic or phonological associations with a non-presented theme (see **Appendix A**). Specifically, for half of the trials, the four words in each sublist were *semantically* related to a non-presented theme word (e.g., '*bed*', '*nap*', '*awake*', '*doze*' for the lure word '*sleep*'). For

the other half of the trials, the four words in each sublist were *phonologically* related to a theme word (e.g., '*leap*, '*steep*, '*bleep*, '*heap* for the lure word '*sleep'*). We operationalized false memories as the recollection of words corresponding to the preselected theme word, or words similar in meaning or sound but not included in the original study list, providing a parsimonious account of false-memory formation consistent with prior work (Atkins & Reuter-Lorenz, 2008; Dimsdale-Zucker et al., 2019; MacDuffie et al., 2012). We hypothesized that this design would allow us to examine false memory generation across list positions both within individuals (i.e., examining differences in semantic vs. phonological false memories across list positions) and between age groups (i.e., comparing performance between younger and older adults). In Experiment 1, younger adults (YAs) and older adults (OAs) studied and later recalled a series of word lists through simultaneous visual and auditory presentation. In Experiment 2, we replicated this in an OA-only sample that studied word lists via visual-only presentation.

Based on extant theories, we hypothesized two possible patterns of results: First, based on accounts that posit memory for early and late study positions is supported by different underlying memory stores (Barnhardt et al., 2006; Brainerd & Reyna, 2002; Deese & Kaufman, 1957; Glanzer & Cunitz, 1966; Murdock, 1962; Reyna & Brainerd, 1995), we anticipated that semantic false memories would be more frequent for items presented in the early list positions—reflecting strong gist traces but relatively weaker verbatim traces. In contrast, we hypothesized that phonological false memories would be more likely for items presented at the end of the list where both gist and verbatim traces are presumed to be strong. Furthermore, we predicted that OAs would exhibit a stronger tendency to generate semantic rather than phonological false memories, even for items presented at later list positions, reflecting age-related declines in memory precision (i.e., relative preservation of gist-level memory representations and decline of

verbatim traces). Alternatively, if representations for early and late sublists instead rely on shared underlying memory stores and processes (Jonides et al., 2008; Nairne, 2002; Öztekin et al., 2010), we instead expected to observe similar rates of semantic and phonological false memories across list positions and age groups. Specifically, we hypothesized that even if there were differences in the number of false memories from different sublist positions (e.g., more veridical and therefore fewer false memories from the most recent positions), that there would be no observable differences in the type of false memories—semantic or phonological—from these different sublists for both YAs and OAs. Altogether, this study aims to contribute to the ongoing discussion about the dual- versus single-store nature of underlying memory stores that can support both successful and erroneous memory performance, thereby providing new insights into our theoretical understanding of false memories and the structure of memory itself.

Methods

Experiment 1

Participants. Native English speakers, free from any reported neurological or psychological disorders, and naïve to the stimulus set were recruited for this study. 48 Boston College undergraduate students (30 = female) participated either for course credit (0.5 credits/half hour) or cash ($5/half hour). All participants completed a battery of neuropsychological assessments that included the Mini-Mental State Examination (MMSE; Folstein et al., 1975), the Shipley Institute of Living Scale (SILS; Shipley, 1986), and the FAS (total, perseverations, and intrusions) to evaluate cognitive function, memory, and verbal fluency. The Digit Symbol Substitution Test, Digits Forward, and Digits Backward tasks were used to assess processing speed and working memory. The Wisconsin Card Sorting Test (WCST) was administered to measure executive function and cognitive flexibility. The Beck Anxiety Inventory (BAI) and

Beck Depression Inventory (BDI; Beck et al., 1996) were used to evaluate levels of anxiety and depression. Given that depression is associated with deficit in episodic memory (for review, see James et al., 2021), six participants were excluded for scores greater than 10 on the BDI (per scores <10 being indicative of mild to moderate depression; Beck et al., 1988), leaving a remaining sample of 42 participants (25 = female; $Mean_{age}$= 19.34 years, SD_{age}= 0.99 years; see **Table 1** for neuropsychological assessment data).

An additional sample of 46 healthy, community dwelling seniors (aged 59 and up) were recruited from the Greater Boston area and compensated with cash ($5/half hour) for their participation. One participant was excluded for scores greater than 10 on the BDI (Beck et al., 1996). An additional 3 participants withdrew themselves following the practice citing task difficulty and personal concerns about memory abilities precluding completion of the task. Results presented below are from the remaining 42 older participants (27 = female; $Mean_{age}$= 76.31 years, SD_{age}= 6.70 years; see **Table 1** for neuropsychological assessment data).

While an a priori sample size calculation was not conducted, the sample size of 42 participants per age group was determined based on precedent from prior work from our group (Dimsdale-Zucker et al., 2019), which successfully detected relevant effects with a comparable design. We additionally conducted a post hoc sensitivity analysis to estimate the smallest effect size detectable with our sample (N = 42). Although such analyses do not substitute for an a priori power analysis, they can help contextualize the study's sensitivity. Using G*Power (Faul et al., 2007), we determined that with 42 subjects per group, six conditions (2×3), $\alpha = .05$, and power = .80, the smallest detectable effect size was Cohen's f = 0.228 ($\eta^2 = .049$), corresponding to a small-to-medium effect. This estimate, based on a conservative one-way ANOVA approximation

rather than the exact repeated-measures design used here, suggests that the study was sufficiently sensitive to detect small-to-medium effects.

Materials. One 12-item list was presented on each of 46 trials (see **Appendix A**). Words were printed on a computer screen while a computerized voice simultaneously read the words aloud. Each list was composed of three sublists where each sublist contained four words that were all related to a common non-presented theme word (see **Appendix B**). On half of the trials, associates were related to the theme semantically (see **Figure 1B**) and for the other half of trials phonologically (see **Figure 1B**). Order of semantic and phonological trials was pulled randomly for each participant. Semantic versus phonological relationship of list items was counterbalanced across participants such that each triplet was only presented once in either the semantic or phonological condition for each participant. Additionally, the order of sublists in each list were counterbalanced so that each theme occurred in each presentation position (primacy, middle, recency) equally often across participants.

Sublists represent a subset of 128 lists taken from previously published work on semantic false memories in STM (Flegal et al., 2010). These lists were supplemented with additional themes taken from neutral associates published in Kensinger and Corkin (2003) and associates were selected from the University of South Florida (USF) Free Association Norms (Nelson et al., 2004) following similar procedures as previous work (Flegal et al., 2010) to yield a total of 138 themes that were organized into 46 possible triplets. Audio files for each word were culled from the Linguistic Data Consortium at UPenn (http://www.ldc.upenn.edu/). Additional stimuli not available from this dataset were constructed in Lips (http://mac-free.com/download/Lips.html) using the “Victoria” voice and further edited in Audacity (http://audacity.sourceforge.net/).

Table 1.

Neuropsychological assessment data

	Experiment 1		Experiment 2
	YA	OA	OA
MMSE	29.43 *(0.83)*	28.76 *(1.36)*	29.16 *(0.71)*
Shipley	32.57 *(2.82)*	35.93 *(3.04)*	36.59 *(3.20)*
Digits symbol	46.79 *(6.80)*	33.21 *(7.67)*	40.50 *(14.97)*
Digits forward	11.57 *(1.99)*	10.88 *(2.37)*	11.23 *(2.37)*
Digits backwards	8.02 *(1.98)*	8.62 *(2.72)*	8.18 *(2.26)*
Digits total	19.63 *(3.23)*	20.12 *(4.72)*	19.30 *(4)*
BAI	6.95 *(6.02)*	4.18 *(4.27)*	3.33 *(3.82)*
BDI	3.05 *(2.37)*	3.43 *(2.83)*	3.15 *(3.19)*
FAS (total)	48.81 *(8.44)*	47.26 *(15.99)*	47.48 *(14.61)*
FAS (perseverations)	1.33 *(0.58)*	1.76 *(2.09)*	2.61 *(2.24)*
FAS (intrusions)	-	-	0.18 *(0.54)*
WCST (categories)	-	5.24 *(1.32)*	5.32 *(1.16)*
WCST (errors)	-	5.45 *(4.57)*	4 *(3.83)*
WCST (perseverations)	-	2.85 *(2.94)*	2.93 *(3.17)*
Mental control	-	24.21 *(5.21)*	26.84 *(6.25)*
Arithmetic	-	14.95 *(3.41)*	14.93 *(4.19)*

Notes. Standard deviation indicated in parentheses; missing FAS data in Experiment 1 was due to data handling errors by the experimenter (HDZ). WCST, Mental Control, and Arithmetic were not administered to YA.

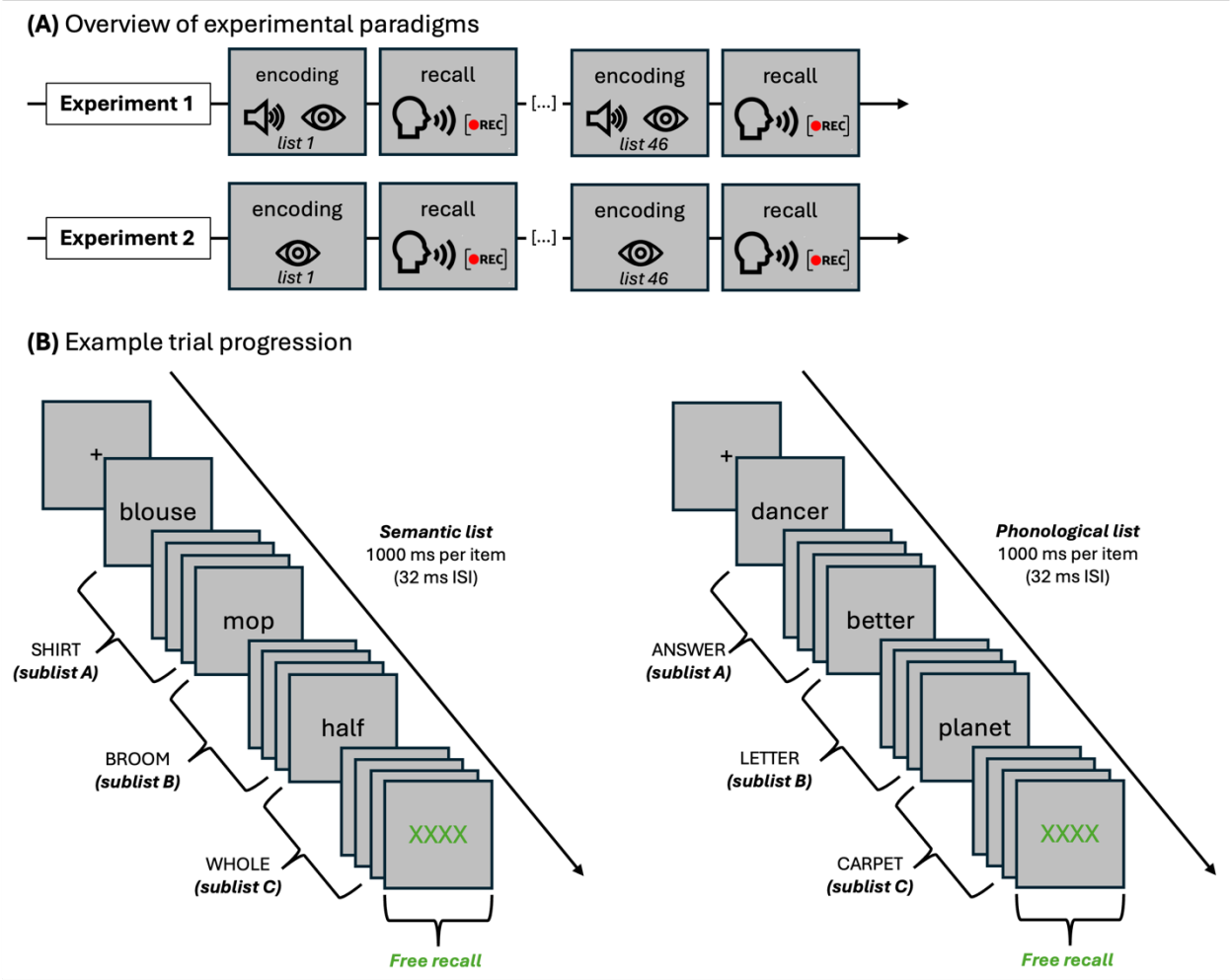

Figure 1. (A) Overview of experimental paradigms used in Experiment 1 and Experiment 2. **(B)** Example trial progression for semantic and phonological lists.

Procedure. Participants gave written consent in accordance with the Boston College IRB. The experimenter explained the memory task and then participants completed five practice trials before proceeding through the remaining 46 task trials. Experimenters remained in the room and transcribed audio responses on laptop computers. Audio responses were additionally recorded using an Olympus WS-210s voice recorder to allow for later analysis. Following the memory task, participants were given feedback about the task, completed neuropsychological tasks, and were debriefed as to the purpose of the study.

Words were presented serially at a rate of 1000 ms per word (after Bartha et al., 1998) with 32 ms between words. Timing and presentation were controlled using ePrime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Stimuli appeared in black Courier New font on a silver background. Presentation of the 12 words was followed by a row of five green "X"s that served as the recall cue and remained the screen for the duration of the free recall. Participants were required to take a minimum of 30 seconds to freely recall the words aloud. Participants pressed the space bar to advance to the next set of words which began 1000 ms later.

Data analysis. In addition to the experimenter's in-room transcriptions, a second experimenter transcribed each audio file. A third coder resolved discrepancies between these transcriptions. Incorrect responses were classified as (1) *semantic:* associates listed in the USAF norms as related to the theme word or judged as related in meaning to the memory set by a trained coder[1], and (2) *phonological:* related in sound to the theme word. All other error types—including within-trial repeats, between-trial repeats, unrelated responses, and non-words or

[1] Although false memory in DRM paradigms is traditionally operationalized using only the critical lure, we also included other semantically related intrusions. This provides a more parsimonious and comprehensive index of false memory by capturing all opportunities for semantic false memories, following the approach used in Atkins & Reuter-Lorenz, 2008; Dimsdale-Zucker et al., 2019; MacDuffie et al., 2012.

mispronunciations—were collapsed into a miscellaneous error category (see Atkins & Reuter-Lorenz, 2008 for similar scoring procedures). As no a priori hypotheses concerned the relation of these miscellaneous errors to sublist or condition, this category was excluded from the primary analyses. Throughout the manuscript, degrees of freedom of the ANOVAs were corrected using the Geisser-Greenhouse method to adjust for violations of sphericity, resulting in non-integer values for the degrees of freedom and Tukey's HSD (Honestly Significant Difference) was used to ensure accurate pairwise comparisons.

Results

Total number of recalled words (combined correct and false)

A repeated measures ANOVA with condition (semantic, phonological) and sublist (A, B, C) entered as a within-subjects factor and age group (YA, OA) specified as a between-subjects factor indicated a significant main effect of age group ($F(1, 82) = 34.66$, $p < 0.001$, $\eta^2_p = 0.30$) such that YAs recalled overall a greater number of words as compared to OAs (see **Table 2** for numeric values). There was a significant main effect of condition ($F(1, 82) = 134.82$, $p < 0.001$, $\eta^2_p = 0.62$) showing that participants recalled a greater number of words in the semantic relative to the phonological condition. The main effect of sublist was also significant ($F(1.63, 133.56) = 297.56$, $p < 0.001$, $\eta^2_p = 0.79$). Participants recalled a greater number of words in Sublist A compared to Sublist B ($t(82) = 2.49$, $p = 0.04$); however, they recalled significantly fewer words in Sublist A ($t(82) = -16.55$, $p < 0.001$) and in Sublist B ($t(82) = -23.83$, $p < 0.001$) compared to Sublist C. This analysis also revealed a significant interaction effect between age group and condition ($F(1, 82) = 35.97$, $p < 0.001$, $\eta^2_p = 0.31$). Both YA ($t(82) = -12.45$, $p < 0.001$) and OA ($t(82) = -3.97$, $p = 0.001$) participants reported a greater number of words in semantic compared to phonological trials, though the magnitude of this effect was stronger in YAs. None of the other

effects—specifically, the interactions between age group and sublist, condition and sublist, nor the three-way interaction between age group, condition, and sublist—reached statistical significance (all p-values > 0.05). See **Figure C1A** in **Appendix C** for an illustration of the data.

Table 2.

Total number of recalled words (correct and false combined; means with standard deviation in parentheses) split by age group, list type, sublist, and experiment.

Experiment 1			
Semantic trials		Younger adults	Older adults
	A	2.32 *(0.59)*	1.53 *(0.69)*
	B	2.25 *(0.52)*	1.27 *(0.58)*
	C	3.46 (0.39)	2.85 *(0.91)*
Phonological trials			
	A	1.73 *(0.57)*	1.32 *(0.69)*
	B	1.73 *(0.40)*	1.22 *(0.53)*
	C	3.08 *(0.45)*	2.66 *(0.88)*
Experiment 2			
Semantic trials			Older adults
	A	-	1.94 *(0.56)*
	B	-	1.14 *(0.45)*
	C	-	2.37 *(0.76)*
Phonological trials			
	A	-	1.47 *(0.54)*
	B	-	0.92 *(0.45)*
	C	-	2.23 *(0.66)*

Note. Only older adults were tested in Experiment 2.

Table 3.

Raw means of correct words and false memories recalled (standard deviation in parentheses) split by age group, list type, sublist, and experiment.

Experiment 1			
Correct responses			
Semantic trials		Younger adults	Older adults
	A	2.14 *(0.55)*	1.36 *(0.67)*
	B	1.98 *(0.51)*	1.15 *(0.57)*
	C	3.39 *(0.41)*	2.78 *(0.91)*
		Phonological trials	
	A	1.55 *(0.51)*	1.13 *(0.63)*
	B	1.56 *(0.36)*	1.05 *(0.49)*
	C	2.91 *(0.47)*	2.49 *(0.87)*
False memories			
Semantic trials		Younger adults	Older adults
	A	0.18 *(0.15)*	0.17 *(0.11)*
	B	0.17 *(0.12)*	0.12 *(0.10)*
	C	0.07 *(0.07)*	0.07 *(0.06)*
Phonological trials			
	A	0.17 *(0.14)*	0.19 *(0.14)*
	B	0.16 *(0.10)*	0.17 *(0.12)*
	C	0.17 *(0.11)*	0.17 *(0.10)*
Experiment 2			
Correct responses			
Semantic trials			Older adults
	A	-	1.77 *(0.57)*
	B	-	1.03 *(0.42)*
	C	-	2.29 *(0.75)*
Phonological trials			
	A	-	1.30 *(0.51)*
	B	-	0.82 *(0.39)*
	C	-	2.09 *(0.63)*
False memories			
Semantic trials			Older adults
	A	-	0.18 *(0.15)*
	B	-	0.11 *(0.08)*
	C	-	0.09 *(0.07)*
Phonological trials			
	A	-	0.18 *(0.11)*
	B	-	0.10 *(0.08)*
	C	-	0.14 *(0.10)*

Note. Only older adults were tested in Experiment 2.

Correct responses (output accuracy)

To account for the overall differences in the total number of items recalled between YAs & OAs, we re-scored the number of correct responses as a proportion of total recalled items. This approach—commonly used when analyzing data across different age groups (e.g., Khanna & Cortese, 2009; McGuire et al., 2015; Sugrue & Hayne, 2006)—enables us to remove age-related differences in overall recall rates. Of note, our calculation reflects the relative balance of correct versus false memories in each recall attempt (i.e., output accuracy) rather than relative to total items presented. This approach enables comparison of the distribution of correct verses false memory responses within participants' output. For completeness, we report raw descriptive statistics of the number of correct and false memory responses without this scoring correction in **Table 3**. Additionally, we computed and report serial position curves (Murdock, 1962) to examine the probability of correct responses as a function presentation position. These curves revealed a recency effect such that there was a greater probability of correct responses from the last list position relative to the first and middle lists (see **Figure 2**); this was replicated with statistical analyses (reported below).

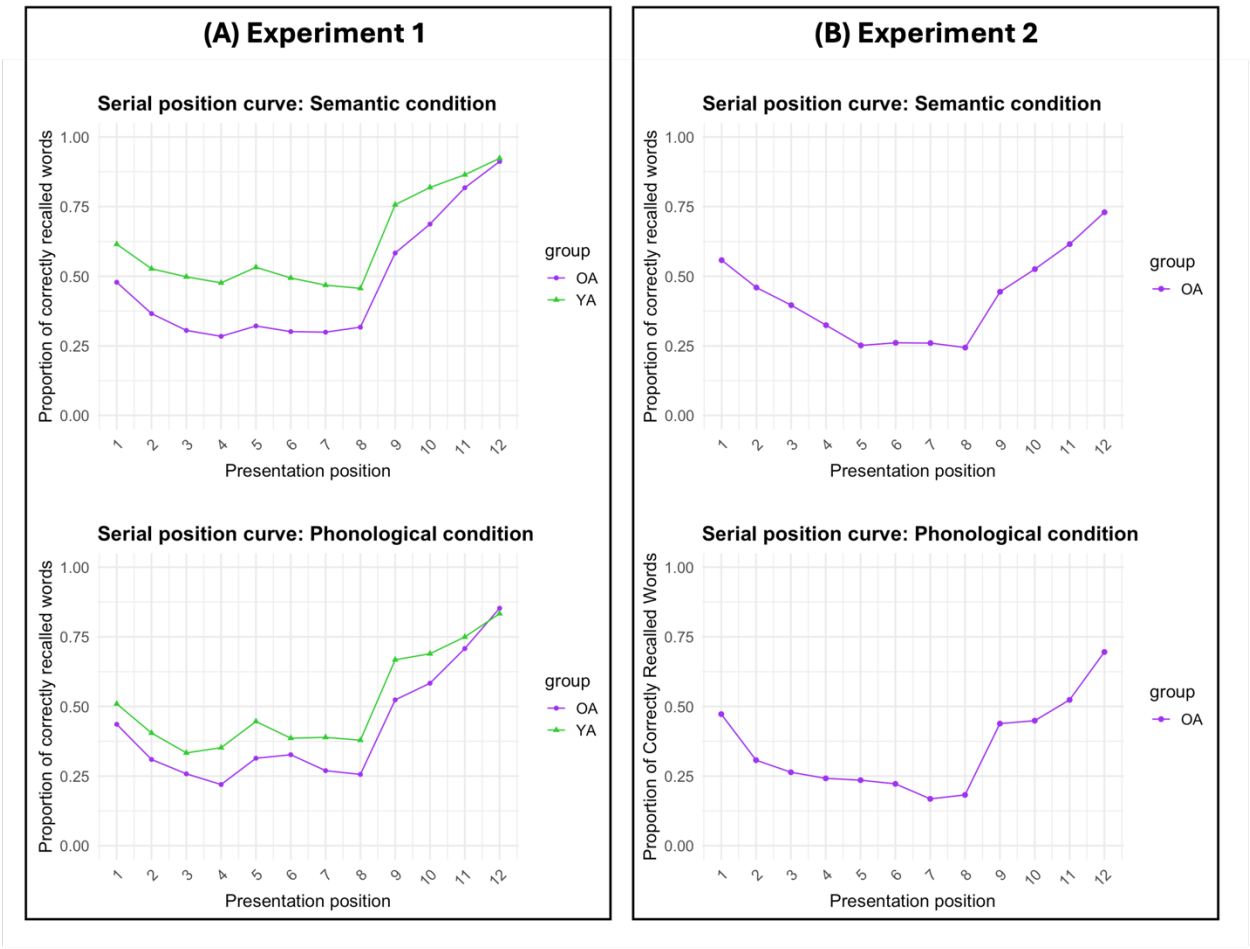

Figure 2. Serial position curves (correct responses) as a function of condition (semantic, phonological) for **(A)** Experiment 1 and **(B)** Experiment 2 (only older adults were tested in Experiment 2).

A repeated measures ANOVA was conducted with condition (semantic, phonological) and sublist (A, B, C) entered as within-subjects factors and age group (YA, OA) as a between-subjects factor. There was a significant main effect of age group ($F(1, 82)= 5.03$, $p = 0.028$, $\eta^2_p = 0.06$) such that YA reported a greater proportion of correct words relative to OA ($t(82) = -2.24$, $p = 0.03$). The main effect of condition was also significant ($F(1, 82) = 21.85$, $p < 0.001$, $\eta^2_p = 0.21$) such that there was a greater proportion of correctly recalled words in the semantic relative to the phonological condition ($t(82) = 4.67$, $p < 0.001$). Lastly, there was a significant main effect of sublist ($F(1.93, 158.26) = 69.62$, $p< 0.001$, $\eta^2_p = 0.46$) indicating significant differences

between sublists A and C ($t(82) = -9.86$, $p < 0.001$) and sublist B and C ($t(82) = -11$, $p < 0.001$) while no significant differences were found between sublists A and B ($t(82) = -0.55$, $p = 0.85$). That is, participants reported a greater proportion of correct words in sublist C relative to both sublists A and B. All other effects, notably including an interaction between condition, sublist, and age, did not reach the threshold for statical significance (all $p > 0.05$). See **Table 4** for descriptive statistics and **Figure C2A** in **Appendix C** for an illustration of the data. Thus, although we observed *numerically* greater memory for the earliest presented items in the list, we failed to find a statistically significant primacy effect.

Table 4.

Proportion of correct words recalled (standard deviation in parentheses) split by age group, list type, sublist, and experiment.

		Experiment 1	
Semantic trials		Younger adults	Older adults
	A	0.92 *(0.06)*	0.83 *(0.24)*
	B	0.92 *(0.05)*	0.85 *(0.25)*
	C	0.98 *(0.02)*	0.91 *(0.25)*
Phonological trials			
	A	0.90 *(0.07)*	0.81 *(0.23)*
	B	0.91 *(0.05)*	0.81 *(0.24)*
	C	0.94 *(0.04)*	0.87 *(0.25)*
		Experiment 2	
Semantic trials			Older adults
	A	-	0.90 *(0.08)*
	B	-	0.90 *(0.06)*
	C	-	0.96 *(0.03)*
Phonological trials			
	A	-	0.87 *(0.07)*
	B	-	0.89 *(0.09)*
	C	-	0.94 *(0.04)*

Note. Only older adults were tested in Experiment 2.

False memories

A repeated measures ANOVA was conducted with condition (semantic, phonological) and sublist (A, B, C) entered as within-subjects factors and age group (YA, OA) entered as the between-subjects factor. This analysis revealed a significant main effect of group ($F(1, 82) = 5.03$, $p = 0.03$, $\eta^2_p = 0.06$) such that, on average, OAs had an increased proportion of false memories as compared to YAs ($t(82) = 2.24$, $p = 0.03$). There was a significant main effect of condition ($F(1, 82) = 21.85$, $p < 0.001$) indicating an overall greater proportion of false memories in the phonological relative to the semantic condition ($t(82) = -4.67$, $p < 0.001$). There was a significant main effect of sublist ($F(1.93, 158.26) = 69.62$, $p < 0.001$, $\eta^2_p = 0.46$), establishing significant differences between sublist A and sublist C ($t(82) = 9.86$, $p < 0.001$), as well as between sublist B and sublist C ($t(82) = 11$, $p < 0.001$). Both sublists A and B showed higher proportions of false memories compared to sublist C. No other significant effects were found, notably including with age (all $p > 0.05$). See **Table 5** for descriptive statistics and **Figure 3A** for an illustration of the data.

Table 5.

Proportion of false memories (standard deviation in parentheses) split by age group, list type, sublist, and experiment.

		Experiment 1	
Semantic trials		Younger adults	Older adults
	A	0.08 *(0.06)*	0.17 *(0.24)*
	B	0.08 *(0.05)*	0.15 *(0.25)*
	C	0.02 *(0.02)*	0.09 *(0.25)*
Phonological trials			
	A	0.10 *(0.07)*	0.19 *(0.23)*
	B	0.09 *(0.05)*	0.19 *(0.24)*
	C	0.06 *(0.04)*	0.13 *(0.25)*
		Experiment 2	
Semantic trials			Older adults
	A	-	0.10 *(0.08)*
	B	-	0.09 *(0.06)*
	C	-	0.04 *(0.03)*
Phonological trials			
	A	-	0.13 *(0.07)*
	B	-	0.11 *(0.09)*
	C	-	0.06 *(0.04)*

Note. Only older adults were tested in Experiment 2.

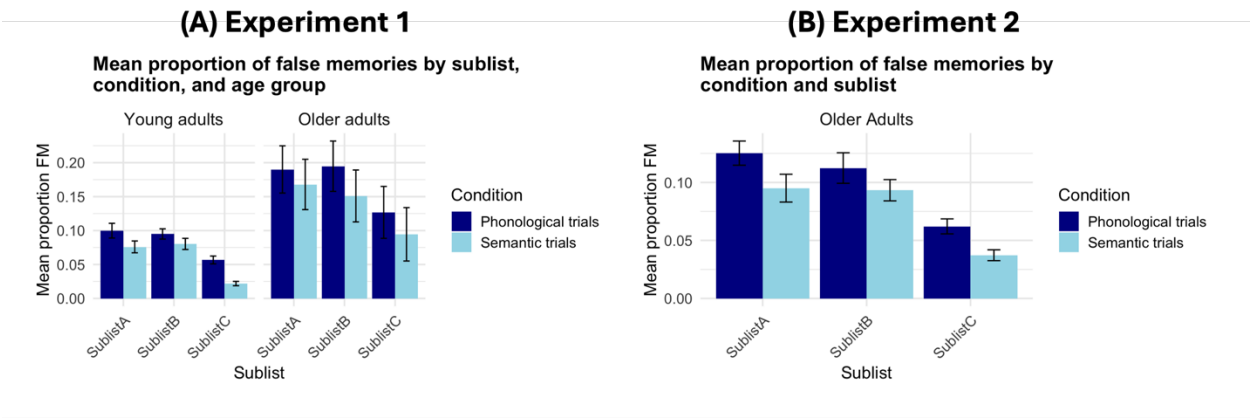

Figure 3. (A) Average proportions of false memories recalled as a function of age group (YA, OA), condition (semantic, phonological), and sublist (A, B, C). **(B)** Average proportions of false memories recalled as a function of condition (semantic, phonological), and sublist (A, B, C). In the semantic condition, false memories refer to semantic false memories, while in the phonological condition, they refer to phonological false memories. Only older adults were tested in Experiment 2. Error bars indicate standard error of the mean.

Experiment 1: Summary

In Experiment 1, we examined performance on an adapted version of the DRM paradigm, focusing on the effects of list type (semantic vs. phonological) and sublist position (A, B, C) between age groups (YAs and OAs). Our findings revealed several key patterns across both age groups: All participants, irrespective of age, recalled a higher proportion of correct words in the semantic condition compared to the phonological condition, with this effect being more pronounced in YAs. A recency effect was also observed, as all participants recalled a higher proportion of correct words from sublist C, regardless of age or condition. Furthermore, all participants generated a lower proportion of false memories from sublist C, suggesting that verbatim traces may have been more robust and thus better able to counteract gist traces from this most recent sublist (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995).

Interestingly, both groups exhibited similar patterns of false memory production. That is, despite OAs remembering fewer items overall relative to YAs, all participants generated a higher proportion of false memories in the phonological relative to the semantic condition and had similar distributions of false memories across the sublists (fewest false memories from the most recent C sublist than either sublist A or B). Importantly, these age-equivalent patterns emerge despite OAs generating more false memories overall, highlighting the robustness of these effects even in the context of significant main effects of age. This finding contradicts the prediction that OAs would be more prone to phonological errors, especially from the most recent C sublist, as compared to YAs, and suggests that false memory production is similar across age groups.

Our findings also challenge the view that the nature of false memories is strictly tied to a specific memory store, suggesting instead that the types of false memories generated are more closely related to the properties of the studied lists. To further investigate whether the surprising

finding of comparable performance observed in OAs and YAs was influenced by the multimodal presentation of stimuli (i.e., both visual and auditory), we conducted Experiment 2, focusing solely on OAs in which we presented words visually only. Prior work has indicated that presenting information in multiple modalities may enhance task performance in older adults (e.g., Mahoney et al., 2011; for a review, see De Dieuleveult et al., 2017). As such, if the enhanced encoding from the multimodal presentation in Experiment 1 masked age-related differences, we expected that removing auditory cues would reduce memory performance in OAs.

Experiment 2

Methods

Participants. A sample of 47 healthy, community dwelling seniors (aged 59 and up) were recruited from the Greater Boston area and compensated at a rate with cash ($5/half hour) for their participation. Three participants were excluded: 2 participants were excluded for having an MMSE score less than 28 and 1 participant was excluded due to errors during data collection. Results presented below are from the remaining 44 older participants (26 = female; $Mean_{age}$= 72.5 years, SD_{age}= 7.11 years; see **Table 1** for neurological assessment data).

Materials. All materials were identical to those used in Experiment 1.

Procedures. All procedures were identical to those in Experiment 1, with the exception that stimuli were presented visually only, rather than presented both visually and auditorily.

Data analysis. All analyses were identical to those in Experiment 1.

Results

Total number of recalled words (combined correct and false)

A repeated measures ANOVA with condition (semantic, phonological) and sublist (A, B, C) entered as within-subjects factors was conducted on the overall number of words recalled, regardless of correct or FM status. This analysis revealed, first, a significant main effect of condition ($F(1, 43) = 105.33$, $p < 0.001$, $\eta^2_p = 0.7$) showing that participants recalled on average a greater number of words in the semantic relative to the phonological condition ($t(43) = 10.26$, $p < 0.001$). Second, there was a significant main effect of sublist ($F(1.37, 58.80) = 59.13$, $p < 0.001$, $\eta^2_p = 0.58$). Pairwise comparisons indicated that participants recalled significantly more words in Sublist A compared to Sublist B ($t(43) = 7.29$, $p < 0.001$). Conversely, they recalled significantly fewer words in Sublist A ($t(43) = -3.90$, $p < 0.001$) and in Sublist B ($t(43) = -13.12$, $p < 0.001$) as compared to Sublist C. Lastly, this analysis revealed a significant interaction effect between condition and sublist ($F(1.96, 84.15) = 7.39$, $p < 0.001$, $\eta^2_p = 0.15$). Pairwise comparisons indicated that within all three sublists (A: $t(43) = 7.32$, $p < 0.001$; B: $t(43)$ 3.81, $p < 0.001$; C: $t(43) = 2.68$, $p = 0.01$), participants recalled more words in the semantic relative to the phonological condition. See **Table 2** for detailed descriptive statistics and **Figure C1B** in **Appendix C** for an illustration of the data.

Correct responses (output accuracy)

A repeated measures ANOVA with condition (semantic, phonological) and sublist (A, B, C) specified as within-subjects factors was conducted. While the interaction between the factors was not significant ($F(1.66, 71.34) = 0.26$, $p = 0.73$, $\eta^2_p = 0.01$), the analysis revealed a significant main effect of condition ($F(1, 43) = 16.23$, $p < 0.001$, $\eta^2_p = 0.27$). Participants recalled a greater proportion of correct words in the semantic relative to the phonological condition ($t(43) = 4.03$, $p < 0.001$). Additionally, the main effect of sublist ($F(1, 98, 85.04) = 26.18$, $p <0.001$, $\eta^2_p = 0.38$) was significant. While there was no significant difference in the

proportion of correctly recalled words between sublist A and B ($t(43) = -0.77$, $p = 0.73$), participants recalled a significant lower proportion of correct words in sublist A relative to sublist C ($t(43) = -6.94$, $p < 0.001$) as well as a significantly lower proportion of correct words in sublist B relative to sublist C ($t(43) = -5.85$, $p < 0.001$). See **Table 6** for descriptive statistics and **C2B** for an illustration of the data.

False memories

A repeated measures ANOVA with condition (semantic, phonological) and sublist (A, B, C) entered as within-subjects factors was conducted. This analysis revealed a main effect of condition ($F(1, 43) = 16.23$, $p < 0.001$, $\eta^2_p = 0.27$) such that participants reported a greater proportion of false memories in the phonological relative to the semantic condition ($t(43) = -4.03$, $p < 0.05$). The main effect of sublist was also significant ($F(1.98, 85.04) = 26.18$, $p < 0.001$, $\eta^2_p = 0.38$). Pairwise comparisons showed that while there was no significant difference in the proportion of false memories between sublists A and B ($t(43)$ 0.77, $p = 0.73$), sublist C had a significantly lower proportion of false memories compared to both sublists A ($t(43) = 6.94$, $p < .001$) and B ($t(43) = 5.85$, $p < .001$). Lastly, the interaction between the factors was not significant ($F(1.66, 71.34) = 0.26$, $p = 0.81$, $\eta^2_p = 0.01$). See **Table 5** for detailed descriptive statistics and **Figure 3B** for an illustration of the data.

Experiment 2: Summary

In Experiment 2, we focused our investigation solely on OAs and replicated key findings from Experiment 1: First, OAs recalled a higher proportion of correct words in the semantic condition compared to the phonological condition, mirroring the pattern observed in Experiment 1. A recency effect was also evident, as in Experiment 1, with OAs recalling a higher proportion of correct words from sublist C. When analyzing false memory production, OAs generated a

higher proportion of phonological false memories compared to semantic false memories. Additionally, a higher proportion of false memories were observed from sublist A relative to sublist C, complementing the finding that more items were correctly recalled from the end of the presented lists supporting the recency effect seen in correct recall, and from sublist B relative to sublist C; however, there was no interaction between the types of false memories generated and list position.

One notable pattern across the two experiments was that OAs exhibited numerically higher performance (i.e., increased accurate recall) in Experiment 2 than in Experiment 1. Although there were some demographic differences between the samples—e.g., OAs in Experiment 1 were somewhat older than those in Experiment 2—these are unlikely to fully account for the observed differences in recall performance. Instead, it appears that the bimodal (i.e., simultaneous auditory and visual) presentation format used in Experiment 1 may have been unexpectedly detrimental for OAs. While prior research suggests that multimodal presentation of information can enhance task performance in older adults (for a review, see De Dieuleveult et al., 2017), it is possible that the bimodal presentation in this case introduced additional processing demands that negatively affected OA performance on our task. Nevertheless, the patterns observed in Experiment 2 were largely consistent with those found in Experiment 1. This aligns with other research showing that presenting visual stimuli, with or without auditory input, results in similar effects on learning and memory recall in healthy older adults (Constantinidou & Baker, 2002).

Discussion

This study examined the generation of false memories across putative short- and long-term memory stores, and how age-related changes influence these processes. Using a modified

DRM paradigm with semantic and phonological sublists, we found that both younger adults (YAs) and older adults (OAs) were more likely to generate phonological than semantic false memories, regardless of sublist position. Although OAs recalled fewer items overall, the relative pattern of false memory generation was remarkably similar across age groups. We found a largely similar pattern of results in Experiment 2 in an exclusively OA sample when we restricted presentation of items to written format only, removing the potential additional support of the bimodal presentation used in Experiment 1. Once again, OAs generated a greater number of phonological as compared to semantic false memories, irrespective of list position. Critically, across both experiments, there was no interaction between sublist and list type, suggesting that semantic and phonological false memories can arise from all list positions. Altogether, our findings suggest that the nature of false memories is not, as the traditional view suggests, tied to the memory store from which they originate, but, instead, may be related to properties of the studied items in a list.

Shared cognitive processes underlying false memory generation

Given the striking similarities across age groups, it is important to consider the factors that may have contributed to the generation of false memories. Fuzzy Trace Theory (FTT; Brainerd & Reyna, 2002; Reyna & Brainerd, 1995) suggests that information is simultaneously stored as both a verbatim and a gist representation. The verbatim representation captures specific surface-level details of encoded information whereas the gist representation captures the general meaning and semantic content of the encoded information. For veridical memory, both the verbatim and gist representations are strong, facilitating accurate recall. However, as verbatim traces decay over time, individuals are thought to rely increasingly on gist representations, which, while more durable, are also more prone to errors, and, consequently, may contribute to

the generation of false memories. Consistent with this framework, both YAs and OAs in our study exhibited higher veridical recall for the most recent (i.e., sublist C) as compared to the first sublist. This is consistent with FTT's prediction that verbatim traces are stronger for recently presented items, thereby supporting more accurate recall.

However, another prediction that extends from FTT is that semantic false memories should be more frequent for items presented at the beginning of a list, as these items are primarily supported by gist representations, which are more prone to distortions related to broader semantic content (Craik & Levy, 1970). In contrast, phonological false memories should be more frequent for items presented at the end of the list, where the verbatim representation, particularly of fragile surface-level information such as acoustic or phonological details, is strongest (Craik, 1968). Consistent with this, recent work (Coane et al., 2024) has shown that phonological false memories (measured as false alarms) are more frequent at short delays, whereas semantic false memories emerge more strongly after longer delays, suggesting that the contributions of verbatim and gist representations change over time and differentially influence false memory formation. However, in lists that maximized the opportunity to form either semantic or phonological representations of studied items, we found that both semantic and phonological false memories could be generated from all list positions. One methodological difference that may account for this discrepancy is that Coane and colleagues (2024) used a recognition memory paradigm with relatively short lists (four or six items), probing primarily short-term memory processes, whereas our study employed free recall and used substantially longer lists (twelve items), thereby engaging broader memory mechanisms. Thus, the absence of an interaction between list type (semantic, phonological) and sublist position in the current study suggests that the differences in the nature of false memories (i.e., semantic vs. phonological) may

not be solely attributable to differential reliance on putative short-term versus long-term memory stores during recall. Instead, other cognitive factors that operate uniformly across all sublist positions may play a role in the generation of false memories.

One position-invariant factor that may contribute to false memory generation is source monitoring—i.e., the ability to accurately recall the context in which information was first learned (Johnson et al., 1993). Source monitoring is crucial for distinguishing between different memory sources, such as differentiating between information learned from a list of words and information generated internally (e.g., recognizing that you saw the word '*cat*' on a list, rather than having thought of the word *'cat'* after seeing the word *'bat'* on the list). Source monitoring is largely supported by the prefrontal cortex (Johnson et al., 1993; Mitchell & Johnson, 2009) and studies with YAs show increased prefrontal activation in the short-term variant of the DRM task (Atkins & Reuter-Lorenz, 2011), highlighting the role of these regions in mitigating false memory generation. Although both age groups likely used source monitoring to adjudicate studied from falsely recalled items, OAs demonstrate well-documented deficits in source monitoring (Chalfonte & Johnson, 1996; Dehon & Brédart, 2004; Hashtroudi et al., 1989; Johnson et al., 1993; Mitchell et al., 2003; Mitchell & Johnson, 2009; Spencer & Raz, 1995) likely due to both structural and functional changes in frontal and core memory regions (Fjell et al., 2013; Salat et al., 2004). To address potential concerns regarding age-related declines in executive control, OA participants were screened with a battery of neuropsychological measures (see **Table 1**), allowing us to at least partially mitigate concerns about deteriorations in executive control processing. Nevertheless, cognitive deficits in aging are multifactorial, encompassing impairments in memory discrimination (Korkki et al., 2020; Leal & Yassa, 2014; Stark et al., 2013), inhibitory control (Hasher et al., 1999), working memory deficits (Babcock & Salthouse,

1990; D'Esposito & Postle, 2015; Grady et al., 2006; Mattay et al., 2006), as well as greater difficulty in discriminating between externally and internally generated information (i.e., identifying whether a word was heard or thought about; Hashtroudi et al., 1989). Age-related deficits in source monitoring are particularly pronounced when the information being judged is highly similar (e.g., identifying which of two female speakers spoke (Ferguson et al., 1992)). In the present study, both the semantic and phonological conditions were designed to induce false memories through the presentation of intentionally similar items, either conceptually related in the semantic condition or acoustically similar in the phonological condition. As such, source monitoring demands were, importantly, comparable across our conditions. Thus, the absence of age differences across false memory types in our study suggest that cognitive mechanisms supporting source monitoring operate similarly across semantic and phonological false memories, implying that shared cognitive processes contribute to both types of memory errors.

Aligning with this notion of shared cognitive mechanisms underlying both semantic and phonological false memories, the Activation-Monitoring Framework (AMF; Roediger et al., 2001) offers a possible parsimonious theoretical explanation for our findings (for a detailed discussion, see Chang & Brainerd, 2021). The AMF suggests that words may be connected by properties such as semantic or lexical networks as well as other properties such as acoustic or phonological features (Roediger et al., 2001). Accordingly, false memories arise through the automatic activation of related concepts, alongside source monitoring errors that lead individuals to misattribute the source of their memories[2]. This framework underscores how memory retrieval can be influenced by semantic associations and prior knowledge (Gallo & Roediger, 2003), as

[2] Of note, these types of source monitoring errors are thought to stem from recollection-based processes—such as recall-to-reject and the distinctiveness heuristic (Gallo et al., 2006)–rather than from general executive control mechanisms as described by Johnson et al., (1993). However, the design of the present study was not intended to directly test the influence of these specific monitoring processes.

well as phonological similarity (Baddeley, 1964; Conrad, 1963; Wickelgren, 1965). Collins & Loftus (1975) demonstrated that semantic false memories emerge through spreading activation within semantic lexical networks, leading to the intrusion of semantically related but incorrect memories. Similarly, phonological similarity can lead to false memories by creating interference during recall, as demonstrated by Marmurek et al. (2006), who report that phonological false memories arise through misattributions based on acoustic similarity. Taken together, our findings suggest that false memories, whether semantic or phonological, emerge from shared processes, and highlight the importance of considering both phonological and semantic influences on memory retrieval within a unified theoretical framework of false memory.

Cognitive mechanisms underlying age differences in overall false memory rates

Additional cognitive mechanisms may also contribute to age-related differences in overall false memory rates. Post-retrieval monitoring—evaluating the relevance of retrieved information to the goal of retrieval (Rugg, 2004)—declines with increasing age (Henkel et al., 1998; Johnson et al., 1993). Such monitoring processes are essential for verifying the veracity of retrieved information. In the context of the DRM paradigm, such monitoring failures may contribute to the increased susceptibility of older adults to false memories. In other words, older adults may be unable to monitor and reject a falsely remembered item that is similar to one they were originally presented with, thereby leaving them more prone to false memories in general (Dehon, 2020; Gallo & Roediger, 2003; Mitchell & Johnson, 2009). Age-related declines in rehearsal and working memory may further reduce the ability of OAs to verify retrieved information and increasing memory errors. Evidence indicates that OAs generate overall more false memories in working memory tasks relative to YAs (Jantz et al., 2021), possibly due to age-related decline in rehearsal processes (Naveh-Benjamin & Cowan, 2023). Relatedly, Pirolle and

colleagues (2025) demonstrated that the rate of rehearsal is positively associated with performance, such that limiting rehearsal opportunities increases the incidence of false memories. Taken together, these age-related declines in monitoring and rehearsal processes likely contribute to the elevated false memory rates we observed in OAs relative to YAs.

Another possibility that could help to understand the group-level differences in the overall level of false memories are the differences observed in veridical memory rates. That is, verbatim traces can play a dual role by both supporting veridical memory, but also counteracting gist-based traces that can support false memories (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). In Experiment 1, we observed a significant main effect of age group on both correct responses and false memories. Although these age group effects did not interact with list type or sublist position, they raise the possibility that the differences observed in false memories could be explained by the differences in veridical memory performance. Additionally, this relative increase in the proportion of false memories in OAs was partly driven by lower recall of items overall in OAs (both veridical and false memories) compared to their younger counterparts. This aligns with previous findings that have reported overall reduced episodic memory even in cognitively healthy OAs (Brickman & Stern, 2009; Craik & Salthouse, 2011; Rönnlund et al., 2005). When focusing on correctly recalled items, OAs also recalled a lower proportion of correct items than YAs, consistent with prior research using the DRM paradigm in aging samples (Balota et al., 1999; Budson et al., 2003; Dehon, 2020; Dennis et al., 2007; Norman & Schacter, 1997; Tun et al., 1998). Although we attempted to mitigate the impact of these group-level output differences by using proportion scoring (Khanna & Cortese, 2009; McGuire et al., 2015; Sugrue & Hayne, 2006), it could still be that differences in overall levels of false memories between age

groups can at least be partially explained by differences in verbatim traces and concomitant veridical memory (c.f., Atkins & Reuter-Lorenz, 2008).

Importantly, even when using a proportion scoring to account for age-related differences in total output, there were no significant interactions between age and false memories with sublist position. Instead, we saw commonalities: Both YAs and OAs made more false memories in the phonological relative to the semantic condition and were more likely to make false memories from the A and B than C sublist. These findings challenge the hypothesis that OAs are more prone to semantic errors than YAs due to their presumed reliance on gist-based memory traces (Abadie & Camos, 2019; Brainerd et al., 2008; Grilli & Sheldon, 2022; Schacter et al., 1997). Instead, our findings suggest that false memory generation may be more influenced by factors such as the properties of the studied list rather than age-related memory differences.

Role of list characteristics in false memory generation

Research beyond the aging literature also highlights the importance of list features in generating false memories (for an in-depth review, see Coane et al., 2021). For example, manipulating associative strength—the degree to which words co-occur in language—can strongly induce false memories (Kumar, 2021). In this case, similarity is based on conceptual links between words (e.g., '*cat',* '*dog'*, and '*bird'* have a high associative strength because they are all domestic animals). Another relevant factor is categorical similarity which refers to the degree to which items belong to the same category (e.g., '*apple'*, '*banana'*, and '*orange'* all belong to the category of fruits; Buchanan et al., 1999). Other studies have explored orthographic or phonological similarity, showing that false memories can also be induced when words share similar spelling or pronunciation. For instance, Sommers and Lewis (1999) presented participants with 15-item lists and found that lists containing words with greater phonological

similarity (e.g., '*fat*', '*that*', '*cab*' for the word '*cat*') led to more false memories. This effect was attenuated when phonological similarity was reduced (e.g., '*can*', '*coat*', '*scat*', '*cash*' for the word '*cat*'). Our findings extend this literature by demonstrating that list characteristics can fundamentally shape memory errors and the generation of false memories, irrespective of age. Critically, when items are presented in a way that encourages phonological errors, both YAs and OAs are equally susceptible to these types of memory distortions. This builds upon previous research by Dimsdale-Zucker (2018), who demonstrated that phonological errors can occur at all list positions. Importantly, the current study extends this work by showing that this effect occurs within the same participants, across age groups, and under the same study conditions.

Another key feature of the list design in our paradigm is the combination of a long, supra-span list—typically used to isolate memory representations from separate short- and long-term memory stores (Murdock, 1962)—with nested sublists that allow for more precise tagging of early, middle, and late study positions (adapted from Dimsdale-Zucker, et al., 2019). This structure allows us to examine at the relative predominance of both veridical and false memories from different list positions. As expected, we observed a robust recency effect across both age groups: YAs and OAs recalled a greater proportion of items from the end of the list (sublist C) than from earlier positions (sublists A and B), consistent with prior work attributing this effect to short-term memory processes (Deese & Kaufman, 1957; Glanzer & Cunitz, 1966; Howard et al., 2006; Murdock, 1962). The recency advantage likely reflects the greater accessibility of recently encountered items, possibly due to reduced interference or decay (Glanzer & Cunitz, 1966; Murdock, 1962). Complementing this, the proportion of false memories was lowest for items presented at the end of the list (sublist C), in comparison to those from the beginning (sublist A) or middle (sublist B) positions. These results suggest that recent items are not only more

accurately recalled but also less likely to result in false memories, possibly because they are stored as more accessible verbatim traces, minimizing effects of interference and decay. Overall, these findings indicate that the cognitive mechanisms underlying the recency effect remain largely intact across age, even in the presence of broader age-related declines in memory performance.

Conclusion

From a theoretical perspective, our findings challenge the traditional notion that semantic and phonological false memories emerge from distinct memory stores. Instead, they suggest that the cognitive mechanisms responsible for false memories might be more similar across short- and long-term memory systems than previously thought. That is, both semantic and phonological false memories may arise through common underlying cognitive processes, including the activation of related concepts and the misattribution of source information. This challenges the traditional dual-store perspective of memory (Atkinson & Shiffrin, 1968; Glanzer & Cunitz, 1966) and aligns more closely with the idea that memory is a fluid and interconnected system, where different types of information (e.g., phonological and semantic) are not processed or stored in isolation across different memory systems but rather coexist in a unified memory store (Jonides et al., 2008; Nairne, 2002; however see Forsberg et al., 2025 for recent evidence suggesting an alternative perspective).

In conclusion, this study provides important insights into the nature of false memory generation across the lifespan. Our findings demonstrate that while OAs exhibit overall declines in veridical episodic memory performance and an increased susceptibility to false memories, their relative propensity to generate semantic versus phonological false memories remains similar to that of YAs. These findings challenge the notion that semantic and phonological false

memories operate via entirely separate storage mechanisms and support the idea that common underlying cognitive processes drive these memory distortions.

Data availability statement: The data and analysis scripts that support the findings of this study are available on https://github.com/dzmemoryandcontextlab/FMOA.git; none of the experiments was preregistered.

Disclosure statement: The authors have no relevant financial or non-financial interests to disclose.

References

Abadie, M., & Camos, V. (2019). False memory at short and long term. *Journal of Experimental Psychology: General*, *148*(8), 1312.

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition*, *36*(1), 74–81.

Atkins, A. S., & Reuter-Lorenz, P. A. (2011). Neural mechanisms of semantic interference and false recognition in short-term memory. *NeuroImage*, *56*(3), 1726–1734.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*(4), 89–195.

Babcock, R. L., & Salthouse, T. A. (1990). Effects of increased processing demands on age differences in working memory. *Psychology and Aging*, *5*(3), 421.

Baddeley, A. D. (1964). Semantic and acoustic similarity in short-term memory. *Nature*, *204*(4963), 1116–1117.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362–365.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology*, *20*(3), 249–264.

Baddeley, A. D., & Hitch, G. (1974). *Working memory* (G. Bower, Ed.; Vol. 8). New York: Academic Press.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and False Memories in Healthy Older Adults and in

Dementia of the Alzheimer'S Type. *Cognitive Neuropsychology*, *16*(3–5), 361–384. https://doi.org/10.1080/026432999380834

Barnhardt, T. M., Choi, H., Gerkens, D. R., & Smith, S. M. (2006). Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language*, *55*(2), 213–231.

Bartha, M. C., Martin, R. C., & Jensen, C. R. (1998). Multiple interference effects in short-term recognition memory. *The American Journal of Psychology*, *111*(1), 89.

Bartlett, F. C. (1993). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev*, *8*, 77–100.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343.

Brickman, A. M., & Stern, Y. (2009). *Aging and memory in humans*.

Buchanan, L., Brown, N. R., Cabeza, R., & Maitson, C. (1999). False memories and semantic lexicon arrangement. *Brain and Language*, *68*(1–2), 172–177.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*, *51*(3), 251–261.

Chalfonte, B. l, & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403–416.

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language*, *119*, 104210.

Coane, J. H., McBride, D. M., Chang, K., Cam, Y., & Marsh, E. (2024). Comparison of semantic and phonological false memories in short-and long-term tests. *Quarterly Journal of Experimental Psychology*, *77*(12), 2551–2571.

Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: A review of how structural and conceptual similarity affect false memory. *Frontiers in Psychology*, *12*, 668550.

Coane, J. H., McBride, D. M., Raulerson III, B. A., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature*, *197*(4871), 1029–1030.

Constantinidou, F., & Baker, S. (2002). Stimulus modality and verbal learning performance in normal aging. *Brain and Language*, *82*(3), 296–311.

Craik, F. I. M. (1968). Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*, *7*(6), 996–1004.

Craik, F. I. M. (2023). Memory, aging and the brain: Old findings and current issues. *Aging Brain*, *4*, 100096.

Craik, F. I. M., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology*, *86*(1), 77.

Craik, F. I. M., & Salthouse, T. A. (2011). *The handbook of aging and cognition*. Psychology press.

De Dieuleveult, A. L., Siemonsma, P. C., Van Erp, J. B., & Brouwer, A.-M. (2017). Effects of aging in multisensory integration: A systematic review. *Frontiers in Aging Neuroscience*, *9*, 80.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology*, *54*(3), 180.

Dehon, H. (2020). Variations in processing resources and resistance to false memories in younger and older adults. In *Memory Editing Mechanisms* (pp. 692–711). Psychology Press.

Dehon, H., & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging*, *19*(1), 191.

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia*, *45*(14), 3157–3166.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*(1), 115–142.

Dimsdale-Zucker, H. R., Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2018). Serial position-dependent false memory effects. *Memory*, *27*(3), 397–409.

Dimsdale-Zucker, H. R., Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2019). Serial position-dependent false memory effects*. *Memory*, *27*(3), 397–409. https://doi.org/10.1080/09658211.2018.1513039

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and Aging*, *7*(3), 443.

Festini, S. B., & Reuter-Lorenz, P. A. (2013). The short-and long-term consequences of directed forgetting in a working memory task. *Memory*, *21*(7), 763–777.

Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A. M., & Walhovd, K. B. (2013). Critical ages in the life course of the adult brain: Nonlinear subcortical aging. *Neurobiology of Aging*, *34*(10), 2239–2247.

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1331.

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*(3), 189–198.

Forsberg, A., Guitard, D., Greene, N. R., Naveh-Benjamin, M., & Cowan, N. (2025). Differential information transfer and loss between working memory and long-term memory across serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Gallo, D. (2013). *Associative illusions of memory: False memory research in DRM and related tasks*. Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848.

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*(6), 730–741.

Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, *31*(7), 1036–1044.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 351–360.

Glisky, E. L., Rubin, S. R., & Davidson, P. S. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1131.

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, *13*(7), 491–505.

Grady, C. L., Springer, M. V., Hongwanishkul, D., McIntosh, A. R., & Winocur, G. (2006). Age-related changes in brain activity across the adult lifespan. *Journal of Cognitive Neuroscience*, *18*(2), 227–241.

Grilli, M. D., & Sheldon, S. (2022). Autobiographical event memory and aging: Older adults get the gist. *Trends in Cognitive Sciences*.

Hasher, L., Zacks, R. T., & May, C. P. (1999). *Inhibitory control, circadian arousal, and age*.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, *4*(1), 106.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, *127*(3), 251.

Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag recency effects with the temporal context model. *Psychonomic Bulletin & Review*, *13*(3), 439–445.

James, T. A., Weiss-Cowie, S., Hopton, Z., Verhaeghen, P., Dotson, V. M., & Duarte, A. (2021). Depression and episodic memory across the adult lifespan: A meta-analytic review. *Psychological Bulletin*, *147*(11), 1184.

Jantz, T. K., Festini, S. B., & Reuter-Lorenz, P. A. (2021). Failing to forget? Evidence for both impaired and preserved working memory control in older adults. *Aging, Neuropsychology, and Cognition*, *28*(6), 884–906.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.*, *59*(1), 193–224.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*(8), 1169–1180.

Khanna, M. M., & Cortese, M. J. (2009). Children and adults are differentially affected by presentation modality in the DRM paradigm. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(6), 859–877.

Korkki, S. M., Richter, F. R., Jeyarathnarajah, P., & Simons, J. S. (2020). Healthy ageing reduces the precision of episodic memory retrieval. *Psychology and Aging*, *35*(1), 124.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80.

Leal, S. L., & Yassa, M. A. (2014). Effects of aging on mnemonic discrimination of emotional information. *Behavioral Neuroscience*, *128*(5), 539.

Levenshtein, V. (1966). *Binary codes capable of correcting deletions, insertions, and reversals. Cybernetics and control theory 10 (8)*.

Lim, L. C. L., & Goh, W. D. (2019). False recognition modality effects in short-term memory: Reversing the auditory advantage. *Cognition*, *193*, 104008.

Loftus, E. F. (1997). Creating false memories. *Scientific American*, *277*(3), 70–75.

MacDuffie, K. E., Atkins, A. S., Flegal, K. E., Clark, C. M., & Reuter-Lorenz, P. A. (2012). Memory distortion in Alzheimer's disease: Deficient monitoring of short-and long-term memory. *Neuropsychology*, *26*(4), 509.

Mahoney, J. R., Li, P. C. C., Oh-Park, M., Verghese, J., & Holtzer, R. (2011). Multisensory integration across the senses in young and old adults. *Brain Research*, *1426*, 43–53.

Marmurek, H. H. C., Proctor, C., & Javor, A. (2006). Stroop-like serial position effects in color naming of words and nonwords. *Experimental Psychology*, *53*(2), 105–110.

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., Meyer-Lindenberg, A., Goldberg, T. E., Callicott, J. H., & Weinberger, D. R. (2006). Neurophysiological

correlates of age-related changes in working memory capacity. *Neuroscience Letters*, *392*(1–2), 32–37.

McBride, D. M., Coane, J. H., Xu, S., Feng, Y., & Yu, Z. (2019). Short-term false memories vary as a function of list type. *Quarterly Journal of Experimental Psychology*, *72*(12), 2726–2741.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*(7), 1065–1077.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language*, *35*(2), 212–230.

McDermott, K. B., & Roediger III, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*(3), 508–520.

McGuire, K., London, K., & Wright, D. B. (2015). Developmental trends in false memory across adolescence and young adulthood: A comparison of DRM and memory conformity paradigms. *Applied Cognitive Psychology*, *29*(3), 334–344.

Miller, G. (1956). Human memory and the storage of information. *IRE Transactions on Information Theory*, *2*(3), 129–137.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638.

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology: The*

Official Journal of the Society for Applied Research in Memory and Cognition, *17*(1), 107–119.

Monge, Z. A., & Madden, D. J. (2016). Linking cognitive and visual perceptual decline in healthy aging: The information degradation hypothesis. *Neuroscience & Biobehavioral Reviews*, *69*, 166–173.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*(1), 53–81.

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*(5), 1170.

Naveh-Benjamin, M., & Cowan, N. (2023). The roles of attention, executive function and knowledge in cognitive ageing of working memory. *Nature Reviews Psychology*, *2*(3), 151–165.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*(6), 838–848.

Nyberg, L., & Pudas, S. (2019). Successful memory aging. *Annual Review of Psychology*, *70*, 219–243.

Old, S. R., & Naveh-Benjamin, M. (2008). Memory for people and their actions: Further evidence for an age-related associative deficit. *Psychology and Aging*, *23*(2), 467.

Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science*, *21*(8), 1123–1133.

Piroelle, M., Guette, C., & Abadie, M. (2025). The role of articulatory rehearsal in short-term false memories during ageing. *Quarterly Journal of Experimental Psychology*, 17470218241269320.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75.

Roberts, K. L., & Allen, H. A. (2016). Perception and cognition in the ageing brain: A brief review of the short-and long-term links between perceptual and cognitive decline. *Frontiers in Aging Neuroscience*, *8*, 39.

Robinson, K. J., & Roediger III, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*(3), 231–237.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roediger III, H. L., Balota, D. A., & Watson, J. M. (2001). *Spreading activation and arousal of false memories.*

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, *20*(1), 3.

Rugg, M. D. (2004). *Retrieval processing in human memory: Electrophysiological and fMRI evidence.*

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., Morris, J. C., Dale, A. M., & Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex*, *14*(7), 721–730.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, *13*(4), 140–144.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661. https://doi.org/10.1038/nrn2213

Schacter, D. L., Koutstaal, W., Johnson, M. K., Gross, M. S., & Angell, K. E. (1997). False recollection induced by photographs: A comparison of older and younger adults. *Psychology and Aging*, *12*(2), 203.

Schneider, B. A., & Pichora-Fuller, M. K. (2000). *Implications of perceptual deterioration for cognitive aging research.*

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, *9*(1), 20–26.

Shipley, W. C. (1986). Shipley institute of living scale. *The Journal of Psychology: Interdisciplinary and Applied*.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*, *40*(1), 83–108.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, *10*(4), 527.

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449.

Sugrue, K., & Hayne, H. (2006). False memories produced by children and adults in the DRM paradigm. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(5), 625–631.

Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learning & Memory*, *16*(5), 338–342.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*, 381–403. https://doi.org/10.1017/S0140525X00047257

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, *13*(2), 230.

Watson, J. M., Balota, D. A., & Roediger III, H. L. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language*, *49*(1), 95–118.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, *70*(1), 102.

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*(9), 968–979.

Appendix A. The forty-six possible study lists used in Experiments 1 & 2

buy-forever-nose	snake-blood-flash	sweet-doctor-pitch
bread-dirt-funny	whiskey-baby-handle	rough-morning-draw
church-man-urge	chain-vote-lesson	fog-basement-shake
finish-girl-process	beautiful-sleep-heaven	long-sneeze-train
site-boring-teeth	sheet-job-chair	rent-dance-fail
logic-sink-horse	high-math-dense	pen-annoy-hydrogen
shirt-broom-whole	hold-thief-boom	waiter-muscle-trash
window-give-impact	chaos-smart-height	slow-butterfly-justice
clam-fruit-data	smell-present-fence	king-bee-mine
focus-many-alone	back-roof-city	building-square-decrease
admission-ring-river	trip-basket-argue	marry-shovel-plain
soap-pig-nut	scale-fate-lie	army-dry-speaker
gas-piano-opinion	act-corn-hawk	atom-stress-hint
adult-anger-needle	black-cold-again	cheese-flag-lose
rain-mix-plaster	answer-letter-carpet	pan-bake-cabbage
rock-movie-saloon		

Appendix B. The 138 theme words in alphabetical order and their phonological and semantic associates used in Experiments 1 & 2

Theme	Phonological associates	Average Levenshtein distance	Semantic associates	Average backward association strength (BAS)
act	tract, tact, pact, enact	1.5	drama, pretend, portray, perform	0.22
admission	mission, nutrition, audition, condition	4.5	ticket, enter, in, school	0.06
adult	result, consult, insult, fault	3	grown, mature, responsible, kid	0.12
again	remain, detain, sustain, obtain	3.25	twice, never, stop, repeat	0.12
alone	prone, clone, lone, zone	1.5	solo, isolated, lonely, one	0.33
anger	stranger, linger, danger, finger	2	fury, enrage, rage, mad	0.40
annoy	soy, deploy, joy, employ	3.5	irritate, bother, aggravate, disturb	0.24
answer	sweater, center, setter, better	4.25	stamp, mail, mailbox, envelope	0.35
argue	value, revenue, rescue, pursue	3.75	complain, debate, disagree, agree	0.25
army	calmly, smarmy, fondly, darkly	3.25	soldier, navy, marines, infantry	0.35
atom	awesome, custom, bottom, blossom	3.75	neutron, molecule, proton, nucleus	0.26
baby	maybe, flabby, lobby, hubby	3	crib, diaper, carriage, infant	0.59
back	plaque, crack, track, quack	2.75	front, spine, behind, forward	0.35
bake	lake, mistake, fake, brake	1.75	oven, cook, broil, cake	0.15
basement	department, experiment, confident, element	4.75	underground, attic, downstairs, cellar	0.16
basket	jacket, ticket, cricket, picket	3	laundry, waste, wicker, picnic	0.16
beautiful	careful, hopeful, plentiful, dutiful	4.5	gorgeous, ugly, pretty, lovely	0.12
bee	three, free, flee, see	2	sting, hive, buzz, bumble	0.68
black	snack, slack, track, smack	1.75	gray, white, coal, brown	0.41
blood	bud, thud, spud, flood	3	vampire, vein, donor, plasma	0.51
boom	zoom, fume, assume, consume	3.75	bomb, sonic, bang, explosion	0.14
boring	soaring, scoring, wearing, roaring	2.25	dull, lecture, interesting, exciting	0.23
bread	head, pled, tread, read	1.75	rye, butter, toast, loaf	0.52
broom	plume, room, bloom, resume	3	mop, sweep, dustpan, witch	0.30
building	ending, finding, holding, forbidding	3.75	structure, construction, empire, blocks	0.23
butterfly	multiply, ally, apply, supply	6.25	moth, cocoon, insect, wing	0.17
buy	why, bye, guy, sly	1.75	sell, store, purchase, spend	0.34

cabbage	voyage, foliage, package, village	3.75	sauerkraut, patch, slaw, lettuce	0.04
carpet	cabinet, summit, planet, helmet	4	red, floor, magic, rug	0.14
chain	cane, contain, pain, brain	2.5	whip, bicycle, necklace, link	0.17
chair	pair, lair, fair, stair	2	table, swivel, rocking, recliner	0.62
chaos	moss, cross, toss, floss	3.75	hectic, havoc, confusion, anarchy	0.14
cheese	these, breeze, tease, please	2.75	swiss, cheddar, mouse, cracker	0.45
church	lurch, search, perch, birch	2.75	preacher, cathedral, steeple, temple	0.48
city	witty, bitty, pity, ditty	1.75	urban, suburb, town, metropolis	0.42
clam	swam, cram, glam, exam	1.5	mussel, chowder, shell, oyster	0.29
cold	fold, sold, bold, old	1	hot, shiver, arctic, frigid	0.64
corn	born, worn, torn, adorn	1.25	husk, field, flake, cob	0.47
dance	glance, advance, enhance, chance	2.25	song, ballet, aerobics, ballerina	0.22
data	debater, later, beta, crater	3.25	analysis, computer, information, entry	0.04
decrease	niece, fleece, cease, crease	4.25	lower, reduce, less, increase	0.12
dense	dispense, hence, defense, commence	3	thick, stupid, forest, dumb	0.00
dirt	divert, alert, flirt, hurt	2.25	grime, soil, filth, mud	0.62
doctor	roster, lobster, proctor, blotter	3.25	stethoscope, nurse, physician, surgeon	0.59
draw	thaw, jaw, claw, flaw	2	crayon, scribble, sketch, paint	0.25
dry	pry, fly, deny, cry	1.5	moist, thirst, desert, towel	0.15
fail	nail, rail, sail, bail	1	pass, succeed, try, flunk	0.28
fate	hate, late, bate, mate	1	luck, death, destiny, life	0.05
fence	suspense, whence, intense, immense	3.75	yard, wire, gate, barrier	0.13
finish	diminish, nourish, flourish, tarnish	3.5	end, done, complete, start	0.39
flag	tag, lag, rag, bag	1.75	pole, stripes, checkered, banner	0.32
flash	cash, ash, bash, clash	1.75	light, lightning, bright, camera	0.02
focus	bonus, locus, crocus, caucus	2	attention, concentrate, blur, lens	0.01
fog	frog, clog, bog, dog	1.25	mist, smog, haze, unclear	0.33
forever	clever, sever, lever, endeavor	3.5	eternity, always, endless, infinity	0.46
fruit	cute, mute, suit, commute	4	citrus, pear, berry, kiwi	0.45
funny	money, puny, runny, sunny	1.75	comedian, humor, hilarious, clown	0.53

gas	mass, amass, class, brass	2.75	oil, fuel, station, petroleum	0.34
girl	twirl, hurl, swirl, whirl	2	dress, dolls, female, boy	0.27
give	live, active, elective, olive	2.75	take, generous, charity, share	0.30
handle	candle, vandal, sandal, scandal	2.75	suitcase, care, door, drop	0.01
hawk	walk, stalk, chalk, squawk	2.5	bird, eagle, dove, prey	0.01
heaven	eleven, leaven, heathen, seven	2	angel, hell, god, earth	0.17
height	fight, right, night, tight	2	basketball, feet, length, tall	0.01
high	rely, sigh, defy, why	3.25	jump, elevate, tower, low	0.28
hint	tint, lint, mint, squint	1.5	guess, suggestion, clue, secret	0.05
hold	told, gold, scold, mould	1.5	grip, grasp, keep, carry	0.29
horse	source, endorse, enforce, concourse	3.75	gallop, saddle, pony, colt	0.74
hydrogen	conclusion, burgeon, region, occasion	6.75	air, helium, oxygen, peroxide	0.09
impact	compact, contact, exact, abstract	3.5	hit, crash, car, impression	0.01
job	blob, knob, mob, snob	1.75	occupation, career, task, employment	0.55
justice	armistice, solstice, service, practice	3.75	courts, truth, lawyer, liberty	0.08
king	sing, ding, fling, bring	1.5	crown, throne, reign, queen	0.59
lesson	parson, lessen, fasten, listen	2.75	teacher, teach, learn, school	0.00
letter	center, sweater, better, setter	1.75	envelope, mail, mailbox, stamp	0.35
lie	pie, dye, tie, ply	1.75	untruthful, dishonest, fib, deception	0.45
logic	tragic, optic, allergic, garlic	3.5	thought, reason, sense, philosophy	0.02
long	strong, gong, wrong, throng	2.25	far, short, distance, hair	0.13
lose	confuse, choose, cruise, accuse	3.75	gain, find, defeat, win	0.46
man	tan, van, fan, can	1	handsome, male, lady, woman	0.31
many	granny, zany, nanny, penny	2.25	few, plenty, several, much	0.31
marry	tarry, ferry, wary, starry	1.75	engage, hitch, wed, single	0.26
math	heath, hath, bath, path	1.25	equation, algebra, arithmetic, calculus	0.69
mine	line, nine, wine, fine	1	yours, selfish, possession, me	0.22
mix	affix, fix, six, nix	1.5	blend, stir, integrate, combine	0.38
morning	charming, disarming, dawning, lining	3.75	early, dawn, dew, evening	0.19
movie	gravy, savvy, wavy, groovy	4.5	popcorn, film, theater, cinema	0.54

muscle	hassle, castle, rustle, tassel	3.25	strength, weights, tone, flex	0.19
needle	paddle, beatle, meddle, handle	2.5	thread, injection, syringe, haystack	0.51
nose	rose, pose, chose, prose	1.5	face, throat, ear, big	0.07
nut	strut, hut, gut, cut	1.5	squirrel, almond, pecan, cashew	0.47
opinion	onion, minion, dominion, pavilion	2.5	fact, choice, bias, view	0.08
pan	an, ran, lifespan, ban	2	pot, fry, skillet, dish	0.22
pen	ten, hen, when, then	1.5	marker, pencil, quill, write	0.37
piano	rhino, soprano, domino, volcano	3.5	organ, keyboard, ivory, guitar	0.19
pig	rig, jig, gig, fig	1	sty, hog, sow, pork	0.44
pitch	rich, which, ditch, itch	1.75	catch, throw, ball, baseball	0.01
plain	sprain, crane, sane, grain	3	ordinary, plaid, simple, fancy	0.11
plaster	pastor, blaster, master, caster	1.75	wall, concrete, cement, cast	0.00
present	lenient, absorbent, rodent, omniscient	4.75	past, future, absent, gift	0.35
process	compress, profess, progress, address	3	way, elimination, procedure, system	0.03
rain	vain, pain, main, wane	1.5	hail, storm, umbrella, puddle	0.38
rent	dent, extent, bent, sent	1.5	own, lease, apartment, monthly	0.14
ring	thing, cling, swing, spring	2	bell, diamond, jewelry, phone	0.31
river	sliver, driver, giver, quiver	1.5	bridge, stream, creek, flow	0.30
rock	block, smock, lock, clock	1.75	stone, solid, boulder, roll	0.39
roof	aloof, spoof, proof, goof	1.5	shingle, tin, tar, ceiling	0.27
rough	enough, fluff, stuff, gruff	3.5	tough, coarse, smooth, sandpaper	0.33
saloon	cartoon, baboon, lagoon, balloon	2.25	beer, drink, cowboy, bar	0.00
scale	ale, pale, stale, whale	1.75	pound, balance, weigh, measure	0.04
shake	quake, wake, opaque, steak	2.75	hands, vanilla, wiggle, milk	0.03
sheet	seat, cheat, sleet, meet	1.75	pillow, paper, blanket, cover	0.03
shirt	assert, squirt, overt, insert	3	blouse, sleeves, collar, shorts	0.40
shovel	gavel, marvel, level, bevel	3	dig, spade, rake, pail	0.17
sink	pink, blink, kink, think	1.5	drain, float, faucet, bathroom	0.21
site	quite, polite, unite, kite	2	area, place, camp, location	0.02
sleep	leap, steep, bleep, heap	1.75	bed, nap, awake, doze	0.67

slow	glow, grow, crow, blow	1.5	turtle, fast, snail, sluggish	0.45
smart	tart, cart, heart, dart	2	genius, intelligent, knowledge, wis	0.47
smell	cell, spell, dwell, swell	1.5	scent, whiff, aroma, stench	0.61
snake	sake, stake, ache, make	2	slither, serpent, reptile, rattle	0.54
sneeze	wheeze, squeeze, freeze, ease	2.5	allergy, tissue, cough, handkerchief	0.22
soap	scope, slope, microscope, cope	4.25	lather, suds, bubble, shower	0.34
speaker	beaker, sneaker, seeker, weaker	1.75	sound, guest, stereo, microphone	0.05
square	rare, spare, snare, scare	2.25	triangle, circle, shape, round	0.37
stress	bless, assess, confess, chess	3.25	anxiety, pressure, tension, strain	0.19
sweet	heat, street, beet, sheet	2	bitter, sugar, honey, sour	0.43
teeth	beneath, wreath, sheath, heath	2.75	gums, braces, tongue, mouth	0.40
thief	chief, brief, relief, grief	2	robber, burglar, bandit, crook	0.31
train	lane, slain, stain, attain	2.5	track, engine, railroad, caboose	0.37
trash	stash, flash, mash, dash	2	garbage, dump, rubbish, debris	0.33
trip	clip, drip, flip, chip	1.75	baggage, travel, journey, vacation	0.25
urge	merge, purge, scourge, surge	1.75	convince, impulse, need, persuade	0.01
vote	afloat, dote, note, boat	2.5	ballot, campaign, election, register	0.24
waiter	hater, gator, traitor, skater	2.75	server, waitress, food, tip	0.19
whisky	frisky, sticky, tricky, risky	2.5	scotch, gin, bourbon, rum	0.08
whole	sole, role, goal, mole	2.5	half, all, part, piece	0.16
window	bravado, meadow, shadow, avocado	4.5	curtain, shutter, pane, sill	0.55

Notes. The Levenshtein distance is a metric that measures the minimum number of single-character insertions, deletions, or substitutions needed to transform one string into another, with lower values indicating greater orthographic similarity and higher values indicating greater dissimilarity (Levenshtein, 1966). The average across the four phonological associates is reported. Backward association strength (BAS) quantifies the probability that an associate word is produced in response to a given theme word. The average BAS was calculated across the four semantic associates and obtained from the University of South Florida Free Association Norms (Nelson et al., 2004).

Appendix C

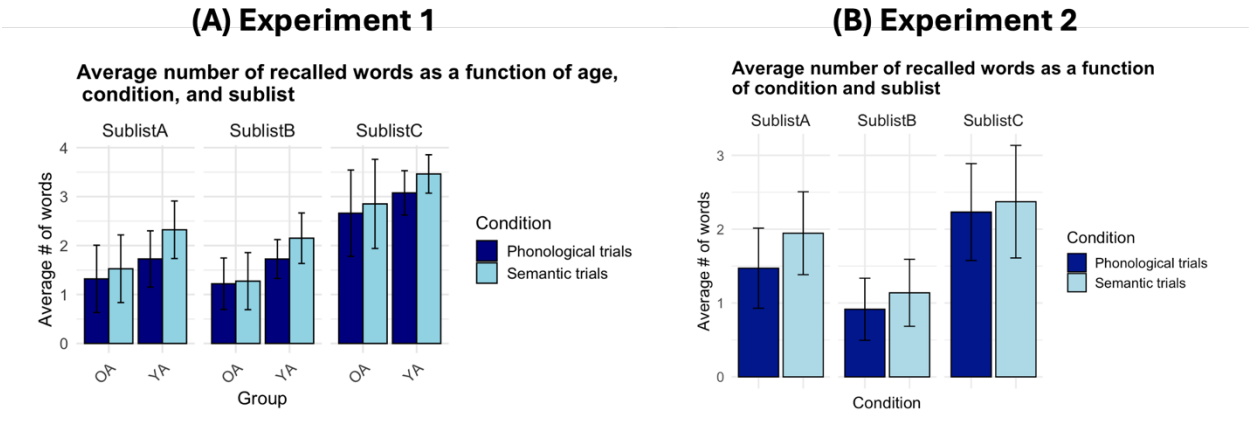

Figure C1. (A) Average number of words (correct and false) recalled by young adult (YA) and older adult (OA) participants as a function of condition (semantic, phonological) and sublist (A, B, C). **(B)** Average number of words recalled by OA participants as a function of condition (semantic, phonological) and sublist (A, B, C). Only older adults were tested in Experiment 2. Error bars indicate ±1 standard deviation from the mean.

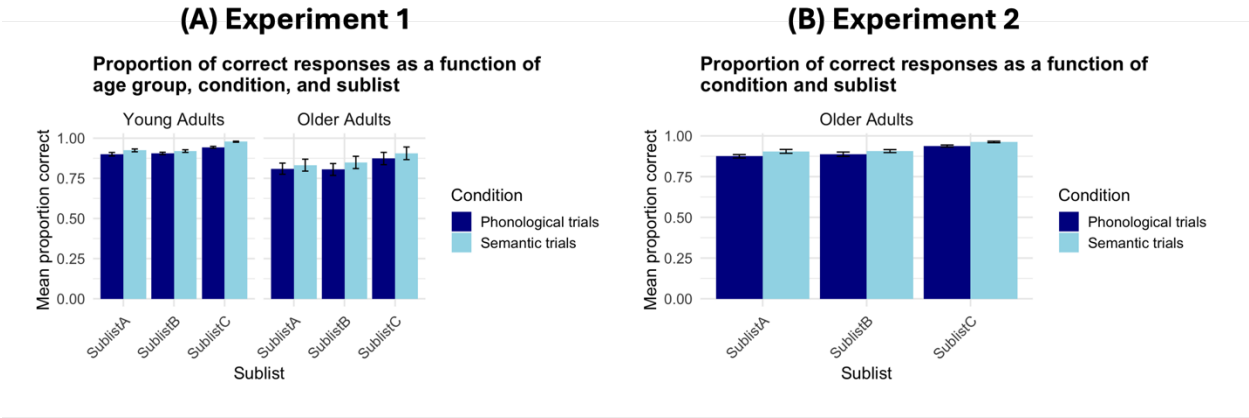

Figure C2. (A) Average proportions of correct responses recalled as a function of age group (YA, OA), condition (semantic, phonological), and sublist (A, B, C). **(B)** Proportion of correctly recalled words by older adult (OA) participants as a function of condition (semantic, phonological) and sublist (A, B, C). Only older adults were tested in Experiment 2. Error bars indicate standard error of the mean.